DISCUSSION PAPER SERIES

DP16084

Selecting the Most Effective Nudge: Evidence from a Large-Scale Experiment on Immunization

Esther Duflo, Abhijit Banerjee, John Floretta, Anna Schrimpf, Anirudh Sankar, Francine Loza, Harini Kannan, Matthew O. Jackson, Arun G Chandrasekhar, Maheshwor Shrestha and Suresh Dalpath

DEVELOPMENT ECONOMICS

CEPR

Selecting the Most Effective Nudge: Evidence from a Large-Scale Experiment on Immunization

Esther Duflo, Abhijit Banerjee, John Floretta, Anna Schrimpf, Anirudh Sankar, Francine Loza, Harini Kannan, Matthew O. Jackson, Arun G Chandrasekhar, Maheshwor Shrestha and Suresh Dalpath

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Development Economics

Selecting the Most Effective Nudge: Evidence from a Large-Scale Experiment on Immunization

Abstract

We evaluate a large-scale set of interventions to increase demand for immunization in Haryana, India. The policies under consideration include the two most frequently discussed tools—reminders and incentives—as well as an intervention inspired by the networks literature. We cross-randomize whether (a) individuals receive SMS reminders about upcoming vaccination drives; (b) individuals receive incentives for vaccinating their children; (c) influential individuals (information hubs, trusted individuals, or both) are asked to act as "ambassadors" receiving regular reminders to spread the word about immunization in their community. By taking into account different versions (or "dosages") of each intervention, we obtain 75 unique policy combinations. We develop a new statistical technique—a smart pooling and pruning procedure—for finding a best policy from a large set, which also determines which policies are effective and the effect of the best policy. We proceed in two steps. First, we use a LASSO technique to collapse the data: we pool dosages of the same treatment if the data cannot reject that they had the same impact, and prune policies deemed ineffective. Second, using the remaining (pooled) policies, we estimate the effect of the best policy, accounting for the winner's curse. The key outcomes are (i) the number of measles immunizations and (ii) the number of immunizations per dollar spent. The policy that has the largest impact (information hubs, SMS reminders, incentives that increase with each immunization) increases the number of immunizations by 44 % relative to the status quo. The most cost-effective policy (information hubs, SMS reminders, no incentives) increases the number of immunizations per dollar by 9.1%.

Esther Duflo - eduflo@mit.edu
Massachusetts Institute of Technology and CEPR

Abhijit Banerjee - banerjee@mit.edu
MIT Department of Economics and CEPR

John Floretta - jfloretta@povertyactionlab.org
J-PAL

Anna Schrimpf - schrimpf@povertyactionlab.org
MIT

Anirudh Sankar - asankar@povertyactionlab.org
MIT

Francine Loza - f.n.loza@gmail.com
MIT

Harini Kannan - harini.kannan@ifmr.ac.in
J-PAL

Matthew O. Jackson - jacksonm@stanford.edu
Stanford

Arun G Chandrasekhar - arungc@stanford.edu
Stanford

Maheshwor Shrestha - mshrestha1@worldbank.org
World Bank

Suresh Dalpath - sureshdalpath@yahoo.com
Health Department, Government of Haryana

SELECTING THE MOST EFFECTIVE NUDGE: EVIDENCE FROM A LARGE-SCALE EXPERIMENT ON IMMUNIZATION

ABHIJIT BANERJEE[†,*,§], ARUN G. CHANDRASEKHAR[‡,*,§], SURESH DALPATH[#], ESTHER DUFLO[†,*,§], JOHN FLORETTA[*], MATTHEW O. JACKSON[‡,*], HARINI KANNAN[*], FRANCINE LOZA[*], ANIRUDH SANKAR[‡], ANNA SCHRIMPF[*], AND MAHESHWOR SHRESTHA[§]

ABSTRACT. We evaluate a large-scale set of interventions to increase demand for immunization in Haryana, India. The policies under consideration include the two most frequently discussed tools—reminders and incentives—as well as an intervention inspired by the networks literature. We cross-randomize whether (a) individuals receive SMS reminders about upcoming vaccination drives; (b) individuals receive incentives for vaccinating their children; (c) influential individuals (information hubs, trusted individuals, or both) are asked to act as "ambassadors" receiving regular reminders to spread the word about immunization in their community. By taking into account different versions (or "dosages") of each intervention, we obtain 75 unique policy combinations. We develop a new statistical technique—*a smart pooling and pruning procedure*—for finding a best policy from a large set, which also determines which policies are effective and the effect of the best policy. We proceed in two steps. First, we use a LASSO technique to collapse the data: we *pool* dosages of the same treatment if the data cannot reject that they had the same impact, and *prune* policies deemed ineffective. Second, using the remaining (pooled) policies, we estimate the effect of the best policy, accounting for the winner's curse. The key outcomes are (i) the number of measles immunizations and (ii) the number of immunizations per dollar spent. The policy that has the largest impact (information hubs, SMS reminders, incentives that increase with each immunization) increases the number of immunizations by 44 % relative to the status quo. The most cost-effective policy (information hubs, SMS reminders, no incentives) increases the number of immunizations per dollar by 9.1%.

1. INTRODUCTION

Immunization is recognized as one of the most effective and cost-effective ways to prevent illness, disability, and death. Yet, worldwide, close to 20 million children every year do not receive critical immunizations (UNICEF and WHO, 2019). Resources devoted to routine immunization have risen substantially over the past decade (WHO, 2019). There is mounting evidence, however, that despite efforts to make immunization more widely available, insufficient parental *demand* for immunization has contributed to a stagnation in immunization coverage. This has motivated experimentation with "nudges," such as small incentives in cash or kind,[1] symbolic social rewards,[2] SMS reminders,[3] and the use of influential individuals in society or in the social network as "ambassadors."[4]

There is evidence, gathered from varied contexts, that all of these strategies may improve the take up of immunization at low cost (in some cases even reducing the cost *per immunization*). But to guide policy, we need to know which of these nudges is the most effective (i.e., leads to the largest increase in immunization), and which is the most cost-effective (i.e., leads to the largest increase in immunization per dollar spent). Moreover, there may be different dosage variants within each strategy. For example monetary incentives could be high or low, and they could also be constant or increasing with each shot. Finally, policies may work best in tandem or counteract each other. Hence, what we really need to know is which *combination* of nudges works best. In this paper, we run an experiment and develop a methodology to answer this question.

The most common way to compare policies is to perform meta-analyses of the literature: collect estimates from different papers, put them on a common scale, and compare them. This is the type of exercise routinely performed by Campbell Collaborative, the Cochrane review, and J-PAL, to name some examples. This is useful, but an issue with this approach is that both the populations and the interventions vary across studies, which makes it difficult to interpret any difference purely as the result of the intervention rather than a different experimental population. Moreover, if different interventions are tested in different contexts, identifying interactions between them is impossible. Thus, where possible, running a single large scale experiment in a relevant context and directly comparing different treatments, is desirable before launching a program at scale.

However, as the number of options increases, one runs into two issues. First, there are often a large number of possible interventions or variants of interventions and an even larger number of interactions between them. Since sample sizes are generally limited,

---

[1]See Banerjee et al. (2010); Bassani et al. (2013); Wakadha et al. (2013); Johri et al. (2015); Oyo-Ita et al. (2016); Gibson et al. (2017).
[2]See Karing (2018).
[3]See Wakadha et al. (2013); Domek et al. (2016); Uddin et al. (2016); Regan et al. (2017).
[4]See Alatas et al. (2019); Banerjee et al. (2019).

the researchers face an awkward choice. On the one hand, they can severely restrict the number of interventions (or combinations) that they test (McKenzie, 2019). This presumes that the researchers a priori know which small subset of policies are likely to have an effect, which is what they want to study in the first place. On the other hand, they can include all unique policy combinations but then end up with very low statistical power. In practice, what researchers often do to test multiple variants on interventions and mixes between them, is to only report the un-interacted specifications or pool different versions of the policies. But Muralidharan, Romero, and Wuthrich (2019) point out that this popular method of selection can be seriously misleading, for both conceptual and statistical reasons. Second, even if the number of policy options is sufficiently small to be estimated in the available sample, any estimate of the "best" policy risks being biased upwards, since it was selected for being the best (Andrews, Kitagawa, and McCloskey, 2019).

In this paper, we propose and implement a new approach—*a smart pooling and pruning procedure*—to deal with these problems, under bounds on the number of combinations that have an impact and on how many policies are such that intensities matter.

First, we select candidate policies and estimate their impact. The first step is to represent the policy combinations in a manner amenable to pooling. To do so, we incorporate information about the structure of the policies: some involve different dosages – or versions – of the same treatment, while some are fundamentally different. Only different dosages of the same underlying intervention may be pooled. This imposes important limitations on the possible collapsing of policies and ensures their interpretability. While this involves careful manipulation when there are many intensities and interaction, the intuition for this approach is simply to represent the treatment variables under the form "any dose of treatment A" and "high dose of treatment A", and all the relevant interactions, rather than "low dose of A" and "high dose of A", (assuming two possible dosages for treatment A).

We then use a version of LASSO to determine which variants can be pooled and which policies are irrelevant and can, therefore, be pruned. Before applying LASSO, we must however further transform the data, to address the issue that the candidate treatment variables are correlated with each other. For example, the variable "any dose of treatment A" is mechanically correlated with "high dose of treatment A." More generally, multiple dummies will be "on" for the same observation, which introduces correlation in the design matrix. The amount of correlation violates the condition of irrepresentability necessary for LASSO to work (Zhao and Yu, 2006): in essence, the correlation implies that regardless of the number of observations, there will always be a substantial risk that LASSO picks the wrong variable.

Fortunately, this problem can be addressed by pre-conditioning the design matrix using a *Puffer transformation*, developed in Rohe (2014) and Jia and Rohe (2015). We show that our setting, a cross-randomized RCT with varying dosages, is one where the Puffer transformation works especially well, since while the variables are correlated with each other, we can prove that their correlation is sufficiently bounded. Finally, we apply the post-LASSO procedure of Belloni and Chernozhukov (2013) to estimate the effect of the selected policy bundles. Their results (and related literature) imply that we have consistent estimates of this restricted set of relevant (potentially pooled) policies.

Second, we estimate the effect of the *best* policy in this set. This is subject to a winner's curse: namely, precisely because the best policy has the maximum effect relative to all other alternatives, it is also likely to be selected as best per the data when there is a positive random shock. Therefore, a naive estimate of the effect of a best policy will be too large and hence an unadjusted estimator of it will be upward biased (Andrews et al., 2019). Using a method proposed by Andrews et al. (2019), and leveraging our pruning of the set of policies from the first stage, we correct for this bias and deliver appropriate estimates for the best policy.

Although our approach combines existing techniques from statistics and econometrics for model selection and subsequent inference, the overall procedure is new and policy relevant, and the proposed estimators enjoy the properties of Belloni and Chernozhukov (2013) (consistency) and Andrews et al. (2019) (approximate unbiasedness) under reasonable assumptions. Both the steps in our procedure are important. Smartly pooling and pruning the various policy options aids both with estimation directly, as well as with problems when estimating the effect of the best policy. In particular, since the winner's curse adjustment penalizes the best policy effect more when the set of policies being compared is larger (as the conditional expectation of the positive shock such that the policy with the maximum effect comes out on top in the data increases in the number of alternatives), both pooling and pruning reduce the number of policies in the horse race and guard against over-penalizing the effect of the best policy. In particular, the penalization can be large when the second-largest effect is close to the maximal effect, and pooling these different dosages avoids this issue.[5]

Our approach thus provides a theoretically sound method to pick the most effective policy out of a large number of candidates without sacrificing power, without cherry-picking and without imposing strong priors. The intention to deploy the two-step algorithm, as well as the policies that could turn out to be candidates for pooling, can be specified in a pre-analysis plan, which avoids specification search. The ultimate result

---

[5]When a pooled policy is recommended as the best policy, one can in principle deploy any dosage among the menu of dosages pooled together. This offers flexibility in the policy recommendation: for example, the cheapest dosage on the menu may be chosen.

is a reliable estimate of the impact of the most effective or cost-effective policy, which can be conveyed to policymakers.

Our empirical application is a large-scale experiment covering seven districts, 140 Primary Health Centers (PHCs), 2,360 villages involved in the experiment including 915 at risk for all the treatments, and 295,038 children in the resulting data base, which we conducted in collaboration with the government of Haryana, India. For several years, the government of Haryana had developed various strategies to make reliable supply of immunization services available to rural areas, but the take-up of immunization remained low.

Part of the low demand reflects deep-seated doubts about the effectiveness of immunization, its side effects, or the motives of those trying to immunize children (Sugerman et al., 2010; Alsan, 2015; Martinez-Bravo and Stegmann, 2018). However, another part of the low take-up, even when immunizations are free and available, reflects a combination of relative indifference, inertia, and procrastination. In surveys we conducted in Haryana, many parents reported being in favor of immunization (in our context in rural India, 90% believed it was beneficial and 3% believed that it was harmful). Nonetheless, a large fraction of children received the first vaccine but did not complete the schedule, which is consistent with high initial motivation but difficulty with following through. We therefore experimented with nudges that have been shown to be effective in other contexts. The goal was to find the best combination and dosage of those nudges.

The experiment was a cross-randomized design of three main nudges: providing monetary incentives, sending SMS reminders, and seeding ambassadors. The ambassadors were either selected randomly or through a nomination process. In the latter case, a small number of randomly selected villagers are asked to identify those members of their village who are either particularly trusted or are information hubs for their community, or both. To identify information hubs, we asked these respondents to name the community members best positioned to spread information most wildly. In previous work, Banerjee et al. (2019), we have called this person a "gossip," and we have shown that they are, indeed, effective in spreading information.

For each of these nudges, the experiment included several variants. We varied the level and schedule of the incentives, the number of people receiving reminders, and mode of selecting the ambassadors, leading to a large number (75) of finely differentiated policy bundles.

We first prune and pool this large set of policies using our "smart" selection machinery to identify a best policy, and then derive the winner's curse-adjusted estimate of the best policy. For the immunization outcome, the policy set stemming from smart selection contains four distinct policies. The best policy is that which combines three nudges:

first, there are incentives for the child's caregiver that increase with each administered vaccine; second, SMS reminders are sent to the caregiver about the next scheduled vaccination; and third, the information about when an immunization camp will happen is diffused through community ambassadors who are information hubs. Correcting for the winner's curse, the best policy is estimated to increase the number of immunizations by 44% ($p < 0.05$). We find that low and high incentives are equally effective, and high or low level of SMS coverage are equally effective. Picking the cheapest of these options, our data recommends information hub seeding, "low-value" (INR 250) but increasing incentives and SMS reminders to 33% of caregivers.

The budget-constrained policymaker may care more about the number of immunizations per dollar, a measure of cost-effectiveness, although of course, a policy maker may be willing to pick a policy that is more expansive than the status quo per shot given, as long as it increases immunizations, and remain cheaper, per live saved, than other possible use of the funds. Smart pooling robustly selects one policy as more cost effective than the status quo: the one that combines either information hubs or information hubs with SMS reminders without incentives. Accounting for winner's curse, this policy increases the number of immunizations per dollar by 9.1% ($p < 0.05$). Once again, our data suggests that low and high levels of SMS coverage, as well as information and trusted information hubs, be pooled.

A substantive finding from this analysis is that using information hubs magnifies the effect of other interventions. Neither incentives nor SMS reminders are selected on their own, but are selected in combination with information diffusion via information hubs. Conversely, the information hubs are not selected for efficacy on their own, but only when combined with SMS reminders with or without incentives that grow with the number of shots (we speculate some reasons why this might be the case in the Conclusion). This underscores the danger in designing experiments that only include the un-interacted treatments (as suggested by Muralidharan et al. (2019)) in a setting where is no strong a priori reason to rule out interaction effects: in our setting one would have concluded from this kind of experiment that no intervention works, when there are in fact very effective interventions.

2. Context, Experimental Design, and Data

2.1. **Context.** This study took place in Haryana, a populous state in North India, bordering New Delhi. In India, a child between 12 and 23 months is considered to be fully immunized if he or she receives one dose of BCG, three doses of Oral Polio Vaccine (OPV), three doses of DPT, and at least one dose of a measles vaccination. India is one of the countries where immunization rates are puzzlingly low. According to the

2015-2016 National Family Health Survey, only 62% of children were fully immunized (NFHS, 2016). This is not because of lack of access to vaccines or health personnel. The Universal Immunization Program (UIP) provides all vaccines free of cost to beneficiaries, and vaccines are delivered in rural areas–even in the most remote villages. Immunization services have made considerable progress over the past few years and are much more reliably available than they used to be. During the course of our study we found that the monthly scheduled immunization session was almost always run in each village.

The central node of the UIP is the Primary Health Centre. PHCs are health facilities that provide health services to an average of 25 rural and semi-urban villages with about 500 households each. Under each PHC, there are approximately four sub-centres (SCs). Vaccines are stored and transported from the PHCs to either sub-centerss or villages on an appointed day each month, to a mobile clinic where the Auxiliary Nurse Midwife (ANM) administers vaccines to all eligible children. A local health worker, the Accredited Social Health Activist (ASHA), is meant to help map eligible households, inform and motivate parents, and take them to the immunization session. She receives a small fee for each shot given to a child in her village.

Despite this elaborate infrastructure, immunization rates are particularly low in North India, especially in Haryana. According to the District Level Household and Facility Survey, the full immunization coverage among 12-23 months-old children in Haryana fell from 60% in 2007-08 to 52.1% in 2012-13 (DLHS, 2013).

In the district where we carried out the study, a baseline study revealed even lower immunization rates (the seven districts that were selected were chosen because they have low immunization). About 86% of the children (aged 12-23 months) had received at least three vaccines. However, the fraction of children whose parents had reported they received the measles vaccine (the last in the sequence) was 39%, and only 19.4% had received the vaccine before the age of 15 months, whereas the full sequence is supposed to be completed in one year.

After several years focused on improving the supply of immunization services, the government of Haryana was interested in testing out strategies to improve household take-up of immunization, and in particular, their persistence over the course of the full immunization schedule. With support from USAID and the Gates Foundation, they entered into a partnership with J-PAL to test out different interventions. The final objective was very much to pick out the best policy for a possible scale up throughout the state.

Our study took place in seven districts where immunization was particularly low. In four districts, the full immunization rate in a cohort of children older than the one we consider, was below 40%, as reported by parents (which is likely a large overestimation

of the actual immunization rate, given that children get other kinds of shots and parents often find it hard to distinguish between them, as noted in Banerjee et al. (2021)). Together, the districts cover a population of more than 8 million (8,280,591) in more than 2360 villages, served by 140 PHCs and 755 SCs. The study covered all these PHCs and SCs, and are thus fully representative of the seven districts. Given the scale of the project, our first step was to build a platform to keep a record of all immunizations. Sana, an MIT-based health technology group, built a simple m-health application that the ANMs used to register and record information about every child who attended at least one camp in the sample villages. Children were given a unique ID which made it possible to track them across visits and centers. Overall, 295,038 unique children were recorded in the system, and 471,608 vaccines were administered. Data from this administrative database is our main source of information on immunization. We discuss the reliability of the data below. More details on the implementation are provided in the publicly available progress report (Banerjee et al., 2021).

2.2. **Interventions.** The study evaluates the impact of several nudges on the demand for immunization: small incentives, targeted reminders, and local ambassadors, all implemented in 2017.

2.2.1. *Incentives.* When households are indifferent or have a propensity to procrastinate, small incentives can offset any short term cost of getting to an immunization camp and lead to a large effect on immunization. Banerjee et al. (2010) shows that small incentives for immunization in Rajasthan (a bag of lentils for each shot and a set of plate for completing the course) led to a large increase in the rates of immunization. Similar results were subsequently obtained in other countries, suggesting that incentives tend to be effective (Bassani et al., 2013; Gibson et al., 2017). In the Indian health system, households receive incentives for a number of health behavior, including hospital delivery, pre-natal care visits, and, in some states (like Tamil Nadu), immunization.

