Improving Full Immunization Rates in Haryana: Evaluating Incentives and Communication Methods

Updated Pre-Analysis Plan

Abhijit Banerjee, Arun Chandrasekhar, Esther Duflo, John Floretta, Matthew Jackson, Harini Kannan, Francine Loza, Anna Schrimpf, Maheshwor Shrestha

June 18th, 2018

This document is intended to accompany the original pre-analysis plan submitted for this project in August 2017 and included in the appendix, the purpose of which is to present our revised analysis plan. Broadly, there are two main modifications to the original implementation and evaluation plan. In section I, we present a detailed description of these changes, and proceed in Section 2 to present the revisions to the empirical strategy in light of these changes.

1 Modifications to the Original Data Collection & Implementation Plan

1.1 Revised Endline Strategy

At the onset, we were unsure of the quality of the administrative data being collected through the tablets, and therefore planned to conduct a comprehensive endline survey[1] for analysis purposes. Further, to assess the quality of the tablet data, we carried out a "child verification survey", as described in section 4.2.IV of the original plan. The first round of this systematic audit suggested that the tablet data was far more accurate than we had anticipated; in particular, we were able to locate 91% of thhe children successfully. Amongst the children we located, identifiers matched 96% of the time [2] , and on average, 88% of the cases had matching immunization records (BCG – 91.20%,Penta-1 – 91.66%, Penta-2 – 90.36%, Penta-3 – 88.81% and Measles-1 – 78.51%) [3].

Given that these rates far exceed the 80% threshold we set in the original plan, we are comfortable with the validity of the tablet data, and decided use the administrative data as our primary data source: the sample is much larger and the data is actually much more precise than what households can tell us in survey.

[1] see section 4.2.III of the original plan for a detailed description.

[2] child name – 87.88% (excluding the 5.57% child not named at time of vaccination), child's gender – 98.38%, father's name – 96.72%, mother's name – 97.20% and child's birthday – 96.31%.

[3] One possible explanation for the lower measles match rate is that it is due much later in the immunization schedule, and caregivers often lose the vaccination cards.

We thus decided to redesign the household endline survey entirely, to collect the data that is the most useful to complement and interpret the administrative data as opposed to serving as a substitute for it. Instead of collecting the same information in the tablet data in a random subset of children, we redesigned this exercise to address three main questions that cannot be addressed using just the tablet data and verification audit:

1. One obvious limitation of the tablet data, is that we only observe children who visit an immunization camp held by the nurses; so we don't observe children who receive vaccines elsewhere (ex. hospitals), or children who have never received a vaccine. The first goal of this exercise is therefore to establish a denominator for the fraction of all children we actually observe through the tablet data.

2. Related, a second concern with a simple analysis of the tablet data is that any treatment effect might reflect substitution to camps as a source of immunization from other sources as opposed to a true increase in immunization rates; the second goal of this exercise is thus to assess the extent to which the children not in our tablet data are immunized. And the extent to which they differ from children in our data on observables.

3. The tablet data is relatively limited in terms of it's ability to shed light on the mechanisms driving the effectiveness of the intervention; a third goal of this exercise is therefore to collect information on attitudes and knowledge towards immunization.

Towards these goals, we replaced the planned endline exercise with two data-collection exercises: a (second) census and a complementary endline instrument which will focus on collecting the additional information needed to interpret the administrative data, in particular:

- First, we conducted a *door-to-door census exercise* in 200 randomly selected village-units[4] The 200 village-units were randomly sampled in two stages: in the first stage, we sampled one village-unit per SC, resulting in 755 village-units. In the second stage, we randomly selected 200 of those village-units stratifying by district and incentive sub-treatment arm.

  We identified 18,963 eligible children identified during this census. For these children, we also collected information on whether a given eligible child has an MCP card (immunization card), in addition to key identifying information: child's name, date of birth, gender, parents' names, barcode, and the child's aadhar number.

- Using this information, we proceeded by attempting to *match* this set of children with the tablet data. In theory, if a child has visited a session camp they will have an immunization card with a barcode, and thus this information would be sufficient to distinguish between children in our data ("matched children") and children not in our data ("unmatched children"). This simple dichotomous categorization is somewhat complicated by a practicality: some of the community-workers store immunization cards on behalf of parents, so if a parent is unable to present an immunization card it does not necessarily mean that they have never taken their child to an immunization camp.

  Using a simple gated algorithm, we classified each child into one of three categories:

[4]Due to the large variation in the size of village populations, we divided villages into roughly equally sized "village-units" of 600 households, to ensure uniform sampling.

1. Matched (47.2%): This group contains the set of children that are successfully matched on barcode or the child's aadhar number, regardless of whether or not they have an MCP card.
2. Unmatched (31.85%): Those are children that either (1) Don't have an MCP, and we were unable to match using their aadhar number, or (2) have an MCP, but no barcode and we were unable to match using their aadhar number.
3. Uncertain (21.13%): Children that have an MCP card, but whose primary caregivers are not able to present it at the time of the survey.

- Next, we randomly selected approximately 3000 children from the full population of eligible children identified through the census to complete the *revised endline survey*.

  Depending on the child's match status, we collect their full immunization history in addition to administering the instrument. In addition, to complement the tablet data, the instrument varies by the child's match status: to avoid collecting redundant information, we only collect the full immunization history for unmatched children.

  Note that for "uncertain" children, since we also collect all the identifiers, we are able to determine their match status ex-post[5].

  Given that we had already collected information on the subset of matched children through our rolling tablet data audit, we oversampled the unmatched and uncertain categories such that the final sample is composed of 40% of unmatched children, 30% of uncertain children and 30%. In addition, due to large variation in the match rate across districts, the sampling was stratified by district.

  Aside from the basic identifiers and immunization history for unmatched children, we also included a module on caregiver's knowledge and attitude towards immunization, including their sources of immunization-related information. In villages that are part of the network experiment, we are able to match sources of information against our list of seeds.

1.2 Reminder Implementation Glitch

The second significant modification to the original implementation plan stems from a technical glitch, due to which the implementation of the targeted reminders intervention was discontinued starting November 2017.

In the remainder of this document we present the revised empirical strategy in light of these changes.

2 Revised Empirical Strategy

In this section, we present the revised empirical strategy in light of these changes.

2.1 Tablet Data

We plan to use the administrative tablet data as our main data source for estimating the effect of the various components of the program, unless we observe a large substitution effect in the sample of unmatched children.

[5] We mitigate the said problem of unavailable immunization cards by requesting the community health workers to return to the cards to the caregivers.

Since the reminders were only implemented in a sub-set of the data, and the networks experiment is limited to the set of 980 villages "at risk for seeds"; we will estimate treatments effects for each of the experiments separately.

Outcomes

- For aggregated data (at the village-month or SC-month level depending on the specification), we estimate treatment effects on the following outcomes:
  - The number of *fully immunized* children: the vaccine package administered in this study is the WHO/UNICEF Extended Package of Immunization (EPI), which is the package provided by the Indian government. For children, the EPI includes one dose of BCG vaccine, three doses of Pentavelent vaccine , three doses of oral polio vaccine (OPV), and one dose of measles vaccine. Since the OPV doses are almost always given together with the Penta doses, we use pentavelent shots as a proxy for both pentavelent and OPV. For this measure, we consider both vaccines administered at a session camp during the course of the program, and vaccines previously administered or administered elsewhere, and recorded by the nurses as part of the child's "vaccination history".
  - The number of children who visited a session to receive each of the following vaccines: Pentavalent-1, Penta2, Penta3, Measles1 (the key EPI vaccines- omitting BCG since it is usually administered at birth).
  - (Note: since most children receive BCG at birth, it is often not recorded by the nurses. For this reason, we will often use measles-1 as a proxy for fully immunized, since it is the final shot in the sequence.)

  We intend to estimate treatment effects for all these outcomes in both logs and levels.
- For individual-level data, we construct the following outcomes:
  - An indicator for the child having received the vaccine during the program, for each of the key EPI vaccines identified above (excluding BCG) .
  - An indicator for whether the child is fully immunized.

Village-Level Controls

We intended to use the baseline survey, administered in a subset of the villages, as a source to construct village-level controls. However, since we are using the tablet data as our main source instead of the endline data, we have no reason to restrict our analysis to the baseline sample.
Instead, were successful in matching 70% of the villages in our data with the 2011 Indian National Census, we will use the census data instead, since it is available for a much larger fraction of the villages in our data. We intend to select the controls using the Belloni-Cherznukhov-Hansen Double Post Lasso procedure (Belloni et al. ,2014).

Missing Data

- For log outcomes, we replace 0s with 1 prior to taking logs conditional on a session being held. (Although there are few at this level off aggregation.)

- For control variables that do not take on the value 0, we replace missing values with zeros and include an indicator for when the variable is missing at baseline.

2.1.1 Incentives Experiment

For the incentives experiment, we will estimate the following specification using the tablet data aggregated at the SC-Month level, since the incentives experiment was randomized at the SC-Level. In particular, we will estimate the following specification:

$$Y_{st} = \alpha + \sum_j \beta_j T^j{}_s + \mu_d + \mu_t + X_s + \epsilon_{s,t} \tag{1}$$

where $T^j{}_s$ denotes the type incentive scheme SC s is assigned to: $j \in \{$ high slope, low slope , high flat, low flat$\}$. μ_d denotes a set of district fixed effects, and μ_t denotes a set of month fixed effects. Y_{st} represents a given outcome for SC s in month t, for the set of outcomes specified above. Standard errors are clustered at the SC-level.

Although the type of incentive was randomized at the SC level, the provision/omission of incentives was randomized at the PHC level; clustering at the SC-level therefore potentially underestimates the standard errors. Since there is no straightforward analytical solution to this problem, we address this concern by conducting a randomization inference test for the same specification, which will respect the experimental design.

While we will use this as our baseline specification, will also estimate versions of this specification:

- Pooling the different incentive groups: incentive vs. no incentive, slope vs. flat, and high vs. low incentive.
- Including SC-level baseline controls (aggregated from the village-level controls described above).
- Including a full set of controls for the reminders experiment.
- We will also estimate the same specification with data aggregated at the SC-Level (summed over months.

2.1.2 Communications Experiment

For the communications experiment, we focus on the subset of 980 villages within which we randomly assigned villages to a communication sub-treatment, the set 'at risk for seeds'. Given the level of randomization, we aggregate the data at the village-month level, and estimate the following specification:

$$Y_{vt} = \alpha + \sum_k \delta_k \text{Seed}^k{}_v + \sum_j \beta_j T^j{}_s + \mu_p + \mu_d + \mu_t + \epsilon_{v,t} \tag{2}$$

where $\text{Seed}^k{}_v$ is an indicator equal to 1 if village v has been assigned to seed type k, $k \in \{$gossip, trusted, trusted gossip, random$\}$. The fixed effects are defined as before. And Y_{vt} represents a given outcome for village v in month t, for the set of outcomes specified above. We estimate cluster-robust standard errors, clustered at the village level. To maximize precision and reduce noise as much as possible, we include a full set of controls for the incentives experiment and a full set of PHC indicators (Note that we omit one of the treatment dummies, since they are collinear given the PHC-level FEs).

In addition to these baseline specifications, we intend to estimate other variations :

- Without the incentive and PHC controls.
- Including a full set of controls for the incentives and reminders experiments.
- Pooling together the different types of nominated seeds: gossip, trusted, trusted-gossip.
- Including village-level baseline controls.
- We also run this specification on the subset of villages who are assigned to some type of seed. In this case, we will omit the indicator for random seed due to collinearity, in order to compare the effects of different types of seeds relative to randomly selected seeds.
- We will also estimate a version of this specification with data aggregated at the village-level (the unit of randomization), and excluding time fixed effects.
- We intend to estimate this specification separately for incentive control and incentive treatment villages, and perhaps separately by incentive type or pooling the different incentive arms, since the nature of the communication through seeds may change depending on the incentives.

2.1.3 Targeted Reminders Experiment

For the targeted reminders experiment, we will restrict our sample for the analysis of the reminders to the implementation period for this intervention, prior to the glitch.

SC-Level Effect of Reminders

Since the fraction of individuals receiving a reminder was randomized at the SC-level, we analyse the data at the SC-month level, using the following specification:

$$Y_{st} = \alpha + \phi_1 Q_s^{33} + \phi_2 Q_s^{66} + \mu_d + \mu_t + \epsilon_{st} \tag{3}$$

where $Y_{st}, \mu_d, \mu_t,$ and $\epsilon_{s,t}$ are defined as above, and ${Q^{\pi}}_{s[v]}$, is an SC level binary variable indicating that SC s belongs to the assigned to the group where $\pi\%$ of individuals receive a reminder.

