Econometrica, Vol. 93, No. 4 (July, 2025), 1183–1223

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

ABHIJIT BANERJEE
Dept. of Economics, MIT, JPAL, and NBER

ARUN G. CHANDRASEKHAR
Dept. of Economics, Stanford, JPAL, and NBER

SURESH DALPATH
Public Health Planning, Policy, and M&E, Health Department, Govt. of Haryana

ESTHER DUFLO
Dept. of Economics, MIT, JPAL, and NBER

JOHN FLORETTA
JPAL

MATTHEW O. JACKSON
Dept. of Economics, Stanford and Santa Fe Institute

HARINI KANNAN
JPAL

FRANCINE LOZA
JPAL

ANIRUDH SANKAR
Dept. of Economics, Stanford

ANNA SCHRIMPF
JPAL

MAHESHWOR SHRESTHA
The World Bank

Abhijit Banerjee: banerjee@mit.edu
Arun G. Chandrasekhar: arungc@stanford.edu
Suresh Dalpath: sureshdalpath@yahoo.com
Esther Duflo: eduflo@mit.edu
John Floretta: jfloretta@povertyactionlab.org
Matthew O. Jackson: jacksonm@stanford.edu
Harini Kannan: harini.kannan@ifmr.ac.in
Francine Loza: f.n.loza@gmail.com
Anirudh Sankar: asankar@stanford.edu
Anna Schrimpf: schrimpf@povertyactionlab.org
Maheshwor Shrestha: mshrestha1@worldbank.org

We are particularly grateful to the Haryana Department of Health and Family Welfare for taking the lead on this intervention and allowing the evaluation to take place. Among many other, we acknowledge the tireless support of Rajeev Arora, Amneet P. Kumar, Sonia Trikha, V.K. Bansal, Sube Singh, and Harish Bisht. We are also grateful to Isaiah Andrews, Adel Javanmard, Adam McCloskey, Tyler McCormick, and Karl Rohe for

Policymakers often choose a policy bundle that is a combination of different interventions in different dosages. We develop a new technique—*treatment variant aggregation* (TVA)—to select a policy from a large factorial design. TVA pools together policy variants that are not meaningfully different and prunes those deemed ineffective. This allows us to restrict attention to aggregated policy variants, consistently estimate their effects on the outcome, and estimate the best policy effect adjusting for the winner's curse. We apply TVA to a large randomized controlled trial that tests interventions to stimulate demand for immunization in Haryana, India. The policies under consideration include reminders, incentives, and local ambassadors for community mobilization. Cross-randomizing these interventions, with different dosages or types of each intervention, yields 75 combinations. The policy with the largest impact (which combines incentives, ambassadors who are information hubs, and reminders) increases the number of immunizations by 44% relative to the status quo. The most cost-effective policy (information hubs, ambassadors, and SMS reminders, but no incentives) increases the number of immunizations per dollar by 9.1% relative to the status quo.

KEYWORDS: Factorial designs, heterogeneous treatment effects, immunization, nudges, regularization, social networks, winner's curse.

1. INTRODUCTION

IN MANY SETTINGS, policymakers have to select the best policy among potential bundles that combine several interventions, each with different possible dosages or varieties. Similarly, in medicine, a particular treatment regimen may combine several drugs in different potential dosages. For example, the management of HIV-AIDS was revolutionized in the mid-1990s by the combination of two or three drugs in subtle dosages, the famous "AIDS Cocktail."

In this paper, we consider the problem of the state government in Haryana, India, which was looking to choose a new bundle of interventions to increase children's immunization coverage.[1] Immunization is recognized as one of the most effective and cost-effective ways to prevent illness, disability, and death. Yet nearly 20 million children under the age of one do not receive critical immunizations each year (UNICEF and WHO (2019)). In 2016, 7 million of these children were in India, where only 60% of children received the basic set of vaccination within one year of life. Though resources directed towards immunization have increased steadily, there is mounting evidence that insufficient parental demand has contributed to stagnating immunization rates (WHO (2019)). Based on the existing research, the options considered were small incentives to parents, social network mobilization, and SMS reminders. The government also needed to determine the level and slope of incentives, the set of people to mobilize, and the intensity of the SMS reminder campaigns.

The ideal strategy, if time and implementation capacities were not constraints, would be to experiment iteratively in the context until the best bundle is found. There is a growing

helpful discussions. We thank Emily Breza, Denis Chetvetrikov, Paul Goldsmith-Pinkham, Ben Golub, Nargiz Kalantarova, Shane Lubold, Douglas Miller, Francesca Molinari, Suresh Naidu, Eric Verhoogen, and participants at various seminars for suggestions. Financial support from USAID DIV, 3iE, J-PAL GPI, and MIT-Givewell, and NSF Grant SES-2018554 is gratefully acknowledged. Chandrasekhar is grateful to the Alfred P. Sloan foundation for support. We thank Rachna Nag Chowdhuri, Diksha Radhakrishnan, Mohar Dey, Maaike Bijker, Niki Shrestha, Tanmayta Bansal, Sanjana Malhotra, Aicha Ben Dhia, Vidhi Sharma, Mridul Joshi, Rajdev Brar, Paul-Armand Veillon, and Chitra Balasubramanian, as well as Shreya Chaturvedi, Vasu Chaudhary, Shobitha Cherian, Louis-Mael Jean, Anoop Singh Rawat, Elsa Trezeguet, Simon Yaspo, and Meghna Yadav for excellent research assistance.

[1]Appendix A is part of this article, while all Supplemental Appendices can be found in the working paper (Banerjee et al. (2021b)).

literature on how to conduct and analyze adaptive trials (Hadad, Hirshberg, Zhan, Wager, and Athey (2021), Kasy and Sautmann (2021), Zhan, Ren, Athey, and Zhou (2021)). However, it is often not possible to conduct such sequential trials: the window for experimentation may be short before a policy must be chosen, or a decision on a treatment regimen or vaccine must be made quickly because of an urgent health threat. In such cases, the only option may be to conduct a large-scale experiment that simultaneously tests many different policy bundles. This was certainly true in the case of HIV-AIDs, where there were huge pressures to rapidly identify and approve a treatment. In Haryana, it was possible to conduct a single experiment with multiple treatments, in over 900 villages.

However, with several alternative interventions and multiple possible dosages, there is an enormous number of potential combinations, each of which is a unique policy bundle. Altogether, in Haryana, there were 75 possible bundles of interventions with different dosages. There is no clear guidance in the literature on how to design and analyze such trials when the number of potential options is large.

We fill this gap by developing a methodology for *treatment variant aggregation* (TVA): a principled algorithm that pools together policy variants that have similar impact and prunes ineffective policy options. This reduces the dimensionality of the problem, and enables identification of the best overall combination and consistent estimation of its impact. This methodology allows us to solve two problems that arise when evaluating large numbers of candidate policy bundles (henceforth we use "policy" as a short-hand for "policy bundle").

First, the researcher must decide how many and which potential policies to include in an experimental design, and how to analyze the trial. One approach, recommended in McKenzie (2019), is to only include a limited number of bundles. For example, the ACTG-320 trials compared the three-drug cocktail (protease inhibitor plus two nucleosides), in one specific dosage, to the two nucleosides in the same dosage (Hammer et al. (1997)). However, the optimality of such an approach presumes an "oracle property": that the researcher or policymaker *already knows* which policies are worth comparing. We consider situations in which any of many policies could turn out to be optimal, and the researcher would like to choose among all unique policy bundles in a fully-saturated factorial design. This reduces power since each policy may only be observed on a small sample; therefore, to increase precision, researchers often attempt to pool policy bundles ex post based on observed outcomes. Without specific structure to the problem, however, this can be misleading in finite samples, especially when interaction effects are small, so that a test against zero has limited power, but are not quite zero (this is the "local-to-zero" problem; see Muralidharan, Romero, and Wuthrich (2019)). Thus, we need to find minimal and realistic assumptions on the inferential environment that enable a principled, data-driven approach to reducing the dimensionality of the problem.

The second problem is that the impact of a policy that is estimated to be the best out of a set of K unique policies can be overestimated when K is large, due to the "winner's curse" (Andrews, Kitagawa, and McCloskey (2021)). Some policy $k^\star$ could have the highest estimated impact partially due to a high true effect, but could also partially be due to randomness. Conditional on being the best in the data, some of the estimated impact is likely due to randomness. As a result, the expected impact of policy $k^\star$ is overestimated and the statistician must adjust for it.

Our main methodological contribution takes place before any estimation: it is a method for mapping a large number of unique policies into a small number of bundles in a data-driven way. We develop a tool (a specific Hasse diagram) to represent a complicated factorial design while incorporating what the researcher knows about the structure of the

problem, and enable treatment variant aggregation. We argue that this satisfies conditions required to apply appropriate existing regularization estimators (the Puffer LASSO) to collapse dosages into a smaller number of intervention bundles. This innovation thus allows us to combine existing methods to develop estimators that are consistent and asymptotically normally distributed for a large factorial design. In the case of the interventions to maximize the number of measles immunizations in Haryana, this step reduces it from 75 candidate policies to 4.

The statistical setting that we analyze is as follows. There are M possible intervention arms, R possible "dosages" per arm (including zero), and therefore $K = R^M$ possible policies. The policymaker is uncertain about which policies are effective. However, it can be that, in some circumstances, an *incremental* dosage change on some arm does not have a meaningful effect on the outcome, for some other combination of other arms. For example, if there are three interventions to increase immunization demand (x, y, z), and two adjacent dosages z and z' for the third intervention (e.g., SMS are sent to 33% of the parents, or to 66% of the parents), then it is possible that the bundles (x, y, z) and (x, y, z') are equally effective for some particular choices of x and y. We can thus pool those two policies, (x, y, z) and (x, y, z'), and treat them as one for all practical purposes. The policymaker conducts *treatment variant aggregation* (TVA). This *pools* together policy variants that are not meaningfully different (e.g., (x, y, z) and (x, y, z') are pooled together as above) and *prunes* all the combinations that are ineffective (those that pool with the null policy). TVA allows us to restrict attention to aggregated policy variants and only those that matter, which can improve estimation. We discuss how we can use TVA to consistently estimate policy effects and estimate the best policy effect adjusting for the winner's curse. We proceed in several steps.

The first step is to represent the fully-saturated factorial regression of the outcome on unique policies in terms of another, equivalent specification that tracks the effects of incremental dosages. TVA utilizes a Hasse diagram lattice of policy variants to deduce how zeros in the marginal effects determines pruning and pooling of variants.

To fix ideas, consider a simple example with two arms ($M = 2$) and two nonzero dosages for each ($R = 3$), yielding $K = 9$ unique policies. So each arm can either be used or not, and used in either a low or high dose. Let us represent these by $[T_1, T_2] = [0, 0]$, $[0, 1]$, $[0, 2]$, $[1, 0]$, etc., where the entries are the corresponding treatment levels with 0 being not used, and 1 being low and 2 high dosage. A standard regression would just have a dummy variable for each particular policy combination $[T_1, T_2]$, and then a corresponding coefficient $\beta_{[T_1,T_2]}$. An alternative representation breaks these into marginal effects:

$$\begin{aligned} y = {} & \alpha_{[0,0]} + \alpha_{[1,0]} \cdot 1\{T_1 \geq 1\} \cdot 1\{T_2 = 0\} + \alpha_{[0,1]} \cdot 1\{T_1 = 0\} \cdot 1\{T_2 \geq 1\} \\ & + \alpha_{[2,0]} \cdot 1\{T_1 = 2\} \cdot 1\{T_2 = 0\} + \alpha_{[0,2]} \cdot 1\{T_1 = 0\} \cdot 1\{T_2 = 2\} \\ & + \alpha_{[1,1]} \cdot 1\{T_1 \geq 1\} \cdot 1\{T_2 \geq 1\} + \alpha_{[2,1]} \cdot 1\{T_1 = 2\} \cdot 1\{T_2 \geq 1\} \\ & + \alpha_{[1,2]} \cdot 1\{T_1 \geq 1\} \cdot 1\{T_2 = 2\} + \alpha_{[2,2]} \cdot 1\{T_1 = 2\} \cdot 1\{T_2 = 2\} + \epsilon. \end{aligned} \tag{1}$$

In this specification, the $\alpha_{[r_1,r_2]}$ are all marginal effects, and hence, inspecting the vector α and checking which $\alpha_{[r_1,r_2]} = 0$ tells us which adjacent policies can be pooled together, and which ones can be pruned (pooled with the null policy; for instance, if $\alpha_{[1,0]} = 0$). In a general factorial design of K unique policies, we have regressors of the form $1\{T_1 \geq r_1, T_2 \geq r_2, \ldots, T_M \geq r_M\}$ for treatment arm intensities T_m and thresholds r_m for arm m, with $K - 1$ regressors plus an intercept. At every stage, we ask whether an incremental increase in dosage for a given arm of some policy causes a marginal change. That is, we check for zero effects: $\alpha_{[r_1,\ldots,r_M]} = 0$ for some or multiple $[r_1, \ldots, r_M]$.