The Haryana government was interested in experimenting with incentives. The incentives that were chosen were mobile recharges for pre-paid phones, which can be done cheaply and reliably on a very large scale. Almost all families have at least one phone and the overwhelming majority of the phones are pre-paid. Mobile phone credits are of uniform quality and fixed price, which greatly simplify procurement and delivery.

A small value of mobile phone credit was given to the caregivers each time they brought their child to get immunized. Any child under the age of 12 months receiving one of the five eligible shots (i.e., BCG, Penta-1, Penta-2, Penta-3, or Measles-1), was considered eligible for the incentives intervention. Mobile recharges were delivered directly to the caregivers' phone number that they provided at the immunization camp. Seventy (out of the 140) PHCs were randomly selected to receive the incentives treatment.

In Banerjee et al. (2010), only one reward schedule was experimented with. It involved a flat reward for each shot plus a set of plates for completing the immunization program. This left many important policy questions pending: does the level of incentive make a difference? If not, cheaper incentives could be used. Should the level increase with each immunization to offset the propensity of the household to drop out later in the program?

To answer these questions, we varied the level of incentives and whether they increased over the course of the immunization program. The randomization was carried out within each PHC, at the subcenter level. Depending on which sub-center the caregiver fell under, she would either receive a:

(1) Flat incentive, high: INR 90 ($1.34 at the 2016 exchange rate, $4.50 at PPP) per immunization (INR 450 total).
(2) Sloped incentive, high: INR 50 for each of the first three immunizations, 100 for the fourth, 200 for the fifth (INR 450 total).
(3) Flat incentive, low: INR 50 per payment (INR 250 total).
(4) Sloped incentive, low: INR 10 for each of the first three immunizations, 60 for the fourth, 160 for the fifth (INR 250 total).

Even the high incentive levels here are small and therefore implementable at scale, but they still constitute a non-trivial amount for the households. The "high" incentive level was chosen to be roughly equivalent to the level of incentive chosen in the Rajasthan study: INR 90 was roughly the cost of a kilogram of lentils in Haryana during our study period. The low level was meant to be half of that (rounded to INR 50 since the vendor could not deliver recharges that were not multiple of 10). This was meaningful to the households: INR 50 corresponds to 100 minutes of talk time on average. The provision of incentives was linked to each vaccine. If a child missed a dose, for example Penta-1, but then came for the next vaccine (in this case, measles), they would receive both Penta-1 and measles and get the incentives for both at once, as per the schedule described above.

To diffuse the information on incentives, posters were provided to ANMs, who were asked to put them up when they set up for each immunization session. The village ASHAs and the ANMs were also supposed to inform potential beneficiaries of the incentive structure and amount in the relevant villages. However, there was no systematic large scale information campaign, and it is possible that not everybody was aware of the presence or the schedule of the incentives, particularly if they had never gone to a camp.

2.2.2. *Reminders.* Another frequently proposed method to increase immunization is to send text message reminders to parents. Busy parents have limited attention and reminders can put the immunization back at the "top of the mind." Moreover, parents do not necessarily understand that the last immunization in the schedule (measles) is for a different disease and is at least as important as the previous ones. SMSs are also

extremely cheap and easy to administer in a population with widespread access to cell phones. Even if not everyone gets the message, the diffusion may be reinforced by social learning, leading to even faster adoption.[6]

The potential for SMS reminders is recognized in India. The Indian Academy of Pediatrics rolled out a program in which parents could enroll to get reminders by providing their cell phone number and their child's date of birth. Supported by the Government of India, the platform planned to enroll 20 million children by the end of 2020.

Indeed, text messages have already been shown to be effective to increase immunization in some contexts. For example, a systematic review of five RCTs finds that reminders for immunization increase take up on average (Mekonnen et al., 2019). However, it remains true that text messages could potentially have no effect or even backfire if parents do not understand the information provided and feel they have no one to ask (Banerjee et al., 2018). Targeted text and voice call reminders were sent to the caregivers to remind them that their child was due to receive a specific shot. To identify any potential spillover to the rest of the network, this intervention followed a two step randomization. First, we randomized the study sub-centers into three groups: no reminders, 33% reminders, and 66% reminders. Second, after their first visit to that sub-center, children's families were randomly assigned to either get the reminder or not, with a probability corresponding to the treatment group for their sub-centers. The children were assigned to receive/not receive reminders on a rolling basis.

The following text reminders were sent to the beneficiaries eligible to receive a reminder. In addition, to make sure that the message would reach illiterate parents, the same message was sent through an automated voice call.

(1) Reminders in incentive-treatment PHCs:
> "Hello! It is time to get the «name of vaccine» vaccine administered for your child «name». Please visit your nearest immunization camp to get this vaccine and protect your child from diseases. You will receive mobile credit worth «range for slope or fixed amount for flat» as a reward for immunizing your child."

(2) Reminders in incentive-control PHCs:
> "Hello! It is time to get the «name of vaccine» vaccine administered for your child. Please visit your nearest immunization camp to get this vaccine and protect your child from diseases."

---

[6]See, e.g., Rogers (1995); Krackhardt (1996); Kempe, Kleinberg, and Tardos (2003); Jackson (2008); Iyengar, den Bulte, and Valente (2010); Hinz, Skiera, Barrot, and Becker (2011); Katona, Zubcsek, and Sarvary (2011); Jackson and Yariv (2011); Banerjee, Chandrasekhar, Duflo, and Jackson (2013); Bloch and Tebaldi (2016); Jackson (2017); Akbarpour, Malladi, and Saberi (2017).

2.2.3. *The Immunization Ambassador: Network-Based Seeding.* The goal of the immunization ambassador intervention was to leverage the social network to spread information. In particular, the objective was to identify influential individuals who could relay to villagers both the information on the existence of the immunization camps, and, wherever relevant, the information that incentives were available. Existing evidence shows that people who have a high centrality in a network (e.g., they have many friends who themselves have many friends) are able to spread information more widely in the community (Katz and Lazarsfeld, 1955; Aral and Walker, 2012; Banerjee et al., 2013; Beaman et al., 2018; Banerjee et al., 2019). Further, members in the social network are able to easily identify individuals, whom we call information hubs, who are the best placed to diffuse information as a result of their centrality as well other personal characteristics (social mindedness, garrulousness, etc.)(Banerjee et al., 2019).

This intervention took place in a subset of 915 villages where we collected a full census of the population (see below for data sources). Seventeen respondents in each village were randomly sampled from the census to participate in the survey, and were asked to identify people with certain characteristics (more about those later). Within each village, the six people nominated most often by the group of 17 were recruited to be ambassadors for the program. If they agreed, a short survey was conducted to collect some demographic variables, and they were then formally asked to become program ambassadors. Specifically, they agreed to receive one text message and one voice call every month, and to relay it to their friends. In villages without incentives, the text message was a bland reminder of the value of immunization. In villages with incentives, the text message further reminded the ambassador (and hence potentially their contacts) that there was an incentive for immunization.

While our previous research had shown that villagers can reliably identify information hubs, a pertinent question for policy unanswered by previous work is whether the information hubs can effectively transmit messages about health, where trust in the messengers may be more important than in the case of more commercial messages.

There were four groups of ambassador villages, which varied in the type of people that the 17 surveyed households were asked to identify. The full text is in Appendix I.

(1) *Random* seeds: In this treatment arm, we did not survey villages. We picked six ambassadors randomly from the census.
(2) *Information hub* seed: Respondents were asked to identify who is good at relaying information.
(3) *Trusted* seed: Respondents were asked to identify those who are generally trusted to provide good advice about health or agricultural questions

(4) *Trusted information hub* seed: Respondents were asked to identify who is both trusted and good at transmitting information

2.3. **Experimental Design.** The government was interested in selecting the best policy, or bundle of policies, for possible future scale up. We were agnostic as to the relative merits of the many available variants. For example, we did not know whether the incentive level was going to be important, nor did we know if the villagers would be able to identify trusted people effectively and hence, whether the intervention to select trusted people as ambassadors would work. However, we believed that there could be significant interactions between different policies. For example, our prior was that the ambassador intervention was going to work more effectively in villages with incentives, because the message to diffuse was clear. We therefore implemented a completely cross-randomized design, as illustrated in Figure 1.

We started with 2,360 villages, covered by 140 PHCs, and 755 sub-centers. The 140 PHCs were randomly divided into 70 incentives PHCs, and 70 no incentives PHCs (stratifying by district). Within the 70 incentives PHCs, we randomly selected the sub-centers to be allocated to each of the four incentive sub-treatment arms. Finally, we only had resources to conduct a census and a baseline exercise in about 900 villages. We selected about half of the villages from the coverage area of each subcenter, after excluding the smallest villages. Only among the 915 villages did we conduct the ambassador randomization: after stratifying by sub-center, we randomly allocated the 915 villages to the control group (no ambassador) or one of the four ambassador treatment groups.

In total, we had one control group, four types of incentives interventions, four types of ambassador interventions, and two types of SMS interventions. Since they were fully cross-randomized (in the sample of 915 villages), we had 75 potential policies, which is large even in relation to our relatively large sample size. Our goal is to identify the most effective and cost-effective policies and to provide externally valid estimates of the best policy's impact, after accounting for the winner's curse problem. Further, we want like to identify other effective policies and answer the question of whether different variants of the policy had the same or different impacts.

2.4. **Data.**

2.4.1. *Census and Baseline.* In the absence of a comprehensive sampling frame, we conducted a mapping and census exercise across 915 villages falling within the 140 sample PHCs. For conducting the census, we visited 328,058 households, of which 62,548 households satisfied our eligibility criterion (children aged 12 to 18 months). These exercises were carried out between May and November 2015. The data from the census was used to sample eligible households for a baseline survey. We also used the census to sample

the respondent of the ambassador identification survey (and to sample the ambassadors in the "random seed" villages). Around 15 households per village were sampled, resulting in data on 14,760 households and 17,000 children. The baseline survey collected data on demographic characteristics, immunization history, attitudes and knowledge and was conducted between May and July 2016. A village-level summary of baseline survey data is given in Appendix Table H.

2.4.2. *Outcome Data.* Our outcomes of interest are the number of vaccines administered for each vaccine every month, and the number of fully immunized children every month. We focus the main analysis of this paper on the number of children who received the measles vaccines in each village every month: this is the last vaccine in the immunization schedule and the ANMs check the immunization history and administer missing vaccines when a child is brought in for this vaccine. Thus, this is a good proxy for a child being fully immunized.

For our analysis, we use administrative data collected by the ANM using the e-health application on the tablet, and stored on the server, to measure immunization. At the first visit, a child was registered using a government provided ID (or in its absence, a program-generated ID) and past immunization history, if any. In subsequent visits, the unique ID was used to pull-up the child's details and update the data. Over the course of the program, about 295,038 children were registered, yielding a record of 471,608 immunizations. We use the data from December 2016 to November 2017. We do this because of a technical glitch in the system–the SMS intervention was discontinued from November 2017, although the incentives and information hub interventions were continued a little longer, through March 2018.

Since this data was also used to trigger SMS reminders and incentives, and for the government to evaluate the nurses' performance,[7] it was important to assess its accuracy. Hence, we conducted a validation exercise, comparing the administrative data with random checks, as described in Appendix G. The data quality appears to be excellent. Finally, one concern (particularly with the incentive program) is that the intervention led to a pattern of substitution, with children who would have been immunized elsewhere (in the PHC or at the hospital) choosing to be immunized in the camp instead. To address this issue, we collected data immediately after the intervention on a sample of children who did not appear in the database (identified through a census exercise), to ascertain the status of their immunization. In Appendix F, we show that there does not appear to be a pattern of substitution, as these children were not more likely to be immunized elsewhere.

---

[7]Aggregated monthly reports generated from this data replaced the monthly reports previously compiled by hand by the nurses.

Below, the dependent variable is the number of measles shot given in a village in a month (each month, one immunization session is held at each site). On average, in the entire sample, 6.16 measles shot were delivered per village every month (5.29 in the villages with no intervention at all). In the sample at risk for the ambassador intervention (which is our sample for this study) 6.94 shots per village per month were delivered.

2.5. **Interventions Average Effects.** In this section, we present the average effects of the interventions using a standard regression without interactions.

We begin by comparing average results for the incentive and SMS interventions in the entire sample and the ambassador sample. In the entire sample, we run the following regression:

$$y_{dsvt} = \alpha + \beta'\text{Incentive}_s + \gamma'\text{SMS}_s + \delta'\text{Ambassador}_v + \lambda\text{Ambassador Sample}_v + \upsilon_{dt} + \epsilon_{dsvt}$$

where y_{dsvt} is the number of measles shot given in month t in village v in sub-center (SC) s, and district d. Ambassador Sample$_v$ is a dummy indicating that a village is part of the Ambassador sample, and Ambassador$_v$ is a vector of the four possible ambassador interventions (randomly chosen, nominated as "information hub," nominated as "trusted information hub," and nominated as "trusted"). Incentive$_s$ is a vector of incentive interventions (low slope, high slope, low flat, high flat), and SMS$_s$ is a vector of SMS interventions (33% or 66%). υ_{dt} is a set of district-time dummies (since the intervention was stratified at the district level), and ϵ_{dsvt} represents the error term.

In the sample of census villages that will be used for the rest of the analysis (which are the villages where the ambassador intervention was also administered), we run the same specification, but leave out the Ambassador Sample dummy:

$$y_{dsvt} = \alpha + \beta'\text{Incentive}_s + \gamma'\text{SMS}_s + \delta'\text{Ambassador}_v + \upsilon_{dt} + \epsilon_{dsvt}.$$

In all specifications, we weight these village-level regressions by village population, and standard errors are clustered at the SC level.[8]

The results are presented graphically in Figure 2 for the expanded sample (Panel A) and the census villages (Panel B). Incentives and SMS had very similar impacts in both samples. The only intervention that appears to have a significant impact is the "high slope" incentives, which increases the number of immunization relative to control by 1.74 in the full sample, and 1.97 in the ambassador study sample. The low slope has a smaller positive effect, but always insignificant, and the SMS interventions have no impact.

---

[8]This is the highest level at which a treatment is administered, so clustering at this level should yield the most conservative estimate of variance. In practice clustering at the village level or SC level does not make an appreciable difference.

The results of the ambassador intervention, in panel B (already reported in Banerjee et al. (2019)) show that, on average, using information hubs ("gossips" in that paper) as ambassadors has positive effects on immunization: 1.89 more children receive a measles vaccine on a base of 7.32 in control in this sample ($p = 0.04$). This is near-identical to the effect of the high-powered, sloped incentive, though this intervention is considerably cheaper. In contrast, none of the other ambassador treatments–random seeding, seeding with trusted individuals, or seeding with trusted information hubs–have benefits statistically distinguishable from zero ($p = 0.42$, $p = 0.63$, and $p = 0.92$ respectively) and the point estimates are small, as well.

The conclusion from this first set of analyses is that financial incentives can be effective to boost demand for immunization, but only if they are large enough and increase with each immunization. Of the two cheaper interventions, the SMS interventions, promoted widely in India and elsewhere, seem disappointing. In contrast, leveraging the community by enrolling local ambassadors, selected using the cheap procedure of asking a few villages who are good information hubs, seems to be as effective as using incentives. It leads to an increase of 26% in the number of children who complete the schedule of immunization every month. This alone could increase full immunization rate in those districts from 39% (our baseline full immunization rate, as reported by parents) to nearly 49%.

This analysis does not fully answer the policymaker's question, however. It could well be that the interventions have powerful interactions with each other, which has two implications: first, the main effect, as estimated, does not tell us what the impact of the policy would be in Haryana if implemented alone (because as it is, they are a weighted average of a complicated set of interacted treatments). Second, it is possible that the government could do better by combining two (or more) interventions. For example, our prior in designing the information hub ambassador intervention (described in our proposal for the project)[9] was that it would have a positive interaction effect with incentives, because it would be much easier for the information hubs to relay hard information (there are incentives) than a vaguer message that immunization is useful. The problem, however, is that there are a large number of interventions and interactions: we did not—nor was it feasible to—think through ex-ante all of the interactions that should or should not be included, which is why in Banerjee et al. (2019), we only reported the average effects of each different type of seeds in the entire sample, without interactions. In the next section, we propose a disciplined approach to select which ones to include, and to then estimate the impact of the "best" policy.

---

[9]https://doi.org/10.1257/rct.1434-4.0

3. ESTIMATION

3.1. **Environment.** We have a randomized controlled trial with M cross-randomized arms. Each arm has R ordered dosage intensities: $\{\text{none}, \text{intensity } 1, ..., \text{intensity } R-1\}$. Although R is here considered fixed, the logic extends to the case where R varies across treatment arms. The total number of treatment combinations is $K := R^M$. Every one of n observational units, villages in our setting and henceforth referred to as such, is randomized to one treatment combination.

Our first assumption bounds the growth rate of the number of treatment combinations, and therefore arms and dosage intensity variants, relative to the number of observations.

ASSUMPTION 1. $R \geq 3$, and n, $K < n$ and $K = O(n^\gamma)$ for some $0 < \gamma < \frac{1}{2}$.

The condition $R \geq 3$ makes sure that there are at least two non-zero dosage variants of each treatment arm.[10] $K < n$ ensures there are no more treatment combinations than observations in every finite sample under consideration.

Let $T_{i,k}$ be a dummy for whether treatment combination k was assigned to village i. Thus, $T_{i,k}$ is 1 for exactly the treatment combination k that was assigned to it, and 0 otherwise. $T \in \{0,1\}^{n,K}$ is then the matrix capturing the treatment status. $T_{\cdot,k}$ (without the i) denotes the n-length column vector (dummy variable) corresponding to the treatment combination k.

The regression of interest is

$$(3.1) \qquad\qquad y = T\beta + \epsilon$$

where $y \in \mathbb{R}^{n \times 1}$ is the outcome of interest and $\beta \in \mathbb{R}^{K,1}$ is the vector of treatment effects.

For ease of exposition, we assume homoskedastic errors, which is restrictive but in keeping with the literature on the techniques utilized below (Rohe, 2014; Jia and Rohe, 2015).

ASSUMPTION 2. $\epsilon_i \sim \mathcal{N}(0, \sigma^2 I_n)$ with $\sigma^2 > 0$ fixed in n.

Only a subset S_β of treatment combinations have a non-zero effect on the outcome of interest:

$$S_\beta := \{k : \beta_k \neq 0\}.$$

Treatments are assumed to have either no effect or have sufficiently large (positive or negative) influence on the outcomes. That is, the non-zero efficacies are assumed to exceed a threshold.[11]

---

[10]This is obviously for notational convenience, as will become clearer below. The discussion of pooling dosages is moot if there are no dosages to pool, so this is effectively without loss of generality.

[11]The uniform lower bound is stronger than needed–see Zhao and Yu (2006) or Jia and Rohe (2015) for a weaker requirement allowing slowly declining effects of treatments as n grows: $n^{\frac{1-c}{2}} \min_{k \in S_\beta} |\beta_k| \geq M$

ASSUMPTION 3. $|S_\beta| < K$ and $\min_{k \in S_\beta} |\beta_k| > c > 0$ for c fixed in n.

In what follows we have two goals:

(1) Consistently estimate the effects of the relevant policies.
(2) Find the best policy $k^\star = \mathrm{argmax}_{k \in S_\beta} \beta_k$, and estimate the best policy effect $\beta_{k^\star}$.

We proceed in two steps. First, we estimate the set of relevant policies S_β. Second, we use post-estimation to both consistently estimate the policy effects (Belloni and Chernozhukov, 2013) and estimate the effect of the best policy (Andrews et al., 2019).

3.2. **Estimating Relevant Policies.** To estimate the set of relevant policies, we develop a *smart pooling and pruning* approach. We pool dosages if they have no differential effect on outcomes and remove irrelevant treatment combinations. This improves performance of the best policy, which tends to underperform when there are many potential alternatives, and especially, many alternatives that are very similar to it.