Aside from this baseline specification, we also intend to estimate the following:

- Including village-level controls.
- Including a full set of PHC fixed effects.
- We will also run this on data aggregated at the SC level (the level of randomization), and obviously excluding time fixed effects.

Individual Effect of Reminders

In addition to estimating the SC-level effect of assigning a fraction of individuals to receive reminders, we estimate the individual effect of reminders, which is well identified since we randomly assigned children to receive reminders conditional on having received the previous vaccine, based on the fraction $\pi\%$. To this

end, we use the following as our base specification on individual-level data, restricting to SCs assigned to a non-zero-fraction reminder treatment group:

$$Y_{ivt} = \alpha + \theta R_i + \mu_v + \mu_t + \epsilon_i \tag{4}$$

where we estimate the intent-to-treat effect by regressing outcome Y_{ist} for child i in SC s and month t, for the set of child-level outcomes defined above; on an indicator R_i for whether or not child i has been assigned to receive reminders. We include a full set of village (μ_v) and time (μ_t) fixed effects. Since we restrict to SCs where the probability of receiving a reminder is non-zero, we report heteroskedasticity robust standard errors .

Note that we only randomize children into a treatment group once they visit for the first time, so the eligible for this experiment are the set of children who visit a session-site at least once.

Further to estimating this base specification, we also estimate variations of it:

- including individual-level controls.
- including baseline village-level controls.
- including a full set of controls for the incentives and communications experiments.
- We will also separately estimate this specification for villages with and without incentives, for villages with the various seed treatments, and villages with high or low level of reminder treatments.
- Because the reminders were conditional on whether the child received the previous vaccine, for the individual vaccines outcomes we also run a 2SLS specification where we estimate the effect of whether they we were actually scheduled to receive a reminder for the vaccine in question, conditional on their assigned treatment.
- We will also estimate the treatment effect on the treated for the vaccine specific outcomes using an instrumental variables strategy, where the endogenous regressor is an indicator for whether or not the child's caregiver received the reminder (which depends both on their group and endogenous past action).

Spillover Effects of Reminders

Finally, to estimate the spillover effect of reminders, we estimate a version of specification (4) on the entire individual-level data for the implementation period of the reminders experiment. In particular, we estimate the following specification:

$$Y_{ist} = \alpha + \gamma_1 R_i * Q_s^{33} + \gamma_2 R_i * Q_s^{66} + \phi_1 Q_s^{33} + \phi_2 Q_s^{66} + \mu_d + \mu_t + \epsilon_{st} \tag{5}$$

where all the variables are defined as before, and we estimate this for the full set of child-level outcomes. Standard errors are clustered at the SC level.

In addition to this, we also estimate versions of this:

- including a full set of PHC fixed effects.
- including individual-level controls.
- Pooling together the SC-level reminder groups.

2.2 Endline Data

In this sub-section, we outline the analysis plan for the endline data, given the analysis plan outlined from the tablet data.

1. **Substitution between camps and other immunization places:** To address the question of substitution between camps as a source of immunization from other sources, we will estimate the effect of the intervention on immunization outcomes in the set of unmatched children. If the treatment effect observed in the tablet data reflects a substitution as opposed to a true treatment effect, we expect to observe lower immunization rates for unmatched children in treated villages. To this end, we estimate the following specification:

$$Y_{iv} = \alpha + \sum_j \beta_j T^j{}_{[v]} + \mu_d + \epsilon_{iv} \tag{6}$$

   where $T^j{}_{[v]}$ denote the treatments received by the village v among the incentive treatments, the communication treatments and the reminder treatments. We report standard errors clustered at the village level to account for possible correlation between children in the same village. Y_{iv} denotes the immunization outcomes of child i in the village v, specifically a set of indicators for each of the key vaccine, and an indicator for successful completion of the entire schedule. We will test each of the β_j for equality against 0, and fail to reject the substitution hypothesis for a given treatment arm if β_j is significantly different from 0. If we reject the substitution hypothesis, then given the quality of the data as established by the child verification survey, we can rely exclusively on the tablet data to measure the treatment effects on the immunization rate.

2. **Proportion of eligible children in our system:** To compute the proportion of children (in/out) of our tablet data for the 200 villages in the census, we plan to: first, use the additional data and identifiers collected in the revised endline to conclude whether a child is in or out of the system ("obtain a true match status"). Second, to extend this measure from the endline population to the entire population of eligible children in the census data, we estimate the true match rate within each of the initial census match-categories (matched unmatched, uncertain) and use that to compute the proportion of eligible children identified in the census data that are in our system, separately by *treatment cell* and by *district X treatment cell*. Thus, for a treatment cell, j, the proportion of children in the tablet data, $\hat{\mu}_j^{In}$ is given by:

$$\hat{\mu}_j^{In} = \sum_{k \in \{\text{matched, unmatched, uncert.}\}} p_{k,j} * \frac{N_{k,j,In}^{Endline}}{N_{k,j}^{Endline}} \tag{7}$$

   where k denotes the initial census match status. $p_{k,j}$ is the proportion of children in treatment group j with match status k in the census sample, $N_{k,j,In}^{Endline}$ is the number of tablet data children in the endline with census status k and in treatment cell j, $N_{k,j}^{Endline}$ is the number of children in the endline with census status k and in treatment group j

3. **Average immunization rates:** For the set of villages sampled for the census, we will also compute the average immunization rate across all children. Using the observed initial match-distribution in the census data, we will compute the immunization rate for the entire sample of eligible children identified

in the census, separately by *treatment cell* and by *district X treatment cell*. Thus, for a treatment cell, j, the average immunization rate for the vaccine i, $\hat{\mu}_j^i$ is given by:

$$\hat{\mu}_j^i = \hat{\mu}_j^{In} * \hat{\mu}_j^{i,In} + (1 - \hat{\mu}_j^{In}) * \frac{N_{Out,j,i}^{Endline}}{N_{Out,j}^{Endline}} \tag{8}$$

where $\hat{\mu}_j^{In}$ is the proportion of children in the tablet data for treatment cell j in the Census sample, $\hat{\mu}_j^{i,In}$ is the immunization rate for children in treatment cell j and in the tablet data and given directly by the tablet data, $N_{Out,j,i}^{Endline}$ is the number of endline children not in the tablet data and in treatment cell j who received vaccine i, $N_{Out,j}^{Endline}$ is the number of children not in the tablet data and in treatment cell j. We will get the standard errors by simulation.

Using the resulting estimates in this set of 200 villages, we can estimate treatment effects using the same specifications as before.

4. **Immunization Rates, by Village:** To obtain estimates of the immunization rate for all the villages in our sample, we proceed as follows:

   - First, we use information from the 2011 National Census and the tablet data to train an algorithm predicting the *immunization rate* computed in step 3, within the census sample.
   - We then use the resulting algorithm to extend this measure to the full set of villages in the tablet data, using the available tablet data and National Census data to predict immunization rates in the full sample.

   Note that this measure may turn out to be very noisy, which is why we don't intend to use as our headline number unless we observe large substitution rates.

5. **Comparisons of characteristics between immunized and not immunized:** To compare immunized and not immunized children, we will estimate the following specification:

$$Y_i = \alpha + \sum_j \beta_j M_i * T^j{}_{[v]} + \mu_d + \epsilon_i \tag{9}$$

   where M_i is an indicator equal to 1 if the child is immunized, and 0 otherwise, where $T^j{}_{[v]}$ denote the treatments received by the village v, and Y_i denotes child level outcomes. We also include a set of district fixed effects μ_d, since the sampling was done separately for each district. We will estimate this specification to compare knowledge and attitude toward immunization and demographic information, between children in our data and children not in our data. We report standard errors clustered the village level.

6. **Treatment effects on new outcomes:** we will use the endline data to estimate treatment effects on outcomes not observed in the tablet data such as primarily knowledge and attitudes towards immunization. The outcome variables will be:

   - **Reasons of non-immunization and inconveniences of immunization**: the reasons why your child have missed some vaccinations, the reasons why your child has never been given any vaccinations, factors which make it inconvenient to get your child vaccinated, reason why time an

inconvenience, reason why money an inconvenience, reason child's discomfort an inconvenience, inconvenient to get your child immunized (on a 1 to 5 five scale)

- **Knowledge about immunization camps**: aware of immunization camps, date of the last immunization camps, means through they receive informations about the immunization camps, someone has ever informed her about immunization camps, who has ever informed her about immunization camps, names of the persons who informed you about immunization, ability of these persons to diffuse information, trust toward these persons, ever informed people about immunization camps
- **Knowledge about immunization**: number of vaccines a child should get in the first year, places where she gets informations on benefits and harms of vaccines
- **Attitudes toward immunization**: benefits of vaccines, long-term harmful effects of vaccines, total effects of vaccines on child's health (on 1 to 5 scale, from very harmful to beneficial), neighbours share your view about vaccines, friends share your view about vaccines, other villagers share your view about vaccines, people you know get their children vaccinated

To this end, we will estimate the following weighted regression:

$$Y_{ivt} = \alpha + \sum_j \beta_j {T^j}_{s[v]} + \nu \text{ SeedRisk }_v + \sum_k \delta_k \text{seed}_v^k + \phi_1 Q_{s[v]}^{33} + \phi_2 Q_{s[v]}^{66} + \gamma_1 X_v + \gamma_2 X_{iv} + \mu_d + \epsilon_{iv} \quad (10)$$

where ${T^j}_{s[v]}$ denote the type incentive scheme SC $s[v]$ of village v is assigned to $j \in \{$ high slope, low slope , high flat, low flat$\}$. SeedRisk_v is an indicator equal to 1 if the village was eligible to be assigned to a seeds group in (part of the sample for the communications experiment), ${\text{seed}^k}_v$ for $k = 1, \ldots, 4$ indicate the type of seed in the communication experiment (gossip, random, trusted, trusted gossip). For the reminders experiment, we denote ${Q^\pi}_{s[v]}$, is an SC level dummy variable indicating an assignment rate of $\pi\%$, for village v belonging to SC s. Since we oversampled the fraction of unmatched children, we re-weight such that the sample is representative of the overall population.

Standard errors are clustered at the village level; since we only sampled one village per SC, this is effectively equivalent to clustering at the SC level, which is the relevant unit of clustering for the incentives experiment.

2.3 Additional Analysis

In addition to the main hypotheses related to program effects and cross-program effects, we intend to perform the following additional analyses most of which is as described in the original plan:

2.3.1 Mechanisms

We intend to use the endline data , child verification data, as well as the baseline data to carry out complementary analyses (not fully specified here), to shed light on potential mechanisms through which the various treatments affect immunization. For example, we can look at the correlation between attitudes towards

immunization and immunization status. Further, we can look how people heard about incentives , whether they knew about them, and the sources of information related to immunization. Finally, we can also correlate outcomes with the implementation quality measures described in detail in the next section.

2.3.2 Heterogeneous Effects

We will test for the differential impact of the program of various factors, by including main effects and interaction terms with treatment variables in all specifications above (and possibly, intermediate versions pooling sub-treatments, depending on effect sizes for power considerations, since the goal is to understand what drives heterogeneity in effect sizes). We may estimate the heterogenous effects of incentives (and if power allows, of sub-treatments) on all outcomes for the following:

- *Gender*: we will test for differential program effects by gender.
- *Village-level baseline/national census variables*: including assets, beliefs, knowledge, and attitudes towards immunization.
- *Exposure to the program*: we will estimate program effects separately for the sub-sample of the population that was born one month before the program started, since this is the sub-sample of individuals who were affected by the program their entire lives. We will also estimate exposure effects by interacting with a variable of months of exposure to the program (as defined from the child's date of birth, and the date of program roll-out). (Note: this varies by treatment (ex. the relevant subsample in the incentives experiment is children who were born at the start of the program, whereas in the targeted reminders it is children who were at at least 6 weeks old.[6] In addition, this may also be further restricted depending on the outcome (for example, the for the fraction of children having received > 5 vaccines, the relevant age group is children who were at most 11 months old at the start of the program.)
- *Implementation Quality*: We will look at heterogeneity by implementation quality. We will construct an index for implementation quality using principle component analysis (PCA) or other variable selection methods to select amongst the numerous proxies for quality measures obtained from the the three sources of monitoring data detailed above:
  - From the session site monitoring these measures proxy for two things: (1) The nurse's compliance with stated protocol that affect the provision of the program, using the different measures collected in our instrument, for example: the fraction of monitored ANMs adhering to protocol for outcomes collected in the survey, for example: whether they are using tablets, asking beneficiaries for phone numbers, entering data correctly, had the program "banner" up, etc.). (2) Supply side issues such as a lack of antigens in stock, a lack of barcodes (used to identify children), the presence of the government posters used to advertise the program etc.
  - We will make use of the following measures from the hotline data in the "issue log": the percentage of nurses reporting issues, disaggregated by type of issue, as well as the percentage of issues which were unresolved[7].