This approach makes use of the researcher's a priori knowledge of which policies can be pooled: these are policies that are dosage variations of the same treatment *profile*, or underlying policy type. Therefore, it places discipline on the problem. It ensures we are not mis-naming pooled choices by pooling non-comparable policy bundles, which is the issue implicitly raised in Muralidharan, Romero, and Wuthrich (2019).[2] We assume that when there are nonzero marginal effects, those effect sizes are large enough—assuming away the local-to-zero range—so that we may actually discover and make inferences about the best policy combinations. Our approach works when these assumptions to allow regularization are palatable. When the assumptions, or reasonable relaxations, cannot be justified, sequential testing, rather than simultaneous testing with post-selection processing, is unavoidable.[3]

Our goal is to identify the support (the set of nonzero coefficients) of the regression equation (1). Under our maintained assumptions, a natural way to do this is to use LASSO. This requires an extra step, however, since the regressors in equation (1) are typically strongly correlated. For instance, $1\{T_1 = 2\} \cdot 1\{T_2 = 0\}$ implies $1\{T_1 \geq 1\} \cdot 1\{T_2 = 0\}$. In fact, the marginal effects specification may fail the necessary and sufficient condition for LASSO support consistency, of "irrepresentability," which requires that regressors are not too correlated (Zhao and Yu (2006)). Thus, the second step is to apply the *Puffer transformation* to the variables to which LASSO is being applied (Rohe (2014), Jia and Rohe (2015)). This de-correlates the design matrix that comes from (1). We show that the specific structure of the RCT makes it particularly suitable for this technique.

Once LASSO has been applied on the Puffer-transformed variables to consistently estimate the marginal effects support, the third step is to reconstruct a set of unique policies taking into account the pooling and pruning implied by the LASSO results.

The fourth step is to estimate OLS on the new set of unique policies, post-selection. Using an argument adapted from Javanmard and Montanari (2013), we show that this estimator is consistent and asymptotically normally distributed.[4]

This vector is of independent interest: these are the estimates and confidence interval of all the relevant policies. There is an optional fifth step, in case the policymaker is interested in identifying and getting estimate for a single "best" policy. This step is to re-estimate the effect of the best pooled and pruned policy, adjusting for the winner's curse (Andrews, Kitagawa, and McCloskey (2021)). There are three advantages of conducting this adjustment post-TVA rather than on the full factorial design. First, when there are fewer potential alternatives to the best policy $k^\star$, the odds of picking the best by chance are lower. In fact, in many cases, the winner's curse adjustment may not be necessary, when there are sufficiently few policy bundles that survive at the TVA step. Second, with fewer alternatives, it is less likely that the second best alternative has an effect that is similar to the $k^\star$ effect, which will reduce the shrinkage penalty. Third, there is the benefit of

[2]Specifically, Muralidharan, Romero, and Wuthrich (2019) took issue with "short models" such that, for example, what is claimed as the effect of $(x, 0, 0)$ actually also includes some of the effect of $(x, y, 0)$. In this sense, the treatment is "mis-named." In TVA, the policy $(x, y, 0)$ is considered to be a categorically different treatment type from $(x, 0, 0)$ for $x, y > 0$. More generally, the pooled policy names always unambiguously indicate which unique policy combinations are pooled together.

[3]In practice, we show in Supplemental Appendix E.3 through simulations that we may relax the local-to-zero assumption in several directions and still retain strong performance for this final objective.

[4]The convergence in distribution is not uniform (in the parameter space) (Leeb and Pötscher (2005)). Nevertheless, asymptotic normality holds pointwise (in the parameter space)—essentially, in our setting, the non-uniformity does not have much bite since incorrect selection of the high-effect policies happens with probability tending rapidly to zero.

coherence: if two policies can be pooled, there is no point in applying a strong shrinkage penalty because of a competition between them.

We apply this method to the large-scale immunization experiment that we conducted in Haryana, India, from December 2016 to November 2017 in collaboration with the government of Haryana, which was interested in selecting the best policy for full-scale adoption in the state. To stimulate demand for immunization, a large literature has found the effectiveness of "nudges," including conventional ones such as small cash or in-kind incentives,[5] SMS reminders,[6] as well as more novel interventions such as symbolic social rewards[7] or using influential individuals in a social network as "ambassadors."[8] We cross-randomized three arms with different nudges that had shown some promise in earlier work: (1) monetary incentives, (2) SMS reminders, and (3) seeding ambassadors. Incentives came in two types (linear and convex) with two dosages each (low and high in terms of value). SMS reminders had two dosages. Either 33% or 66% of caregivers received SMS reminders (and voicemails) about the next scheduled vaccination. Ambassadors were either randomly selected or selected through a nomination process. The nomination process was done in three ways, one of which came in two dosages (Information Hub). All together, we have 75 unique policies, and 915 villages were at risk for all three treatments taken together.

Applying TVA, we find that when the outcome is the number of measles shots administered, four policies survive as candidate policies, including two with coefficients that are significantly different from zero, both of which involve the combination of ambassadors nominated by the social network, SMS, and incentives. The best policy is to use information hubs and either low or high SMS coverage, in combination with convex incentives that can be either low or high. This increases the number of immunizations by 44% ($p < 0.05$), after accounting for the winner's curse. Choosing the cheapest among these suggests that the policymaker should choose low convex incentives, send SMS to 33% of caregivers, and identify information hubs to relay the message. To maximize the number of immunizations per dollar spent, the best policy is using information hubs along with SMS reminders at 33% or more of caregivers covered. It increases the number of immunizations per dollar by 9.1% ($p < 0.05$) compared to the status quo with no additional intervention. It is the only policy that strictly increases the number of immunizations per dollar spent, and hence in this case the winner's curse adjustment makes essentially no difference (and this step could be omitted).

The results highlight the importance of complementarities that may get lost had a factorial design not been used. Information hubs magnify the effect of other interventions and spark diffusion: neither incentives nor reminders are selected on their own, but are selected when combined with information hubs. Similarly, information hubs are not selected on their own, but are selected when combined with the conventional strategies. This suggests that in cases where there are no strong reasons to rule out interactions a priori, it is important to accommodate them in the design and the statistical analysis.

[5]See Banerjee, Vinayak, Duflo, Glennerster, and Kothari (2010), Bassani, Arora, Wazny, Gaffey, Lenters, and Bhutta (2013), Wakadha et al. (2013), Johri et al. (2015), Oyo-Ita, Wiysonge, Oringanje, Nwachukwu, Oduwole, and Meremikwu (2016), Gibson et al. (2017).

[6]See Wakadha et al. (2013), Domek et al. (2016), Uddin et al. (2016), Regan, Bloomfield, Peters, and Effler (2017).

[7]See Karing (2018).

[8]See Alatas, Chandrasekhar, Mobius, Olken, and Paladines (2019), Banerjee, Chandrasekhar, Duflo, and Jackson (2019).

2. TREATMENT VARIANT AGGREGATION

2.1. *Overview and Setup*

We have a randomized controlled trial of M arms and R ordered dosages ({none, intensity 1, … , intensity $R-1$}). This yields $K := R^M$ unique treatment combinations or *unique policies*. Let $T_{ik} \in \{0, 1\}$ be a dummy variable indicating that unit i is assigned to unique policy k. Unique policies are described as *variants* of each other when they differ only in the (nonzero) dosages of the treatments applied. This implies that two policies differing only in whether some arm is active or inactive (dosage is zero) are not considered variants, as formalized below in Section 2.1.1.

Assuming the same number of dosages per arm is for notational ease and without substantive loss of generality. In practice, the number of dosages per arm can vary.

The *unique policy regression* is given by

$$y_i = T\beta^0 + \epsilon_i. \tag{2}$$

The support of this regression is given by the set of unique policies that have nonzero effect relative to control,

$$S_\beta := \left\{k \in [K] : \left|\beta_k^0\right| \neq 0\right\}.$$

Some of the variants have equivalent effects and ought to be considered as one policy. Some arms may be altogether ineffective and ought to be pruned (i.e., pooled with control). We construct a method of *treatment variant aggregation* (TVA) in order to *pool and prune variants* systematically.

2.1.1. *Treatment Profiles and Policy Variants*

A fundamental concept is a *treatment profile*. This clarifies which unique policies are *variants* of each other and could potentially be pooled with one another (without being pooled with the control).

The treatment profile $P(k)$ of a unique policy k designates which of the M arms are active (having positive dosages), without regard to how high the dosage is. Two unique policies k, k' are variants of each other if and only if $P(k) = P(k')$, that is, exactly the same arms are active for both policies. Thus, K unique policies are categorized into 2^M treatment profiles.

EXAMPLE 1: Consider observation i that has an assigned policy $k =$ (No Ambassador, 33% SMS, low-value flat incentives) and observation j that has an assigned policy $k' =$ (No Ambassador, 66% SMS, low-value flat incentives). Though k and k' are distinct treatment combinations, they share the same treatment profile—$P(k) = P(k')$—of (No Ambassador, Some SMS, Some incentives). Therefore, k and k' are variants. They would not be variants if instead $k' =$ (No Ambassador, 66% SMS, *No incentives*).

2.1.2. *Treatment Variant Aggregation: Pooling and Pruning*

Increasing the dosage in a treatment arm may be inconsequential after a point, and more generally policy variants may have the same impact. Here, we consider a respecification of (2) that explicitly tracks the marginal effect of increasing dosages by

grouping together policy variants that have the same effect on the outcome.[9] When these marginal effects are zero, this means that a set of variants are to be either pooled or pruned (pooled with control).

Let $\mathcal{P}$ denote the set of all partitions of the K policies. Elements of $\mathcal{P}$ comprise every conceivable pooling of the K policies, with generic partition denoted Π. Whether two given policies k and k' are pooled corresponds to whether they are members of the same element of the partition, $\pi \in \Pi$.

Out of the full universe of all conceivable poolings, only some make sense, and we refer to those as the *admissible* poolings. Informally, admissibility says that only policies affected by the same set of nonzero marginal effects *may* be pooled. The admissible pools are then a strict subset $\mathcal{P}_{|\Lambda} \subset \mathcal{P}$. The target $S_{\text{TVA}} \in \mathcal{P}_{|\Lambda}$ is defined to be the maximally admissible pooled and pruned set of policies (i.e., the coarsest partition).

Letting $Z_{S_{\text{TVA}}}$ denote the matrix of indicator variables for the pooled policies, our goal is to estimate the *pruned and pooled policy regression*:

$$y = Z_{S_{\text{TVA}}} \eta^0_{S_{\text{TVA}}} + \epsilon. \tag{3}$$

Comparing (3) with (2), η^0 is the projection coefficients of $T\beta^0$ onto $Z_{S_{\text{TVA}}}$, that corresponds to grouping certain policies, and estimating the parameters for the grouped policies.

Let us make admissibility precise. For a treatment combination k, α^0_k is the marginal effect of the dosages in k within its treatment profile relative to incrementally lower dosages. Formally, the marginal effect α^0 may be defined implicitly so that a policy's effect is the sum of marginal effects from increasing dosages up to its particular dosage profile:

$$\beta^0_k = \sum_{k' \leq k; P(k') = P(k)} \alpha^0_{k'}. \tag{4}$$

Equation (4) can be inverted to recover α^0 in terms of β^0. An explicit expression for its terms α^0_k is more unwieldy in its full generality, but depending on the policy k, it can be a difference between two variants' effects or reflect a complementarity, that is, the interaction effect from combining dosages in different arms. This is consistent with the interpretation that a policy's effect is the main effects of the highest dosages in each arm, considered separately, plus the relevant interaction effects.

For each policy k, consider the set of marginal effects on the right-hand side of β^0_k in (4) that are nonzero: these are the *active marginal effects for* k, $A(k)$. These are the set of marginal effects that "influence" the policy k, either as a main effect or as a complementarity.

ASSUMPTION 1: Π *is an admissible pooling—that is*, $\Pi \in \mathcal{P}_{|\Lambda}$*—if and only if* $k, k' \in \pi$ *implies* $A(k) = A(k')$. *That is, only policies influenced by the same set of active marginal effects may be pooled.*

[9]While sometimes what is "dosage" and "dosage ordering" is readily apparent from the arm, as in the SMS arm of our intervention with saturation levels 33% and 66%, in other cases the researcher has to decide this (of course, this can be pre-specified). For example, in the seeds arm of our intervention, we decided that the information hub ambassador comes in two dosages, with those that are trusted for health advice as the higher dosage.

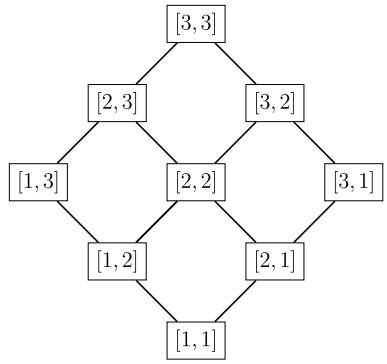


FIGURE 1.—Hasse diagram for $M = 2$, $R = 4$ for the treatment profile where both arms are active. A line upwards from treatment combinations $[r_1, r_2]$ to $[r_1', r_2']$ means that $[r_1, r_2] \leq [r_1', r_2']$ and $[r_1, r_2] \neq [r_1', r_2']$ in the intensity ordering.