3.2.1. *Smart Pooling and Pruning.* One way to estimate equation (3.1) is to use LASSO. Under the above assumptions, the estimated support set $\hat{S}_\beta$ will equal S_β with probability tending to one. However, in the basic specification, two different dosages of the same treatment arm are treated exactly like any two arbitrary treatments. So, LASSO may select both (or both of their interactions with the other treatments) since they have very similar effects. For instance, in our setting, it may be the case that information hubs work equally well whether a 33% or a 66% reminder rate is used, nor does it matter whether high or low sloped incentives are used; the higher dosages of reminders and incentives do not increase efficacy. However, all the four variants of the information hub treatment (i.e., where it is combined with high or low reminder rate, high or low sloped incentives) will have equal claims to be chosen by LASSO.

In this case, if information hubs work equally well, irrespective of the reminders or incentives used, the policymaker would like to know this for two reasons. First, if the granular policies are essentially the same—an information hub policy—then insisting on granularity reduces power. Second, when adjusting for the winner's curse in estimating the effect of the best policy, both the number of alternatives and the gap in treatment effect between the best and second-best policy determines how conservatively we need to shrink our estimated effect.

A more sophisticated approach may be to run LASSO on (3.1) and then attempt to pool policies ex-post. This presents its own challenges. The researcher needs to organize their selected unique policy combinations into collections of pairwise and groupwise

---

for some constants $c \in (0, 1)$ and $M > 0$ independent of n. In our case $\gamma = c$. Even that requirement is to ensure that all relevant policies are sufficiently relevant enough relative to the rate at which information accumulates. It is cleaner to study the case with uniform bound; we proceed accordingly.

comparisons consistent with pooling goals. There can be an enormous number of comparisons (see the Hasse diagram in Appendix B, Figure B.1 to see the complexity of even a simplified example). Therefore, many hypothesis tests, with adjustments for multiple comparisons and false discovery rates, must be conducted. This may be complex to implement and, further, it is not immediately clear what the statistical properties of post-estimation from this procedure would be.

Our approach is to transform the problem in such a way that the specification natively executes the pooling and pruning in the process of estimation, and the procedure is consistent. To provide a structure that will help LASSO eliminate irrelevant variations between interventions (if that is what the data supports), it is useful to transform (3.1) to separate the marginal impacts of different levels of intensity from the base effect of a particular combination of treatments. It is important that we only collapse policy variants that differ only in treatment dosages, something which can be pre-specified. We do not collapse fundamentally different treatments into the "short" models discussed in Muralidharan et al. (2019).

First, every treatment combination k has an associated *treatment profile $P(k)$*, which is a unique element of the 2^M combinations of treatment arms without regard to intensity. In other words, it captures which treatment arms are "active" for a treatment combination. We say that treatment combinations k and k' share the same treatment profile if $P(k) = P(k')$.

EXAMPLE **1.** *Consider an example where village i is assigned the treatment combination $k =$ (No Seeding, low-value flat incentives, 33% reminders), while village j is assigned $k' =$ (No Seeding, low-value flat incentives, 66% reminders). Then $P(k) = P(k')$, because k and k' share the same treatment profile (No Seeding, any flat incentives, any reminders).*

Second, we can consider a partial ordering of treatment combinations with respect to their intensities. Specifically, for treatment combinations k, k', we can say $k \geq k'$ if the intensities of k *in each arm* weakly dominate that of k'. In other words, k has a weakly higher dosage in each arm than k'.[12] Then, we define a *profile and dosage* matrix $X \in \{0,1\}^{n,K}$ as follows:

$$X_{i\ell} = \mathbf{1}(k(i) \geq \ell \text{ and } P(k(i)) = P(\ell))$$

where $k(i)$ is the unique treatment combination that i is assigned to. In other words, i gets a 1 for all treatment combinations that share $k(i)$'s treatment profile and are weakly dominated in intensity by $k(i)$, and zero otherwise.

---

[12]Clearly, some treatment combinations are not comparable, which is why this is a partial order

EXAMPLE **1** (continued)**.** *The treatment combination*

$$k(i) = \textit{(No Seeding, low-value flat incentives, 33\% reminders)}$$

is weakly dominated by the treatment combination

$$k(j) = \textit{(No Seeding, low-value flat incentives, 66\% reminders)}$$

For village i, $X_{ik} = 1$ exactly for $k = k(i)$. However, for village j, $X_{ik} = 1$ for both $k = k(j)$ and $k = k(i)$.

*The column vector $X_{\cdot,k(i)}$ stands for all villages assigned treatments satisfying (No Seeding, **at least** low-value flat incentives, **at least** 33% reminders), while the column vector $X_{\cdot,k(j)}$ stands for all villages assigned treatments satisfying (No Seeding, **at least** low-value flat incentives, 66% reminders)*

We study the following *smart pooling and pruning* specification

$$(3.2) \qquad\qquad y = X\alpha + \epsilon.$$

It is an invertible linear transformation of (3.1).

This transformation is useful for two reasons. First, as defined here, α_k is either the baseline effect of that particular treatment profile (for the lowest intensity treatment combination within that profile) or the marginal impact of treatment k's intensity profile as a dosage relative to the next lower intensity within this treatment profile (for treatments with higher than the minimum intensity). In the experiment we only have two nonzero dosages, so there is only the baseline impact or the marginal impact of the higher dosage. If we estimate $\hat{\alpha}_k = 0$ for some k that is not the baseline level for that treatment, this indicates that this particular marginal dosage has zero impact and that the two policy variants may not have distinct effects and therefore may be pooled. Second, $\hat{\alpha}_k = 0$ may also imply that one or more $\hat{\beta} = 0$, meaning that even the baseline level of the treatment has no impact and thus, potentially, this specification also prunes in addition to pooling.

EXAMPLE **1** (continued)**.** *Returning to the example above, this formulation implies that if the marginal value of more reminders, given this treatment profile, is non-zero then we would expect for there to be a non-zero coefficient $\alpha_{k(j)}$ on $X_{k(j)}$. Otherwise, the coefficient is expected to be zero and the two intensities can be pooled for this treatment profile. We can substitute for $T_{\cdot,k(i)}, T_{\cdot,k(j)}$ a new treatment variable $T_{\cdot,k(i)\cup k(j)} = T_{\cdot,k(i)} + T_{\cdot,k(j)}$. This new treatment pools together villages i and j, i.e., $T_{i,k(i)\cup k(j)} = T_{j,k(i)\cup k(j)} = 1$.*

We apply a LASSO-based procedure to (3.2) to estimate S_α, the support of this equation (i.e., the variables that are not eliminated by LASSO). We then operate a final

transformation of S_α to generate S_{pool}, the collection of mutually exclusive pooled policy combinations that are represented by α.

An important practical detail is how to make sure S_{pool} is obtained correctly—that is, pool only those treatment combinations deemed to have identical treatment effects, and take a pooling decision on every single treatment combination. With just a few treatment intensities and arms, these can be eye-balled from S_α, because it is easy to see what is the "main effect" and what is "marginal." When the number of dosages and treatment arms increases, however, the partial ordering of intensities gets more involved and one might unintentionally mis-pool by simply eye-balling it. In Appendix B we propose a general algorithm (Algorithm 2) for recovering S_{pool} for any R, M and S_α, and prove that when S_α is correctly estimated, the derived S_{pool} correctly pools and covers the support of S_β.

3.2.2. *Estimation of S_α: Puffer-preconditioned LASSO.* A natural place to start to estimate S_α would be to apply LASSO to (3.2). However, this approach is not sign consistent.[13] Sign consistency fails because the matrix X fails an irrepresentability criterion, a necessary condition for consistent estimation (Zhao and Yu, 2006). Irrepresentability bounds the acceptable correlations in the design matrix. Intuitively, it requires that regressions of the variables that are not in the support on those that are have small coefficients. Formally, the L_1 norm on those coefficients must be less than 1. Otherwise, an irrelevant variable is "representable" by relevant variables, which makes LASSO erroneously select it with non-zero probability, regardless of sample size.

The smart pooling specification (3.2) fails irrepresentability by construction because of correlation within some treatment profiles. For example the smart pooling covariate where all the M arms are "on" with highest intensity, i.e., X_k for $k = (R-1, ..., R-1)$, is representable by other covariates. A simulation in Appendix C provides a proof by example, in particular, in a computationally reasonable range of R and M.

In our case, the failure of irrepresentability is because of the way in which the treatments are represented, but this does not preclude transforming the data into a form that is consistently estimable. In particular, to estimate S_α consistently, we appeal to a technique from Jia and Rohe (2015). The procedure is analogous to weighted least squares, where the weighting is what the authors call a Puffer transformation. It eliminates the correlation in the design matrix and recovers irrepresentability. It does so at the expense of inflating variance in the error, but the efficiency loss is the cost of being able to implement LASSO. We demonstrate that this cost is not too high owing

_____

[13]Sign consistency refers to estimating the signed support. It is technically a slightly stronger condition than support consistency, but this is standard in the literature.

to the structure of the cross-randomized RCT in the sense that the procedure delivers consistent estimates of the support.

The weighting is constructed as follows. Let $X = UDV'$ denote the singular value decomposition of X. The *Puffer transformation*[14] is $F = UD^{-1}U'$ and the regression is

$$(3.3) \qquad\qquad FY = FX\alpha + F\epsilon$$

where if the original $\epsilon \sim \mathcal{N}(0, \sigma^2)$ per Assumption 2, $F\epsilon \sim \mathcal{N}(0, \sigma^2 UD^{-2}U^T)$. As Jia and Rohe (2015) note, the new matrix FX satisfies irrepresentability because it is orthonormal: $(FX)'(FX) = I$, which is sufficient (Jia and Rohe (2015), Bickel et al. (2009)). The relevant and irrelevant variables do not exhibit excess correlation by construction.

To understand the reason it works recall that the orthonormal matrices U and V' can be viewed as rotations of $\mathbb{R}^n$ and $\mathbb{R}^K$ respectively, while D rescales the principal components of X. D is the diagonal matrix of singular values, ordered from largest to smallest. The transformation F preserves the rotational elements of X without the rescaling D. Thus, FX has a singular value decomposition $FX = UV'$. The new singular values are all set to unity—the transformation normalizes or "cancels out" D.

Now, the i-th singular value of X captures the residual variance of X explained by the i-th principal component of X after partialing out the variance explained by the first $i-1$ principal components. When there is correlation inside X, less than K principal components effectively explain the variation in X, and so the later (and therefore lower) singular values shrink toward zero. By normalizing the singular values to unity, the Puffer transformation F effectively inflates the lowest singular values of X so that each principal component of the transformed FX explains the variance in FX equally. In this sense, FX is de-correlated, and for $K < n$, mechanically irrepresentable. The cost is that this effective re-weighting of the data also amplifies the noise associated with the observations that would have had the lowest original singular values.[15,16]

The reason why LASSO is particularly amenable to the Puffer transformation in our specific setting of the cross-randomized experiment with varying dosages is that the smart pooling design matrices are highly structured. In particular, the assignment probabilities to the various unique treatments are given, and as a result, the correlations within the X matrix are bounded away from 1. This has the implication that the minimal

---

[14]This is named after the pufferfish as it inflates an otherwise ellipsoid loss function contour set to be more spherical, much like the fish.

[15]As Jia and Rohe (2015) point out, from the perspective of LASSO, if the amplification is too great it "can overwhelm the benefits [of the transformation]." Depending on the assumptions of the data generating process, it can hinder LASSO efficiency in the finite sample at best and destroys LASSO sign consistency, at worst.

[16]In $K > n$ cases–not studied here and not having a full characterization in the literature–even irrepresentability is not immediate and the theory developed is only for special cases (a uniform distribution on the Stiefel manifold) and a collection of empirically relevant simulations (Jia and Rohe, 2015).

singular value is bounded below so that under standard assumptions on data generation, LASSO selection is sign consistent. While this is guaranteed for a sample size that grows in fixed K, the more important test is whether it works when K goes up with n; we need to show that the Puffer transformation does not destroy the sign consistency of LASSO selection as the minimal singular value of X goes to zero as a function of K. We show that the Puffer transformation continues to work even when K grows without bound with n (but K is still less than n), subject to the limit on its growth rate captured by Assumption 1. Lemma A.1 bounds the rate at which the minimal singular value of X can go to zero as a function of K. Proposition 3.1 below relies on this lemma to then prove that the Puffer transformation insures irrepresentability and consistent estimation by LASSO in our context.[17]

ASSUMPTION 4. *A sequence $\lambda_n \geq 0$ is taken such that $\lambda_n \to 0$ and $\lambda_n^2 n^{1-2\gamma} = \omega(\log(n))$.*

PROPOSITION 3.1. *Let $\hat{\alpha}$ be the estimator of* (3.3) *by LASSO:*

$$\hat{\alpha} := \underset{a \in \mathbb{R}^K}{\operatorname{argmin}} \|Fy - FXa\|_2^2 + \lambda_n \|a\|_1.$$

Assume 1, 2, 3, and 4. Then $\mathrm{P}(\mathrm{sign}(\hat{\alpha}) = \mathrm{sign}(\alpha)) \to 1$.

Thus, with probability tending to one, using LASSO we correctly recover the support S_α which tells us which marginal differences across intensities are relevant and therefore how to prune and pool for post-estimation.

3.3. **Post-Estimation: Policy Effects and the Effect of the Best Policy.** Having estimated the support S_α wpa1, we return to our original goals: (1) estimating the effects of the relevant policies; (2) estimating the effect of the best policy.

3.3.1. *Consistent Estimation of Policy Effects.* The first step is to estimate policy effects. We do this in the usual post-LASSO way, mapping back to the unique treatment specification with a unique dummy for each relevant policy. This is similar to the unique policy specification (3.1) except that (1) treatment combination variables may be pooled (union of two or more variables in T), and (2) $|S_{pool}| < K$ (reflecting the pruning), where S_{pool} is the collection of pooled policies inverted from estimating S_α. Let T_{pool} be the unique treatment variables from the pruned set S_{pool} of pooled policies.

We are interested in the regression

(3.4) $$y = \hat{T}_{pool}\eta + \epsilon$$

---

[17]In practice Rohe (2014) recommends a variation of the Puffer transformation called Puffer$_N$ that better accommodates heteroskedasticity induced by the transformation, which we omit from the exposition for parsimony but use in our estimation.

We can proceed to post-model selection estimation with OLS following Belloni and Chernozhukov (2013).[18] Let $\hat{\eta}$ be the post-LASSO estimator, so OLS on the estimated support.

COROLLARY **3.1.** *Assume 1, 2, 3, and 4. Then $\hat{\eta} \to_p \eta$.*

3.3.2. *The Effect of the Best Policy.* Another policy relevant issue is the recommendation of a "best policy" together with an estimate of the effect of the best policy. To select the best policy, we scan the post-LASSO estimates of policies in $\hat{S}_{pool}$. While intuitively $\max_{k \in \hat{S}_{pool}} \hat{\eta}_k$ appears to be the effect of the best policy, Andrews et al. (2019) points out that this suffers from a winner's curse in the finite sample. There are two reasons why a policy may be deemed best. First, it may have the highest effect. Second, it may have drawn higher random shocks. As a result, the expected effect of the best policy using the naive OLS (in this case post-LASSO) estimator will be upward biased.

The estimation strategy in Andrews et al. (2019) corrects the conventional estimate $\hat{\eta}_{\hat{k}}$ ex-post to construct an approximately median unbiased estimator (for the policy chosen to be the best) and appropriate confidence intervals with desired coverage. Loosely speaking, the winner's curse adjusted estimator takes the estimated best policy and downward adjusts the estimate based on the effect of the second-best policy. Our smart pooling and pruning procedure helps avoid needlessly conservative estimates from this correction.

Specifically, from the corresponding estimated set of pooled policies $\hat{S}_{pool}$, select $\hat{k} = \text{argmax}_{k \in \hat{S}_{pool}} \hat{\eta}_k$ based on post-LASSO estimates, which have an asymptotically normal distribution under our assumptions (Belloni, Chernozhukov, Chetverikov, and Wei, 2018) and therefore, provide a starting point for the application of the Andrews et al. (2019) technique wpa1. For this $\hat{k}$, construct the hybrid estimator $\hat{\eta}_{\hat{k}}^{\text{hyb}}$ described in Section 5.2 of Andrews et al. (2019) with nominal size α and median bias tolerance $\beta/2$. Let $\xi_{\min} = \xi_{\min}(\frac{X}{\sqrt{n}})$ be the minimum singular value of the $\sqrt{n}$-normalized design matrix of smart pooling variables.

COROLLARY **3.2.** *Assume 1, 2, 3, and 4. Apply the Puffer-preconditioned LASSO to estimate the smart pooling specification* (3.4). *Then with probability at least*

$$1 - 2K \exp\Big(-\frac{n\lambda^2 \xi_{\min}^2}{2\sigma^2} \Big),$$

$\hat{\eta}_{\hat{k}}^{hyb}$ is approximately median unbiased (with absolute median bias bounded by $\frac{\beta}{2}$) with confidence intervals with coverage $\frac{1-\alpha}{1-\beta}$, conditional on $\eta_{\hat{k}}$ falling within a simultaneous confidence interval of level $1 - \beta$.

---

[18]The possible redundancy in S_{pool} amounts to including a few zero efficacy policies which do not affect the theoretical validity of this estimator.

A summary of the overall procedure is presented in Algorithm 1.

Algorithm 1: Estimating Treatment Effects by Smart Pooling and Pruning

    (1) Given treatment assignment matrix T, calculate the treatment profile and marginal dosage intensity matrix X.

    (2) Estimate $\hat{S}_\alpha := \{k : \ \hat{\alpha}_k > 0\}$ by estimating (3.2) through a Puffer transformed LASSO.

    (3) Calculate pooled and pruned support $\hat{S}_{pool}$ from $\hat{S}_\alpha$ using Algorithm 2 in Appendix B.

    (4) Estimate pooled and pruned treatment effects of unique (relevant) policies, η, using regression (3.4).

    (5) For the best policy $\hat{S}_{pool}$, select $\hat{k} = \mathrm{argmax}_{k \in \hat{S}_{pool}} \hat{\eta}_k$, construct the hybrid estimator $\hat{\eta}_{\hat{k}}^{\mathrm{hyb}}$ with nominal size α and median bias tolerance $\beta/2$.

3.4. **Simulations.** In Appendix C, we conduct several simulations to both demonstrate the effectiveness of the smart pooling and pruning estimator (in Algorithm 1) and compare it to various natural alternatives to demonstrate the value of each step.

We begin by looking at the first step: whether the estimator consistently recovers the support S_α of the specification (3.2). We find that the smart pooling and pruning estimator does consistently recover the support. In contrast, applying a naive LASSO to this equation yields an inconsistent estimate of the support and, further, the support accuracy appears to be bounded from above irrespective of the number of observations (in this case 75%). This is because of the failure of irrepresentability, which is a necessary condition for LASSO to be sign consistent, and without the Puffer transformation, the procedure fails.

Next, we turn to the second step: identification and estimation of the effect of the best policy. We show that our estimator consistently recovers the best policy—this is clearly true by the above—but also has a near-perfect rate of selecting the best policy even with few observations. The smart pooling and pruning procedure estimates (3.4) while the alternative is using a naive LASSO on (3.1)—the regression with all unique policy combinations. The distinction is that when studying unique policies, the former has already pooled dosages that are not distinct in addition to pruning irrelevant ones, whereas the latter only does the pruning. Given these estimates, we apply the winner's curse adjustment and look at the estimates of the effect of the best policy. Once again we show that the smart pooling and pruning procedure uniformly outperforms the naive LASSO on the unique policy specification: for all observation levels the MSE of the effect of the best policy is lower for our estimator.