[6] The minimum age required to receive the second vaccine (they have had to have visited a session camp at least once to receive a reminder.)

[7] We define it this way, so that higher indicates worse for all measures

– We will use our transactional data to construct PHC level measures of the quality of the program's implementation in terms of delivering the recharges, SMS' and voice calls. (ex. delivery status, call duration etc.)

In addition, we will also make use of the tablet meta-data which can be used to proxy for ANM maleficence (ex. enters her own phone number instead), poor performance (ex. duplicates forms because she mistypes), and treatment delivery delays[8].

We will aggregate the measures in each of these sources at the PHC (or PHCxMonth) level before implementing our selection algorithm.

We will look at whether the program was more effective in places with high/poor quality implementation both by partitioning the sample of PHCs into different implementation quality percentile groups.

2.3.3 Robustness Checks

- *Program Effects on Older Children*: we will estimate these outcomes on older children, this will allow us to test whether there are any effects on children in treatment areas who were not affected by the program, and children in control areas who would not have been affected by the program.
- *Baseline Balance Checks*: We will use the baseline data, and the national census data to confirm that the various treatment groups don't differ on ex-ante observables.

Appendix A Original Pre-Analysis Plan

[8]Delays in the provision of treatments could either be due to network issues/late uploading by the ANMs, or because of a problem on the transaction processing side, the vendor data should account for the former, whereas upload delay time should account for the latter.

Improving Full Immunization Rates in Haryana: Evaluating Incentives and Communication Methods

Pre-Analysis Plan

Abhijit Banerjee, Arun Chandrasekhar, Esther Duflo, John Floretta, Matthew Jackson, Harini Kannan, Francine Loza, Anna Schrimpf, Maheshwor Shrestha

August 2017

1 Introduction

This document outlines a pre-analysis plan for evaluating a set of randomized interventions aimed at stimulating demand for vaccines in seven districts in the state of Haryana, India. These interventions are a part of a program that is being implemented at scale, in partnership with the state government and has high potential for expansion if proven successful. Since the authors completed this general plan before the completion of data collection and before the data has been analyzed, the plan can be used for reference in analyzing the final results of the study. Future updates on this plan will be time-stamped, and this version will be conserved.

Section 2 reviews the motivation for the study, and provides an overview of the project. Section 3 describes the setting of the project and the intervention. Section 4 details the sample and data collection exercises, and Section 5 presents the experimental design. Section 6 reviews the empirical strategy, and Section 7 outlines the main hypotheses to be tested in this study.

2 Motivation & Project Overview

2.1 Background

Immunization is one of the most cost-effective ways to improve child health and survival in developing countries. Yet, yearly, it is estimated that approximately 23 million infants worldwide are still not covered by routine immunization services (WHO, 2010). In the Indian state of Haryana alone, there are at least 224,000 children aged between 0 to 12 months not on track to be fully immunized (DLHS-IV, Haryana).

In middle-income countries, conditional cash transfers (CCTs) have successfully stimulated demand for immunization, by conditioning the transfers on receiving vaccinations. In contrast, standard solutions for improving immunization coverage in resource-poor settings address supply-side issues in vaccine delivery, or leverage intensive door-to-door campaigns. Our baseline data, collected in 2016, indicates that coverage rates in 7 districts in Haryana are much higher for vaccines early in the immunization schedule. 94% of the children received at at least two vaccines, whereas 86%, 65%, and 40% of children receive the 3rd, 4th, and

5th vaccines respectively. Higher rates for the first vaccines suggest that the main barrier is neither driven by a supply side problem, nor a deep-seated resistance to immunization, which may happen, for example, based on religious or personal beliefs.

2.2 Overview

In this program, we address the research and policy gap on the efficacy and optimal design of demand-side solutions with a set of three interventions: provision of small incentives to caregivers, a community intervention that leverages communication through social networks, and an information campaign through phone calls and text message reminders. To our knowledge, this is the first large scale evaluation of such program. The program is being implemented in collaboration with the state government, across 7 districts, comprising 2359 villages,in the state of Haryana, India.

To address causes of demand side issues and possible policy responses to them, we explore three interventions: The first experiment rests on the idea that small *incentives* can be effective in offsetting small costs, "nudging" a mother, who may be largely indifferent to immunizing her child, to overcome procrastination or compensating for the decreased salience. This project builds on previous research of the authors that demonstrates the cost-effectiveness of introducing small incentives for immunization, albeit in a small-scale setting (Banerjee et al. 2010). In this program, caregivers of children who visit an immunization session camp to get one of the five key vaccines, receive calling credit on their mobile phones. This happens when the vaccine is administered and the immunization details of the child entered in the m-health application used by frontline health workers.

The second and third experiments, cross-cut with the incentives experiment, and each other, were designed to leverage various communication strategies. Both of these experiments address the issues of misunderstanding and decreased salience as outlined above.

The second experiment tests the role of village *social networks*. Prior work by the authors (Banerjee et al. 2013), found that community members with a higher degree centrality (who are more connected), based on a unique dataset collected which maps out social networks in 43 villages, are more effective at spreading information in their network. Further work (Banerjee et al. 2014), showed that when communities are asked to nominate individuals whom they think spread information "gossips", they tend to nominate individuals with a high degree centrality, and these individuals these individuals are also effective "seeds". Motivated, by this evidence, in our program community members were asked to nominate individuals from their village who they consider good at spreading information "gossips", trustworthy in the village, or a combination of the two. These nominated 'seeds' were asked (through monthly voice calls and text messages) to spread the information that they receive about immunization to other members in their community.

The third, *targeted reminders* experiment, leverages the digital data collection system we put in place to evaluate the impact of targeted text messages and voice calls providing reminders and information. In a subset of the sample, the SMS and voice call reminders were used to attempt to send information to correct misconceptions about immunization. For example, the message to remind parents to send their child to get the measles vaccine in some instances reminded them that this is a new disease for which they have not yet received any shot. Furthermore, we have randomly varied the fraction of individuals within a village receiving reminders, thereby enabling us to identify spillover effects. To our knowledge, this is the first large-scale test of the use of SMS reminders for immunization, and the first to attempt to distinguish reminders from

information and identify spillover effects.

3 Setting and Intervention

The program integrates into the existing health care structure in the state of Haryana. Primary Health Centres (PHCs) are health facilities that cater health services to an average of 25 rural and semi-urban villages with about 500 households each. A second relevant administrative distinction are Sub-Centers (SCs): SCs are the first point of contact between the community and the health system, and are staffed by the Auxiliary Nurse Midwives (ANMs), who are the frontline health workers for outreach services such as immunization, basic curative care services, and maternal and child health services and preventive services.The program was rolled out in 140[1] PHCs (755 SCs).

As is traditionally done in rural India, incentives were announced through posters and relayed by local frontline health workers, who know the incentive scheme that applies to their village. The provision of the incentives is monitored closely to ensure that the mobile credits are delivered successfully, and that the transfer system does not cause corruption.

Data on immunization is collected through an m-health platform installed on the tablets issued to the frontline health workers in all the villages in the sample. In addition to this, out of the 140 PHCs (2359 villages), approximately 7 villages per PHC have been randomly selected to be sampled for an in-depth household level survey for the baseline and endline. These surveys will most importantly collect data on children who did not visit the immunization camps during the program.

4 Sampling & Data Collection

Our study leverages three main types of data:

1. Tablet Data: This is administrative data, collected by the nurses using the Android m-health application [2]. Each time the child visits a camp, the nurse enters and updates their immunization information, so this database covers the entire population of children who visit a session camp in any of the 2359 villages in the program throughout the duration of it's implementation (December 2016-April 2018). Children are identified by a unique barcode and the identity of their guardian is verified using the unique ID of their parent (Aadhaar number). In addition to immunization information, this data includes demographic information, and the phone number of the guardian.

   Furthermore, this database also contains meta-data- recording higher level information about each record: such as the time the record was entered, the nurse by which it was entered, how long it took for the record to be uploaded to the server after being created, etc. These serve as a useful proxy for implementation quality, which will complement the rest of our monitoring data (described below).

2. Survey Data: Since we leverage SCs as a unit of randomization in this experiment, we conducted

[1] Primary Health Centres in the seven worst performing districts that had a cold chain facility were sampled for the study adding to a total sample of 14 PHCs.

[2] Please refer to the appendix 1 for more details on the m-health platform designed specifically for this program.

surveys in a subset of randomly sampled villages from each of the 755 SC. This set of 1048[3] villages, constitutes our evaluation sample. We conducted a total of eight data collection exercises for the purpose of this study, with different sampling methods for each survey- thus, for readability of this document, we reserve a detailed description of the sampling methods for appendix 2 and include the details of each instrument only insofar as they are relevant to the understanding of the analysis plan in other sections of this document.

For now, we describe the four most significant and largest survey activities we conducted:
I.Census: A census survey was conducted to identify eligible households[4] (households with one or more child between 12 and 36 months of age). A detailed mapping exercise of all the villages in the sample was conducted to estimate the number of dwellings with the village boundaries and help in tracking households during the census. 328,058 households were visited as a part of the census exercise over a period of three months, starting September 2015. Out of these, 62,548 households fit our eligibility criterion.

II.Baseline Survey:

From each of the 943 village units[5] (912 villages) in the 7 sample districts, 15 households were randomly selected from the census data for the baseline survey[6]. In total, 14,670 households with 17,000 children were surveyed during the baseline exercise.

III.Endline Survey: The endline and child verification surveys launched in August 2017, and will continue on a rolling basis until April 2018. Both surveys will be conducted in 3 rounds of data collection (roughly 6,9, and 12 months after the start of the program in the district), in 1048 villages. In each round, households will be sampled as follows: surveyors start from a randomly selected entry point to the village indicated on a map and follow a right-hand-rule. Surveyors will administer the survey to the first 4 eligible children they find – children between 0-18 months for Round 1, 0 – 20 months for Round 2 and 0-24 for Round 3. These age-ranges have been specified in order to maximize the chance to collect data for the set of children who could have one or more of the 5 "shots" while the program was in place– children who were 11 months or younger at the time the program was launched in their district.

IV.Child Verification Survey: In the same villages as the endline, we also conduct a child verification survey. The purpose of this survey is to audit the tablet data, and to check whether records might be disproportionately made-up ("ghost records") in the treatment group relative to the control group. This is a concern as the incentive might provide the nurses an incentive to enter fake records, in order to receive recharges on a mobile number affiliated with her. For this survey, we randomly selected up to 5 child records from the tablet data described above, and attempt to locate these children using the help of local health workers (nurses, and other lower level health and social workers), as well as other members in the village.

[3]these include the 912 villages from baseline, and an additional 136 villages (one additional village per SC), added to make sure that each SC is represented in the original sample (the original baseline sample was not stratified by SC.)

[4]A household comprises people who have lived in the same house for at least 30 days in the past year, eat food cooked from the same stove, and contribute to and share household income.

[5]Large villages were divided into village units for more representative sampling and logistical ease.

[6]In PHCs where less than 7 village units were sampled, a higher number of households from each villages was sampled, to maintain the proportional sampling of household across all PHCs.

Finally, the following data are collected, on an ongoing basis, in order to monitor both implementation quality, and the vaccine administration system (the supply-side factors).

3. Monitoring & Implementation Data:

   I.Session Site Monitoring: Monitors are sent to randomly selected vaccination camps, and asked to observe and record the actions of the nurses and the supply of vaccines, amongst other things. Monitors visit at least one session camp per PHC per month. The number of immunization camps visited each month varies depending on the presence of other ongoing field activities, due to budget constraints.

   II.ANM Issue Log: In order to facilitate the smooth running of the program, we created a hotline used by the nurses to report any problems they are having. The hotline directs calls to our field monitors and supervisors in the relevant geographic regions, who record details of the call into an "Issue Log". They record demographic information as well as the type of issue being reported by the nurse: hardware issues (ex. tablet power-on problem), network issues, usability issues, complaints, and software problems[7]. Monitors are also asked to indicate the actions they took to resolve a given problem, and whether or not it was resolved.