It is easy to see, through (4), that admissibility ensures that (i) only policies with equal treatment effects may be pooled, and (ii) only variants may be pooled (or if non-variant policies are being pooled, they must be pooled with the control—the null policy). Of course, given an idiosyncratic instance of unique policy effects β^0, there may be other pools that satisfy (i) and (ii) besides the admissible pools. However, we restrict our attention to admissible pools since these work generally using only the sign of marginal effects.

Per Assumption 1, the more zeros there are in marginal effects, the more pooling choices become admissible. We can depict this in a Hasse diagram for a treatment profile. In a Hasse diagram, a line upwards from variants k to k' implies $k' > k$, and there is no variant k'' such that $k' > k'' > k$ (in the partial order). The running example is the case of a 2-arm treatment of four intensities (three nonzero intensities "low," "medium," "high"); that is, $M = 2$, $R = 4$, and the treatment profile where both arms are "on." Figure 1 depicts the Hasse for this treatment profile. Here, unique policies are named per their intensity representations; that is, $[r_1, r_2]$ where $r_i \in \{1, 2, 3\}$ is the (nonzero) dosage in arm i.

Zeros in the marginals make admissible certain "concatenations" in the Hasse diagram of policies. This is depicted in Figure 2, where the top panels (A–C) depict the zeros in marginal effects and the bottom panels (D–F) depict the maximal admissible policy concatenations they imply (of course, subsets of these concatenations will also be admissible).

In Panel A, $\alpha_{[2,1]} = \alpha_{[3,1]} = 0$, meaning that keeping the intensity fixed as low in arm 2, there is no marginal contribution of increasing the intensity in arm 1. Panel D depicts how this makes admissible the concatenation of policies $\{[1, 1], [2, 1], [3, 1]\}$, and indeed $\beta_{[1,1]} = \beta_{[2,1]} = \beta_{[3,1]}$. The maximal admissible concatenated policy can be called $[1 : 3, 1]$.

In Panel B, $\alpha_{[2,2]} = 0$. $\alpha_{[2,2]} = (\beta_{[2,2]} - \beta_{[2,1]}) - (\beta_{[1,2]} - \beta_{[1,1]})$, there are only main effects in increasing dosages from low to medium intensities in both arms, and no further complementarity. Since furthermore $\alpha_{[1,2]} = 0$, there is no main effect on arm 2 from increasing low to medium intensity. That is, there is only a main effect in arm 1 from increasing low to medium intensities. This main effect is nonzero, since $\alpha_{[2,1]} \neq 0$. In Panel E, the maximal admissible concatenations reflect this: $\{[1, 1], [1, 2]\}$ into $\{1, 1 : 2\}$, and $\{[2, 1], [2, 2]\}$ into $[2, 1 : 2]$.

The main effect of arm 1 makes inadmissible these concatenated blocks from further concatenating. This changes in Panel C, where $\alpha_{[2,1]} = 0$ implies that there is no main

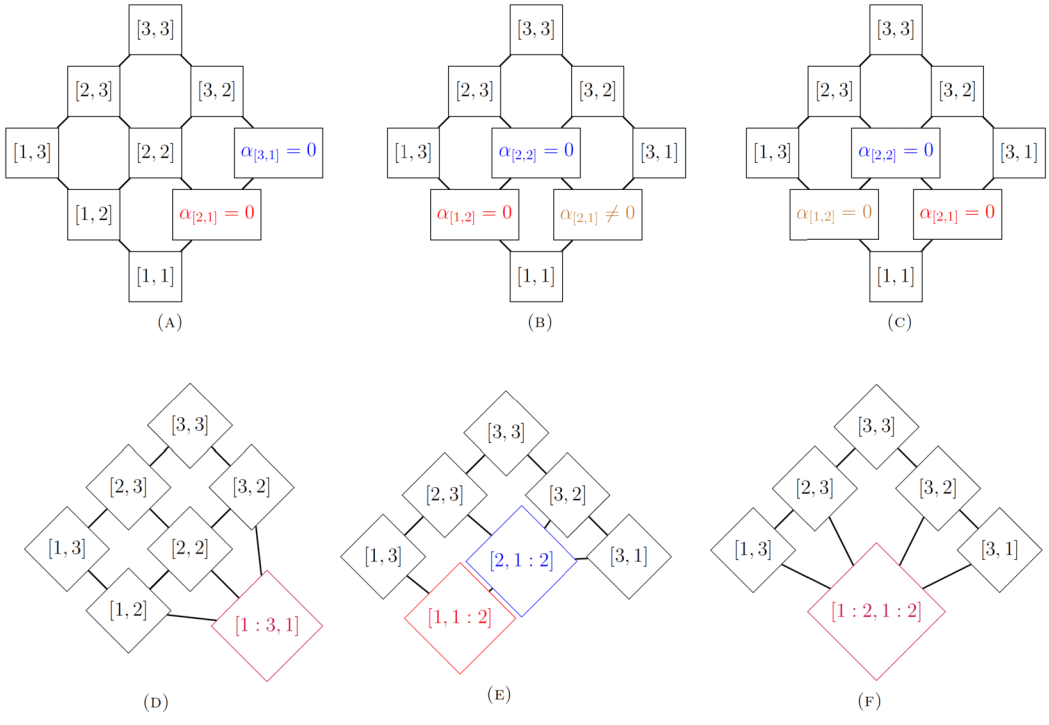


FIGURE 2.—This figure shows how zeros in the marginals induce admissible "concatenations" in the Hasse diagram of policies. Panels A, B, C show examples of zeros in the marginal space and Panels D, E, F show the corresponding policy concatenations in the η-space. On Panel A, $\alpha_{[2,1]} = \alpha_{[3,1]} = 0$. Since $\alpha_{[2,1]} = \beta_{[2,1]} - \beta_{[1,1]}$ and $\alpha_{[3,1]} = \beta_{[3,1]} - \beta_{[2,1]}$, $[1, 1]$, $[2, 1]$, $[3, 1]$ are concatenated on Panel D. On Panel B, $\alpha_{[2,1]} \neq 0 = \beta_{[2,1]} - \beta_{[1,1]}$, $\alpha_{[1,2]} = 0 = \beta_{[1,2]} - \beta_{[1,1]}$, and $\alpha_{[2,2]} = 0 = (\beta_{[2,2]} - \beta_{[2,1]}) - (\beta_{[1,2]} - \beta_{[1,1]})$. The implied poolings are shown on Panel E where policies $\{[1, 1], [1, 2]\}$ and $\{[2, 1], [2, 2]\}$ are concatenated. On Panel C, note that the only change relative to Panel B is that here $\alpha_{[2,1]} = 0$, but the implied concatenation on Panel F yields $\beta_{[1,1]} = \beta_{[1,2]} = \beta_{[2,1]} = \beta_{[2,2]}$.

effect in increasing arm 1 from low to medium either. This makes admissible the concatenation $\{1, 1 : 2\}$ and $[2, 1 : 2]$ into $[1 : 2, 1 : 2]$.

As illustrated through these examples, zeros in α^0 thus show up as admissible policy concatenations in Hasse diagrams. This motivates the *marginal effects regression*:

$$y = X\alpha^0 + \epsilon. \tag{5}$$

This is an invertible transformation of (2). X can be interpreted as indicators

$$X_{i\ell} := \mathbf{1}\{k(i) \geq \ell \cap P(k(i)) = P(\ell)\}.$$

In other words, X assigns for unit i a "1" for all policy variants that share $k(i)$'s treatment profile and are weakly dominated in intensity by $k(i)$ and a "0" otherwise.

The key object of interest is the support of (5):

$$S_\alpha := \{j \in [K] : |\alpha_j^0| \neq 0\}.$$

S_α is the set of all active marginals for any policy, that is, $S_\alpha = \bigcup_k A(k)$. Since $S_{\text{TVA}} \in \mathcal{P}_{|\Lambda}$ is the maximal admissible pooling, it is the tightest pooling using only S_α and nothing else. More precisely, it is the coarsest pooling uniformly consistent over α conditional on S_α.

The idea is to apply a model selection procedure to estimate S_α. In Supplemental Appendix B (Banerjee, Chandrasekhar, Dalpath, et al. (2024)), we show how to construct the unique maximally pooled and pruned set $S_{\text{TVA}} \in \mathcal{P}_{|\Lambda}$ from S_α.[10] The maximality ensures that no contiguous set of intensities thought to have the same treatment effects are left un-pooled.

2.2. *Pooling and Pruning for Support Selection*

The next step is to identify the support S_α. One natural place to start would be to apply LASSO directly to (5). However, this approach can fail to satisfy sign consistency because the marginal effects matrix X may fail an "irrepresentability criterion" which is necessary for consistent estimation (Zhao and Yu (2006)). Irrepresentability bounds the acceptable correlations in the design matrix. Intuitively, it requires that variables that are not in the support are not too strongly correlated with those that are. Otherwise, an irrelevant variable is "representable" by relevant variables, which makes LASSO erroneously select it with nonzero probability irrespective of sample size. We prove by construction that irrepresentability can fail to be satisfied in Supplemental Appendix C, where we also show by simulation that irrepresentability failures can become dramatic with increasing R and M. The structure that we exploit in showing the failure is one in which higher dosage marginals are representable by lower dosage marginals, violating the condition.

A way out is provided by Jia and Rohe (2015). They showed that, under some conditions, one can estimate the LASSO support by transforming the data to recover irrepresentability. They demonstrated that a simple left-multiplication (pre-conditioning) can de-correlate the data (at the expense of inflating variance in the error).

In Proposition 2.1, we demonstrate that in the specific instance of the crossed RCT design with ordered intensities, the pre-conditioning strategy of Jia and Rohe (2015) can be applied because the relevant sufficient conditions are met. Specifically, with an RCT, we can exactly characterize the design matrix and therefore the inflation factor. We can show that the variance inflation cost is tolerable, in the sense that we can consistently recover the support and the treatment effects.

The weighting is constructed as follows. Let us take the singular value decomposition of $X := UDV'$, where U is an $n \times K$ unitary matrix, D is a $K \times K$ diagonal matrix of singular values, and V is a $K \times K$ unitary matrix. The *Puffer transformation*—so named for the fish whose shape is suggested by the geometry of this transformation—is $F := UD^{-1}U'$. The regression of interest is now

$$Fy = FX\alpha + F\epsilon, \tag{6}$$

where $F\epsilon \sim \mathcal{N}(0, UD^{-1}\Sigma D^{-1}U')$. As Jia and Rohe (2015) noted, this satisfies irrepresentability since $(FX)'(FX) = I$, which is sufficient (Jia and Rohe (2015), Bickel, Ritov, and Tsybakov (2009)).

To understand why this works, recall that the matrices U and V' can be thought of as rotations and D as a re-scaling of the principal components. So, the transformation F preserves the rotational elements of X without the re-scaling by D and $FX = UV'$ as its singular value decomposition (with singular values of 1).

The reason this is useful is because when a matrix X has correlation, then the ith singular value of X captures the residual variance of X explained by the ith principal component after partialing out the variance explained by the first $i-1$ principal components.

[10] Following this same procedure with any estimate $\widehat{S}_\alpha$ leads to an estimate $\widehat{S}_{\text{TVA}}$ of pooled and pruned policies.

So, when there is high correlation within X, less than K principal components effectively explain the variation in X and so the later (and therefore lower) singular values shrink to zero. F inflates the lowest singular values of X so that each of the principal components of the transformed FX explains the variance in FX equally. In that sense, FX is de-correlated and, for $K < n$, is 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. Of course, if the amplification is too strong, it can hinder efficiency of LASSO in finite sample and even prevent the sign consistency of LASSO, in the worst case.[11]

Our setting is particularly amenable to the Puffer transformation since the marginal effects design matrices are highly structured. In particular, the assignment probabilities to the various unique treatments are given, and as a result, the correlations with X are bounded away from 1. This has the implication that the minimum 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. In Lemma A.1, we bound the rate at which the minimal singular value of X can go to zero as a function of K in a crossed RCT such as ours, and Proposition 2.2 below relies on this lemma to then prove that the Puffer transformation ensures irrepresentability and consistent estimation by LASSO in our context.

We make the following additional assumptions and discuss their restrictiveness below.

ASSUMPTION 2—Design Growth: $R \geq 3$, $K < n$, *and* $K = O(n^{\gamma})$ *for some* $\gamma \in [0, 1/2)$.[12]

ASSUMPTION 3—Minimal Marginal Effect Size: $|S_{\alpha}| < K$ *and* $\min_{k \in S_{\alpha}} |\alpha_k| > c > 0$ *for* c *fixed in* n.

ASSUMPTION 4—Homoscedasticity: $\epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, *with* $\sigma^2 > 0$ *fixed in* n.