4. Results

4.1. **Identifying effective policies.**

4.1.1. *Method.* We adapt the smart pooling and pruning specification (3.2) for our case. The interventions "information hubs," "slope," "flat," and "SMS" are found in two intensities.[19] The smart pooling specification therefore looks like

$$
\begin{aligned}
y_{dsvt} = \alpha_0 &+ \alpha_{SMS}\text{SMS}_s + \alpha_{H,SMS}\text{High SMS}_s \\
&+ \alpha_{Slope}\text{Slope}_s + \alpha_{H,Slope}\text{High Slope}_s + \alpha_{Flat}\text{Flat}_s + \alpha_{H,Flat}\text{High Flat}_s \\
&+ \alpha_R\text{Random}_v + \alpha_H\text{Info Hub (All)}_v + \alpha_T\text{Trust}_v + \alpha_{TH}\text{Trusted Info Hub}_v \\
&+ \alpha_X'X_{sv} + v_{dt} + \epsilon_{dsvt},
\end{aligned}
$$

where we have explicitly listed some of the variables in "single arm" treatment profiles. X_{sv} is a vector of the remaining 64 smart pooling variables in "multiple arm" treatment profiles, and v_{dt} is a set of district-time dummies.

Our estimation follows the recommended implementation in Rohe (2014), which uses a sequential backward elimination version of LASSO (variables with p-values above some threshold are progressively deselected) on the Puffer$_N$ transformed variables (this aids in correcting for the heteroskedasticity induced by the Puffer transformation). We select penalties λ for both regressions (number of immunizations and immunizations per dollar) to minimize a Type I error, which is particularly important to avoid in the case of policy implementation.[20] This is because it is extremely problematic to have a government introduce a large policy based on a false positive.

This gives $\hat{S}_\alpha$, an estimate of the true support set S_α of the smart pooling specification. We then generate a use of unique pooled policy set $\hat{S}_{pool}$ (following the procedure we outline in Algorithm 2 in Appendix B). Next, we run the pooled specification (3.4) to obtain post-LASSO estimates $\hat{\eta}$ of the pooled policies as well as $\hat{\eta}_{\hat{k}}^{\text{hyb}}$, the winner's curse adjusted estimate of the best policy.

4.1.2. *Results.* The results are presented in Figures 3 and 4. Figure 3 presents the post-LASSO estimates where the outcome variable is the number of measles vaccines per month in the village. Figure 4 presents the post-LASSO estimates where the outcome

---

[19]In the case of information hubs, "trust" adds intensity to the information hub.

[20] Rohe (2014) notes a bijection between a backwards elimination procedure based on using Type I error thresholds and the penalty in LASSO. We take $\lambda = 0.48$ and $\lambda = 0.0014$ for the number of immunizations and immunizations per dollar outcomes, respectively. Both of these choices map to the same Type I error value ($p = 5 \times 10^{-13}$) used in the backwards elimination implementation of LASSO selected to essentially eliminate false positives. Appendix D repeats the exercise for a number of alternative penalties.

variable is the number of measles vaccines per dollar spent. In each, a relatively small subset of policies is selected as part of $\hat{S}_{pool}$ out of the universe of 75 granular policies (16% of the possible options in Figure 3 and 35% in Figure 4).

In Figure 3, two of the four selected pooled policies are estimated to do significantly better than control: information hubs seeding with sloped incentives (of both low and high intensities) and SMS reminders (of both 33% and 66% saturation) are estimated to increase the number of immunizations by 55% relative to control ($p = 0.001$), while trusted seeds with high-sloped incentives and SMS reminders (of both saturation levels) are estimated to increase immunizations by 44% relative to control ($p = 0.009$). The policy of high sloped incentives with SMS reminders has a positive effect but is noisy, while the policy of trusted information hubs with sloped incentives (of any intensity) and SMS reminders (either level of saturation) is solidly zero ($p = 0.515$). The selection of this last policy in $\hat{S}_{pool}$ is an example of $\hat{S}_{pool}$ choosing a superset of the true support S_β; this particular policy shares the same treatment profile as the best policy in this data but has zero impact. Finally, while incentives help, a very robust result is that flat incentives never emerge as an effective policy, in any combination (this is consistent with the fact that they did not have a positive effect, on average).

These two effective policies increase the number of immunizations, relative to the status quo, at the cost of a greater cost for each immunization (compared to standard policy). These policies induce 36.0 immunizations per village per month per \$1,000 allocation (as compared with 43.6 immunizations per village per month in control). The reason is that that the gains from having incentives in terms of immunization rates is smaller than the increase in costs (especially because the incentives must be paid to all the infra-marginal parents). Two things are worth noting to qualify those results, however. First, in (Chernozhukov et al., 2018), we show that in the places where the full package treatment is predicted to be the most effective (which tends to be the places with low immunization), the number of immunizations per dollar spent is not statistically different in treatment and control villages. Second, immunization is so cost effective, that this relatively small increase in the cost of immunization may still mean a much more cost-effective use of funds than the next best use of dollars on policies to fight childhood disease (Ozawa et al., 2012).

Nevertheless, a government may be interested in the most cost-effective policy, if they have a given budget for immunization. We turn to policy cost effectiveness in Figure 4. The most cost-effective policy (and the only policy that reduces per immunization cost) compared to control is the combination of information hub seeding (trusted or not) with SMS reminders (at both 33% or 66% saturation) and no incentives, which leads to a 9.1% increase in vaccinations per dollar ($p = 0.000$).

4.2. **Estimating the Impact of the Best policy.** To estimate the impact of the best policy, we first select the best policy from $\hat{S}_{pool}$ based on the post-LASSO estimate. Then, we attenuate it using the hybrid estimator with $\alpha = 0.05$ and $\beta = \frac{\alpha}{10} = 0.005$, which this is the value used by Andrews et al. (2019) in their simulations. The hybrid confidence interval has the following interpretation: conditional on policy effects falling within a 99.5% simultaneous confidence interval, the hybrid confidence interval around the best policy has at least 95% coverage.

Table 1 presents the results. In column 1, the outcome variable is the number of measles vaccines given every month in a given village. We find that for the best policy in the sample (information hub seeds with sloped incentives at any level and SMS reminders at any saturation) the hybrid estimated best policy effect relative to control is 3.26 with a 95% hybrid confidence interval of [0.032,6.25]. This is lower than the original post-LASSO estimated effect of 4.02. The attenuation is owing to a second best policy (trusted seeds with high sloped incentives with SMS reminders at any saturation) "chasing" the best policy estimate somewhat closely.[21] Nevertheless, even accounting for winner's curse through the attenuated estimates and the adjusted confidence intervals, the hybrid estimates still reject the null. Thus, the conclusion is that accounting for winner's curse, this policy increases immunizations by 44% relative to control.

While policymakers may chose this policy if they are willing to bear a higher cost to increase immunization, there may be settings where cost effectiveness is an important consideration. In column 2, the outcome variable is the number of vaccinations per dollar. Accounting for winner's curse through hybrid estimation, for the best policy of information hubs (all variants) and SMS reminders (any saturation level), the hybrid estimated best policy effect relative to control is 0.004 with a 95% hybrid confidence interval of [0.003,0.004]. Notably, this appears almost unchanged from the naive post-LASSO. This is because no other pooled policy with positive effect is "chasing" the best policy in the sample; the second-best policy is the control (status quo), which is sufficiently separated from the best policy so as to have an insignificant adjustment for winner's curse. Thus, adjusting for winner's curse, this policy increases the immunizations per dollar by 9.1% relative to control.

One concern with these estimates is that they are sensitive to the implied LASSO penalty λ chosen. To check the robustness of our results, we consider alternative values of the λ. Note that as the λ decreases, we incur increasing probability of Type I error (false non-zero estimates) in model selection. This error can manifest in two ways in model selection: (a) spurious policies can be selected as the second-best policy, or (b) spurious policies can be selected as the best policy. Of these (a) is less serious in that it

---

[21]The increased attenuation from a more closely competing second-best policy emerges from the formulas for conditional inference given in Section 3 of Andrews et al. (2019).

may only make our winner's curse estimates more conservative. Case (b) of a fluke best policy is the more serious error. Appendix D presents our results for a number of less stringent penalties (allowing for greater Type I error). For the number of immunizations per dollar, the best policy and associated winner's curse adjusted-effect estimates are extremely robust. With decreasing λ, we find that the best selected policy is always the same, and while the winner's curse estimates do go down, following the possibility (a) mentioned above, for almost all the λ except the smallest $\lambda = 0.00045$, the hybrid estimator consistently rejects the null ($p < 0.05$). For the number of immunizations, however, we do see sensitivity to model selection parameters to the extent that a granular policy of no seeds combined with high sloped incentives and low SMS reminders emerges as the best policy for $\lambda \leq 0.42$. Although we cannot be sure, we have reason to believe that this is a spurious best policy arising from model selection error of type (b)–we observe that when this policy first emerges, the hybrid estimator (and even the naive OLS estimator) already rejects the null ($p < 0.05$).

5. Conclusion

While immunization is one of the most effective and cost-effective method to prevent illness, disability, and disease, millions of children continue to go without it every year. The COVID-19 epidemic risks making the situation even worse: during the pandemic, vaccine coverage has dipped to levels not seen since the 1990s (Bill and Melinda Gates, 2020). Swift policy action will be critical to ensure that this dip is temporary, and children who missed immunizations during the pandemic get covered soon.

To study effective policies to encourage immunization, we conducted a large policy experiment in 2360 villages in India covering 295,038 children. Strategies available to policymakers include conventional instruments such as reminders and incentives, each of which can be designed in a number of ways (e.g., with different coverage rates, levels of incentives, shape of the incentive curve), as well as a new policy, derived from insights from social network analysis (ambassadors from the community to encourage immunization take-up). Again, there are several variants such as recruiting information hubs, trusted individuals, individuals in the intersection of both, and random members of society. All told, our experiment covers 75 policies which exhausts all combinations of these arms. That is, we look at every possible policy combination available to the policymaker contemplating choosing one of these instruments with the goal of scaling it up.

We develop a blueprint, a *smart pooling and pruning* procedure, to perform policy analysis in this context, in a data-driven manner. First, we assume that only a sparse set of policy combinations meaningfully affects the outcome of interest (here the number

of immunizations or the number of immunizations per dollar). Furthermore, there may not necessarily be appreciable differences in the effects of variants in policies differing only in their intensity profiles. By applying the appropriate transformation to represent the data in a manner amenable to pooling and pruning, and then using the Puffer transformation, we are able to consistently recover the collection of relevant policies and obtain consistent estimates of these pooled policies (and confidence intervals) following the post-LASSO procedure described in Chernozhukov et al. (2015). This allows us to identify which policies matter and what kind of flexibility it affords the policymaker. For example, we can see in a data-driven way if high and low incentives tend to have the same effect, which would allow the policymaker to choose lower and cheaper incentives.

Second, we estimate the impact of the best policy in the sense of maximizing either the number of vaccines or number of vaccines per dollar. To do so requires overcoming the winner's curse. Because the policy deemed to be the best is one that, by definition, in the sample has to have an effect bigger than all the other policies, it is more likely to have benefitted from a larger random shock as well and therefore the estimate of its effect will be upward biased. We use the techniques in Andrews et al. (2019) to overcome this and construct (nearly) median unbiased estimates for the number of vaccines (or number of vaccines per dollar) for the best policy. We estimate the best policy in either case to be one where the policymaker uses information hubs and sends SMS reminders to 33% of the community. If we give up on cost effectiveness, we should add to these two sloped incentives at the low amount. From the cost effectiveness perspective simply using information hubs and a low saturation of SMS reminders emerges as the best policy, and is more cost effective than the status quo of no policy.

One possible interpretation, especially given that the most effective ambassador is the information hub (recall this is the person best placed to circulate information according to the community) is that the ambassador ensures widespread diffusion about the presence of the incentives (in incentives villages) and is able to explain and de-mystify the content of the personalized reminders (in SMS villages, even without incentives).[22] In either case, the ambassador has something quite concrete to discuss with the people they talk to (which is not the case in villages without SMS or incentives, perhaps explaining why they have no effect in this case). All told, this suggests that the social network can be used in creative and cost-effective ways to amplify the effect of other policies.

There are three main takeaways. First, from the perspective of public health policy, standard tools that have previously been championed (e.g., SMS reminders) may not be particularly effective and others such as high sloped incentives may not be cost-effective

---

[22]Banerjee, Breza, Chandrasekhar, and Golub (2018) find that complex information circulated to influential people can lead to better understanding and action than complete broadcasting, because people may feel insecure asking questions.

at large scale, such as at a state or national level (although Chernozhukov et al. (2018) find it may be cost effective in pockets of low immunization villages, where it is predicted to be most effective). But using such instruments in combination, particularly with policy insights from social network analysis, yields effective, and cost-effective policies.

Second, the results are consistent with recent literature pointing to the importance of leveraging networks to diffuse information in a variety of economic contexts. Here, this principle suggests that policymakers can benefit tremendously from identifying information hubs to accelerate take-up. These perspectives typically are not in a policymaker's toolkit, but the literature increasingly points to the necessity to incorporate such lessons.

Third, there is a temptation when doing a policy experiment to pare down the number of treatments one is willing to evaluate for power concerns and because selective ex-post pooling can cause biases. However, this has the downside that it requires the policymaker to be somewhat sure of the set of effective policies in the first place. But that assumes the conclusion: if one could already pick the best four policies out of 75 feasible ones, testing out policies may be second order. On the other hand, if there was genuine uncertainty about what works, which is how the problem was presented to us in this case, ex ante paring down the options may get us the wrong answer. In particular, the suggestion of avoiding all interactions in this setting (made in Muralidharan et al. (2019)), would have led to the conclusion that nothing is effective.

To manage the rapidly increasing number of treatment bundles policymakers may consider, we suggest instead a data-driven approach wherein under a natural assumption that most policies are unlikely to be effective, we can use machine learning to identify the sparse set of policies that meaningfully affect the outcome of interest. Given this, we can estimate the effect of the best policy in terms of the outcome of interest, accounting for the winner's curse. This is a straightforward procedure and one that could easily be specified in a pre-analysis plan. The researcher can gain power by incorporating prior knowledge of the policies that are likely to "pool" together (in this instance, these are doses of a treatment) in the smart pooling specification, without making a prior assumption that they have to pool. This structure can easily be pre-specified, and beyond that the researcher does not need to take a stance on the possible effects of a number of interactions that are very hard to predict in advance.

References

Akbarpour, M., S. Malladi, and A. Saberi (2017): "Diffusion, Seeding, and the Value of Network Information," Available at SSRN: https://ssrn.com/abstract=3062830. 6

ALATAS, V., A. G. CHANDRASEKHAR, M. MOBIUS, B. A. OLKEN, AND C. PAL-
    ADINES (2019): "When Celebrities Speak: A Nationwide Twitter Experiment Promot-
    ing Vaccination In Indonesia," Tech. rep., National Bureau of Economic Research. 4

ALSAN, M. (2015): "The effect of the tsetse fly on African development," *American
    Economic Review*, 105, 382–410. 1

ANDREWS, I., T. KITAGAWA, AND A. MCCLOSKEY (2019): "Inference on winners,"
    Tech. rep., National Bureau of Economic Research. 1, 3.1, 3.3.2, 4.2, 21, 5, **??**, A,
    C.2.2, C.3, C.4

ARAL, S. AND D. WALKER (2012): "Creating social contagion through viral product
    design: A randomized trial of peer influence in networks," *Management Science.* 2.2.3

BANERJEE, A., E. BREZA, A. G. CHANDRASEKHAR, AND B. GOLUB (2018): "When
    Less is More: Experimental Evidence on Information Delivery During India's Demon-
    etization," Tech. rep., National Bureau of Economic Research. 2.2.2, 22

BANERJEE, A., A. CHANDRASEKHAR, E. DUFLO, AND M. O. JACKSON (2013):
    "Diffusion of Microfinance," *Science*, 341, DOI: 10.1126/science.1236498, July 26 2013.
    6, 2.2.3

BANERJEE, A., A. G. CHANDRASEKHAR, E. DUFLO, AND M. O. JACKSON (2019):
    "Using Gossips to Spread Information: Theory and Evidence from Two Randomized
    Controlled Trials," *The Review of Economic Studies.* 4, 1, 2.2.3, 2.5

BANERJEE, A., A. CHANDRASHEKHAR, E. DUFLO, S. DALPATH, J. FLORETTA,
    M. JACKSON, H. KANNAN, A. SCHRIMPF, AND M. SHRESTHA (2021): "Evaluating
    the impact of interventions to improve full immunisation rates in Haryana, India," .
    2.1

BANERJEE, A. V., E. DUFLO, R. GLENNERSTER, AND D. KOTHARI (2010): "Improv-
    ing immunisation coverage in rural India: clustered randomised controlled evaluation
    of immunisation campaigns with and without incentives," *Bmj*, 340, c2220. 1, 2.2.1

BASSANI, D. G., P. ARORA, K. WAZNY, M. F. GAFFEY, L. LENTERS, AND Z. A.
    BHUTTA (2013): "Financial incentives and coverage of child health interventions: a
    systematic review and meta-analysis," *BMC Public Health*, 13, S30. 1, 2.2.1

BEAMAN, L., A. BENYISHAY, J. MAGRUDER, AND A. M. MOBARAK (2018): "Can
    network theory-based targeting increase technology adoption?" Tech. rep., National
    Bureau of Economic Research. 2.2.3

BELLONI, A. AND V. CHERNOZHUKOV (2013): "Least squares after model selection in
    high-dimensional sparse models," *Bernoulli*, 19, 521–547. 1, 3.1, 3.3.1, A

BELLONI, A., V. CHERNOZHUKOV, D. CHETVERIKOV, AND Y. WEI (2018): "Uni-
    formly valid post-regularization confidence regions for many functional parameters in
    z-estimation framework," *Annals of statistics*, 46, 3643. 3.3.2, A

BICKEL, P. J., Y. RITOV, A. B. TSYBAKOV, ET AL. (2009): "Simultaneous analysis of Lasso and Dantzig selector," *The Annals of statistics*, 37, 1705–1732. 3.2.2

BILL AND F. MELINDA GATES (2020): "COVID-19: A Global Perspective: 2020 Goalkeepers Report," . 5

BLOCH, FRANCIS AND, M. O. AND P. TEBALDI (2016): "Centrality Measures in Networks," *http://ssrn.com/abstract=2749124*. 6

CHERNOZHUKOV, V., M. DEMIRER, E. DUFLO, AND I. FERNANDEZ-VAL (2018): "Generic machine learning inference on heterogenous treatment effects in randomized experiments," Tech. rep., National Bureau of Economic Research. 4.1.2, 5

CHERNOZHUKOV, V., C. HANSEN, AND M. SPINDLER (2015): "Post-selection and post-regularization inference in linear models with many controls and instruments," *American Economic Review*, 105, 486–90. 5, C.1, 28

DLHS (2013): "District Level Household and Facility Survey-4," . 2.1

DOMEK, G. J., I. L. CONTRERAS-ROLDAN, S. T. OâĂŹLeary, S. BULL, A. FURNISS, A. KEMPE, AND E. J. ASTURIAS (2016): "SMS text message reminders to improve infant vaccination coverage in Guatemala: A pilot randomized controlled trial," *Vaccine*, 34, 2437–2443. 3

GIBSON, D. G., B. OCHIENG, E. W. KAGUCIA, J. WERE, K. HAYFORD, L. H. MOULTON, O. S. LEVINE, F. ODHIAMBO, K. L. O'BRIEN, AND D. R. FEIKIN (2017): "Mobile phone-delivered reminders and incentives to improve childhood immunisation coverage and timeliness in Kenya (M-SIMU): a cluster randomised controlled trial," *The Lancet Global Health*, 5, e428–e438. 1, 2.2.1