   III.Intervention Delivery: The implementation of all three interventions is predicated on the successful delivery of either recharges or SMS'/voice calls. These processes are triggered using software that keeps track of due processes, and interacts with vendors through APIs to execute these processes. For each transaction (recharge, SMS, or call) we record all details of that transaction, including the time of delivery and delivery status, as well as call duration for calls.

5 Experimental Design & Methodology

5.1 Treatment Assignment

Figure 1 outlines the design of the study.

Incentives Experiment

The presence or absence of incentives is randomized at the PHC level, both for administrative simplicity (nurses are supervised by a doctor at the PHC, vaccines and other vaccines are distributed at the PHC level, and the training is at PHCs) and to avoid spillovers between the treatment and the control groups. PHCs tend to be separated by several kilometers and travel is not easy between them. The 70 PHCs that have been randomly assigned to the treatment group receive incentives, and the remaining 70 form the control group. Within the PHCs in the treatment group, the level and slope of the incentives has been randomly assigned at the SC level, to allow for identification of the most effective level and slope of the incentive (albeit in a limited menu). While this means that nurses who meet might have slightly different incentives schedules to explain to the parents, this does not seem to have created confusion (we discuss spillovers below). There is no scope for manipulation of the incentive level at the individual level, or for errors, as this is centrally administered and automatized.

Incentives are provided for each visit to an immunization camp conducted by the ANM in the 70 incentive PHCs. In India, a child is considered to be fully immunized if she receives the following shots in the first 12 months after birth: one dose BCG, 4 doses of the oral polio vaccines, 3 doses of Pentavalent, 3 doses of

[7] We have not analyzed this data in aggregate yet, though the volume of issues is relatively small.

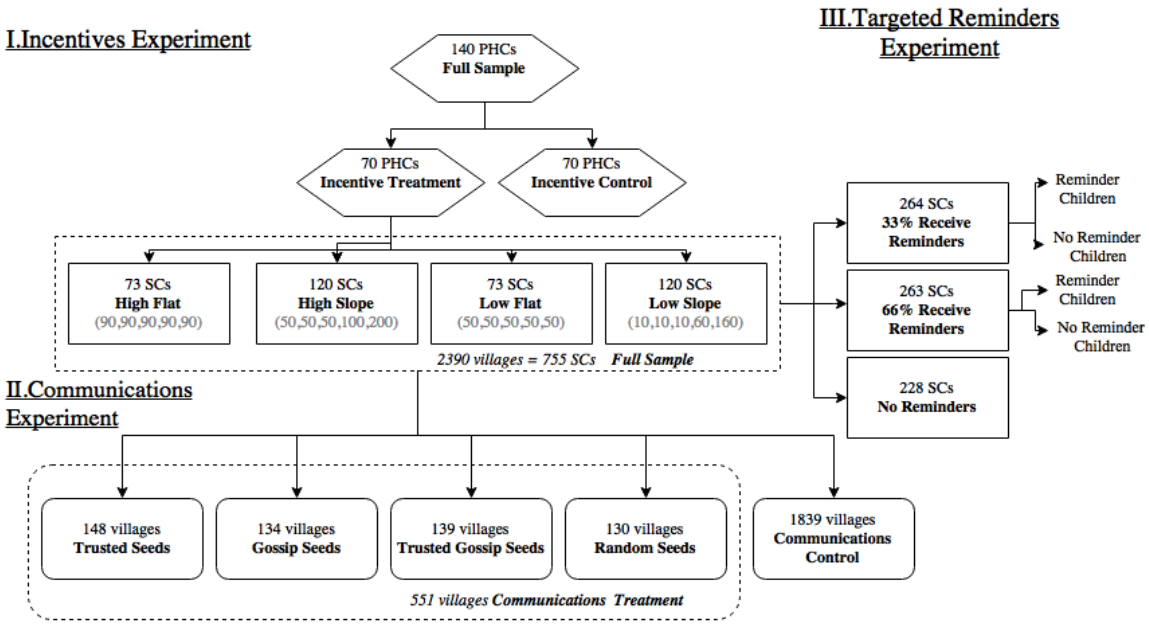

Figure 1: Study Design

rotavirus, and one dose of measles vaccine. This is done in 5 sessions: The BCG vaccine is typically given at birth, then the child receives Pentavalent 1, 2 and 3 at 6 weeks, 10 weeks and 14 weeks respectively. Finally, the first dose of measles vaccine is given at 9 months or older. From our perspective, each of these visits is considered as a "shot".

After each visit, the caregiver automatically receives a mobile credit on the phone number provided at the session camp, which can be used for phone calls or SMS messages. The mobile credit recharge is triggered through the application on the ANM's tablet at each immunization, and the parent also receives a notification of the amount transferred. Incentives amounts vary at the sub-center (SC) level. The variation is on cost (Rs. 250 vs Rs 450 for a fully immunized child) as well as on slope (flat schedule vs increasing schedule) as indicated below (as a reminder, 70 PHCs have on incentive at all):

A High incentives, flat payment: 90 rupees per immunization

B High incentives, increasing payment: 50 rupees for the first three immunization, 100 for the fourth, 200 for the fifth

C Low incentive, flat payment: 50 rupees per payment

D Low incentive, increasing payment. 10 rupees for the first three immunization, 60 for the fourth, 160 for the fifth

These incentive levels were chosen to be small (the high level is comparable to what was offered in the Udaipur experiment (Banerjee et al 2010), inflation adjusted), while the low level is smaller, and therefore politically implementable if feasible, but still a non-trivial amount for the households: 50 rupees corresponds to 100 minutes of talk time on average. Contrasts that will be considered include: A and B [and C and D] have the same total payment for the five immunization, but higher slope at the end. The impact of B [D] on full immunization should be higher if households are forward looking [since they will make sure to get the early shots to qualify for the later ones, and then face a larger incentives on the later ones], but could be lower otherwise [as households are less likely to get the first shots and thus qualify for the later incentives]. B and C start at the same level, but B has a larger increment for the later shots. We can contrast behavior in the later shots to look at the immediate effect of the incentives, and behavior in the early shots to understand investment behavior. Note that the provision of incentives is independent across vaccines in the treatment group: if a child misses a vaccine (ex. penta 1), and comes to receive the next vaccine (in this case, measles), they still receive an incentive for measles[8].

Communications Experiment

The communications experiment was cross-randomized with the incentive experiment, at the village level. We randomly selected 529 of our sample villages to receive one of four communications interventions, and randomized the sub-treatment arms to which they belong (described in detail below). In each of the villages assigned to one of the non-random seeds groups (gossip, trusted people, and trusted gossips), we selected seeds by asking 17 households randomly sampled from the census to nominate 5 individuals matching one of the following descriptions[9], depending on the village's treatment assignment:

- *Gossip seed **(134 villages)**:* Gossip seeds are people who are expected to be good at diffusing information in their village. We used the following script to illicit this:

[8]Note this is not the case in the TR experiment

[9]see appendix 2 for details on the nominations survey

"Who are the people in this village, who when they share information, many people in the village get to know about it. For example, if they share information about a music festival, street play, fair in this village, or movie shooting many people would learn about it. This is because they have a wide network of friends/contacts in the village and they can use that to actively spread information to many villagers. Could you name four such individuals, male or female, that live in the village (within OR outside your neighborhood in the village) who when they say something many people get to know?"

- *Trust seed* ***(148 villages):*** Trust seeds are people who are relied upon in the village and any information emanating from them will be received with much more confidence. Information on health services, immunization in this case, tends to more sensitive than usual gossip information. We used the following script to illicit this:

  "Who are the people in this village that you and many villagers trust, both within and outside this neighborhood, trust? When I say trust I mean that when they give advice on something, many people believe that it is correct and tend to follow it. This could be advice on anything like choosing the right fertilizer for your crops, or keeping your child healthy. Could you name four such individuals, male or female, who live in the village (within OR outside your neighborhood in the village) and are trusted?"

- *Trusted gossip seed* ***(139 villages)*** *:* This is an intersection of trust and gossip seeds i.e. people who are well trusted in the village and effective at spreading information. We used the following script to illicit this:

  "Who are the people in this village, both within and outside this neighborhood, who when they share information, many people in the village get to know about it. For example, if they share information about a music festival, street play, fair in this village, or movie shooting many people would learn about it. This is because they have a wide network of friends/contacts in the village and they can use that to actively spread information to many villagers. Among these people, who are the people that you and many villagers trust? When I say trust I mean that when they give advice on something, many people believe that it is correct and tend to follow it. This could be advice on anything like choosing the right fertilizer for your crops, or keeping your child healthy. Could you name four such individuals, male or female, that live in the village (within OR outside your neighborhood in the village) who when they say something many people get to know and are trusted by you and other villagers?"

- *Random seeds* ***(130 villages):*** In this treatment arm, randomly selected individuals from the census in the village are provided with the same information as in the other treatment arms.

Respondents were also asked to provide demographic characteristics and identifying information for each of their nominees. The nominations were matched and aggregated by village (see appendix 3 for details

on the record linking algorithm), the top 6 nominated individuals in each village were selected to be seeds. Throughout the course of our intervention, seeds received monthly reminders and calls asking them to disseminate information about immunization.

Targeted Reminder and information Experiments

The reminders experiment was randomized in two steps (stratified by treatment status) in order to identify spillovers. First, we randomized the fraction of the village population receiving targeted reminders into 3 groups (in 711 villages individuals receive no reminders, in 886 villages 33% of individuals receive reminders, and in the remaining 793 villages 66% of individuals receive reminders). Within each village, the targeted reminders treatment arms are randomly assigned at the individual level, according to the saturation level assigned to that village in the first step. The randomization was done on a rolling basis, where the treatment status for each child visiting a session camp for the first time since the launch of the program is assigned to receive/not to receive reminders as soon as their record is added to our database. Children only receive reminders conditional on visiting a session camp for the previous vaccine.

Furthermore, if a child in the reminders treatment group is due for measles, we also randomize the content of the message they receive as a reminder to obtain their measles vaccine, as mentioned above- to distinguish the effect of salience from pure information, some individuals received a message that reminded them that this is a new disease for which they have not yet received any shots.

Since the sample is very large, the study is strongly powered to detect even very small effects of the program, particularly when using the data collected through ANM tablets. Note that since the experiments are cross-randomized, there is a "Pure Control" group which will not receive incentives or reminders.

5.2 Missing Data

For missing variables that do not take on the value 0 we'll replace missing values with zeros and include an indicator for when the variable is missing at baseline, effectively pooling towards the mean in order to preserve the number of observations.

6 Empirical Strategy

6.1 General Notes

- Village level controls include baseline variables: the relevant control variables will be picked via a Double Post Lasso (Belloni et. al 2014). We will estimate specifications with and without controls.

6.2 Endline Data

The endline data gives us individual level data, for repeated cross sections of individual children in a panel of villages. It is not a panel of children.

Main Outcomes

To assess the effect of the program, we will follow Banerjee et al (2010) and focus our analysis on the following three outcomes:

- Number of vaccines received per child
- Full immunization status: the vaccine package administered in this study is the WHO/UNICEF Extended Package of Immunization (EPI), which is the package provided by the Indian government. For children, the EPI includes one dose of BCG vaccine, three doses of Pentavelent vaccine, three doses of oral polio vaccine (OPV), and one dose of measles vaccine.
- The immunization status for each vaccine individually: BCG, Penta 1, 2 , 3 and Measles. *Note that we do not expect an impact for BCG, since it tends to be administered at birth, not in session sites.*

Secondary Outcomes

Changes in beliefs, attitudes, knowledge, prevalence of MCP cards, adherence to vaccine schedule (as defined by percentage of prescribed vaccines received on time), adherence to vaccine schedule for vaccine later in the schedule.

6.2.1 Overall Incentive Effect

The basic intent-to-treat (ITT) analysis for the incentives experiment is straightforward, given the empirical design. We will estimate a model of the form:

$$Y_{ip} = \alpha + \beta T_p + \gamma_1 X_v + \gamma_2 X_{ip} + \mu_t + \mu_d + \epsilon_{ip} \tag{1}$$

where Y_{ipt} denotes a given outcome for child i in PHC p at time t. T_p denotes the treatment status of PHC p an X_v represents a set of village level baseline covariates, described above (we will report estimates including and excluding these). We include a vector of average baseline village level covariates $X_{v[i]}$, where $v[i]$ represents village $v[i]$, to which child i belongs. And X_{ip} is a vector of child level covariates. μ_t is a set vector of time fixed effects included in specifications using endline data to indicate the round in which the observation was collected, and μ_d is a set of district level fixed effects to account for stratification.

In this specification, we cluster standard errors at the PHC level.