ASSUMPTION 5—Penalty Sequence: *Take a sequence* $\lambda_n \geq 0$ *such that* $\lambda_n = \omega(n^{-\nu})$ *where* $0 < \nu < \frac{1}{2} - \gamma$.[13]

Assumption 2 restricts the growth of the problem, preventing settings with too many treatments relative to observations. Without this assumption, the correct support may not be estimated with probability tending to 1, and the post-estimators may not necessarily be asymptotically normally distributed. In practice, it means that the RCT cannot have cells in the fully saturated treatment design with very few units assigned to that unique treatment combination. Assumption 3 is the conventional LASSO-sparsity assumption

[11] 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)).

[12] This ensures support consistency (Proposition 2.1) at exponential rates. It also implies that $K^2 \log(K) = o(n)$, which allows for post-LASSO inference under a normal distribution (Proposition 2.3). The latter requirement stems from the growth rate of K needing to be tempered for the central limit theorem to operate in this growing parameter regime.

[13] $\omega(\cdot)$ ("small omega" notation, from the same family of notations as "big O" notation) denotes an asymptotically loose lower bound. Formally, $f(n) = \omega(g(n))$ if and only if $\lim_{n \to \infty} \frac{f(n)}{g(n)} = \infty$.

applied to the marginal effects formulation. It imposes that *adjacent policy variants* are either appreciably different or have no difference (i.e., the so-called "beta-min" assumption in the literature). We do not handle the case of local alternatives among adjacent variants, that is, very small yet nonzero differences, but policies that are not variants of each other or are nowhere adjacent are allowed to be local alternatives as discussed in Section 2.5. Assumption 4 places our theory under homoscedastic errors following the literature on Puffer transformation. Extension to heteroscedasticity is left for future work. Finally, Assumption 5 imposes a restriction on the LASSO-penalties, standard in the regularization literature.

PROPOSITION 2.1: *Suppose Assumptions 1–5 hold. Let* $\widetilde{\alpha}$ *be the estimator of* (6) *by LASSO*:

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

Then $\mathrm{P}(\mathrm{sign}(\widetilde{\alpha}) = \mathrm{sign}(\alpha^0)) = 1 - \exp(-\omega(n^{(1-2(\nu+\gamma))}) \to 1.$

In other words, the correct support of (5) is selected with probability tending to 1 exponentially fast in n.

All proofs are in Appendix A unless otherwise noted.

2.3. *Consistency of the TVA Estimator*

Having constructed an estimator $\widehat{S}_\alpha$ of the support S_α, the next step is to use Algorithm 2 in Supplemental Appendix B to construct $\widehat{S}_{\mathrm{TVA}}$, the estimated set of pooled and pruned unique policies, and then estimate policy effects.[14] The regression of interest is (3). We show this estimator is consistent.[15]

PROPOSITION 2.2: *Suppose Assumptions 1–5 hold. Let* $\hat{\eta}_{\widehat{S}_{\mathrm{TVA}}}$ *be the post-Puffer LASSO OLS estimator of* (3) *on support* $\widehat{S}_{\mathrm{TVA}}$. *Then, with probability at least* $1 - 2e^{-\frac{n^{1-2\gamma}\lambda^2}{2\sigma^2}+\gamma\log n} = 1 - e^{-\omega(n^{(1-2(\nu+\gamma))})} \to 1$,

$$\left\|\hat{\eta}_{\widehat{S}_{\mathrm{TVA}}} - \eta^0_{S_{\mathrm{TVA}}}\right\|_\infty \le \sqrt{\frac{\log n}{n^{1-\gamma/2}}}.$$

2.4. *Asymptotic Normality*

The post-Puffer LASSO estimators are asymptotically normally distributed (pointwise) for the following reason. If the correct support, S_{TVA}, were always selected, mechanically the estimators are asymptotically normal.

So, in practice, we need to worry about two errors: (a) the asymptotic distribution of the estimator with some incorrect support being selected, and (b) the asymptotic distribution of the true estimator when the incorrect support is selected. We show in Appendix A that both of these terms are vanishing in our setup.

Intuitively, the second term can be ignored. After all, the true estimator itself is asymptotically normally distributed, so given the very unlikely event of incorrect selection, this

[14] Algorithm 2 constructing $\widehat{S}_{\mathrm{TVA}}$ generalizes the Hasse concatenation examples in Section 2.1.1.

[15] We thank Adel Javanmard for a helpful discussion of the proof.

term is asymptotically negligible. The first term requires more work. But again, one can show that the amount of potential bias accumulated due to selecting the wrong support is slow relative to the rate of actually estimating the wrong support.[16]

Given that these errors can be controlled, we show that the estimator is asymptotically normally distributed. Note that since the parameter vector is of increasing dimension, the asymptotic normality result must be stated slightly differently than in the usual way. The result states that any linear combination of any of the estimated parameters, when normalized properly, converges to the standard normal, which is the infinite-dimensional analog to the Cramer–Wold device (He and Shao (2000)).

PROPOSITION 2.3: *Suppose Assumptions 1–5 hold. Then, for* $\hat{\eta}_{\widehat{S}_{\mathrm{TVA}}}$, *the post-Puffer LASSO selection OLS estimator of* (3) *performed on support* $\widehat{S}_{\mathrm{TVA}}$, *we have*

$$\sqrt{n}c'\big(\hat{\eta}_{\widehat{S}_{\mathrm{TVA}}} - \eta^0_{\widehat{S}_{\mathrm{TVA}}}\big)/\big(\sigma\|c\|\big) \rightsquigarrow \mathcal{N}(0,1)$$

for any $c \in \mathbb{R}^{|\widehat{S}_{\mathrm{TVA}}|}$.

The proof applies the central limit theorem of He and Shao (2000) for the growing number of parameters regime after controlling the events that the wrong support is selected as in Javanmard and Montanari (2013).[17]

It is well known that one cannot uniformly (over the parameter space) build post-selection asymptotic distributions (Leeb and Pötscher (2005, 2008)). This is the subject of much discussion of a larger literature on post-selection inference—interpretations of the post-estimation procedures and its practical function (Berk, Brown, Buja, Zhang, and Zhao (2013), Tibshirani, Taylor, Lockhart, and Tibshirani (2016), Lee et al. (2016)). In our context, several remarks are worth making. First, our claim is about pointwise inference, not uniformity over the parameter space. Second, we have nothing to say conditional on incorrect selection, hence the non-uniformity. Still, no matter what model is selected—even if an incorrect one—since in our setting the regressors are always orthogonal, there is some valid post-selection interpretation in the sense of Berk et al. (2013), but we do not characterize what occurs in the vanishing probability events. Third, as we recover the support with probability tending to 1, and at an exponential rate, in a practical sense the non-uniformity occurs only for very small (local to zero) alternatives in the space of marginals, which are assumed away per Assumption 3.[18] Loosely, recall that the non-uniformity comes up when the probability of correct selection does not go to 1, along a sequence that is local to the event of failed selection. Given the very high rate of correct selection (tending to 1 exponentially fast in n), these unsupported local alternatives must be exceedingly close to the true parameter (the sequence of alternatives converging to the

[16] An entirely different approach would be to use a recent focus in the literature on exact post-selection inference using the observation that the LASSO procedure to select a model generates a polyhedral conditioning set (Lee, Sun, Sun, and Taylor (2016)). This generates a parameter estimator distribution that is a truncated, rather than complete, normal. In our special environment—a correctly specified linear model, sparse parameters, restrictions on shrinkage rate of minimal values of parameters on the support—the truncation points diverge when conditioning on the event that the true model is the estimated model. In the winner's curse context, an analogous point was made in Andrews, Kitagawa, and McCloskey (2021), Proposition 3. This means that the distribution returns to the usual Gaussian. However, we provide a simpler, direct argument where we can calculate the distribution when the correct support is selected and bound the problematic terms in the event of poor selection.

[17] We again thank Adel Javanmard for a helpful discussion of the proof.

[18] We are grateful to Adam McCloskey for pointing this out.

true parameter at very fast rate in n). See analogous discussion in McCloskey (2020) and the discussion of (A.1) in that paper.

Indeed, consistent with the theoretical results, as we will show in Supplemental Appendix E.1, the estimators look normal in practice, indicating that the non-uniformity concerns are likely to not be large in at least many practical cases, in our specific setting. Further, in our setting, since the elements with the highest effects tend to be selected first, and because of orthogonality, in practice the large parameter estimates almost always perform well.

2.5. *The Effect of the Best Policy*

The TVA procedure generates a set of pruned and pooled policies $\widehat{S}_{\text{TVA}}$ with post-LASSO estimates $\hat{\eta}_{\widehat{S}_{\text{TVA}}}$. This full set of pooled estimates is of direct interest to the policymaker. We now propose an optional last step, in case a policymaker is particularly interested in a single "best" policy (based on the sample estimates) and its estimated performance. This policy is the one in $\widehat{S}_{\text{TVA}}$ with the highest post-LASSO estimate:

$$\hat{\kappa}^{\star} = \text{argmax}_{\kappa \in \widehat{S}_{\text{TVA}}} \hat{\eta}_{\widehat{S}_{\text{TVA}},\kappa}.$$

If the true best policy has a (population-level) treatment effect that far exceeds that of the next best policy, then with high probability, $\hat{\kappa}^{\star}$ is that policy and the post-LASSO estimator $\hat{\eta}_{\widehat{S}_{\text{TVA}},\hat{\kappa}^{\star}}$ (and corresponding confidence interval) is its most efficient unbiased estimator. In that case, the policy that emerges is the best policy, and no further adjustment is needed.

However, if there are other non-pooled policies that perform similarly to the true best policy, there can be a bias in the estimated effectiveness of the "best" among them because of the winner's curse (Andrews, Kitagawa, and McCloskey (2021)). Specifically, when another policy $\kappa \in \widehat{S}_{\text{TVA}}$ has a treatment effect close to $\hat{\kappa}^{\star}$, in the sense that $|\eta^0_{\widehat{S}_{\text{TVA}},\hat{\kappa}^{\star}} - \eta^0_{\widehat{S}_{\text{TVA}},\kappa}|$ is small, there is a chance (due to sampling variation) that they may be incorrectly ranked in the sample: that is, $\hat{\eta}_{\widehat{S}_{\text{TVA}},\hat{\kappa}^{\star}} > \hat{\eta}_{\widehat{S}_{\text{TVA}},\kappa}$ even though $\eta^0_{\widehat{S}_{\text{TVA}},\hat{\kappa}^{\star}} < \eta^0_{\widehat{S}_{\text{TVA}},\kappa}$. In these close races, sampling variation can end up determining the estimated best policy. Importantly, even if the ranking is correct, $\hat{\eta}_{\widehat{S}_{\text{TVA}},\hat{\kappa}^{\star}}$ will be biased upwards relative to the true effect $\eta^0_{\widehat{S}_{\text{TVA}},\hat{\kappa}^{\star}}$.

Given that such a winner's curse arises only in close races, and our assumption of sparse marginal effects (Assumption 3) and steps of pruning and pooling, the winner's curse arises primarily when *non-variant* policies have similar performance. For example, Assumption 3 does not prevent the pooled policy (Any information hubs, No SMS, No incentives) from having a similar impact as (No seeds, Any reminders, Any slope incentives), as these emerge from marginals in different treatment profiles, and cannot be pooled. Were these policies to both have similar and near-maximum impacts, then a winner's curse can arise in their horserace.[19]

The relevant notion of "similar effect" is a *local alternative*:

DEFINITION 2.1: Pooled policies κ, κ' are *local alternatives* if there is a constant $r_{\kappa\kappa'}$ fixed in n such that

$$\eta^0_{S_{\text{TVA}},\kappa} = \eta^0_{S_{\text{TVA}},\kappa'} + \frac{r_{\kappa\kappa'}}{\sqrt{n}}. \tag{7}$$

[19]In fact, even within a treatment profile, two pooled policies can still have similar treatment effects, in spite of Assumption 3. It can occur if they are activated by marginals from sufficiently different parts of the Hasse diagram. All these cases are formalized in Proposition G.1 of Supplemental Appendix G.

This definition holds asymptotically, in that local alternatives have similar effects for *any* sample size. As we have shown, there can be several local alternatives when Assumption 3 holds. We assume a fixed number of local alternatives, and that these are the policies with the most (mutually) similar effects.

ASSUMPTION 6: *There are at most* $q < \infty$, *independent of* n, *pairs of local alternatives, that is, pairs* $\kappa, \kappa' \in S_{\text{TVA}}$ *such that* $\eta^0_{S_{\text{TVA}},\kappa} = \eta^0_{S_{\text{TVA}},\kappa'} + \frac{r_{\kappa\kappa'}}{\sqrt{n}}$, *where* $r_{\kappa\kappa'}$ *is fixed in* n. *All other policy pairs are further separated, that is, separated by at least* $\omega(\frac{1}{\sqrt{n}})$.

This nests $q = 0$, when there are no local alternatives. The best policy is then well separated from the next best policy and the post-LASSO estimates $\hat{\eta}_{\hat{S}_{\text{TVA}},\kappa^*}$ do not need any adjustment. One is free to assume this or assess it ex post. For $q > 0$, there is a chance that the best policy is a local alternative to others, so there may be a winner's curse.