HINZ, O., B. SKIERA, C. BARROT, AND J. U. BECKER (2011): "Seeding Strategies for Viral Marketing: An Empirical Comparison," *Journal of Marketing*, 75:6, 55–71. 6

IYENGAR, R., C. V. DEN BULTE, AND T. W. VALENTE (2010): "Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science*, 30:2, 195–212. 6

JACKSON, M. O. (2008): "Average Distance, Diameter, and Clustering in Social Networks with Homophily," in *the Proceedings of the Workshop in Internet and Network Economics (WINE 2008), Lecture Notes in Computer Science, also: arXiv:0810.2603v1*, ed. by C. Papadimitriou and S. Zhang, Springer-Verlag, Berlin Heidelberg. 6

——— (2017): "A Typology of Social Capital and Associated Network Measures," *SSRN http://ssrn.com/abstract=3073496*. 6

JACKSON, M. O. AND L. YARIV (2011): "Diffusion, strategic interaction, and social structure," *Handbook of Social Economics, San Diego: North Holland, edited by Benhabib, J. and Bisin, A. and Jackson, M.O.* 6

JIA, J. AND K. ROHE (2015): "Preconditioning the Lasso for sign consistency," *Electronic Journal of Statistics*, 9, 1150–1172. 1, 3.1, 11, 3.2.2, 3.2.2, 15, 16, A

JOHRI, M., M. C. PÉREZ, C. ARSENAULT, J. K. SHARMA, N. P. PAI, S. PAHWA, AND M.-P. SYLVESTRE (2015): "Strategies to increase the demand for childhood vaccination in low-and middle-income countries: a systematic review and meta-analysis," *Bulletin of the World Health Organization*, 93, 339–346. 1

KARING, A. (2018): "Social Signaling and Childhood Immunization: A Field Experiment in Sierra Leone," *University of California, Berkeley.* 2

KATONA, Z., P. P. ZUBCSEK, AND M. SARVARY (2011): "Network Effects and Personal Influences: The Diffusion of an Online Social Network," *Journal of Marketing Research*, 48:3, 425–443. 6

KATZ, E. AND P. LAZARSFELD (1955): *Personal influence: The part played by people in the flow of mass communication*, Free Press, Glencoe, IL. 2.2.3

KEMPE, D., J. KLEINBERG, AND E. TARDOS (2003): "Maximizing the Spread of Influence through a Social Network," *Proc. 9th Intl. Conf. on Knowledge Discovery and Data Mining*, 137âĂŞ146. 6

KRACKHARDT, D. (1996): "Strucutral Leverage in Marketing," in *Networks in Marketing*, ed. by D. Iacobucci, Sage, Thousand Oaks, 50–59. 6

MARTINEZ-BRAVO, M. AND A. STEGMANN (2018): "In Vaccines We Trust? The Effects of the CIA's Vaccine Ruse on Immunization in Pakistan," Tech. rep., CEMFI. 1

MCKENZIE, D. (2019): "Be careful with inference from 2-by-2 experiments and ther cross-cutting designs," [Online; accessed Octobrer 19, 2019]. 1

MEKONNEN, Z. A., K. A. GELAYE, M. C. WERE, K. D. GASHU, AND B. C. TILAHUN (2019): "Effect of mobile text message reminders on routine childhood vaccination: a systematic review and meta-analysis," *Systematic reviews*, 8, 1–14. 2.2.2

MURALIDHARAN, K., M. ROMERO, AND K. WUTHRICH (2019): "Factorial designs, model selection, and (incorrect) inference in randomized experiments," . 1, 3.2.1, 5

NFHS (2016): "National Family Health Survey-4 State Fact Sheet for Haryana," . 2.1

OYO-ITA, A., C. S. WIYSONGE, C. ORINGANJE, C. E. NWACHUKWU, O. ODUWOLE, AND M. M. MEREMIKWU (2016): "Interventions for improving coverage of childhood immunisation in low-and middle-income countries," *Cochrane Database of Systematic Reviews.* 1

Ozawa, S., A. Mirelman, M. L. Stack, D. G. Walker, and O. S. Levine (2012): "Cost-effectiveness and economic benefits of vaccines in low-and middle-income countries: a systematic review," *Vaccine*, 31, 96–108. 4.1.2

Regan, A. K., L. Bloomfield, I. Peters, and P. V. Effler (2017): "Randomized controlled trial of text message reminders for increasing influenza vaccination," *The Annals of Family Medicine*, 15, 507–514. 3

Rogers, E. (1995): *Diffusion of Innovations*, Free Press. 6

Rohe, K. (2014): "A note relating ridge regression and ols p-values to preconditioned sparse penalized regression," *arXiv preprint arXiv:1411.7405*. 1, 3.1, 17, 4.1.1, 20

Sugerman, D., A. Barskey, M. Delea, I. Ortega-Sanchez, D. Bi, K. Ralston, P. Rota, K. Waters-Montijo, and C. Lebaron (2010): "Measles outbreak in a highly vaccinated population, San Diego, 2008: role of the intentionally undervaccinated." *Pediatrics*, 125, 747–755. 1

Uddin, M. J., M. Shamsuzzaman, L. Horng, A. Labrique, L. Vasudevan, K. Zeller, M. Chowdhury, C. P. Larson, D. Bishai, and N. Alam (2016): "Use of mobile phones for improving vaccination coverage among children living in rural hard-to-reach areas and urban streets of Bangladesh," *Vaccine*, 34, 276–283. 3

UNICEF and WHO (2019): "Progress and challenges with achieving universal immunization coverage: 2018 estimates of immunization coverage WHO," . 1

user1551 (HTTPS://MATH.STACKEXCHANGE.COM/USERS/1551/USER1551) (2017): "Find the uniform lower bound of the smallest eigenvalue of a certain matrix," Mathematics Stack Exchange, https://math.stackexchange.com/q/2438656 (version: 2017-09-21). 25

Wakadha, H., S. Chandir, E. V. Were, A. Rubin, D. Obor, O. S. Levine, D. G. Gibson, F. Odhiambo, K. F. Laserson, and D. R. Feikin (2013): "The feasibility of using mobile-phone based SMS reminders and conditional cash transfers to improve timely immunization in rural Kenya," *Vaccine*, 31, 987–993. 1, 3

WHO (2019): "Situation Analysis of Immunization Expenditure," . 1

Zhao, P. and B. Yu (2006): "On model selection consistency of Lasso," *Journal of Machine learning research*, 7, 2541–2563. 1, 11, 3.2.2

FIGURES
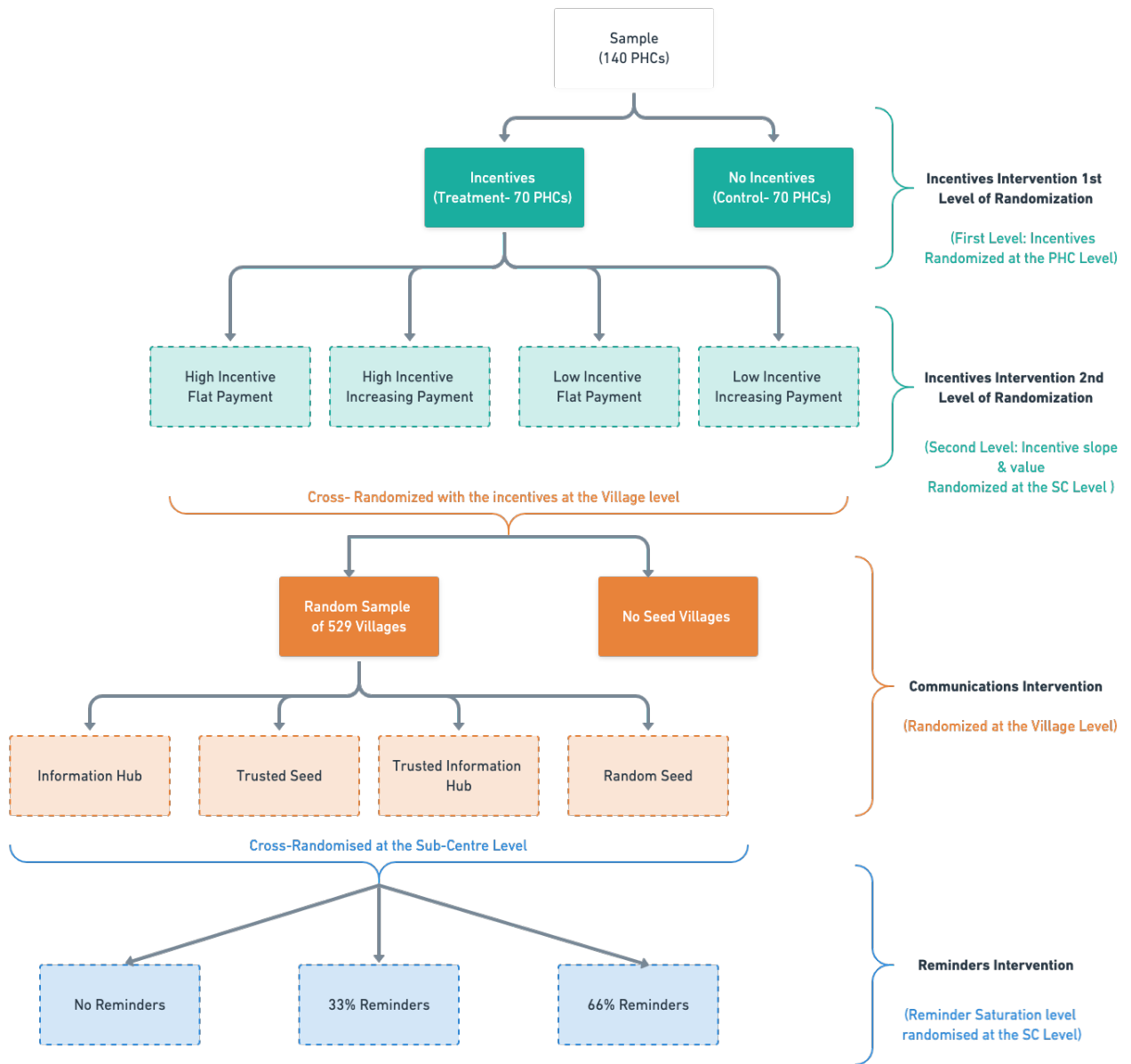
FIGURE 1. Experimental Design
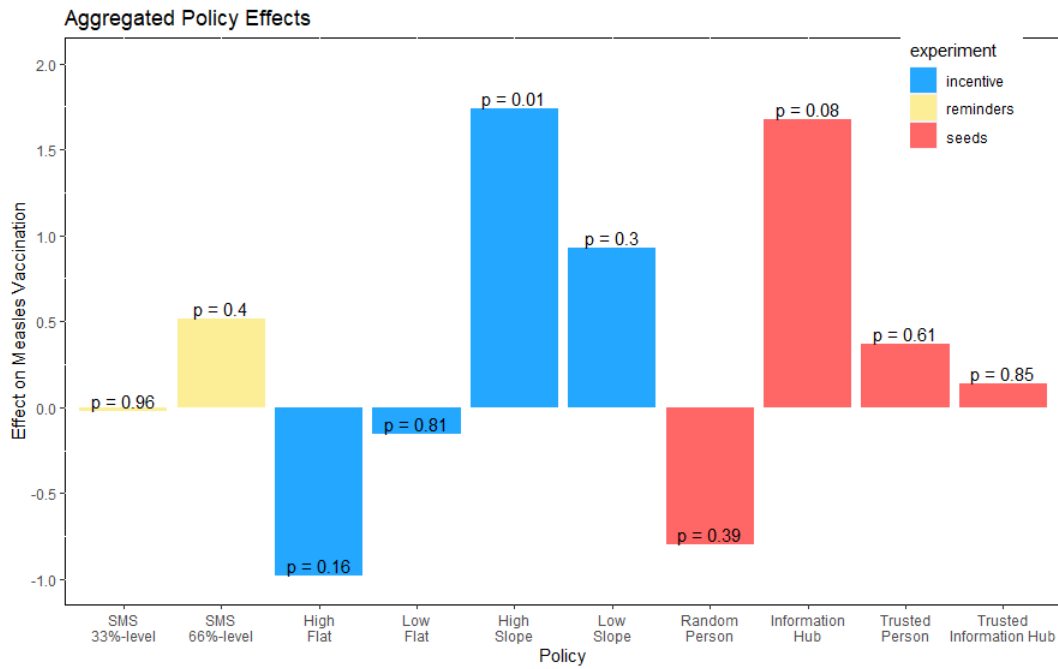
(A) Full Sample
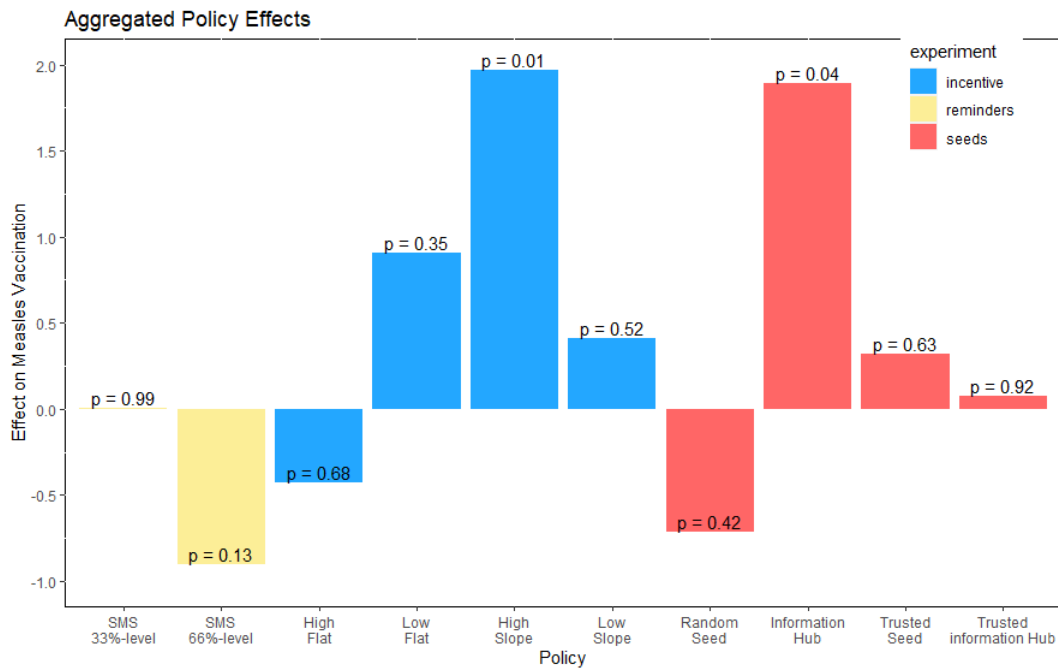
(B) Ambassador Sample

FIGURE 2. Effects on the number of measles vaccinations relative to control (5.29 in Panel A and 7.32 in Panel B) by reminders, incentives, and seeding policies, for entire sample controlling for inclusion in ambassador sample in Panel A and restricted to the ambassador sample in Panel B. The specification is weighted by village population, controls for district-time fixed effects, and clusters standard errors at the sub-center level.

Individual Policy Effects
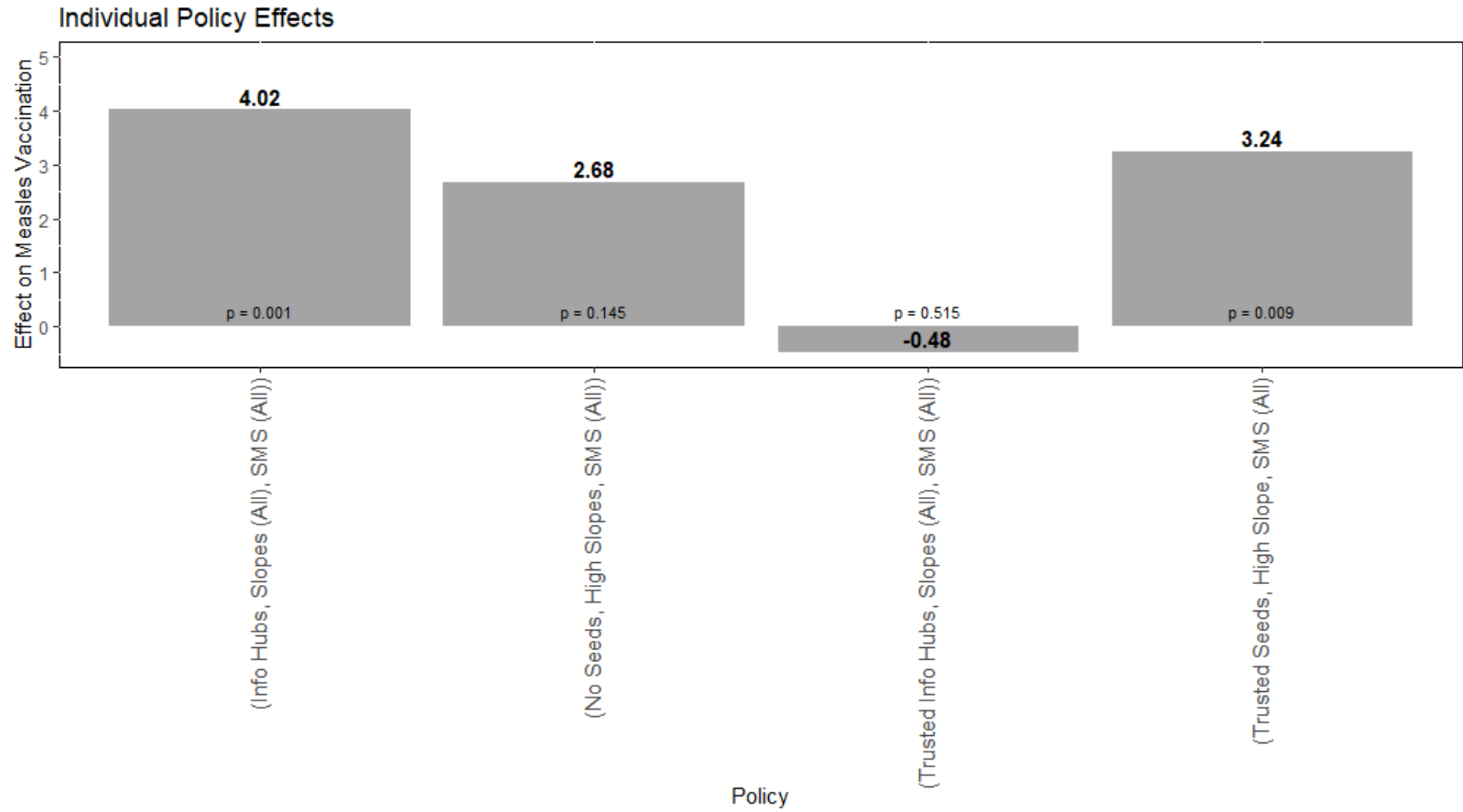
FIGURE 3. Effects of the smartly pooled and pruned combinations of reminders, incentives, and seeding policies on number of measles vaccinations relative to control (7.32). The specification is weighted by village population, controls for district-time fixed effects, and clusters standard errors at the sub-center level.
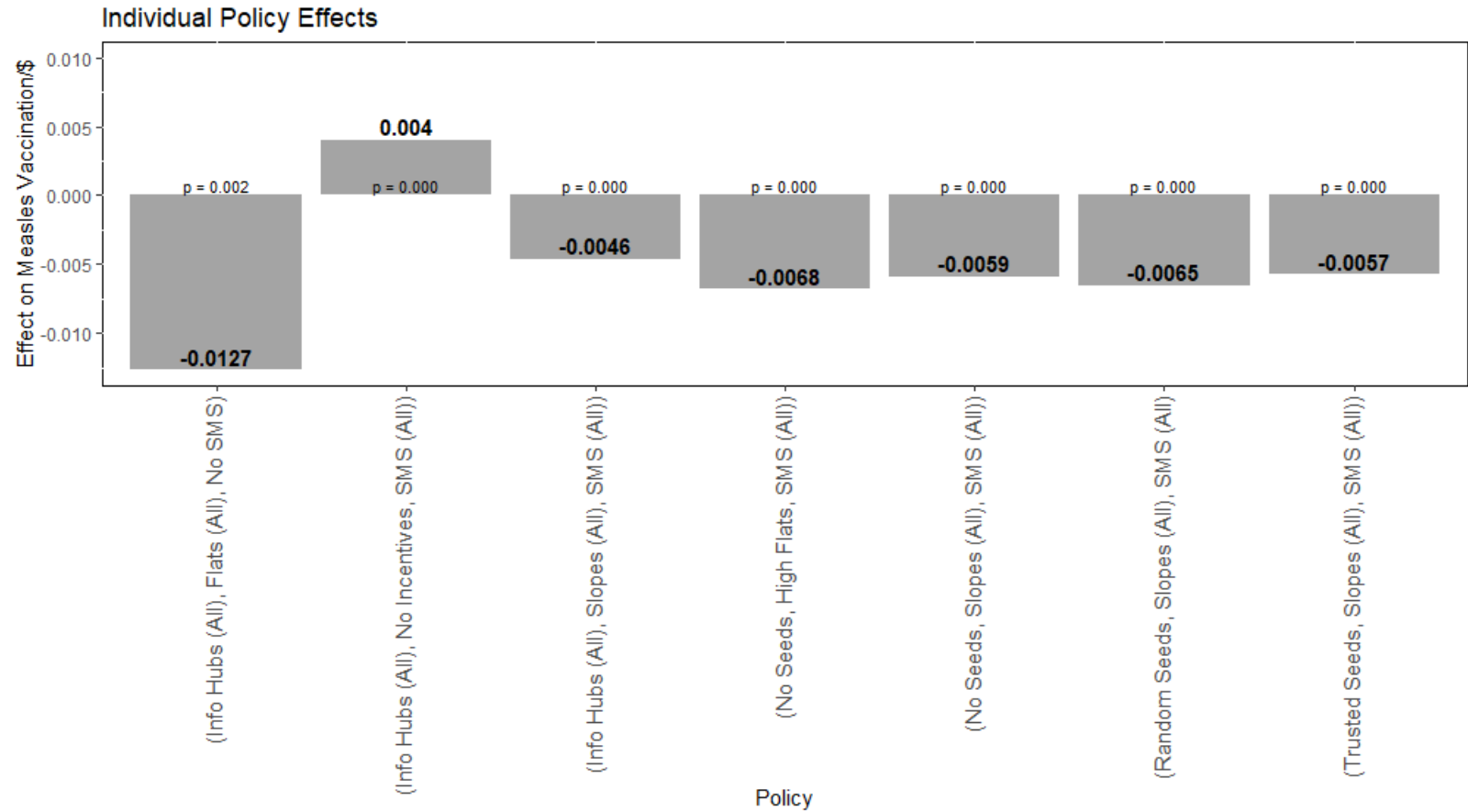
FIGURE 4. Effects of the smartly pooled and pruned combinations of reminders, incentives, and seeding policies on the number of measles vaccines per $1 relative to control (0.0436 shots per $1). The specification is weighted by village population, controls for district-time fixed effects, and clusters standard errors at the sub-center level.