6.2.2 Estimating the effect of the treatment randomized at the village level.

For the other treatments, that were randomized at the SC (or village) level: type of incentive, communication experiment, and the fraction assigned to receive a reminder. To maximize power (to avoid noise created by the other experiments), we estimate all of these effects together, in one specification:

$$Y_{ivt} = \alpha + \sum_j \beta_j T^j{}_{s[v]} + \nu \text{ SeedRisk }_v + \sum_k \delta_k \text{seed}_v^k + \phi_1 Q_{s[v]}^{33} + \phi_2 Q_{s[v]}^{66} + \gamma_1 X_v + \gamma_2 X_{iv} + \mu_t + \mu_d + \epsilon_{is[v]} \tag{2}$$

where ${T^j}_{s[v]}$ denote the type incentive scheme SC $s[v]$ of village v is assigned to $j \in$ high slope, low slope, high flat, low flat. SeedRisk$_v$ is an indicator equal to 1 if the village was eligible to be assigned to a seeds group in (part of the sample for the communications experiment), ${seed^k}_v$ for $k = 1, \ldots, 4$ indicate the type of seed in the communication experiment (gossip, random, trusted, trusted gossip). For the reminders experiment, we denote ${Q^\pi}_{s[v]}$, is an SC level dummy variable indicating an assignment rate of $\pi\%$, for village v belonging to SC s.

Standard errors are clustered at the SC level since this is the relevant level of clustering for the incentives sub-treatments and targeted reminders fraction comparisons, since those were randomized at the SC level[10]. Note that since the first level of the incentive experiment was clustered at the PHC level, clustering by SC potentially underestimates the standard errors for the incentive experiment. There is no straightforward analytical solution to this problem. To address this issue, we will also conduct a randomization inference test for the same specification, which will respect the randomization design.

In addition, we will also estimate a version of equation (2) including PHC fixed effects and dropping the second term- in this specification we will not be able to identify the effect of incentives per se, although we will able to precisely identify the reduced form effects for the second and third experiments.

6.2.3 Estimating the effect of individual reminders

The reminders experiment was randomized at the individual level, conditional on receiving the previous vaccine. We thus turn to an individual level specification, including village level fixed effects.

The first specification will look at whether a child was eligible to receive reminder:

$$Y_{ivt} = \alpha + \theta_1 R_{iv} + \gamma X_{iv} + \mu_v + \mu_t + \mu_d + \epsilon_i \tag{3}$$

where we regress the outcome Y_{ivt} for child i in village v at time t, for all the same outcomes as before; on an indicator for whether or not child i has been assigned to the reminder treatment group R_{iv}. μ_v denotes a set of village fixed effects, and X_{iv} represents a set of individual level controls. We report heteroskedasticity robust standard errors.

For this specification, we limit our sample to the sample of children who were at risk for being randomized, those are children who attend at least one session camp.

We will then separately estimate these specifications for villages with and without incentives, for villages with various seed treatments, and villages with high or low level of reminder treatments.

Because the reminders were conditional on whether the child received the previous vaccine, in the vaccine by vaccine specification, we will also run a 2SLS specification where we will estimate the effect of whether they actually were assigned a reminder for a given vaccine, based on the group they were assigned to.

Third, we will estimate a version of specification 2 at the village level, and exclude children who were randomly assigned to get reminders, which will enable us to identify the pure spillover effect of reminders.

[10] The communications experiment was randomized at the village level, though we included roughly one village per SC in the experiment

In particular, we estimate the following model:

$$Y_{vt} = \alpha + \sum_j \beta_j T^j_{s[v]} + \nu \text{ SeedRisk } + \sum_k \delta_k \text{seed}^k_v + \phi_1 Q^{33}_{s[v]} + \phi_2 Q^{66}_{s[v]} + \mu_t + \mu_p + \epsilon_{s[v]} \quad (4)$$

where we cluster standard errors are the SC level.

6.3 Tablet Data

We will first verify that the tablet data is of good quality by using the Child Verification data to audit it. If we can identify at least 80% of the children in the tablet data in the field, and their information is accurate, we will proceed to use the tablet data as a valid source. If we can locate less than 50% we will treat the tablet data as tainted and won't use it for analysis. In between, we will use it but report the accuracy.

Using the tablet data, we will construct outcomes parallel to those collected in the endline using a combination of child vaccine administration records and immunization history recorded, as entered by the nurses.

We will estimate the same specifications as above, aggregated at the village level[11] We will report results for the following outcomes: log(number of children who receive shot k) for $k = 1, \ldots, 5$, as well as an additional summary measure which has no direct parallel to the individual level data: the number of children attending session camps in any given month.

Specifically, we mirror equation (1) above using the following framework:

$$Y_{vpt} = \alpha + \beta T_p + \mu_t + \mu_d + \epsilon_{vp} \quad (5)$$

where Y_{vp} is a given outcome for village v in PHC p, as explained above. In this case mu_t denotes a set of month fixed effects instead of survey round fixed effects. Standard errors are clustered at the PHC level, the remaining terms are defined as before.

We parallel specification 2 using the following model:

$$Y_{vt} = \alpha + \sum_j \beta_j T^j_{s[v]} + \nu \text{ SeedRisk } + \sum_k \delta_k \text{seed}^k_v + \phi_1 Q^{33}_{s[v]} + \phi_2 Q^{66}_{s[v]} + \mu_t + \mu_d + \epsilon_{s[v]} \quad (6)$$

As before, we cluster standard errors at the SC level.

The tablet data will also be used to estimate the effect of reminders. In this case, we can use the individual level tablet data, since the individuals at risk of being randomized as those who enter the tablet data. The specification will be exactly parallel to the specification above: we will estimate the intention to treat effect of being assigned to the reminder group by regressing the vaccination outcomes defined above on being in the reminder group, and then an IV where the endogenous regressor will be an indicator for whether the individual actually was sent the reminder [which depends on their group and an endogenous past action]. In this case, we will estimate a specification identical to specifications (3) and (4), with the exception that the time fixed effects are month fixed effects. And we report heteroskedasticity robust standard errors.

[11] Note that aggregation at the village level is necessary to get a population level estimate of the effect of the intervention, since children have to attend at least one camp to be immunized.

7 Hypotheses

7.1 Incentives Experiment

Hypothesis 1:The overall effect of incentives will be positive on all outcomes, with smaller effects on vaccines earlier in the schedule.
Methods: t-test for $\beta \neq 0$ and $\beta > 0$ in specifications 1 and 5 outlined above.
Hypothesis 2: Larger incentive amounts will have larger effects.
Methods: Estimate models 2 and 6 for all outcomes, and test β_js are jointly different from 0, are equal, $\beta_{\text{high slope}} + \beta_{\text{high flat}} = \beta_{\text{low slope}} + \beta_{\text{low flat}}$, and $\beta_{\text{high slope}} + \beta_{\text{high flat}} > \beta_{\text{low slope}} + \beta_{\text{low flat}}$.
Hypothesis 3: The flat payment schemes will be more effective than the slopes, assuming people have time inconsistent preferences.
Methods: Estimate models 2 and 6 for all outcomes and test $\beta_{\text{high flat}} + \beta_{\text{low flat}} = \beta_{\text{high slope}} + \beta_{\text{low slope}}$ and $\beta_{\text{high flat}} + \beta_{\text{low flat}} > \beta_{\text{high slope}} + \beta_{\text{low slope}}$.

7.2 Communications Experiment

Hypothesis 4: The dissemination of information through any community-nominated seed will have a positive effect, relative to randomly selected seeds and random seeds.
Methods: Estimate the specifications that are part of models 2 and 6 above, test each of the δs against the null ≤ 0. We will also test them for equality against each other, and in particular against δ_0 all using t-tests. We will also jointly test the coefficients.
Hypothesis 5: The identity of the seed matters (different seeds will have different effects.). Trusted gossips will be most effective in disseminating information.
We will test this by testing the deltas for equality against each other, all using t-tests. We will also jointly test the coefficients.

7.3 Targeted Reminders Experiment

Hypothesis 6: Receiving reminders positively affects immunization outcomes.
Methods: Estimate model 3, using both data sources and test $\theta_1 > 0$, and for equality against 0.

Hypothesis 7: The fraction of beneficiaries assigned to receive a reminder affects the outcome for both reminded and not-reminded beneficiaries. In the absence of externalities, this should not be the case.
Methods: Using model 4, $\phi_1 = \phi_2 = 0$.
Hypothesis 8: There are positive externalities. Immunization outcomes should be more favorable for the control group in SCs where others are assigned to receive reminders.
Methods: Using model 4, $\phi_1 + \phi_2 > 0$, $\phi_1 > 0$ $\phi_2 > 0$.
Hypothesis 9: the positive effect on immunization should be increasing in the fraction π assigned to receive a reminder.
Methods: By running model 3 separately for the group of SCs where a high fraction of individuals were reminded, and the group of SCs where a low fraction of individuals were reminded, and comparing the

coefficients against each other. In particular, we expect the coefficient to be larger for the high fraction .

Hypothesis 10: The impact of the reminder to get the measles vaccine varies by message content.
Methods: We will compare the rates of measles vaccine in both measles message-type treatments using framework (3) using both data sources. We will do this by restricting to the sample of children due for a measles vaccine at any point in our program.

7.4 Cross-Program Effects

Hypothesis 11: We expect the seeding of information to be more effective in incentive areas than in control areas.
Methods: We will do this by estimating a version of equations 2 and 5, and testing the coefficient on the set of interaction terms included in the model are positive, and for equality against 0.
Hypothesis 12 We expect the reminders to be more effective in incentive treatment areas than in control areas.
Methods: Estimating versions of equations 4 and 6 and testing the coefficient are positive and for equality against 0.
Hypothesis 13: We expect the spillover effect of reminders to be larger in incentive treatment areas than in control areas.
Methods: We will do this by estimating a version of equation (4), and testing the coefficient on the interaction is positive and for equality against 0.
Hypothesis 14: We expect the effects of reminders to differ by seed type.
Methods: We will test the coefficients on the interaction in versions of models 2 and 5.

7.5 Heterogenous Effects

We will test for the differential impact of the program of various factors, by including main effects and interaction terms with treatment variables in all specifications above (and possibly, intermediate specifications pooling subtreatments, depending on effect sizes for power considerations, since the goal is to understand what drives heterogeneity in effect sizes). We may estimate the heterogenous effects of incentives (and if power allows, of subtreatments) on all outcomes for the following:

- Exposure to the program: we will estimate program effects separately for the sub-sample of the population that was born one month before the program started, since this is the sub-sample of individuals who were affected by the program their entire lives. We will also estimate exposure effects by interacting with a variable of months of exposure to the program (as defined from the child's date of birth, and the date of program roll-out).*Note: this varies by treatment (ex. the relevant subsample in the incentives experiment is children who were born at the start of the program, whereas in the targeted reminders it is children who were at at least 6 weeks old.*[12] *In addition, this may also be further restricted depending on the outcome (for example, the for the*

[12]The minimum age required to receive the second vaccine (they have had to have visited a session camp at least once to receive a reminder.)

fraction of children having received > 5 *vaccines, the relevant age group is children who were at most 11 months old at the start of the program.*

- Implementation Quality: We will look at heterogeneity by implementation quality. We will construct an index for implementation quality using principle component analysis (PCA) or other variable selection methods to select amongst the numerous proxies for quality measures obtained from the the three sources of monitoring data detailed above:
  - From the session site monitoring these measures proxy for two things: (1) The nurse's compliance with stated protocol that affect the provision of the program, using the different measures collected in our instrument, for example: the fraction of monitored ANMs adhering to protocol for outcomes collected in the survey, for example: whether they are using tablets, asking beneficiaries for phone numbers, entering data correctly, had the program "banner" up, etc.). (2) Supply side issues such as a lack of antigens in stock, a lack of barcodes (used to identify children), the presence of the government posters used to advertise the program etc.
  - We will make use of the following measures from the hotline data in the "issue log": the percentage of nurses reporting issues, disaggregated by type of issue, as well as the percentage of issues which were unresolved[13].
  - We will use our transactional data to construct PHC level measures of the quality of the program's implementation in terms of delivering the recharges, SMS' and voice calls. (ex. delivery status, call duration etc.)

  In addition, we will also make use of the tablet meta-data which can be used to proxy for ANM maleficence (ex. enters her own phone number instead), poor performance (ex. duplicates forms because she mistypes), and treatment delivery delays[14].

  We will aggregate the measures in each of these sources at the PHC (or PHCxMonth) level before implementing our selection algorithm.