We apply the hybrid estimator of Andrews, Kitagawa, and McCloskey (2021) which attenuates the post-LASSO estimate. It balances performance with a small amount of median bias tolerance. Note that the main text of Andrews, Kitagawa, and McCloskey (2021) requires that the estimators are *exactly* jointly normally distributed. They presented two extensions, one for a conditioning event such as model selection and the other for asymptotic normality. They did not work out the case with both issues present, as in our case. So we proceed with the assumption as in their main text, treating $\hat{\eta}_{\hat{S}_{\text{TVA}}}$ as normally distributed (which our simulations support as a reasonable approximation). Extending their work is beyond the scope of this paper.

We can apply Proposition 6 of Andrews, Kitagawa, and McCloskey (2021) directly. This means we can pick two significance levels, $\alpha > \beta$, and use this to characterize confidence intervals and bias for the hybrid estimator. The hybrid estimator will be median unbiased (with a bias bounded by $\beta/2$). The (conditional) confidence interval has coverage $(1-\alpha)/(1-\beta)$. See Supplemental Appendix G and Andrews, Kitagawa, and McCloskey (2021) for details.

The winner's curse-corrected estimators of Andrews, Kitagawa, and McCloskey (2021) are not a cure-all. They cannot correct for ordinal misrankings of policies, and work for case in which the similar effect policies are $\Theta(\frac{1}{\sqrt{n}})$ apart.[20] Nevertheless, we include discussion of such an adjustment because it helps debias estimates and lets us err on the side of conservative inference. The hybrid estimator has the appealing property that it rapidly converges to the post-LASSO estimate whenever the best policy is well separated from the next best policy in the sample. We can thus view it as insurance whenever we estimate the best policy, letting the data decide if the winner's curse risk is appreciable and automatically correcting for it to the extent possible when it arises.

2.6. *Summary of TVA*

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

3. SIMULATION PERFORMANCE

Here, we run simulations in the environment described in Section 2—namely, when a sparse set of policies have meaningful and meaningfully different impacts. In Sections 3.1

[20]Note that conditional on correct support selection $\widehat{S}_{\text{TVA}} = S_{\text{TVA}}$, the post-LASSO estimates of the best policy, even if it is selected incorrectly, must be $\Theta(1/\sqrt{n})$ of the true best policy effect.

Algorithm 1 Estimating Treatment Effects by Treatment Variant Aggregation.

1. Given treatment assignment matrix T, calculate the treatment profile and marginal dosage intensity matrix X.
2. Estimate $\widehat{S}_\alpha := \{j \in [K] : |\widetilde{\alpha}_j| \neq 0\}$ by estimating (5) through a Puffer transformed LASSO.
3. Calculate marginal effects support $\widehat{S}_{\text{TVA}}$ from $\widehat{S}_\alpha$ using Algorithm 2 in Supplemental Appendix B.
4. Estimate pooled and pruned treatment effects of unique (relevant) policies, $\hat{\eta}_{\widehat{S}_{\text{TVA}}}$, using regression (3).
5. Optional: To estimate the best policy in $\widehat{S}_{\text{TVA}}$, select $\hat{\kappa}^\star = \text{argmax}_{\kappa \in \widehat{S}_{\text{TVA}}} \hat{\eta}_{\widehat{S}_{\text{TVA}},\kappa}$. Either report the OLS estimate $\hat{\eta}_{\widehat{S}_{\text{TVA}},\kappa^\star}$, or, if anticipating a winner's curse, construct the hybrid estimator $\hat{\eta}^{\text{hyb}}_{\widehat{S}_{\text{TVA}},\hat{\kappa}^\star}$ with nominal size α and median bias tolerance $\beta/2$.

and 3.2, we outline the simulation setup and performance measures, and show in Section 3.3 that TVA outperforms several other standard approaches. The relative deficiencies of other estimators also highlight the features that give TVA its edge.

3.1. *Simulation Setup*

The idea of the simulation setup is to generate simulated design matrices from marginal specifications (5) that resemble the data, score these on certain metrics, and aggregate these scores into measures of performance for sample size n:

1. Fix $R = (5, 5, 3)$, $M = 3$, and $\sigma = 2.3$: parameters are chosen to loosely mimic our experiment where three treatment arms have asymmetric intensities leading to 75 unique policies and where σ is chosen such that the R^2 of the post-LASSO regression matches the experiment for a similar sample size.
2. The simulation results are plots of performance $\hat{m}(n)$ against sample size n, where n ranges between 1000 and 10,000.[21]
3. These scores $\hat{m}(n)$ are generically computed as follows:
   (a) A set $\mathcal{C}$ of true supports of the marginal specification (5) is randomly chosen. Each member $S^i_\alpha \in \mathcal{C}$ is a particular support or "configuration" and each configuration has fixed support size $|S^i_\alpha| = M$. Specifically, each configuration $\mathcal{C}$ is constructed by randomly sampling M covariates of X. Furthermore, if $S^i_\alpha = (k_1, k_2, \ldots, k_M)$ in some given order, we assign coefficients $\alpha_{k_j} = 1 + 4 \cdot \frac{j-1}{M-1}$. That is, these nonzero coefficients are linearly spaced between 1 and 5. Thus, each configuration fully specifies the set of coefficients α for (5).
   (b) For each $S^i_\alpha \in \mathcal{C}$, a set $\mathcal{S}_{S^i_\alpha}(n)$ of simulations (design matrices) is generated based on the coefficients specified by the configuration, and the Gaussian noise, with sample size n. For each simulation $\hat{s}(n) \in \mathcal{S}_{S^i_\alpha}(n)$, it is scored by a metric $m(\hat{s}(n))$ that will be specified.
   (c) These scores are aggregated over simulations $\mathcal{S}_{S^i_\alpha}(n)$, and then aggregated again over configurations $\mathcal{C}$, to produce the aggregated performance score $\hat{m}(n)$.

[21] For some computationally intensive simulations, n is logarithmically spaced.

3.2. *Performance Measures*

Denoting by $\widehat{S}^i_\alpha(\hat{s}(n))$ the model selection estimator for S^i_α for simulation $\hat{s}(n)$, we use the following performance metrics throughout our simulations.

Support selection accuracy:

$$m\big(\hat{s}(n)\big) := \frac{\big|\widehat{S}^i_\alpha\big(\hat{s}(n)\big) \cap S^i_\alpha\big|}{\big|\widehat{S}^i_\alpha\big(\hat{s}(n)\big) \cup S^i_\alpha\big|}.$$

This is a value between 0 and 1 that increases with support coverage, and is 1 if and only if the support is correctly selected. To construct the aggregated metric $\hat{m}(n)$, it is averaged over the simulations per configuration, and then averaged again over configurations.

"Some" best policy inclusion accuracy:

$$m\big(\hat{s}(n)\big) = \begin{cases} 1 & \text{if } \hat{\kappa}^{i\star}\big(\hat{s}(n)\big) \cap \kappa^{i\star} \neq \emptyset, \\ 0, & \text{otherwise,} \end{cases}$$

where $\kappa^{i\star} = \operatorname{argmax}_{\kappa \in S^i_{\mathrm{TVA}}} \eta_{S_{\mathrm{TVA}},\kappa}$ denotes the true best pooled policy in the marginal effects support S^i_{TVA} (uniquely determined from S^i_α). This measure is again averaged over simulations per configuration, and then averaged over configurations. The final metric therefore gives the share of simulations per n where at least one true best policy was pooled into the estimated best pooled policy.

Minimum dosage best policy inclusion accuracy:

$$m\big(\hat{s}(n)\big) = \begin{cases} 1 & \text{if } k^{i\star\min} \in \hat{\kappa}^{i\star}\big(\hat{s}(n)\big), \\ 0, & \text{otherwise,} \end{cases}$$

where $k^{i\star\min}$ denotes the true minimum dosage best policy.[22] Once aggregated, this measure captures the share of simulations per n, where the minimum dosage best policy was included in the estimated best pool.

Mean squared error (of best policy effect): For each simulation $\hat{s}(n)$, the estimated best policy treatment effect is scored by its error with respect to the true treatment effect:

$$m\big(\hat{s}(n)\big) := \hat{\eta}^{\mathrm{hyb}}_{\widehat{S}_{\mathrm{TVA}},\hat{\kappa}^\star} - \eta_{S_{\mathrm{TVA}},\kappa^\star}.$$

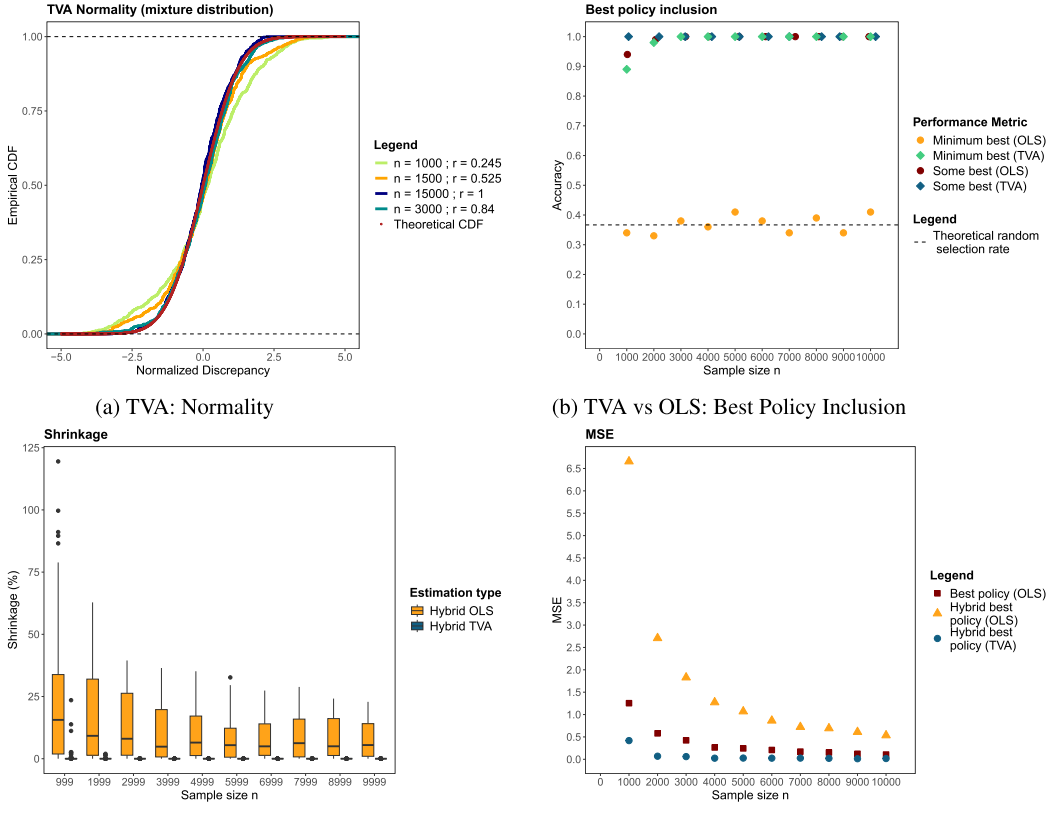
And thus $\hat{m}(n)$ is simply the estimated MSE:

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

3.3. *Performance of TVA*

First of all, simulation performance of TVA attests to its main theoretical properties: support consistency, best policy estimation consistency, and normally distributed coeffi-

[22]The "intersection" and "inclusion" operators for the best policy inclusion measures are to be understood in the following way: suppose the true best policy $\kappa^{i\star} \in S^i_{\mathrm{TVA}}$ pools together m policies as per S^i_α that we can organize into a set $S_1 = \{k^{i\star}_1, \ldots, k^{i\star\min}, \ldots, k^{i\star}_m\}$. Equivalently, we organize into S_2 the n policies composing the estimated best pool as per $\widehat{S}^i_\alpha(n)$. Then $\hat{\kappa}^{i\star}(\hat{s}(n))$ stands for S_2 and $\kappa^{i\star}$ for S_1.



(a) TVA: Normality

(b) TVA vs OLS: Best Policy Inclusion

(c) TVA vs OLS: Shrinkage

(d) TVA vs OLS: MSE

FIGURE 3.—A plot comparing the performance of the TVA estimator to applying OLS on the unique policy specification (2) for a range of measures. On Panel A, we first expose the normality of TVA estimates (r is correct support selection rate). Panel B then uses the best policy inclusion measures defined in Section 3.2 and points are slightly jittered for better readability. For OLS, this measure is set to 1 whenever the highest treatment effect policy is part of the true best pool (some best) or equal to the minimum dosage best policy (min best). Panel C compares the amount of shrinkage imposed by the winner's curse adjustment as percentage of the initial coefficient. Panel D compares the MSE of the best policy estimation, between the TVA and the OLS estimator of the unique policy specification (2) before and after adjusting for the winner's curse. In all panels, there are 20 simulations per configuration and 5 configurations per n.

cient estimates. This is depicted in Figure 3 (Panels A, B, D) and discussed in Supplemental Appendix E.1. In what follows, we primarily make the case for the strong performance of TVA relative to its most straightforward alternative, a direct application of OLS. In our Supplemental Appendix E.2, we provide comparisons for further LASSO-based alternative estimation strategies, the results of which are summarized below.