Tables

Table 1. Best Policies

|	(1) # Measles Shots	(2) # Measles Shots per $1
WC Adjusted Treatment Effect	3.26	0.004
Confidence Interval (95%)	[0.32,6.25]	[0.003, 0.005]
Control Mean	7.32	0.0435
Observations	204	814
Optimal Policy	(Information Hubs, SMS, Slope)	(Information Hubs POOLED, SMS)

Notes: Estimation using Andrews et al. (2019); hybrid estimation with $\alpha = 0.05, \beta = 0.005$. The specifications are weighted by village population and account for district-time fixed effects as well as variance clustered at the sub-center level.

Appendix A. Proofs

Proof of Proposition 3.1. According to Theorem 1 of Jia and Rohe (2015), if $\min_{j \in S_\alpha} |\alpha_j| \geq 2\lambda_n$, then $\hat{\alpha} =_s \alpha$ with probability greater than

$$f(n) := 1 - 2K \exp\left(-\frac{n\lambda^2 \xi_{\min}^2}{2\sigma^2}\right),$$

where, recall, $\xi_{\min} = \xi_{\min}(\frac{X}{\sqrt{n}})$ is the minimum singular value of the $\sqrt{n}$-normalized design matrix. By Assumption 3, the uniform lower bound (in absolute value) of the nonzero $\{\beta\}$ determines a lower bound to the nonzero parameters $\{\alpha\}$ as well, since these specifications are related by an invertible linear transformation. Since by Assumption 4, $\lambda_n \to 0$, for sufficiently high n $\min_{j \in S_\alpha} |\alpha_j| \geq 2\lambda_n$. Theorem 1 applies and $\text{sign}(\hat{\alpha}) = \text{sign}(\alpha)$ with probability greater than $f(n)$.

It will be convenient to re-express $f(n)$ as follows:

$$f(n) = 1 - 2 \exp\left(\log(K) - \frac{n\lambda_n^2 \xi_{\min}^2}{2\sigma^2}\right).$$

And applying Lemma A.1, for sufficiently high n:

$$f(n) \geq 1 - 2 \exp\left(\log(K) - \frac{n\lambda_n^2}{2\sigma^2 K^2}\right).$$

Per Assumption 1, $K = O(n^\gamma)$, it follows that for sufficiently high n:

$$f(n) \geq 1 - 2 \exp\left(\gamma \log(n) - \frac{n^{1-2\gamma}\lambda_n^2}{2\sigma^2}\right) \geq 1 - 2 \exp\left(\gamma \log(n) - \frac{n^{1-2\gamma}\lambda_n^2}{2\sigma^2}\right).$$

By Assumption 1, $0 < \gamma < \frac{1}{2} \implies n^{1-2\gamma} = \omega(\log(n))$ and by Assumption 4, since $\lambda_n^2 n^{1-2\gamma} = \omega(\log(n))$, it follows that $\lim_{n\to\infty} f(n) \geq 1$. Since also $f(n) \leq 1$, it follows that $f(n) \to 1$ and the proof is complete. ∎

LEMMA **A.1.** *For the smart pooling design matrix X, for $R \geq 3$, wpa1 the lowest singular value of $\sqrt{n}$ normalized design matrix, i.e., $\xi_{\min}(\frac{X}{\sqrt{n}})$, has the value $\xi_{\min}(\frac{X}{\sqrt{n}}) = \left(4R\sin^2\left(\frac{R-\frac{3}{2}}{R-\frac{1}{2}}\frac{\pi}{2}\right)\right)^{-\frac{M}{2}}$. Thus, wpa1, $\xi_{\min}(\frac{X}{\sqrt{n}}) > (\frac{1}{K}).$*[23]

Proof of Lemma A.1. It will be useful to index the design matrix X by R and M, i.e. $X = \mathbf{X}_{R,M}$. Let $\mathbf{C}_{R,M} = \lim_{n\to\infty} \frac{1}{n} \mathbf{X}'_{R,M} \mathbf{X}_{R,M}$. Then $\lim_{n\to\infty} \xi_{\min}^2(\frac{\mathbf{X}_{R,M}}{\sqrt{n}}) = \lambda_{\min}(\mathbf{C}_{R,M})$, i.e., the lowest eigenvalue of $\mathbf{C}_{R,M}$. We will characterize this eigenvalue.

The combinatorics of the limiting frequencies of "1"s in smart pooling variables imply that $\mathbf{C}_{R,M}$ is a block diagonal matrix with structure

---

[23]This is a conservative bound; the optimal uniform lower bound is $\left(\frac{1}{K}(\frac{1}{4^M})\right)^{\frac{1}{2}}$.

$$\mathbf{C_{R,M}} = \frac{1}{K} \begin{bmatrix} \mathbf{B}_{R,M} & 0 & \cdots & 0 \\ 0 & \mathbf{B}_{R,M-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{B}_{R,1} \end{bmatrix}$$

where $\mathbf{B}_{R,M-1}$ implies this block is found in $\mathbf{C}_{R,M-1}$ (pertaining to an RCT with one less cross-treatment arm), etc. More than one block of $\mathbf{B}_{R,M-1}$, $\mathbf{B}_{R,M-2}$, ... $\mathbf{B}_{R,1}$ is found in $\mathbf{C_{R,M}}$, but only $\mathbf{B_{R,M}}$ determines the minimum eigenvalue.

The combinatorics of variable assignments also implies that

(1) $\mathbf{B}_{R,M}$ is an $(R-1)^M \times (R-1)^M$ matrix with recursive structure $\mathbf{B}_{R,M} = \mathbf{B}_{R,1} \otimes \mathbf{B}_{R,M-1}$, where $\otimes$ is the Kronecker product.[24]

(2) $\mathbf{B_{R,1}}$ is an $(R-1) \times (R-1)$ matrix with recursive structure

$$\mathbf{B_{R,1}} = \begin{bmatrix} R-1 & R-2 & \cdots & 1 \\ R-2 & & & \\ \vdots & & \mathbf{B_{R-1,1}} & \\ 1 & & & \end{bmatrix} \quad \text{and } \mathbf{B_{2,1}} = [1].$$

SUBLEMMA 1. $\lambda_{\min}(\mathbf{B_{R,1}}) = \left(4 \sin^2 \left(\frac{R-\frac{3}{2}}{R-\frac{1}{2}} \frac{\pi}{2} \right) \right)^{-1}$

Proof. The key insight of the argument[25] is that $\mathbf{B}_{R,1}^{-1}$ is the $(R-1) \times (R-1)$ tridiagonal matrix:

$$\mathbf{B}_{R,1}^{-1} = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & \ddots & \ddots & \\ & & \ddots & \ddots & -1 \\ & & & -1 & 2 \end{bmatrix}$$

Which has known eigenvalues $\mu_j = 4 \sin^2 \left(\frac{j-\frac{1}{2}}{R-\frac{1}{2}} \frac{\pi}{2} \right)$ for $j = 1, 2, ..., R-1$. Thus, given that the inverse of a matrix's eigenvalues are the inverse matrix's eigenvalues, $\lambda_{\min}(\mathbf{B_{R,1}}) = \left(4 \sin^2 \left(\frac{R-\frac{3}{2}}{R-\frac{1}{2}} \frac{\pi}{2} \right) \right)^{-1}$. ∎

(Resuming the proof of Lemma A.1) Per the multiplicative property of the eigenvalues of a Kronecker product, together with the fact that all matrices in question are positive definite, it immediately follows that $\lambda_{\min}(\mathbf{B}_{R,M}) = \lambda_{\min}(\mathbf{B}_{R,1}) \lambda_{\min}(\mathbf{B}_{R,M-1})$, which in

---

[24]Thanks to Nargiz Kalantarova for noticing this Kronecker product and its consequent implication for $\lambda_{\min}(\mathbf{B}_{R,M})$.

[25]The argument is provided on Mathematics Stackexchange (user1551 (https://math.stackexchange.com/users/1551/user1551), 2017).

turn implies $\lambda_{\min}(\mathbf{B}_{R,M}) = \lambda_{\min}(\mathbf{B}_{R,1})^M$. Since by Sublemma 1, $\lambda_{\min}(\mathbf{B_{R,1}}) < 1$, $\mathbf{B}_{R,M}$ is the block determining the rate with the smallest eigenvalue, and therefore, given that the eigenvalues of a block diagonal matrix are the eigenvalues of the blocks:

$$\lambda_{\min}(\mathbf{C}_{R,M}) = \frac{1}{K}\lambda_{\min}(\mathbf{B}_{R,M}) = \frac{1}{K}(\lambda_{\min}(\mathbf{B}_{R,1}))^M = \left(4R\sin^2\left(\frac{R-\frac{3}{2}}{R-\frac{1}{2}}\frac{\pi}{2}\right)\right)^{-M}$$

where the last equality uses Sublemma 1. The Lemma follows. ∎

Proof of Corollary 3.1. By Proposition 3.1, $\hat{S}_\alpha = S_\alpha$ wpa1, which by inverting the linear map, implies that $\hat{S}_{pool} = S_{pool}$ wpa1 and by Proposition B.1, we have that $S_\beta \subset S_pool$. The assumptions of Corollary 2 (to Theorem 5) of Belloni and Chernozhukov (2013) apply and therefore $\hat{n} \to_p \eta$ (where we take the convention that the treatment effect is set to 0 for any unique treatment combination that is excluded from $\hat{S}_{pool}$ but in S_β, and that event happens wpa0). ∎

Proof of Corollary 3.2. By Theorem 2.1 of Belloni, Chernozhukov, Chetverikov, and Wei (2018), the post-regularized estimator is asymptotically normally distributed. Therefore, wpa1,

$$1 - 2K\exp\left(-\frac{n\lambda^2\xi^2_{\min}}{2\sigma^2}\right),$$

the joint normality of the estimates of the pruned and pooled treatment effects hold. By Corollary 2 of Andrews et al. (2019), the result follows. What remains is to check the remaining Assumptions 2-4 required for the corollary to hold. Assumption 2, concerning the uniform Lipschitz structure, follows from the bounded moments (mechanical, as these are independent treatments) in the OLS structure of the problem. Assumption 3, concerning the uniform consistency of the variance estimator, again follows since the sample variance can be used. Assumption 4 bounds from above and below the entries of the variance-covariance matrix and holds given independent treatment assignment. ∎

Appendix B. Pooling Procedure

Recall that every treatment combination k (of a set $\mathcal{K}$ of size $K = R^M$) has a treatment profile $P(k)$, which captures which treatment arms are "active" irrespective of intensity (dosage) of the arm. Recall, furthermore, the partial ordering of treatment combinations where $k \geq k'$ is used when the intensities of k in each arm weakly dominate that of k'. The invertible transformation between the smart pooling specification (3.2) and unique policies (3.1) implies the following relationship between parameters:

$$(B.1) \qquad \beta_k = \sum_{k' \leq k, P(k) = P(k')} \alpha_{k'}$$

As this suggests, the zeros in α_k equate two or more coefficients β_j, and therefore pool treatment combinations within a treatment profile (i.e., pools dosages).[26] Alternatively, one can adopt the perspective that the nonzero α_k distinguishes two or more coefficients β_j, and therefore disaggregates treatment combinations within a treatment profile. In this section, we construct the pooling/disaggregation generically using S_α, the support of the smart pooling specification (3.2). The end result is a set S_{pool} of pooled policies.

B.1. **A Guiding Example.** To lend intuition for this construction, let us fix a guiding example, with a particular R, M, a treatment profile, and the subset of the smart pooling support $S \subseteq S_\alpha$ relevant to this profile.

EXAMPLE **2.** *Consider a two-arm experiment of four intensities each (three non-zero intensities), i.e., $M = 2$, $R = 4$. Consider the treatment profile where both treatment arms are "active." Each treatment combination k within this profile (i.e., dosage) can be denoted $[r_1, r_2]$, where $r_i \in R$ is the intensity in the i-th arm. The partial ordering of treatment combinations can be depicted in the following Hasse diagram in Figure B.1, where a line upwards from x to y implies the latter marginally dominates the former ($y > x$ and there is no z s.t. $y > z > x$). Thus for example $[2, 1]$ has a line upwards to both $[2, 2]$ and $[3, 1]$, while $[2, 2]$ and $[3, 1]$ are incomparable.*

Now we can consider an example of the subset of support $S \subseteq S_\alpha$ of treatment combinations in the smart pooling specification (3.2) relevant to this treatment profile.

---

[26]Note that this includes pooling with the omitted category, which is of course just finding that one more $\beta_j = 0$ and pruning away this set
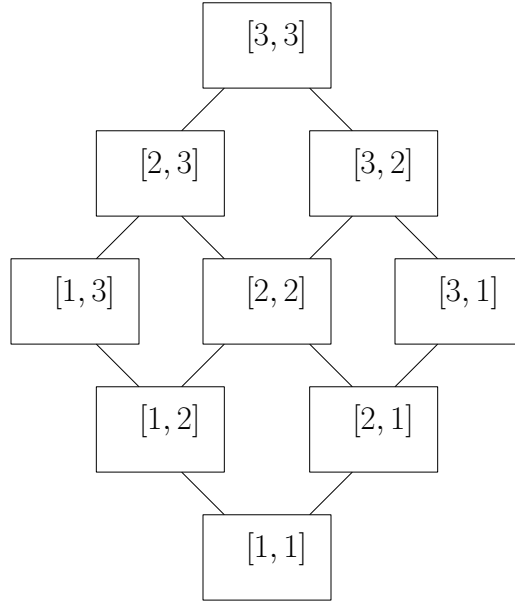
FIGURE B.1. Hasse Diagram

EXAMPLE **2** (continued). *Let $S \subseteq S_\alpha = \{[1,2], [2,1]\}$ be the supported smart pooling vectors within this treatment profile, and let $\alpha_{[1,2]} = 100$, $\alpha_{[2,1]} = 200$ (chosen large for clarity in exposition). Now, consider the sets $A_{[1,2]}$ and $A_{[2,1]}$ of treatment combinations that weakly dominate each of $[1,2]$ and $[2,1]$ respectively. We can depict these directly on the Hasse diagram in Figure B.2. With reference to the parameter relationship* (B.1),

FIGURE B.2. Hasse Diagram with $A_{[2,1]}$ and $A_{[1,2]}$

these will determine treatment effects β_j within this treatment profile. Here are three examples:

(1) *For $j = [1,3]$, per the parameter relationship* (B.1), $\beta_j = 100$. *This is seen visually by noting that only $A_{[1,2]}$ "encircles" $[1,3]$*

(2) *For $j = [3,1]$, per the parameter relationship* (B.1), $\beta_j = 200$. *This is seen visually by noting that only $A_{[1,2]}$ "encircles" $[3,1]$*

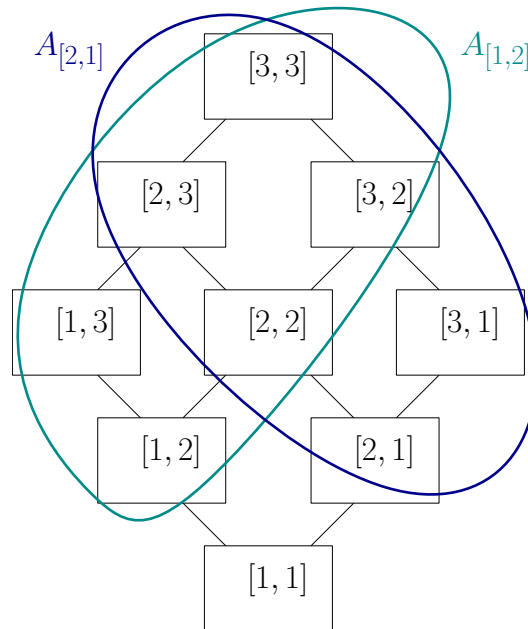(3) *For $j = [2,2]$, per the parameter relationship* (B.1), $\beta_j = 100 + 200 = 300$. *This is seen visually by noting that both $A_{[1,2]}$ and $A_{[2,1]}$ "encircle" $[2,2]$*

We showed that the calculation of three coefficients β_j depends on what "enricles" the treatment combination, and it is now clear that the distinct (mutually disjoint) "regions of encirclement" determine *all* the coefficients β_j within this profile (and therefore all pooling/disaggregation within this profile).