  We will look at whether the program was more effective in places with high/poor quality implementation both by partitioning the sample of PHCs into different implementation quality percentile groups.
- Gender: we will test for differential program effects by gender.
- Village-level baseline variables: including assets, beliefs, knowledge, and attitudes towards immunization.
- We will also look separately at the results by SMS vs. voice calls.

7.6 Robustness Checks

For validity and robustness checks, we will do the following:

[13] We define it this way, so that higher indicates worse for all measures

[14] Delays in the provision of treatments could either be due to network issues/late uploading by the ANMs, or because of a problem on the transaction processing side, the vendor data should account for the former, whereas upload delay time should account for the latter.

– Conduct baseline balance checks. (shown in appendix 4) item Comparend the rate of children located in the child verification exercises to assess the validity of our tablet data.

7.7 Additional Results

First, for the communications experiment we will also estimate impacts on several other measures of interest: for example, we will ask villagers at the endline survey whom they have heard the information on session camp/immunization from, and describe any difference in the type of seeds and villages.

In addition, we will estimate these outcomes on older children, this will allow us to test whether there are any effects on children in treatment areas who were not affected by the program, and children in control areas who would not have been affected by the program.

Finally, there is a possibility that we will continue to collect data using tablets, even after the interventions stop- in that case, we will also leverage the additional data using the analyses framework and tools described above.

References

[1] Abhijit Vinayak Banerjee et al. "Improving immunisation coverage in rural India: clustered randomised controlled evaluation of immunisation campaigns with and without incentives". In: *Bmj* 340 (2010), p. c2220.

[2] Abhijit Banerjee et al. *Gossip: Identifying central individuals in a social network*. Tech. rep. National Bureau of Economic Research, 2014.

[3] Abhijit Banerjee et al. "The diffusion of microfinance". In: *Science* 341.6144 (2013), p. 1236498.

[4] Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. "Inference on treatment effects after selection among high-dimensional controls". In: *The Review of Economic Studies* 81.2 (2014), pp. 608–650.

Appendices

Appendix 1: Overview of the m-health platform

In line with the increasing use of m-health platforms, the Tapplication was developed as an innovative solution to overcome issues in accuracy and availability of data that are often faced in public health systems. Using a tablet-based m-health application, instead of existing paper-based registers, results in: 1) less errors with access to real time data to implement the intervention and informed decision-making; 2) little or no duplication of data by linking unique identifiers (e.g. barcode IDs and government identity cards, 'Aadhaar' cards); 3) reduced workload of front line health workers, through generating automated reports.

There are 6 key components or steps of the TPK app, which front line health workers (ANMs) complete in order to enter a child into the programme:

- Log in to TPK software: ANMs enter unique ID to log in to the TPK app, and selects the relevant session area she is conducting the session camp in from a personalised drop-down list of options.
- Child registration: ANMs scan the QR code on the child's Aadhaar, and scan the unique barcode ID (identifier for the study) using the camera of the tablet; where scanning does not work, ANMs enter ID numbers manually. If the child is new, the ANM fills out a form with the child's name, parent's name, date of birth and gender. For children already in the system, who are returning, this form will already be pre-filled.
- Vaccination details: ANMs enter the child's vaccination history (for first visit), based on their MCP cards, as well as the vaccines given at the present session camp.
- Beneficiaries' phone number and key messages: ANMs enter the record guardian's phone number – where there is no number available, ANMs enter '0000000000'. A pop-up box prompts ANMs to deliver vaccine specific messages and reminding them to take their Aadhaar card and MCP card on the next visit
- Look-up: ANMs can look up records based on Aadhaar or bar codes (this will happen in the case of a returning child)
- Reports: ANMs can generate and view two types of reports from the app: a) child-wise daily report in the format of their register; b) month-wise antigen specific report in DHIS format.

Once form is submitted, the data is sent to a server set up by the project. For all eligible children (12 months or younger) who attended session camps in the relevant SCs in treatment arms, this server sends requests directly to the servers of relevant mobile phone service providers, to process recharges or tailored reminders. Note that for all forms submitted, a congratulatory text message and voice call is sent.

Appendix 2: Surveys and Sampling Methods

PHC Survey

This was a facility based survey conducted to obtain information on the staff and infrastructure availability, current set-up of the vaccine delivery system and its monitoring and the magnitude of population served by each PHC and associated sub-centres.

The PHC Survey targeted 161 PHCs across the seven sample districts to select a final sample of 140 PHCs. There were a total of 154 PHCs in the list provided by NHM. Based on the reconciliation of the PHC list from the PHC survey and the list that we received from NHM, we conducted surveys in 7 extra PHCs, bringing the total to 161 PHCs.

The following criterion was used to mark the eligibility for sampling PHCs. PHC should not receive vaccines from a different PHC. This criterion acted as a proxy for presence of cold storage. This was an important criterion as only PHCs with cold storage units were considered as planning units, that is delivery of vaccines, reporting and fund disbursal was restricted to such PHCs. PHCs which did not have cold storage units, reported to the closest cold storage PHC and all the planning of these PHCs was also delegated to the cold storage PHCs; thus in a way the non-cold storage PHC functioned more like an SC rather than an autonomous PHC. 151 PHCs met this criterion. In addition to this, given the geographical spread of the PHCs in Bhiwani, 11 PHCs that were closest to the Rajasthan border were dropped to provide logistical ease for data collection and implementation.

In order to sample villages for the survey, the first step was to assign *dhanis*[15] to their parent villages. Villages with less than 100 households were dropped as they did not meet the eligibility criterion[16] for the household sampling and villages with more than 1000 households were split into multiple units to generate a weighted sample. 207 villages were split following this decision. In addition, villages with more than 3000 households were also dropped for logistical reasons. With this, we arrived at a final list of villages for sampling. For the baseline survey, seven villages units were randomly selected from each of these 140 PHCs. In all, 912 villages were sampled for the baseline survey. The number of village units in each district vary with the size of the district.

Census

A census survey was conducted to identify eligible households[17] (households with one or more children between 12 and 36 months of age[18]) in sample villages in order to sample them for the baseline household surveys. A detailed mapping exercise of all the villages in the sample was conducted to estimate the number of dwellings with the village boundaries and help in tracking households during the census.

328,058 households were visited as a part of the census exercise over a period of three months. Out of these, 62,548 households fit our eligibility criterion.

[15] Settlements with a very small population, usually an appendix of a bigger village.

[16] 15 households with children between 12 to 36 months of age.

[17] A household comprises people who have lived in the same house for atleast 30 days in the past year, eat food cooked from the same stove, and contribute to and share household income.

[18] This range of age was chosen because these children had completed the immunization cycle and understanding their immunization history provide a reasonable estimate for village-wise baseline immunization rates.

Nominations Survey

A specific set of questions were used to collect nominations for the different types of seeds. The questions are similar, in terms of what is asked of the nominator, to the ones used in Banerjee et. al. (2014) and other gossip studies; however, they have been adapted to suit the local context of Haryana. This was done after a series of nine well-planned pilots, each with a new variation, conducted in Haryana. The nominations from these pilots were carefully studied to arrive at the final questions. To arrive at the final questions, it was necessary to understand the problems that each variation of the instrument faced. For example, it was seen that the gossip question suffered from a 'formal bias' as people were only nominating officials. To address this, in the subsequent pilot, a variation was added to check for this issue. The same steps were followed to address other issues that came up in the pilots.

For the nominations survey, 17 households in every sample village were randomly selected from the census that was conducted in all these villages. A short survey was conducted in these households wherein the respondent was asked to nominate four persons they thought would best fit the description in the question asked. Additional information about the nominee that would help us identify unique nominees and verify overlaps was also collected.

Baseline Survey

The 62,548 households from the census that fit our eligibility constituted our sampling frame, from which we sampled households for the baseline survey. 15 households[19] from each of the 970 village units (912 villages) were randomly selected to be administered the baseline survey[20].

The baseline survey comprised an extensive questionnaire containing eight modules of various questions on vaccines administered to children and their mothers, knowledge and attitudes towards immunization, government's SMS program and health practices for all eligible children in the households that were sampled. The survey also collected information on attitudes towards immunization and knowledge of the government's immunization program.

Responses about household assets, income, education and other demographic information were preferred from the head[21] of the household as they are best placed to answer these questions whereas questions on vaccines and immunization status of the children were compulsorily answered by the primary caregiver[22] of child. In total, 14670 households with 17,000 children were surveyed during the baseline exercise.

Seeds survey

The seeds survey was conducted for the following reasons: (i) to ask the nominated nodes for consent to be a part of the experiment, (ii) to collect phone numbers of the recruited seeds to send them information via text messages and recorded phone calls, and (iii) to collect other demographic information of the recruited seeds.

[19] 15 households in each village unit were also randomly selected as replacements-to be used if a household in the main sample cannot be interviewed for any reason.

[20] In PHCs where less than 7 village units were sampled, a higher number of households from each villages was sampled, to maintain the proportional sampling of household across all PHCs.

[21] The primary decision-maker of the household, mostly financial ones. Usually the eldest male member of the household.

[22] The person primarily responsible for taking care of the child and (in this case) who takes the child to get immunized. Usually the mother or the grandmother of the child

Both in-person and telephonic surveys were conducted to ensure that we recruit as many seeds of total nominated nodes generated from the algorithm. The surveys were conducted in the seven sample districts. As soon as the baseline surveys were completed in a particular district, the seeds survey was started after four days- two days of rest and two days of refresher training. A linear decision rule could not be established because seeds are probabilistic and the identifying information is from third person nominations. Thus, the enumerators were trained to use their discretion in cases where only few characteristics matched or in cases where there was more than one match. If two people were equally likely to be a seed, the enumerators were trained to toss a coin and randomly choose one. There was no scope for replacements except when a seed did not exist in the village.

2601 seeds (about 79 percent) consented to be a part of the program and receive information to disseminate in the village. Of these, 2117 consented for an survey to collect information on their demographics among other things.

Endline survey

The endline survey will be conducted in 3 rounds of data collection (roughly 6, 9, and 12 months after programme implementation) in 1048 sample villages. The target productivity is for one surveyor to survey 4 children per village per day, resulting in an estimated 4192 children surveyed per round, and 12567 children surveyed cumulatively after 3 rounds – this constitutes our target sample size for our outcome analysis.

Unlike baseline, there is no census to guide the identification of households for the endline survey. As such, surveyors use a random entry point into the village, adopt the right hand rule and go door-to-door to survey the first 4 eligible children they find[23]. Surveyors are provided with maps of the villages on which entry points have been clearly marked. These entry points were added retrospectively to villages which already had maps from the census exercise, using Google Maps. New maps of just the boundaries were created for the additional villages that were not originally part of our evaluation sample during baseline. Using a randomization table, that takes into account the total number of entry points in a village, surveyors are assigned a random entry point to use. This entry point will be excluded from the randomization table in the subsequent rounds of endline. Surveyors also use a highlighter to mark on maps what approximate area they have covered in the village, in order to minimize overlap in subsequent rounds.

The endline survey itself is a shortened version of the baseline survey. From baseline it emerged that the measurement burden was high (survey duration ranged from 45 minutes to 2 hours), and the findings from some modules were not as informative and significant as original thought. As such, the modules on polio and mother's vaccination during pregnancy were removed, as well a number of questions in the household assets and income module which showed very high agreement from baseline data. Overall, the endline survey now contains modules on household assets, child vaccination history and knowledge and attitudes towards immunization, and takes 1 hour at most per household. The household head is the preferred respondent for questions relating to the first module, while the primary caregiver or mother is the preferred respondent for questions about the subsequent two modules, respectively. Note that one module has been added – the 'child verification' module. This has been done in order to improve efficiency in surveying those children

[23]Note that the target is number of children rather than number of households, to ensure surveyors are able to complete the endline within the given time. Where there is more than one eligible child in a household, all children will be surveyed. Where the remaining number of children required to survey is lower than the number of eligible children in a household, the children closest to 12 months will be surveyed.

whom the surveyor comes across during endline that are also on his list for child verification. Questions in this module are identical to the questions in the child verification survey (see next section).

Child Verification Survey

In the same villages sampled for endline, surveyors will be completing the child verification survey on the same day. Up to 5 children will be sampled from the tablet data, linked to the particular village in question, and the surveyor will follow a detailed protocol to identify and survey the household to which the child belongs. The protocol includes 4 key steps each surveyor must follow, in order, including: asking the community health workers (ASHAs and/or anganwadi workers) where these children live; calling the ANMs to ask whether these children live; asking around in the village; calling the beneficiary number provided during the session camp.