3.3.1. *Direct OLS*

An intuitive route for inference in this setting is estimating the unique policy specification (2) using only an application of OLS and nothing else (a strategy we call "direct OLS"). Since this is a fully saturated regression, this estimator has no theoretical issues with convergence nor with interpretation. Rather, this is about performance in the finite sample in the environment we describe. Specifically, it is worth noting that without ad-

ditional restrictions on the parameter space, OLS yields uniformly most powerful tests. However, since our approach is different from deriving optimal tests and involves model selection, and because additional restrictions are imposed, OLS may not be the best possible method, as will be demonstrated in the simulations. Most obviously, there is a loss of power in separately estimating the impact of 75 distinct interventions.

Figure 3, Panels B–D, documents three main inadequacies of direct OLS on the question of selecting and estimating best policy effects: (1) it fails at consistently identifying the minimum dosage best policy, (2) the estimates of best policy exhibit a stronger winner's curse, (3) the attenuations from applying Andrews, Kitagawa, and McCloskey (2021) are large. More specifically, Panel A shows that direct OLS (orange) does almost as well as TVA (blue) in estimating as best policy *some* policy that is part of the true best pooled policy, but it effectively picks the dosage at random; thus, selection of the minimum dosage hovers at around 36.67%. Panel C exhibits a strong winner's curse, which is to be expected in a situation with numerous candidates for "best policy," since the odds that a particularly large shock was drawn and thrust one to the top is quite high. The fact that the resulting coefficients are over-attenuated can be inferred from Panel D, where the attenuated best policy MSE (yellow triangles) remains large (of 0.53) even for large n.[23] In contrast, the winner's curse attenuations for TVA are much more modest, because of reduction of the number of competing policies and therefore greater separation between them.

Besides the specific issue of best policy estimation, direct OLS has low power. This is depicted in Figure E.1, where simulated OLS estimates of all the unique policies (2) contrast with the pruned and pooled estimation (3) (refer to Supplemental Appendix E).

3.3.2. *Naive LASSO*

There are two ways we could "naively" apply LASSO. The first is to disregard pooling, and apply LASSO on the unique policy specification (2) because sparse dosages might also mean a sparse set of policies. While there is no theoretical issue with this procedure in terms of model consistency, using this for policy estimation leads to much the same performance limitations as direct OLS with regard to best policy estimation, namely, a persistently high best policy MSE stemming from overly severe correction from Andrews, Kitagawa, and McCloskey (2021)'s winner's curse adjustment. Figure E.2 from our Supplemental Appendix contrasts this "No pooling, only pruning" version of LASSO to TVA on best policy estimation and documents these patterns in detail. The second way to "naively" apply LASSO is to consider both pooling and pruning as important, but adopt a sign-inconsistent model selection procedure by applying LASSO directly on (5) without a Puffer transformation. As expected, simulations attest to inconsistent support selection though MSE on the best policy is comparable to TVA. Importantly, it fails to select the minimum dosage best policy with substantial probability relative to TVA (refer to Figure E.3 in Supplemental Appendix E).

3.3.3. *Debiased LASSO*

Because we are interested in high-dimensional inference,[24] one alternative to a two-step process of model selection and inference is the so-called "debiased LASSO" (Zhang and

[23] Recall that while Andrews, Kitagawa, and McCloskey (2021) estimators are consistent, this assumes some local separation of parameters, which is not guaranteed in these neck-and-neck competitions.

[24] Albeit, we are still in a $K < n$ regime (mechanically since the number of treatments cannot exceed the number of units), sometimes called low dimensional with diverging number of parameters.

Zhang (2014), Javanmard and Montanari (2014), Javanmard and Montanari (2018), Van De Geer (2019)). The basic idea is that since the downward bias in LASSO is estimable, we can reverse it. A feature, however, is that these debiased coefficients are almost surely never exactly zero, so that there is no question of sparsity. We thus only need to consider applying debiased LASSO to (2). In Figure E.4 of our Supplemental Appendix, we show that the debiased LASSO procedure suffers from the same limitations as direct OLS estimation, especially with regard to best policy estimation (high MSE due to over-attenuation of the winner's curse).

3.3.4. *"Off the Shelf" Bayesian Approaches: Spike and Slab LASSO (Nie and Ročková (2022))*

The rules governing admissible pooling encode the econometrician's prior about the environment as does regularization. This raises the possibility of a Bayesian framework. Indeed, LASSO estimates have a Bayesian interpretation with Laplace priors. One can ask whether a more sophisticated, "explicitly" Bayesian approach can address our final objectives. This paradigmatically different route is the topic of future work. In Section E.2.4 of our Supplemental Appendix, we just show that "off the shelf" Bayesian approaches are unlikely to help. In particular, we show that a direct application of spike and slab formulations—the most intuitively relevant method—underperforms relative to our TVA procedure with a performance pattern similar to that of applying naive LASSO to the marginal specification (5).

Taken together, these simulation results make the case that TVA is both a powerful and robust candidate for our setting. Moreover, even under different sparsity and effect size relaxations, TVA is still strong, as demonstrated in Supplemental Appendix E.3. In particular, it does better than the next best practical alternative of applying naive LASSO to the marginal specification (5).

4. INCREASING IMMUNIZATION IN HARYANA: CONTEXT, EXPERIMENTAL DESIGN, AND DATA

We now apply this method to a large-scale experiment conducted in collaboration with the government of Haryana to help them select the most effective policy bundle to stimulate demand for immunization. The objective of the experiment was explicitly to select the best policy to scale up, after one year-long experiment with 75 potentially distinct treatments, making it an excellent application. Indeed, we developed the method in order to analyze these data.

4.1. *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 due to 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 sessions were almost always run in each village.

The central node of the UIP is the Primary Health Center (PHC). 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-centers (SCs). Vaccines are stored and transported from the PHCs to either sub-centers or villages on an appointed day each month, where there is 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-month-old children in Haryana fell from 60% in 2007–2008 to 52.1% in 2012–2013 (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 share 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, while 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 completion 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 to pick the best policy to possibly 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 ones we consider was below 40%, as reported by parents (which is likely a large overestimate 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. (2021a)). 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 is 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 that 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 are our main source of information on immunization and we discuss their reliability below. More details on the implementation are provided in the publicly available progress report (Banerjee et al. (2021a)).

4.2. *Interventions*

The study evaluates the impact of several nudges on the demand for immunization: small incentives, targeted reminders, and local ambassadors.

4.2.1. *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) showed that small incentives for immunization in Rajasthan (a bag of lentils for each shot and a set of plates 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), Chandir, Siddiqi, Abdullah, Duflo, Khan, and Glennerster (2022)). In the Indian health system, households receive incentives for a number of health behaviors, 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 of rewards 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 sub-center level. Depending on which sub-center the caregiver fell under, she would receive one of the following:

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 multiples 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.

4.2.2. *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 faster adoption.[25]

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, Gelaye, Were, Gashu, and Tilahun (2019)). However, it remains true that text messages could have no effect or even backfire if parents do not understand the information provided and feel they have no one to ask (Banerjee, Breza, Chandrasekhar, and Golub (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."

[25]See, for example, Rogers (1995), Krackhardt (1996), Kempe, Kleinberg, and Tardos (2003), Jackson (2008), Iyengar, Van 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, Jackson, and Tebaldi (2016), Jackson (2017), Akbarpour, Malladi, and Saberi (2017).

4.2.3. *The Immunization Ambassador: Network-Based Seeding*

The goal of the ambassador intervention was to leverage the social network to spread information. 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, BenYishay, Magruder, and Mobarak (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 as 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 Supplemental Appendix L.

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

4.3. *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. 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 our Supplemental Appendix Figure H.1.

We started with 2360 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 sub-center, 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 would like to identify other effective policies and answer the question of whether different variants of the policy had the same or different impacts.

4.4. *Data*

4.4.1. *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. To conduct 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 were 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 Supplemental Appendix K.

4.4.2. *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. The main analysis of this paper focuses on the number of children who received the measles vaccines in each village every month. The measles vaccine 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. As a result, it 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, 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 these data were also used to trigger SMS reminders and incentives, and for the government to evaluate the nurses' performance,[26] it was important to assess its accuracy. Hence, we conducted a validation exercise, comparing the administrative data with random checks, as described in Supplemental Appendix J. 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 Supplemental Appendix I, we show that there does not appear to be a pattern of substitution, as these children were not more likely to be immunized elsewhere.

Below, the dependent variable is the number of measles shots 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 shots 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.

4.5. *Interventions Average Effects*

In this section, we present the average effects of the interventions using a standard regression without interactions.

We focus on the sample of census villages used throughout our analysis—which are the villages where the ambassador intervention was also administered—and run the following specification:

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

We weight these village-level regressions by village population, and standard errors are clustered at the SC level.[27]

The results (reported in Banerjee et al. (2019)) are depicted graphically in Figure 4 and 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 nearly 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. To ensure that conclusions are not simply an artifact of this particular sub-sample, we show in Supplemental Appendix H that these results are robust to running the analysis on the full sample.

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

[27] 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.

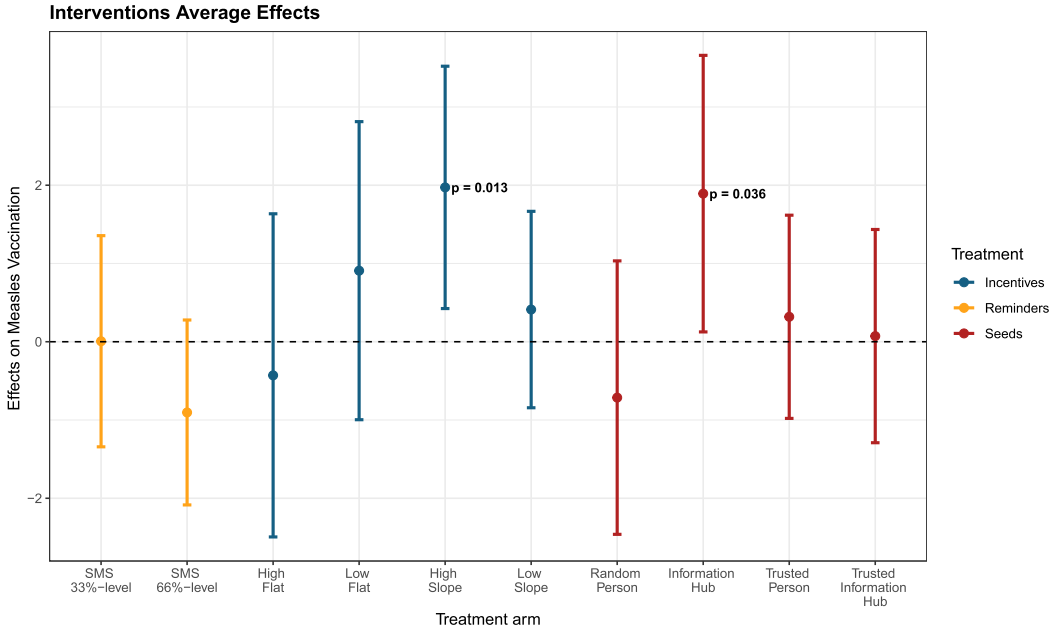


FIGURE 4.—Effects on the number of measles vaccinations relative to control (7.32) by reminders, incentives, and seeding policies, restricted to the villages where the ambassador intervention was administered. The specification is weighted by village population, controls for district-time fixed effects, and clusters standard errors at the sub-center level.

The conclusion from this analysis 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. These results are similar to those found in Pakistan by Chandir et al. (2022) which also test various models of incentives and SMS, with small differences. There, SMS alone has a statistically significant, but small, impact on full immunization. Incentives are more effective, and the level of incentives matter (although in that setting, the slope does not matter).

In our setting, leveraging the community by enrolling local ambassadors, selected using the cheap procedure of asking a few village members to nominate the 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)[28] was that it would have a

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

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 adapt our disciplined approach to select which ones to include, and to then estimate the impact of the "best" policy.

5. RESULTS

5.1. *Identifying Effective Policies*

5.1.1. *Method*

We adapt the TVA procedure for our case. We allow only some pooling within arms depending on the nature of the sub-treatment. In the incentive arms, slope and flat incentives are not allowed to pool, but the amount of money (high or low) is considered to be a dosage. In the ambassador arms, we do not allow pooling between random selection of ambassadors, trusted ambassador, and information hub. Within information hubs, however, we consider that the "trusted information hub" is an increased dosage of information hub, so these may pool with one another.

To summarize, interventions "information hubs," "slope," "flat," and "SMS" are found in two intensities. The marginal specification (5) therefore looks like

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

where we have explicitly listed the variables in "single arm" treatment profiles. X_{sv} is a vector of the remaining 64 marginal effects variables in "multiple arm" treatment profiles, and v_{dt} is a set of district-time dummies. Here SMS refers to "any SMS."