EXAMPLE **2** (continued). *The three "regions of encirclement" are $A_{[1,2]} \cap A_{[2,1]}^c$ ("$A_{[1,2]}$ alone encircles"), $A_{[1,2]}^c \cap A_{[2,1]}$ ("$A_{[2,1]}$ alone encircles"), and $A_{[1,2]} \cap A_{[2,1]}$ ("both $A_{[1,2]}$ and $A_{[2,1]}$ encircle"). They are depicted in Figure B.3. These regions of encirclement*
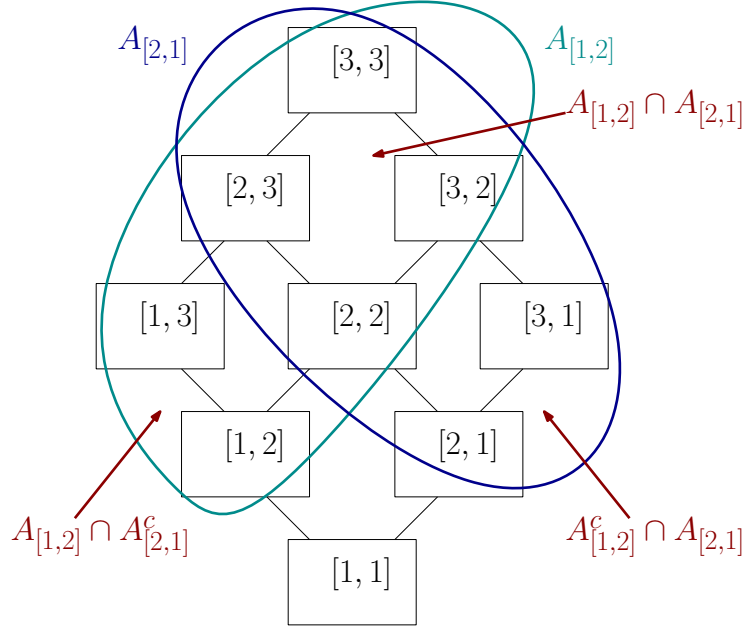

FIGURE B.3. Hasse Diagram with $A_{[2,1]}$ and $A_{[1,2]}$ with complements and intersections.

determine regions of equitreatment effects β_j, and therefore the pooled policies.

(1) *For any $j \in A_{[1,2]} \cap A_{[2,1]}^c$, $\beta_j = 100$. Thus $A_{[1,2]} \cap A_{[2,1]}^c = \{[1,2],[1,3]\}$ is a pooled policy.*

(2) *For any $j \in A^c_{[1,2]} \cap A_{[2,1]}$, $\beta_j = 200$. Thus $A^c_{[1,2]} \cap A_{[2,1]} = \{[2,1],[3,1]\}$ is a pooled policy.*

(3) *For any $j \in A_{[1,2]} \cap A_{[2,1]}$, $\beta_j = 100 + 200 = 300$. Thus $A_{[1,2]} \cap A_{[2,1]} = \{[2,2],[2,3],[3,2],[3,3]\}$ is a pooled policy.*

And thus we generate the set of pooled policies for this treatment profile using the relevant subset $S \subseteq S_\alpha$ of the smart pooling support.

B.2. **The General Construction.** The approach in the guiding example fully generalizes for any treatment profile and for any R, M. Given the estimated support $\hat{S}_\alpha$ of the smart pooling specification (3.2), let $[\hat{S}_\alpha]$ denote its partition into sets of support vectors with the same treatment profile. Each $S \in [\hat{S}_\alpha]$ is thus a set of treatment combinations $\{k_1, ...k_n\}$. For each $k_i \in \{k_1, ...k_n\}$, define the set:

$$\text{(B.2)} \qquad A_{k_i} = \{k \in \mathcal{K} | P(k) = P(k_i) \text{ and } k \geq k_i\}$$

The pooled policies are the mutually disjoint "regions of encirclement" generated through intersections:

$$\text{(B.3)} \qquad \mathcal{A} = A^{a_1}_{k_1} \cap ... \cap A^{a_n}_{k_n}$$

where $a_i \in \{1, c\}$, and A^c denotes the complement of the set A. We are only interested in those intersections $\mathcal{A}$ which are non-empty, and we furthermore exclude consideration of the intersections of complements $A^c_{k_1} \cap ... \cap A^c_{k_n}$.

The estimated set of pooled policies $\hat{S}_{pool}$ is the collection of all such sets $\mathcal{A}$. Algorithm 2 describes this generation of $\hat{S}_{pool}$ procedurally. The following proposition verifies its main properties, namely that when $\hat{S}_\alpha$ is selected correctly (1) all treatment combinations that are pooled have equitreatment effects and (2) all non-zero policies are accounted for.

PROPOSITION **B.1.** *Assume the support from (3.2) is correctly selected, i.e., $\hat{S}_\alpha = S_\alpha$. Call the implied pooling $\hat{S}_{pool} = S_{pool}$. Then*

(1) *For every $\mathcal{A} = A^{a_1}_{k_1} \cap ... \cap A^{a_n}_{k_n}$ in S_{pool}, and for every $k \in \mathcal{A}$,*

$$\text{(B.4)} \qquad \beta_k = \sum_{i|a_i=1} \alpha_{k_i}$$

where the parameters β are from the original specification (3.1) and the parameters α are from the smart pooling specification (3.2). This justifies the statement that $\mathcal{A}$ pools treatment combinations with the same treatment profile and with the same treatment effects.

(2) *S_{pool} is a superset of the support S_β of granular policies from (3.1). That is, if treatment combination k is such that $\beta_k \neq 0$, then $\exists \mathcal{A} \in S_{pool}$ s.t. $k \in \mathcal{A}$.*

Proof of Proposition B.1. Consider an arbitrary $S \subseteq [\hat{S}_\alpha]$ where $S = \{k_1, ...k_n\}$, an arbitrary nonempty $\mathcal{A} = A_{k_1}^{a_1} \cap ... \cap A_{k_n}^{a_n}$ in S_{pool}, and an arbitrary $k \in \mathcal{A}$. Membership in $\mathcal{A}$ immediately implies that:

(1) $\forall i \in \{1, ..., n\}, P(k) = P(k_i)$
(2) $\forall i$ such that $a_i = 1$, $k \geq k_i$
(3) $\forall i$ such that $a_i = c$, $k \not\geq k_i$

Furthermore, if k^* is a vector such that $P(k^*) = P(k)$ but $k^* \notin \{k_1, ..., k_n\}$, then by definition of S_α, $\alpha_{k^*} = 0$. Thus from the parameter relationship (B.1) we have:

$$(\text{B.5}) \qquad \beta_k = \sum_{i|a_i=1} \alpha_{k_i}.$$

Thus, part (1) follows.

For part (2), consider any $\beta_k \neq 0$. By (B.1), there must be a vector k' s.t. $P(k) = P(k')$ and $\alpha_{k'} \neq 0$. Necessarily $k' \in S_\alpha$, so in particular, $k' \in S = (k_1, ...k_n) \in [\hat{S}_\alpha]$. Then $k \in \mathcal{A} = A_{k_1}^{a_1} \cap ... \cap A_{k_n}^{a_n}$ for some $(a_1, ..., a_n)$, since these sets together form a disjoint union of $A_{k_1} \cup ... \cup A_{k_n}$. ∎

Algorithm 2: Pooling Procedure

---

input : Estimated support $\hat{S}_\alpha$ from the smart pooling specification (3.2)
output: Estimated pooled policies $\hat{S}_{pool}$ for pooled specification (3.4)
Partition $\hat{S}_\alpha$ into $[\hat{S}_\alpha]$ per the treatment profile mapping $P(.)$;
Initialize $\hat{S}_{pool} \longleftarrow \emptyset$;
for $S \in [\hat{S}_\alpha]$ **do**
    discover $S = \{k_1, ..., k_n\}$;
    generate $\{A_{k_1}, ..., A_{k_n}\}$;
    for *each* $(a_1, ..., a_n)|a_i \in \{1, c\}$ **do**
        generate $\mathcal{A} = A_{k_1}^{a_1} \cap ... \cap A_{k_n}^{a_n}$;
        if $\mathcal{A} \neq \emptyset$ *and* $\mathcal{A} \neq A_{k_1}^c \cap ... \cap A_{k_n}^c$ **then**
            $\hat{S}_{pool} \longleftarrow \hat{S}_{pool} \cup \mathcal{A}$;
        end
    end
end
return $\hat{S}_{pool}$

APPENDIX C. SIMULATIONS

C.1. **Smart pooling without Puffering, (3.2), fails irrepresentability.** Consider the smart pooling covariate where all M arms are "on" with highest intensity, i.e., X_{k^*} for $k^* = (R-1, ..., R-1)$. We will show that this covariate is "representable" by the other covariates. Intuitively, this means too much of this covariate is explained by the others. Formally, the L_1 norm of the coefficients (excluding intercept) from a regression of this covariate over the others is too great (it exceeds > 1). That is, if an OLS regression finds:

$$X_{k^*} = \hat{\gamma}_0 + \sum_{k \in K, k \neq k^*} \hat{\gamma}_k X_k.$$

X_{k^*} is representable by the other covariates if $\sum_{k \in K, k \neq k^*} |\hat{\gamma}_k| > 1$.

We demonstrate this through a proof by example. A simulation establishes that X_{k^*} is representable (and therefore that the specification fails irrepresentability) for a computationally reasonable range of R and M. We see that the patterns imply that irrepresentability fails even more dramatically for larger R and M.

In this simulation, we choose large $n = 10,000$ so that the propensities of "1" within each covariate have stabilized. We consider two kinds of regressions: an "unstandardized" regression where the raw smart pooling covariates are regressed, and a "standardized" regression where the smart pooling covariates are first standardized by the L_2 norm. The latter corresponds to a preprocessing step that LASSO packages typically apply before LASSO selecting; we would like to know if irrepresentability fails even in this case. Indeed, we see the L_1 norms are greater than 1 in both cases, and irrepresentability fails.

R	M	L_1 norm (standardized vars)	L_1 norm (unstandardized vars)
3	2	1.73	1.26
3	3	3.67	2.32
3	4	7.66	4.2
4	2	1.77	1.24
4	3	3.98	2.43
4	4	8.27	4.14
5	2	1.87	1.28
5	3	3.78	2.29
5	4	7.90	4.70

C.2. **Comparisons of Smart Pooling and Pruning with alternatives.** We show that the smart pooling and pruning estimator selects the support S_α of (3.2) consistently (whereas the naive LASSO does not) and then show that the overall estimator outperforms (in terms of estimating the best policy effect) an estimator using naive LASSO on the set of unique policies (with out pooling) as in (3.1).

In what follows, we consider results on simulated design matrices of (3.2) with the following common setup:

(1) Fix $R = 3, M = 4$ and $\sigma := \sqrt{\text{var}(\epsilon)} = 1$

(2) The simulation results are plots of performance $\hat{m}(n)$ against sample size n, where n is logarithmically spaced between 1000 and 10000.[27]

(3) These scores $\hat{m}(n)$ are generically computed as follows.

   (a) A set $\mathcal{C}$ of true supports of (3.2) is drawn based on a conditionally random logic that will be specified. Each member $S_\alpha^i \in \mathcal{C}$ is a particular support or "configuration" (3.2). Each configuration has fixed support size $|S_\alpha^i| = M$. Furthermore, if $S_\alpha^i = (k_1, k_2, .., k_M)$ in some given order, we assign coefficients $\alpha_{k_j} = \left(1 + \frac{j-1}{M-1}\right)\sigma$. That is, these nonzero coefficients are linearly spaced between σ and 2σ. Thus each configuration fully specifies the set of coefficients α for (3.2).

   (b) For each $S_\alpha^i \in \mathcal{C}$, a set $\mathcal{S}_{S_\alpha^i}(n)$ of simulations (design matrices) is generated per the coefficients specified by the configuration, and the Gaussian noise, with sample size n. For each simulation $\hat{s}(n) \in \mathcal{S}_{S_\alpha^i}(n)$, it is scored by a metric $m(\hat{s}(n))$ that will be specified.

   (c) These scores are aggregated over simulations $\mathcal{S}_{S_\alpha^i}(n)$, and then aggregated again over configurations $\mathcal{C}$, to produce the aggregated performance score $\hat{m}(n)$.

C.2.1. *Smart Pooling and Pruning outperforms LASSO in pooling policies.* We demonstrate that the first step in the smart pooling and pruning procedure consistently recovers the support S_α whereas the naive LASSO fails to do so. This is owing to the aforementioned violation of irrepresentability, which is overcome using the Puffer transformation.

Here we draw randomly chosen support configurations $\mathcal{C}$ where S_α^i is entirely within the treatment profile where all arms are "active": this is where we expect excess correlation, as the irrepresentability failures in Appendix Section C.1 shows. Ex-ante, we do not assume anything about the locus of the true support, so any model selection procedure must be consistent for this "worst case" possibility.

---

[27]We logarithmically space for computational ease, as simulations become computationally intensive at high n

We demonstrate that the preprocessed smart pooling and pruning estimator (with sequential elimination implementation of LASSO) consistently selects the model S_α^i but the naive strategy of directly applying to LASSO to (3.2) does not. To evaluate each simulation $\hat{s}(n)$, given model selection estimator $\hat{S}_\alpha(\hat{s}(n))$, it is scored by the *support selection accuracy*

$$m(\hat{s}(n)) := \frac{|\hat{S}_\alpha^i(\hat{s}(n)) \cap S_\alpha^i|}{|\hat{S}_\alpha^i(\hat{s}(n)) \cup S_\alpha^i|}.$$

This is a value between 0 and 1 that increases with support coverage, and is 1 iff the support is correctly selected. It is aggregated by taking averages over the simulations per configuration, and then averaged again over the configurations.

$$\hat{m}(n) := \frac{1}{|\mathcal{C}|} \sum_{\mathcal{C}} \frac{1}{|\mathcal{S}_{S_\alpha^i}(n)|} \sum_{\mathcal{S}_{S_\alpha^i}(n)} m(\hat{s}(n)).$$

In the below simulation result, we draw five support configurations, and 20 simulations per configuration, i.e., $|\mathcal{C}| = 5$ and $|\mathcal{S}_{S_\alpha^i}(n)| = 20, \forall S_\alpha^i \in \mathcal{C}, \forall n$.
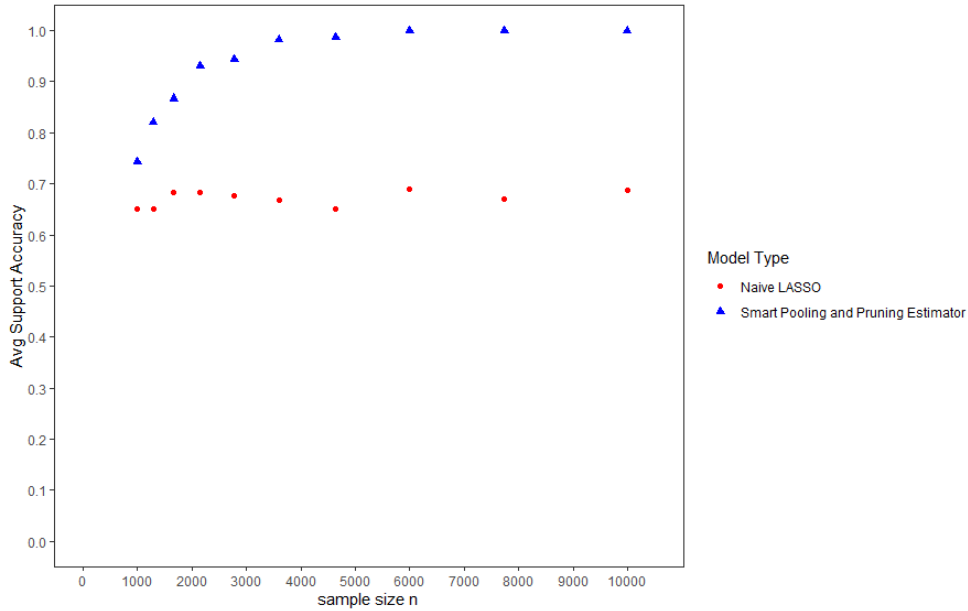
FIGURE C.1. A comparison of average support accuracies between the smart pooling and pruning estimator and a direct implementation of LASSO on (3.2) using Chernozhukov et al. (2015). There are 20 simulations per support configuration per n, for five support configurations.

From Figure C.1, we can clearly see that only our preprocessed smart pooling and pruning estimator is support consistent, converging to 100% average accuracy, while a

direct application of LASSO is not, as $\hat{m}(n)$ does not exhibit a clear monotonic pattern and never exceeds 75% average accuracy.[28]

Besides being consistent, even when the smart pooling and pruning estimator fails to select the correct support (for lower n), it still tends to select the correct best policy. To see this, we define a *best policy inclusion* scoring function:

$$m(\hat{s}(n)) = \begin{cases} \frac{c_1}{c_2} & \text{if } \hat{S}_\alpha^i(n) \text{ pools together a subset of the true best policy per } S_\alpha^i, \\ & \text{and } \hat{S}_\alpha^i(n) \text{ pools together } c_1 \text{ policies in the best policy while } S_\alpha^i \text{ pools } c_2 \text{ policies} \\ 0 & \text{otherwise.} \end{cases}$$

This is again a value between 0 and 1 that increases with best policy inclusion, and is 1 iff the best policy is selected and pooled perfectly. $\hat{m}(n)$ is computed the same as before (with the same $\mathcal{C}$ and $\mathcal{S}_{S_\alpha^i}(n)$. Plotting together the average best policy inclusion and average support accuracy for the pooling and pruning estimator:
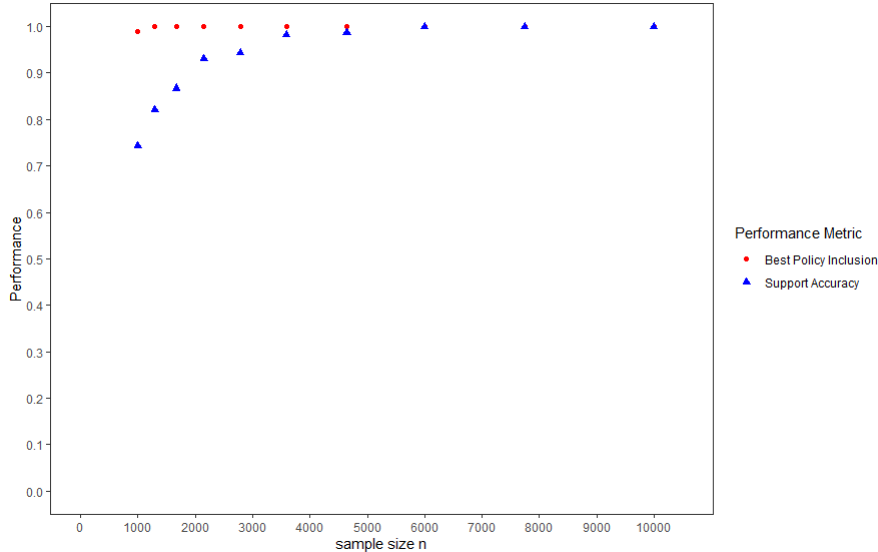
FIGURE C.2. A plot of average support accuracy and best policy inclusion for the smart pooling and pruning estimator. There are 20 simulations per support configuration per n, for 5 support configurations.

We see that average best policy inclusion is near perfect even when there are support errors at low n, and indeed reaches 100% coverage in simulations very quickly for modest increases in n.

C.2.2. *Smart Pooling and Pruning outperforms LASSO on unique policies.* The previous simulations show that our model selection procedure with the Puffer transformation is

---

[28]Note that in both LASSO implementations, the LASSO parameter λ increases with n. In the case of a direct application of LASSO using Chernozhukov et al. (2015), λ is calculated via the formula $\lambda = 2c\sqrt{n}\hat{\sigma}\Phi^{-1}(1 - 0.1/(2K))$.

the right way to pool. We can also demonstrate the importance of pooling in the first place, which is the second step. After all, the alternative is to simply apply the naive LASSO on (3.1).

Indeed, the relevant counterfactual is applying LASSO as a pruning (but not pooling) step on the specification (3.1) of unique, finely differentiated policies, which are determined by (3.2) by an invertible linear transformation.