For each child on the list, the surveyor must complete the survey and state whether or not he could find the child/child's household – if so, how he found the household; if not, what steps of the protocol he followed to find the child. For children that have been located, the surveyor will: ask questions to confirm the child's identity (name, parents' name, DOB); ask for and copy details of Aadhaar card and barcode ID; ask for and copy details of the MCP card (immunization card); ask question around recall of whether there was a tablet present at the session camp, whether the beneficiary provided a mobile phone number, and whether beneficiaries received any messages or voice calls on the number provided. The survey duration ranges between 10 to 20 minutes per child.

The primary analysis of the child verification data will be a 'verification rate' – what percentage of the children listed have been verified. Note that we are still defining what 'verified' means, and this will likely translate into a number of categories or degrees of certainty about a child's existence. Secondary analyses will look at: the accuracy of the data – e.g. the match between vaccination details from the MCP card and the tablet data; and implementation indicators (e.g. the proportion of valid phone numbers provided, the proportion of beneficiaries who recall receiving recharges, text messages and voice calls, etc)

For children who are identified as living outside the village, the child verification survey collects details on the correct village. Depending on the frequency of this situation, these children will be surveyed over the phone or, time permitting, in person, in between rounds of endline. In addition, since the programme covers more villages than just the evaluation sample for endline, the child verification survey will likely also cover a sample of these non-evaluation villages in between rounds of endline.

Appendix 3: Seeds Record Linking Algorithm

This section details the methods used to aggregate and match nominations collected from the nominations survey.

Data

The data was collected by a survey administered to 17 households in treatment villages. Each household was asked to nominate 4 individuals who matched a particular description (Gossip, Trusted Person, Trusted Gossip)- depending on the treatment group the village was assigned to.

For each nomination, respondents were asked to provide information on the nominated person. Specifically, data was collected on the following characteristics:

1. Name
2. Nickname
3. Gender
4. Age
5. Occupation
6. Household Head's Name
7. Father's Name
8. Mother's Name
9. Spouse's Name
10. Child's Name
11. Other Family Member's Name
12. Caste
13. Landmark near nominee's home
14. Where can we usually expect to find the nominee
15. Household location range on a map

The following is a plot of the distribution of the number of nominations per village.

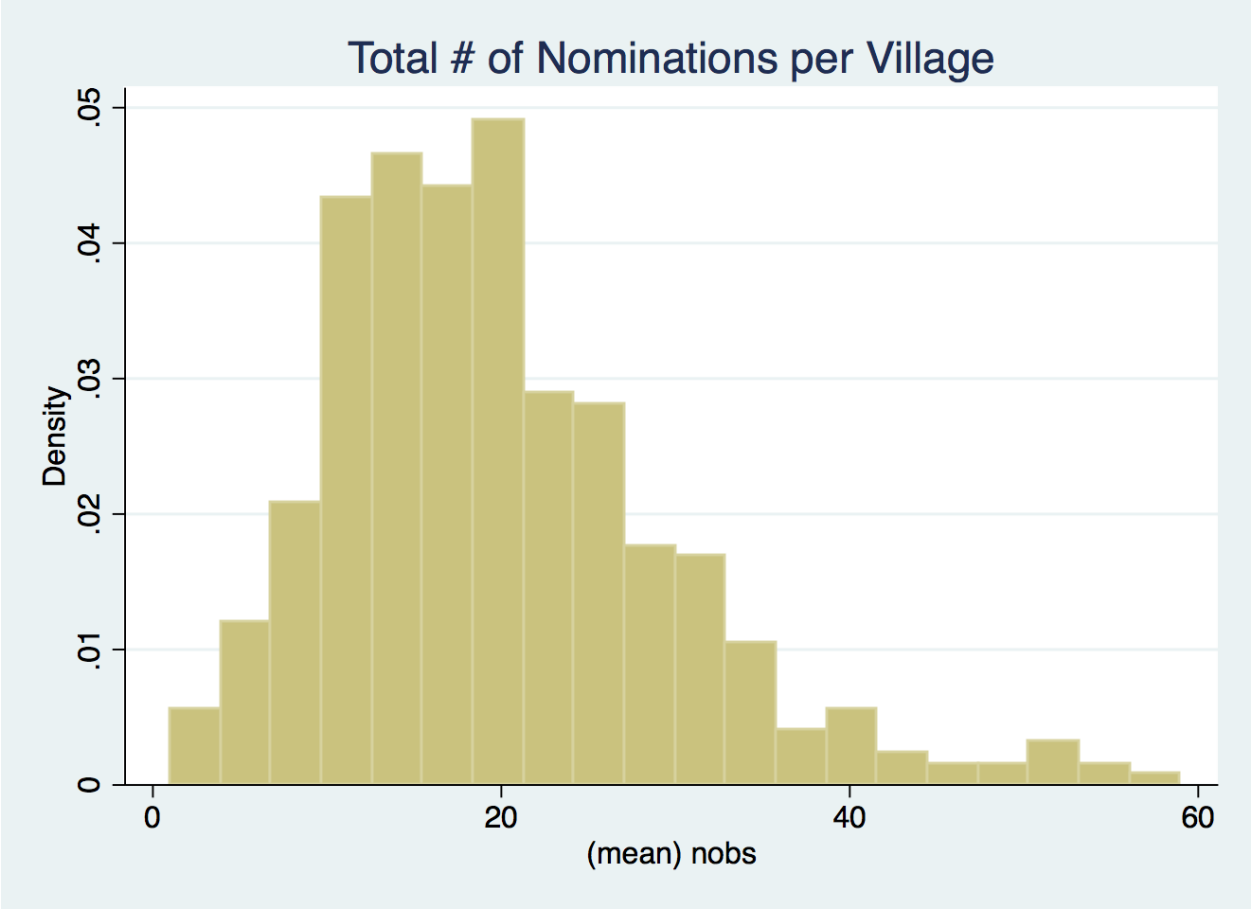

Figure 2: Total Number of Nominations per Village

The median number of nominations per village is 23.

The Algorithm

The goal of the algorithm is to aggregate these nominations by individual. However, since two people may have the same name, or two people who have nominated the same person may provide a different set of characteristics- this is not a straightforward problem. The algorithm is probabilistic, though every effort has been made to make this as accurate and precise as possible. The following is a general outline of the algorithm:

1. First, we standardized the names and occupations within each village- to harmonize the spellings etc. The data was also cleaned etc. to recode missing values and so forth.

2. For each village, we generated the universe of pairs for all the nominations within each village- so that nominations are compared on a pair-wise basis.

3. To this goal, we generated a "similarity score" $c_{i,j}$ for characteristic j within pair i, $\forall$ 13 characteristics, $\forall$ pairs. The similarity score was computed differently depending on the characteristic:

   Names: Names were scored using a string comparison algorithm based on the Levenshtein distance between two strings- though this was modified to account for Indian names. The details of this are included in a separate appendix. For nominee name, we took the minimum of the similarity scores of name to name, nickname to name, name to nickname, and nickname to nickname.
   These scores were multiplied by -1, so that they go in the same direction as the total score. (higher is better)

   Age: The age similarity score, was computed as
   $1 - 0.1 *$ the difference in the ages between the two elements of the pair if the difference is less than 10 years and 0 otherwise.

   Map Location: The approximate location of a nominee on the map was given as a range. Given that the accuracy of a range doesn't necessarily the likelihood of two nominations referring to the same person, we decided to simply give the location a similarity score of 1 if there is an overlap in the ranges, and 0 otherwise.

   Categorical Variables: These include gender, occupation, caste, landmark, and expected location. Categorical variables were standardized, so we assigned a similarity score of 0in cases where the variables didn't match, and a similarity score of 1 in cases whether the variables matched.

4. We then generated a "total score" T_i intended to capture the degree of match between two elements of a pair. Letting $mc_{i,j}$ is an indicator equal to 1 if characteristic i is missing for pair j (i.e $c_{i,j}$ is missing), and $c_{i,j}$ has been replaced with 0 in these cases. The total score was generated as a weighted sum of the similarity score for each characteristic, where we subtract a penalty for missing characteristics. i.e

$$T_i = \sum_{j=1}^{13} w_j c_{i,j} - \sum_{j=1}^{13} p_j mc_{i,j} \tag{7}$$

   where w_j denotes the weight applied to the similarity score for characteristic j.

5. The similarity score weights and penalty weights were obtained from a randomly selected subset of 50 villages, which we have designated as the "training data". Pairs in the training data were matched manually- so that pairs where both elements were thought to refer to the same individual were classified as a match, and pairs where both elements were thought to refer to different individuals, were classified as a non-match. Letting M_i denote an indicator equal to 1 if pair i is a match, and 0 otherwise- we obtained the weights using the coefficients obtained by estimating the following:

$$M_i = \alpha + \sum_{j=1}^{13} \beta_j c_{i,j} + \sum_{j=1}^{13} \gamma_j mc_{i,j} + \epsilon_i \tag{8}$$

   where ϵ_i denotes the error term, and is assumed to be independently identically normally distributed.

6. In order to avoid over-weighting a similarity score for a particular characteristic, we do not incorporate similarity scores for characteristics whose coefficients are not significant at the 5% level. Furthermore, we also exclude similarity scores for characteristics whose coefficients are negative, since it does not

make sense to discount the quality of a match of a particular pair by an amount that is increasing in the degree of match of a particular characteristic. i.e

$$w_j = \begin{cases} \beta_j, & \text{if } |\frac{\beta_j}{SE(\beta_j)}| \geq 1.96 \text{ and } \beta_j \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

Similarly, for the weights on penalties if the similarity score for a characteristic is missing, we apply the same logic:

$$p_j = \begin{cases} \gamma_j, & \text{if } |\frac{\gamma_j}{SE(\gamma_j)}| \geq 1.96 \text{ and } \gamma_j \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

7. We then qualified a given pair i as a match, if it satisfied *all* of the following criterion:

   (a) Total Score $T_i <= T^*$, where T^* is the threshold for the total score.

   (b) Name Score $N_i <= N^*$, where N_i is the name score for pair i, computed as outlined above. And N^* is a threshold for a name score. [24] [25]

   (c) Gender is the same for both elements of the pair, if gender is non-missing for both of them.

   (d) The elements of the pair were *not* nominated by the same respondent. (Since respondents were asked to nominate *distinct* individuals.)

   The thresholds T^* and N^* were chosen based on the training data, to maximise the True Positive Rate (TPR, the ratio of true number of accurately predicted matches to the total number of matches) and minimize the False Positive Rate (FPR, the ratio of incorrectly predicted matches to the total number of non-matches). This is a commonly used technique in Machine Learning to determine the optimal discrimination threshold for a binary classifier. A plot of the TPR against the FPR for different thresh hold values, is known as the ROC curve.
   Figures 3 and 4 show the ROC curves for the total score, and name score respectively. These were obtained by calculating the TPR and FPR for 100 different thresh holds between the maximum and minimum for each score (total and name). Each point on the plot represents the quality of the algorithms' classification for a given thresh hold level. The point $(0, 1)$ represents perfect classification. The 45 degree line represents the outcome of random guessing. For more info, see: ROC-Wikipedia.

[24] We filtered by name, because we found it to be a necessary but not sufficient requirement to qualify a match.

[25] We also include an indicator for whether the name is a match (whether the name score is below the said threshold- in the regressions.)