Our Puffered LASSO model selection estimation follows the recommended implementation in Rohe (2014), which uses a sequential backwards elimination version of LASSO (variables with p-values above some threshold are progressively deselected) on the Puffer_N transformed variables (the N refers to a further right multiplication of Puffer transformed variables with a diagonal matrix that aids in correcting for the heteroscedasticity 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.[29] This makes

[29] Rohe (2014), a supplementary note to Jia and Rohe (2015), deduces an algorithmic equivalence between a backwards elimination procedure based on using Type I error thresholds and LASSO on Puffer_N transformed variables. In particular, the variables deselected by a LASSO penalty level λ are exactly those variables with classical OLS p-value above the Type I error threshold. 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, a threshold selected to essentially eliminate false positives. Supplemental Appendix D elaborates on this choice.

sense because it is extremely problematic to have a government introduce a large policy based on a false positive. This reasoning is elaborated in Supplemental Appendix D.

This gives $\widehat{S}_\alpha$, an estimate of the true support set S_α of the marginal effects specification (5). We then generate a use of unique pooled policy set $\widehat{S}_{\text{TVA}}$ (following the procedure we outline in Algorithm 2 in Supplemental Appendix B). Next, we run the pooled specification (3) to obtain post-LASSO estimates $\hat{\eta}_{\widehat{S}_{\text{TVA}}}$ of the pooled policies as well as $\hat{\eta}^{\text{hyb}}_{\widehat{S}_{\text{TVA}},\hat{\kappa}^\star}$, the winner's curse adjusted estimate of the best policy.

5.1.2. *Results*

The results are presented in Figure 5. Panel A presents the post-LASSO estimates where the outcome variable is the number of measles vaccines per month in the village. Panel B presents the post-LASSO estimates where the outcome variable is the number of measles vaccines per dollar spent. In each, a relatively small subset of policies is selected as part of $\widehat{S}_{\text{TVA}}$ out of the universe of 75 granular policies (16% of the possible options in Panel A and 35% in Panel B).

In Figure 5, Panel A, 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 ($p = 0.009$).

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 $1000 allocation (as compared with 43.6 immunizations per village per month in control). The reason is that the gains from having incentives in terms of immunization rates is smaller than the increase in costs (e.g., the incentives must be paid to all the infra-marginal parents).

Two things are worth noting to qualify those results, however. First, in Chernozhukov, Demirer, Duflo, and Fernandez-Val (2018), we showed that in the places where the full package treatment is predicted to be the most effective (which tend 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, Mirelman, Stack, Walker, and Levine (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 5, Panel B. 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 ($p = 0.000$).

The algorithmic equivalence thus permits a common interpretation of the two models on the two outcomes as admitting the same Type I error. Its other advantage is that we can implement Rohe's recommendation of *sequentially* deselecting variables until this threshold rather than eliminating them in a single step, because, as Rohe noted, this better handles variable correlation.

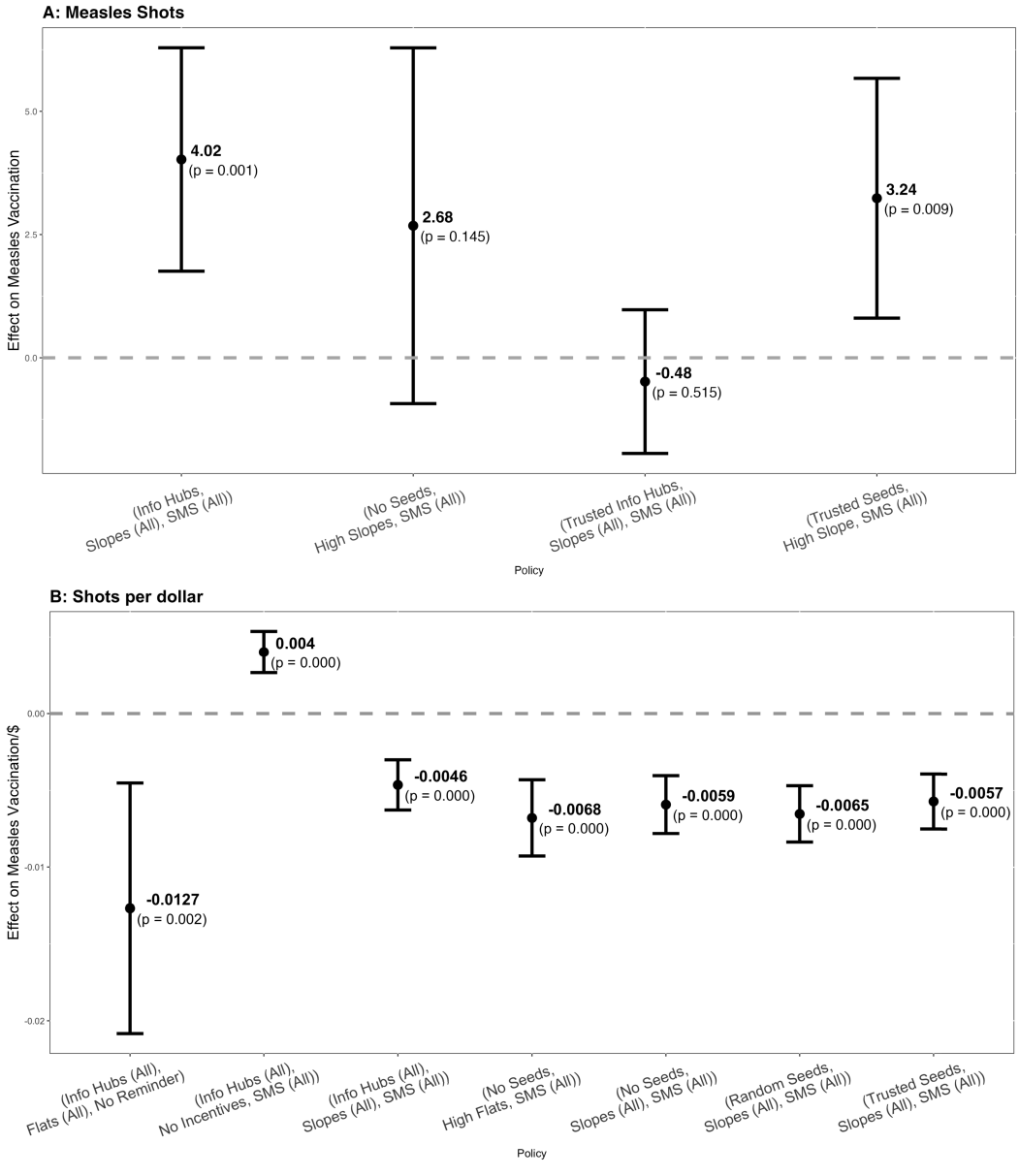


FIGURE 5.—TVA estimates of combinations of reminders, incentives, and seeding policies on the number of measles vaccines (Panel A) and the number of measles vaccines per $ (Panel B) relative to control (7.32 and 0.0436, respectively). The specifications are weighted by village population and include controls for district–time fixed effects. Standard errors are clustered at the sub-center level. 95% confidence intervals displayed.

5.2. *Estimating the Impact of the Best Policy*

To estimate the impact of the best policy, we first select the best policy from $\widehat{S}_{\text{TVA}}$ 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 is the value used by Andrews, Kitagawa, and McCloskey (2021) in their simulations. The hybrid confidence interval has the following

TABLE I

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	0.0436
Observations	204	814
Optimal Policy	(Information Hubs, SMS, Slope)	(Information Hubs POOLED, SMS)

Note: Estimation using Andrews, Kitagawa, and McCloskey (2021); 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.

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. It also has at least 95% coverage unconditionally.[30]

Table I 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.[31] 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 choose 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 may be that they are sensitive to the implied LASSO penalty λ chosen. To check the robustness of our results, we consider alternative values of λ. However, we also need a criterion for evaluating results under various λ since a marginal effects support will never be robust for the whole range of λ. Supplemental

[30]Per Proposition 6 of Andrews, Kitagawa, and McCloskey (2021), it has unconditional coverage between $1 - \alpha = 95\%$ and $\frac{1-\alpha}{1-\beta} = 95.58\%$.

[31]The increased attenuation from a more closely competing second best policy emerges from the formulas for conditional inference given in Section 3 of Andrews, Kitagawa, and McCloskey (2021).

Appendix D spells out this criterion, which amounts to formulating a set of "admissible" λ. In a nutshell, the criterion is to avoid including in $\widehat{S}_\alpha$ first and second best policies that are very likely to be false positives. Including a false positive as the first best is a serious error in the context of policy advising, but it also matters for the second best, since including these in the support may overly attenuate the best policy estimate for winner's curse. In our case, we find that for both immunizations and immunizations per dollar, the winner's curse adjusted estimates are robust for their respective sets of admissible λ. To exemplify this robustness, we can take the union of confidence intervals within their admissible sets. This is [0.32,6.25] for immunizations and [0.001,0.006] for immunizations per dollar. Neither is much wider than the single confidence interval for the choice of λ we highlight.

Though admissible λ are on a different scale for the two outcomes, there is a sense in which the admissible set is larger for immunizations per dollar. This suggests a different kind of robustness concern which is more about the relative fragility of the TVA estimator for each of the outcomes. We can speak to this fragility using a bootstrapping analysis described in detail in our Supplemental Appendix F. Intuitively, it captures stability of best policy estimation in terms of observation leverage, where conclusions driven by outliers will fare worse. In this analysis, the best policy for cost-effectiveness holds for 96% with highly concentrated estimates around the main one in actual data. Meanwhile, the best policy for immunizations holds for 77% of bootstrapped samples with estimates more widely dispersed. This speaks to the relative stability of the best policy for cost-effectiveness over that for immunizations.

6. CONCLUSION

Despite immunization being one of the most effective and cost-effective methods to prevent disease, disability, and mortality, millions of children each year go unvaccinated. The COVID-19 epidemic has made the situation worse: vaccine coverage has dipped to levels not seen since the 1990s (Bill and Foundation Melinda Gates (2020)). Swift policy action is critical to ensure that this dip is temporary and children who missed immunizations during the pandemic get covered soon.

In rural India, there was a priori reason to believe that nudges may work. After all, many children get their first vaccines but caregivers rarely follow through. This is consistent with the vast majority of caregivers reporting that vaccines are helpful. Yet, it was a priori unclear as to which nudge, let alone which policy bundle out of the 75 candidates, would be effective.

Respecting this genuine uncertainty was critical. If we had simply done parallel treatments of incentives, reminders, and ambassadors, we might have found no effects. Our key finding is that combined interventions work better than each in isolation. Though there is temptation of paring down the number of treatments a priori for power, there is a danger in not doing this in a data-driven way. The suggestion of avoiding all interactions in this setting (made in Muralidharan, Romero, and Wuthrich (2019)) would have led to the conclusion that nothing is effective.

From the point of view of public health policy, the interaction effects identified by TVA tell us that it is valuable to add network-based insights (information hubs), which are not in a typical policymaker's toolkit, to catalyze the effects of conventional instruments. From a basic research perspective, it also suggests that the information hubs, that is, the person best placed in a village to circulate information, may be more effective when they have something concrete to talk about, such as incentives or something to explain such as SMSs. Such questions merit future research.

The method suggested here is applicable to many domains where policymakers have several arms with multiple potential doses, do not have the time or capacity to adaptively experiment, and have genuine uncertainty about which policy bundles should be effective. Rather than guessing, we suggest that policymakers consider a data-driven approach of treatment variant aggregation. The proposed method relies on strong assumptions that rule out some of the cases where model selection leads to invalid inferences. Provided these assumptions are palatable, our findings show that TVA prunes and pools effectively, and that this pays dividends when the policymaker wishes to adjust for the winner's curse without falling into the trap of over-conservatism. The algorithm can be easily prespecified and does not require the researcher to take a stance on the possible effects of myriad interactions which are likely difficult to predict in advance.