A key reason this alternate strategy is worse is that the efficiency of the hybrid estimator of Andrews et al. (2019) can degrade, which happens because the attenuation from winner's curse increases the closer the second-best estimate is to the best. By not pooling, it can over-penalize the estimator by running a best policy against a negligibly smaller second-best dosage variant, leading to needlessly conservative estimates. A simulation attests to this.

Here, there will be a single configuration $\mathcal{C}$, which we can call $S_\alpha^\star$. $M-1$ covariates of X are randomly sampled where at least one treatment arm is "inactive"; these will be less relevant for winner's curse adjustments, because the coefficients are bounded above by $\left(1 + \frac{M-2}{M-1}\right)\sigma < 2\sigma$. The interesting action is the M-th coefficient α_k, which we will assign 2σ. We assign this to $k = (1, 1, .., 1)$, i.e., where X_k indicates "at least some nonzero dosage". By the invertible transformation between (3.2) and (3.1), *all* unique policies with the treatment profile where all arms are "active" will have true treatment effect exactly 2σ. Again, ex-ante we cannot assume the locus of $S_\alpha^\star$, so it suffices to consider this the "worst case" possibility.

For each simulation $\hat{s}(n)$, it is scored (conditional on a model selection procedure) by its error with respect to the true treatment effect:

$$m(\hat{s}(n)) := \hat{\eta}_{\hat{k}}^{\text{hyb}} - 2\sigma.$$

And thus $\hat{m}(n)$ is simply the estimated MSE:

$$\hat{m}(n) := \frac{1}{|\mathcal{S}_{S_\alpha^\star}(n)|} \sum_{\mathcal{S}_{S_\alpha^\star}(n)} m^2(\hat{s}(n)).$$

In the below simulation result, we fix the number of simulations per n as $|\mathcal{S}_{S_\alpha^\star}(n)| = 20$ for all n.
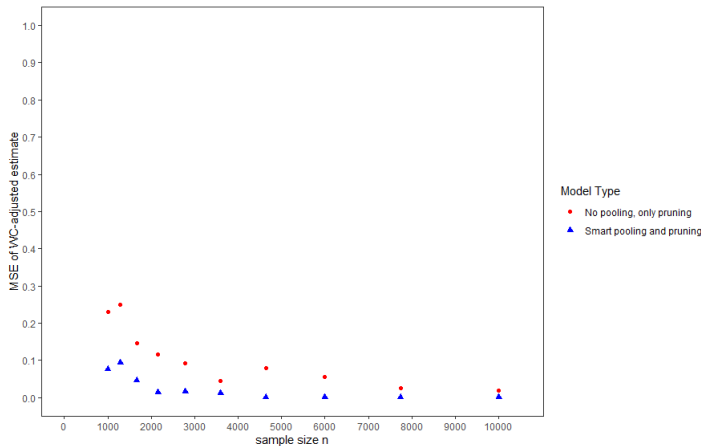
FIGURE C.3. A comparison of the MSE of the hybrid estimator of Andrews et al. (2019) for winner's curse adjusted best policy estimation, between the smart pooling and pruning estimator and a naive strategy of applying LASSO to the unique policy specification (3.1). There are 20 simulations per n.

Clearly, although both estimators are consistent, the hybrid estimator that pools as well as prunes outperforms the other. That this is primarily from the relative penalizations from winner's curse, and not model selection issues, can be verified via a secondary simulation, with the exact same set up but where we condition on the true support of (3.2) and (3.1). The winner's curse estimate without pooling increases MSE because we are effectively running best policies "against themselves" (minor dosage variant with the same effect).
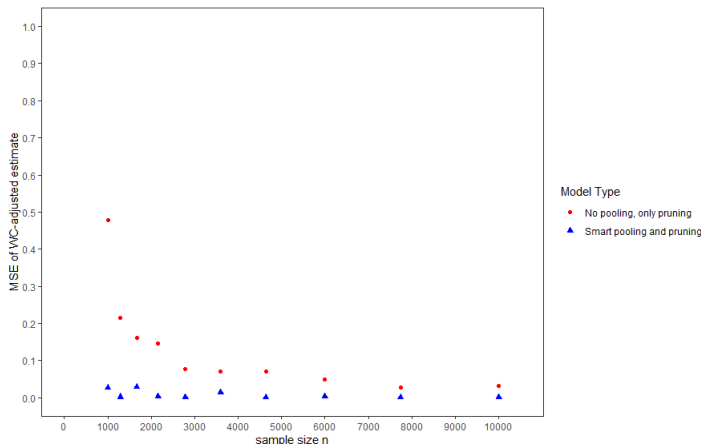
FIGURE C.4. A comparison of the MSE of the hybrid estimator of Andrews et al. (2019) for winner's curse-adjusted best policy estimation, conditional on selecting the true support in (3.2) and (3.1). There are 20 simulations per n.

Appendix D. Robustness
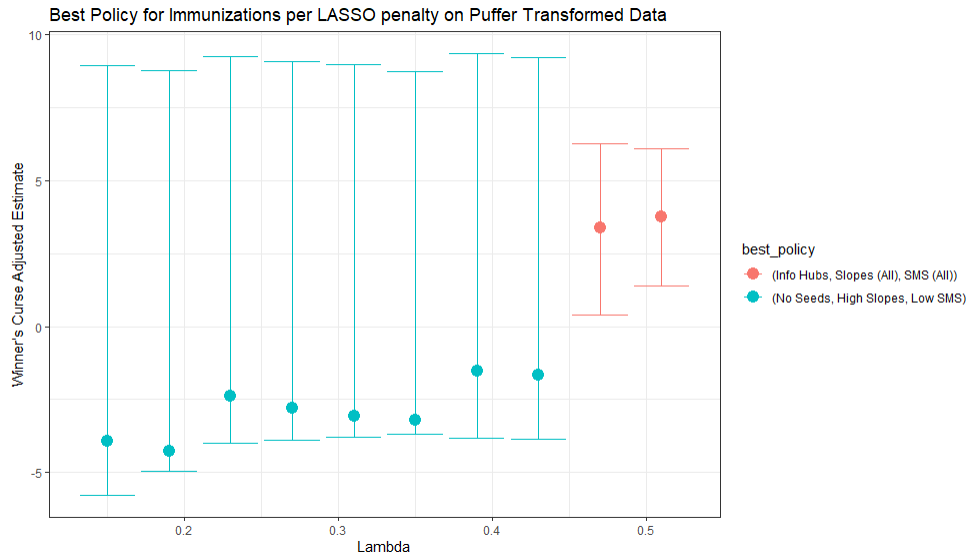
FIGURE D.1. Sensitivity of best-policy estimation for #immunizations per LASSO penalty λ for a sequential elimination version of LASSO on $\text{Puffer}_N(X, Y)$

FIGURE D.2. Sensitivity of best-policy estimation for #immunizations per \$ per LASSO penalty λ for a sequential elimination version of LASSO on $\text{Puffer}_N(X, Y)$

ONLINE APPENDIX: NOT FOR PUBLICATION

Appendix E. Appendix Figures

National Immunization Schedule (NIS) for Infants, Children and Pregnant Women

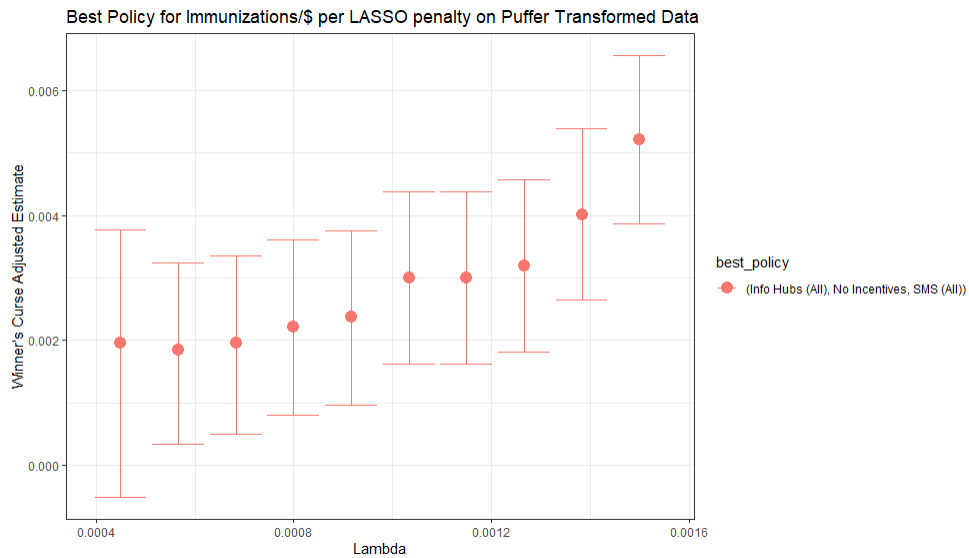| Vaccine	When to give	Dose	Route	Site
For Pregnant Women				
TT-1	Early in pregnancy	0.5 ml	Intra-muscular	Upper Arm
TT-2	4 weeks after TT-1*	0.5 ml	Intra-muscular	Upper Arm
TT- Booster	If received 2 TT doses in a pregnancy within the last 3 yrs*	0.5 ml	Intra-muscular	Upper Arm
For Infants				
BCG	At birth or as early as possible till one year of age	0.1ml (0.05ml until 1 month age)	Intra-dermal	Left Upper Arm
Hepatitis B - Birth dose	At birth or as early as possible within 24 hours	0.5 ml	Intra-muscular	Antero-lateral side of mid-thigh
OPV-0	At birth or as early as possible within the first 15 days	2 drops	Oral	Oral
OPV 1, 2 & 3	At 6 weeks, 10 weeks & 14 weeks (OPV can be given till 5 years of age)	2 drops	Oral	Oral
Pentavalent 1, 2 & 3	At 6 weeks, 10 weeks & 14 weeks (can be given till one year of age)	0.5 ml	Intra-muscular	Antero-lateral side of mid-thigh
Rotavirus#	At 6 weeks, 10 weeks & 14 weeks (can be given till one year of age)	5 drops	Oral	Oral
IPV	Two fractional dose at 6 and 14 weeks of age	0.1 ml	Intra dermal two fractional dose	Intra-dermal: Right upper arm
Measles /MR 1st Dose$	9 completed months-12 months. (can be given till 5 years of age)	0.5 ml	Sub-cutaneous	Right upper Arm
JE - 1**	9 completed months-12 months.	0.5 ml	Sub-cutaneous	Left upper Arm
Vitamin A (1st dose)	At 9 completed months with measles-Rubella	1 ml (1 lakh IU)	Oral	Oral
For Children				
DPT booster-1	16-24 months	0.5 ml	Intra-muscular	Antero-lateral side of mid-thigh
Measles/ MR 2nd dose $	16-24 months	0.5 ml	Sub-cutaneous	Right upper Arm
OPV Booster	16-24 months	2 drops	Oral	Oral
JE-2	16-24 months	0.5 ml	Sub-cutaneous	Left Upper Arm
Vitamin A*** (2nd to 9th dose)	16-18 months. Then one dose every 6 months up to the age of 5 years.	2 ml (2 lakh IU)	Oral	Oral
DPT Booster-2	5-6 years	0.5 ml.	Intra-muscular	Upper Arm
TT	10 years & 16 years	0.5 ml	Intra-muscular	Upper Arm

- *Give TT-2 or Booster doses before 36 weeks of pregnancy. However, give these even if more than 36 weeks have passed. Give TT to a woman in labour, if she has not previously received TT.
- **JE Vaccine is introduced in select endemic districts after the campaign.
- *** The 2nd to 9th doses of Vitamin A can be administered to children 1-5 years old during biannual rounds, in collaboration with ICDS.
- #Phased introduction, at present in Andhra Pradesh, Haryana, Himachal Pradesh and Orissa from 2016 & expanded in Madhya Pradesh, Assam, Rajasthan, and Tripura in February 2017 and planned in Tamil Nadu & Uttar Pradesh in 2017.
- $ Phased introduction, at present in five states namely Karnataka, Tamil Nadu, Goa, Lakshadweep and Puducherry. (As of Feb' 2017)

FIGURE E.1. National Immunization Schedule for Infants, Children, and Pregnant Women.

Data collection activity	Survey Sample	Timeline	Purpose
PHC Survey	161 PHCs across 7 districts	March – April 2015	Sampling PHCs for the study
Census for Baseline	328,058 households in 912 sample villages	September – November 2015	Identifying sampling frame for baseline survey
Baseline Survey	17,000 children in 912 sample villages	May – July 2016	Collecting baseline data
Nominations Survey	15.504 households in 912 sample villages	December 2015 – February 2016	Identifying and sampling seeds
Seeds Survey	2,117 nominated respondents	June – August 2016	Obtaining consent and information on seeds
Child verification survey		June 2017 – March 2018	Verify information entered by ANMs on tablets
New Census	~120,000 households, in 200 sample villages	February – April 2018	Identifying sampling frame for endline survey
Revised Endline Survey	4000 children in 200 sample villages	May – June 2018	Collecting outcome data

FIGURE E.2. Overview of Survey Data Collection Activities.

Appendix F. Substitution Patterns

Table F.1. Incentive Treatment Effects for Non-Tablet Children from Endline Data

	Dependent variable:						
	At Least 2	At Least 3	At Least 4	At Least 5	At Least 6	At Least 7	Measles 1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High Slope	−0.158	−0.052	−0.076	−0.196	−0.187	−0.027	−0.135
	(0.062)	(0.072)	(0.093)	(0.106)	(0.101)	(0.105)	(0.108)
High Flat	−0.021	−0.024	−0.091	−0.078	−0.053	0.102	0.185
	(0.088)	(0.063)	(0.078)	(0.155)	(0.152)	(0.167)	(0.143)
Low Slope	0.090	0.175	0.104	−0.026	−0.152	−0.079	0.051
	(0.064)	(0.060)	(0.085)	(0.099)	(0.080)	(0.077)	(0.100)
Low Flat	0.004	0.069	−0.010	−0.110	−0.005	−0.079	−0.102
	(0.076)	(0.096)	(0.122)	(0.176)	(0.160)	(0.173)	(0.148)
Control Mean	0.69	0.54	0.4	0.31	0.17	0.11	0.39
Total Obs.	1179	1165	1165	1042	1042	706	613
Zeros Replaced	0	0	0	0	0	0	0

Note: Specification includes District Fixed Effects, and a set of controls for seeds and reminders. Control mean shown in levels, and standard errors are clustered at the SC Level

Table F.2. Seeds Treatment Effects for Non-Tablet Children from Endline Data

	Dependent variable:						
	At Least 2	At Least 3	At Least 4	At Least 5	At Least 6	At Least 7	Measles 1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Random	−0.101	−0.045	0.0003	−0.143	0.015	0.058	−0.017
	(0.059)	(0.066)	(0.088)	(0.122)	(0.102)	(0.102)	(0.101)
Information Hub	−0.040	−0.121	−0.025	−0.112	0.018	−0.105	−0.092
	(0.082)	(0.080)	(0.113)	(0.135)	(0.123)	(0.073)	(0.119)
Trusted	0.034	−0.033	0.111	0.027	0.147	−0.011	0.106
	(0.070)	(0.073)	(0.100)	(0.128)	(0.118)	(0.107)	(0.111)
Trusted Information Hub	−0.103	−0.082	−0.031	−0.209	−0.099	−0.057	−0.344
	(0.075)	(0.079)	(0.100)	(0.113)	(0.086)	(0.082)	(0.099)
Control Mean	0.78	0.66	0.5	0.58	0.3	0.22	0.62
Total Obs.	469	461	461	389	389	251	231
Zeros Replaced	0	0	0	0	0	0	0

Note: Specification includes District Fixed Effects, and a set of controls for incentives and reminders. Contro shown in levels, and standard errors are clustered at the SC Level

TABLE F.3. Reminders Treatment Effects for Non-Tablet Children from Endline Data

	Dependent variable:						
	At Least 2	At Least 3	At Least 4	At Least 5	At Least 6	At Least 7	Measles 1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
33%	0.079	0.093	0.070	0.033	0.019	−0.138	0.011
	(0.058)	(0.066)	(0.085)	(0.104)	(0.094)	(0.080)	(0.086)
66%	0.031	0.096	0.042	−0.074	−0.073	−0.044	−0.097
	(0.053)	(0.055)	(0.069)	(0.091)	(0.079)	(0.067)	(0.084)
Control Mean	0.64	0.48	0.34	0.28	0.19	0.13	0.46
Total Obs.	1179	1165	1165	1042	1042	706	613
Zeros Replaced	0	0	0	0	0	0	0

Note: Specification includes District Fixed Effects, and a set of controls for seeds and incentives. Control mean shown in levels, and standard errors are clustered at the SC Level

Appendix G. Data Validation

A household survey was conducted to monitor program implementation at the child-level—whether the record entered in the tablet corresponded to an actual child, and whether the data entered for this child was correct. This novel child verification exercise involved J-PAL field staff going to villages to find the households of a set of randomly selected children which, according to the tablet data, visited a session camp in the previous four weeks. Child verification was continuous throughout the program implementation, and the findings indicate high accuracy of the tablet data. We sampled children every week to ensure no additional vaccine was administered in the lag between them visiting the session camp and the monitoring team visiting them. Data entered in the tablets was generally of high quality. There were almost no incidences of fake child records, and the child's name and date of birth were accurate over 80% of the time. For 71% of children the vaccines overlapped completely (for all main vaccines under age of 12 months). Vaccine-wise, on average, 88% of the cases had matching immunization records. Errors seem genuine, rather than coming from fraud: they show no systematic pattern of inclusion or exclusion and are no different in any of the treatment groups.

Appendix H. Baseline Statistics

Table H.1. Selected Baseline Statistics of Haryana Immunization

	Population-Weighted Average
Baseline Covariates–Demographic Variables	
(Village Level Averages)	
Fraction participating in Employment Generating Schemes	0.045
Fraction Below Poverty Line (BPL)	0.187
Household Financial Status (on 1-10 scale)	3.243
Fraction Scheduled Caste-Scheduled Tribes (SC/ST)	0.232
Fraction Other Backward Caste (OBC)	0.21
Fraction Hindu	0.872
Fraction Muslim	0.101
Fraction Christian	0.001
Fraction Buddhist	0
Fraction Literate	0.771
Fraction Unmarried	0.05
Fraction of Adults Married (living with spouse)	0.504
Fraction of Adults Married (not living with spouse)	0.002
Fraction of Adults Divorced or Seperated	0.001
Fraction Widow or Widower	0.039
Fraction who Received Nursery level Education or Less	0.17
Fraction who Received Class 4 level Education	0.086
Fraction who Received Class 9 level Education	0.158
Fraction who Received Class 12 level Education	0.223
Fraction who Received Graduate or Other Diploma level Education	0.081
Baseline Covariates–Immunization History of Older Cohort	
(Village Level Averages)	
Number of Vaccines Administered to Pregnant Mother	2.271
Number of Vaccines Administered to Child Since Birth	4.23
Fraction of Children who Received Polio Drops	0.998
Number of Polio Drops Administered to Child	2.989
Fraction of Children who Received an Immunized Card	0.877
Number of Observations	
Villages	903

Appendix I. Information Hub Questions

(1) Random seeds: In this treatment arm, we did not survey villages. We picked six ambassadors randomly from the census.

(2) Information hub seed: Respondents were asked to identify who is good at relaying information.

   We used the following script to ask the question to the 17 households:

   "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 neighbourhood in the village) who when they say something many people get to know?âĂİ

(3) "Trust" seed: Respondents were asked to identify those who are generally trusted to provide good advice about health or agricultural questions (see appendix for script)

   We used the following script to elicit who they were:

   "Who are the people in this village that you and many villagers trust, both within and outside this neighbourhood? 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 neighbourhood in the village) and are trusted?"

(4) "Trusted information hub" seed: Respondents were asked to identify who is both trusted and good at transmitting information

   "Who are the people in this village, both within and outside this neighbourhood, 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 neighbourhood in the village) who when they say something many people get to know and are trusted by you and other villagers?"