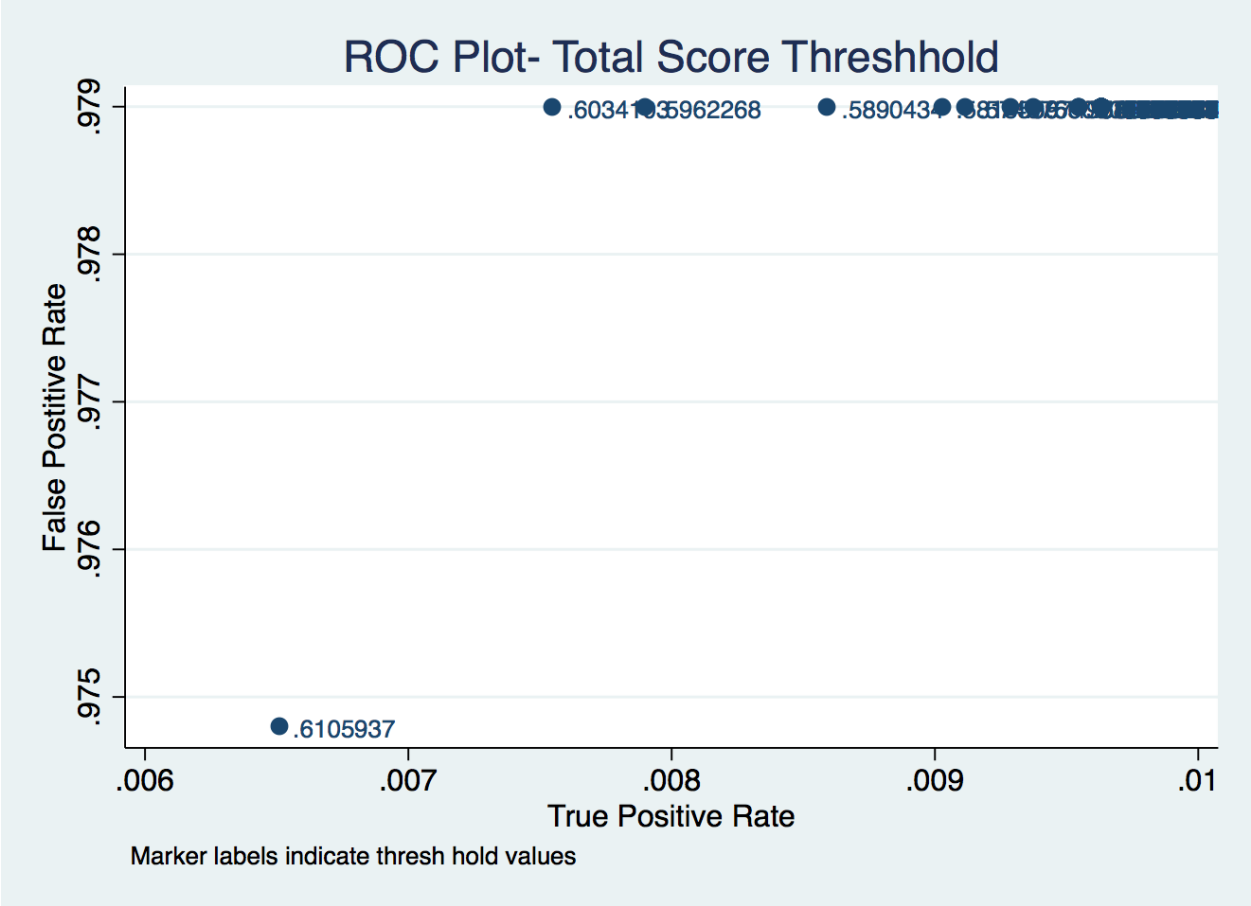

Figure 3: Total Score ROC Plot

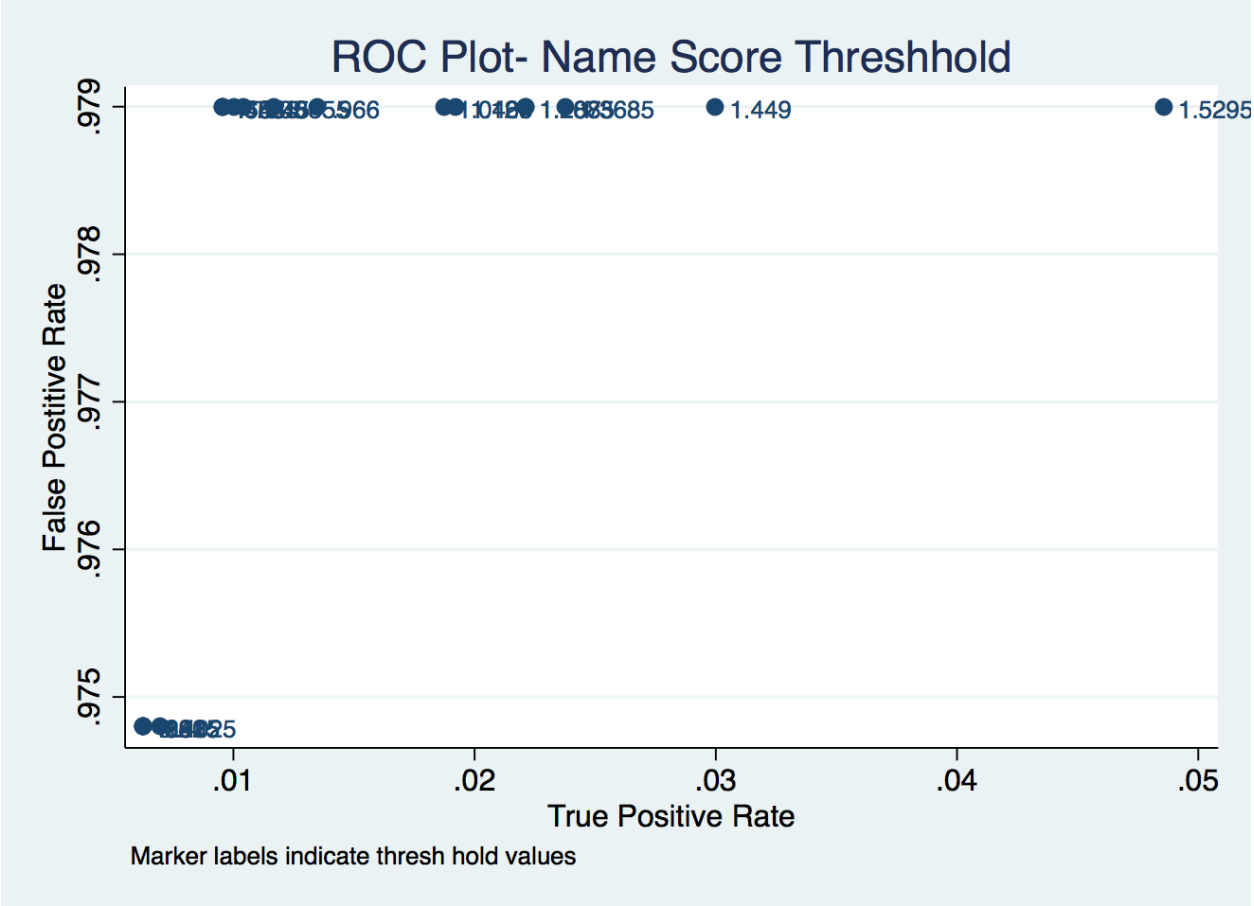

Figure 4: Name Score ROC Plot

8. Once the pairs were matched, we partitioned the set of nominations into equivalence classes, such that the "matching relation" is both transitive, symmetric, and reflexive. Given this, each equivalence class represents an individual.

9. Finally, we sort individuals by the total number of nominations the person received within each village. We select the top 6 individuals as seeds to disseminate information, ties are broken randomly. Figure 5 below shows the distribution of distinct individuals per village.

10. The tracking sheets sorted the information within individuals by completes (the number of non-missing entries), enumerators received a list of individuals with unique information separated by "OR".

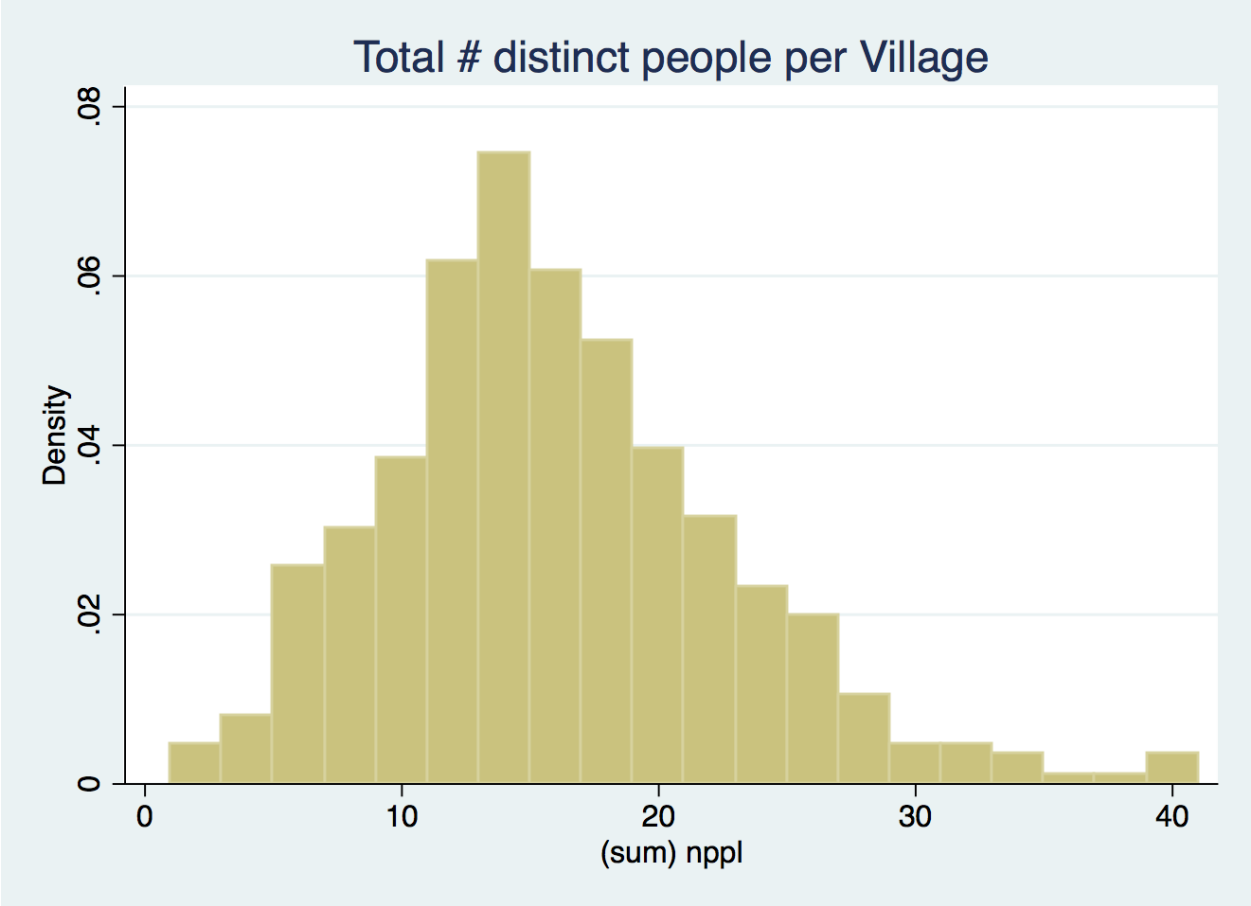

Figure 5: Total Number of Nominated People per Village

Appendix 4: Baseline Balance

A randomization balance check was carried out to check if the randomization was conducted properly, to see if imbalance in baseline characteristics causes any chance bias and to see whether the analysis requires any adjustment for one or more baseline variables. The exercise checks if any characteristic is significantly different between the incentive treatment and control group.

Table 1 reports the coefficients on a treatment indicator, from regressions of each of the baseline characteristics on the treatment indicator. For a given outcome, the coefficients can thus be interpreted as the difference in the mean of that outcome for the treatment and control groups. None of the variables are statistically significant at the 5% level. To be sure that this is no more than spurious correlation, we also report the p-value on the F-test for joint significance: 0.15. This is further confirmation that observable characteristics are jointly uncorrelated with treatment assignment.

Table 1: Baseline Balance Checks

	Treatment Mean	Control Mean	$X_i = \alpha + \beta T_i + \epsilon_i$	Obs
Self-reported financial status (Scale of 1-10)	3.268	3.296	-0.029	14665
			(0.072)	
Land owned	0.449	0.485	-0.036	14669
			(0.019)	
Land area (acres)	6.146	6.942	-0.796	14405
			(0.575)	
Wealth index	-0.054	0.052	-0.106	13790
			(0.129)	
General	0.412	0.418	-0.006	11720
			(0.028)	
Agricultural labour	0.030	0.023	0.007	14447
			(0.005)	
Other labour	0.340	0.326	0.014	14447
			(0.021)	
Maximum education	10.235	10.513	-0.279	14668
			(0.237)	
Number of vaccines administered	3.785	3.833	-0.048	16709
			(0.108)	
Full immunization rate	0.401	0.396	0.005	16709
			(0.026)	
Immunization card present	0.848	0.866	-0.017	16709
			(0.023)	
Immunization card verified	0.496	0.492	0.004	16709
			(0.026)	
Mother's age	26.045	26.061	-0.016	16875
			(0.139)	
Mother's education	7.441	7.833	-0.392	16833
			(0.396)	
Father's age	30.032	30.088	-0.056	16735
			(0.142)	
Father's education	9.161	9.462	-0.301	16669
			(0.261)	

Notes: 1. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

2. The p-value on the joint F-test is 0.1479.

3. All standard errors are clustered at the PHC level.