APPENDIX A: PROOFS

PROOF OF PROPOSITION 2.1: According to Theorem 1 of Jia and Rohe (2015), if $\min_{j\in S_\alpha}|\alpha_j| \geq 2\lambda_n$, then $\widetilde{\alpha} =_s \alpha$ with probability greater than[32]

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

where $\xi_{\min} = \xi_{\min}(\frac{X}{\sqrt{n}})$ is the minimum singular value of the $\sqrt{n}$-normalized design matrix. By Assumption 3, there is a uniform lower bound $c > 0$ on the magnitude of the nonzero $\{\alpha\}$. Since by Assumption 5, $\lambda_n \to 0$, 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 and Assumption 2, for sufficiently high n:

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

By Assumption 5, $\lambda_n^2 n^{1-2\gamma} = \omega(n^{1-2(\nu+\gamma)})$. Since $0 < \nu < \frac{1}{2} \implies 1-2\gamma > 1-(2(\nu+\gamma)) > 0$, and logarithm growth is dominated by polynomial growth, it follows that $-(\gamma\log(n) - \frac{n^{1-2\gamma}\lambda_n^2}{2\sigma^2})$ grows at polynomial rate $\omega(n^{1-2(\nu+\gamma)})$ and therefore $\lim_{n\to\infty} f(n) \geq 1$ at exponential rate. Since also $f(n) \leq 1$, it follows that $f(n) \to 1$ at exponential rates. *Q.E.D.*

LEMMA A.1: *For the marginal effects design matrix X, for $R \geq 3$, wpa1 the lowest singular value of $\sqrt{n}$ normalized design matrix, that is, $\xi_{\min}(\frac{X}{\sqrt{n}})$, has the value $\xi_{\min}(\frac{X}{\sqrt{n}}) = (4R\sin^2(\frac{R-\frac{3}{2}}{R-\frac{1}{2}}\frac{\pi}{2}))^{-\frac{M}{2}}$. Thus, with probability approaching 1, $\xi_{\min}(\frac{X}{\sqrt{n}}) > (\frac{1}{K})$.*[33]

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

[32]The "$=_s$" notation stands for equality in sign following Definition 1 in Zhao and Yu (2006).

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

The combinatorics of the limiting frequencies of "1"s in marginal effects variables implies that $\mathbf{C}_{R,M}$ is a block diagonal matrix with structure $\mathbf{C}_{\mathbf{R,M}} = K^{-1}\text{diag}(\mathbf{B}_{R,M}, \mathbf{B}_{R,M-1}, \ldots, \mathbf{B}_{R,1})$, 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 $\mathbf{B}_{R,M-1}, \mathbf{B}_{R,M-2}, \ldots \mathbf{B}_{R,1}$ is found in $\mathbf{C}_{\mathbf{R,M}}$, but only $\mathbf{B}_{\mathbf{R,M}}$ 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.[34]
2. $\mathbf{B}_{\mathbf{R,1}}$ is an $(R-1) \times (R-1)$ matrix with recursive structure

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

SUBLEMMA 1: $\lambda_{\min}(\mathbf{B}_{\mathbf{R,1}}) = (4\sin^2(\frac{R-\frac{3}{2}}{R-\frac{1}{2}}\frac{\pi}{2}))^{-1}$.

PROOF: The key insight[35] 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 $\mu_j = 4\sin^2(\frac{j-\frac{1}{2}}{R-\frac{1}{2}}\frac{\pi}{2})$ for $j = 1, 2, \ldots, R-1$. Thus, given that the inverse of a matrix's eigenvalues are the inverse matrix's eigenvalues, $\lambda_{\min}(\mathbf{B}_{\mathbf{R,1}}) = (4\sin^2(\frac{R-\frac{3}{2}}{R-\frac{1}{2}}\frac{\pi}{2}))^{-1}$. *Q.E.D.*

(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 turn implies $\lambda_{\min}(\mathbf{B}_{R,M}) = \lambda_{\min}(\mathbf{B}_{R,1})^M$. Since by Sublemma 1, $\lambda_{\min}(\mathbf{B}_{\mathbf{R,1}}) < 1$, $\mathbf{B}_{R,M}$ 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:

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

where the last equality uses Sublemma 1. The lemma follows. *Q.E.D.*

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

[35]The argument is provided on Mathematics Stackexchange (user1551 (2017)).

PROOF OF PROPOSITION 2.2: The proof is found in Javanmard and Montanari (2013), proof of Theorem 2.7, with minor modifications, which we reproduce for completeness. Label the events $\mathcal{E} := \{\widehat{S}_{\text{TVA}} = S_{\text{TVA}}\}$ (that the treatment variants were aggregated correctly). Define the pseudo-true value $\eta_S^0 := \operatorname{argmin}_\eta \mathrm{E}[\|y - Z_S\eta\|_2^2]$, noting that $\eta^0_{S_{\text{TVA}}}$ satisfies this for $S = S_{\text{TVA}}$. Finally, let $\mathcal{F} := \{\|\hat{\eta}_{\widehat{S}_{\text{TVA}}} - \eta^0_{\widehat{S}_{\text{TVA}}}\|_\infty > \epsilon\}$, so it is the event that the estimator exceeds the pseudo-true value on the estimated support by ϵ.

Then, we can write

$$\mathrm{P}(\mathcal{F}) = \mathrm{P}(\mathcal{F} \cap \mathcal{E}) + \mathrm{P}\big(\mathcal{F} \cap \mathcal{E}^c\big) \le \mathrm{P}(\mathcal{F} \cap \mathcal{E}) + \mathrm{P}\big(\mathcal{E}^c\big).$$

By the proof of Proposition 2.1, we have

$$\mathrm{P}\big(\mathcal{E}^c\big) \le 2K \exp\left(-\frac{n(\lambda/K)^2}{2\sigma^2}\right) = 2\exp\left(\gamma \log n - \frac{n^{1-2\gamma}\lambda}{2\sigma^2}\right).$$

Turning to $\mathrm{P}(\mathcal{F} \cap \mathcal{E})$, on the event $\mathcal{E}$,

$$\hat{\eta}_{\widehat{S}_{\text{TVA}}} - \eta^0_{\widehat{S}_{\text{TVA}}} = \big(Z'_{\widehat{S}_{\text{TVA}}} Z_{\widehat{S}_{\text{TVA}}}\big)^{-1} Z'_{\widehat{S}_{\text{TVA}}} \epsilon,$$

since $\widehat{S}_{\text{TVA}} = S_{\text{TVA}}$. Therefore, for every $j \in \{1, \ldots, K\}$, $\hat{\eta}_{S_{\text{TVA}},j} - \eta^0_{S_{\text{TVA}},j}$ is normally distributed with variance order bounded above by $\frac{\sigma^2}{n \cdot C_{\min}}$, where $C_{\min} := \sigma_{\min}(n^{-1} Z'_{S_{\text{TVA}}} Z_{S_{\text{TVA}}})$ is the minimum singular value of the design matrix. But $C_{\min} \ge \frac{1}{\sqrt{K}}$ by definition since each unit is assigned to a disjoint pooled policy, and each policy pools one or more variants. So

$$\mathrm{P}(\mathcal{F} \cap \mathcal{E}) = \mathrm{P}\big(\big\|\hat{\eta}_{\widehat{S}_{\text{TVA}}} - \eta^0_{\widehat{S}_{\text{TVA}}}\big\|_\infty > \epsilon \cap \mathcal{E}\big) \le \mathrm{P}\Big(\sup_j \big|\hat{\eta}_{S_{\text{TVA}},j} - \eta^0_{S_{\text{TVA}},j}\big| > \epsilon\Big) \le 2e^{-\frac{n \cdot \epsilon^2 \cdot K^{-1/2}}{2\sigma^2}},$$

using a Gaussian tail bound and a union bound for uniform control over $j \in \{1, \ldots, |S_{\text{TVA}}|\}$.

Putting the pieces together, we have

$$\mathrm{P}(\mathcal{F}) \le 2\exp\left(-\frac{n \cdot \epsilon^2 \cdot K^{-1/2}}{2\sigma^2}\right) + 2\exp\left(\gamma \log n - \frac{n^{1-2\gamma}\lambda^2}{2\sigma^2}\right).$$

This establishes that $\mathrm{P}(\mathcal{F}) \to 0$ for every ϵ, that is, the consistency of the estimator to the pseudo-true values. With probability $1 - 2\exp(\gamma \log n - \frac{n^{1-2\gamma}\lambda}{2\sigma^2}) \to 1$, the event $\mathcal{E}$ is active and the pseudo-true value will be the true value. In this case, consider $\epsilon = \sqrt{\frac{\log n}{n^{1-\gamma/2}}}$. Then

$$\mathrm{P}(\mathcal{F} \cap \mathcal{E}) = 2e^{-\frac{n\epsilon^2 C_{\min}}{2\sigma^2}} \le 2e^{-\frac{\log n}{n^{1-\gamma/2}} \cdot \frac{nK^{-1/2}}{2\sigma^2}} \le 2e^{-\frac{\log n}{2\sigma^2}} = \frac{2}{n^{1/2\sigma^2}} \to 0,$$

that is, $\|\hat{\eta}_{\widehat{S}_{\text{TVA}}} - \eta^0_{S_{\text{TVA}}}\|_\infty < \sqrt{\frac{\log n}{n^{1-\gamma/2}}}$ with high probability, completing the proof. *Q.E.D.*

PROOF OF PROPOSITION 2.3: As in the proof of Proposition 2.2, let $\mathcal{E} := \{\widehat{S}_{\text{TVA}} = S_{\text{TVA}}\}$. Let $c \in \mathbb{R}^{|\widehat{S}_{\text{TVA}}|}$ whose length depends on the estimated support (and is therefore random), though we suppress this dependence. We can decompose[36] $\sqrt{n}c'(\hat{\eta}_{\widehat{S}_{\text{TVA}}} - \eta^0_{\widehat{S}_{\text{TVA}}})$, a real-

[36] We thank Adel Javanmard for a helpful discussion.

valued random variable, as

$$\begin{aligned}\sqrt{n}c'\big(\hat{\eta}_{\widehat{S}_{\mathrm{TVA}}} - \eta^0_{\widehat{S}_{\mathrm{TVA}}}\big) &= \mathbf{1}\{\mathcal{E}\}\cdot\sqrt{n}c'\big(Z'_{S_{\mathrm{TVA}}}Z_{S_{\mathrm{TVA}}}\big)^{-1}Z'_{S_{\mathrm{TVA}}}\boldsymbol{\epsilon} + \mathbf{1}\{\mathcal{E}^c\}\\ &\quad\cdot\sqrt{n}c'\big(Z'_{\widehat{S}_{\mathrm{TVA}}}Z_{\widehat{S}_{\mathrm{TVA}}}\big)^{-1}Z'_{\widehat{S}_{\mathrm{TVA}}}\boldsymbol{\epsilon}\\ &= \sqrt{n}c'\big(Z'_{S_{\mathrm{TVA}}}Z_{S_{\mathrm{TVA}}}\big)^{-1}Z'_{S_{\mathrm{TVA}}}\boldsymbol{\epsilon} - \mathbf{1}\{\mathcal{E}^c\}\cdot\sqrt{n}c'\big(Z'_{S_{\mathrm{TVA}}}Z_{S_{\mathrm{TVA}}}\big)^{-1}Z'_{S_{\mathrm{TVA}}}\boldsymbol{\epsilon}\\ &\quad+\mathbf{1}\{\mathcal{E}^c\}\cdot\sqrt{n}c'\big(Z'_{\widehat{S}_{\mathrm{TVA}}}Z_{\widehat{S}_{\mathrm{TVA}}}\big)^{-1}Z'_{\widehat{S}_{\mathrm{TVA}}}\boldsymbol{\epsilon}.\end{aligned}$$

It is helpful to note that c is *any* vector of the same length as the size of the estimated support. So conditional on the event of the estimated support being the true support, for instance, it has the true length. Conditional on any realization of an alternative length support, one considers a conformal vector corresponding to the estimated support size.

The proof strategy is to see that the CLT ensures the asymptotic normality of the first term in this sum in the usual way, following He and Shao (2000) since we have a growing number of parameters, while the remaining terms will asymptotically vanish in probability.

Let us take the first term, so we are looking at the case *on* S_{TVA}. Here we show asymptotic normality. In order to show this, we need to show that every linear combination of the vector of regression coefficients (of which there are a growing number) is asymptotically normally distributed when properly normalized. We show the sufficient conditions for Corollary 2.1 in He and Shao (2000). First, by Assumption 2, note that $K^2\log(K) = O(n^{2\gamma}\log(n)) = o(n)$, using that $2\gamma < 1$ and that $\log(n)$ grows more slowly than any polynomial in n. K thereby satisfies the hypothesized growth rate condition in Corollary 2.1 for smooth scores, which is our case since we study linear regression. Second, we check (D1)–(D3) after which the corollary applies. (D1) follows since $n^{-1}Z'_{S_{\mathrm{TVA}}}Z_{S_{\mathrm{TVA}}} = I$ by definition since every observation is assigned to one unique treatment dummy. (D2) follows since in the case of linear regression, the score and its derivative are bounded. As noted in He and Shao (2000), a sufficient condition for (D3) is that the regression vector, here $Z_{S_{\mathrm{TVA},i}}$ for observation i, is such that $\mathrm{E}|c'Z_{S_{\mathrm{TVA},i}}|^4$ for any c in the unit sphere of length $|S_{\mathrm{TVA}}|$. But this is mechanically true since $\|Z_{S_{\mathrm{TVA},i}}\| = 1$ since it is a saturated vector of treatment bundle assignments and therefore has a single 1 corresponding to the entry for the treatment assigned.

Therefore, Corollary 2.1 of He and Shao (2000) applies. This means that

$$\sqrt{n}c'\big(\hat{\eta}_{S_{\mathrm{TVA}}} - \eta^0_{S_{\mathrm{TVA}}}\big)/\sigma(c) \rightsquigarrow \mathcal{N}(0,1),$$

for any $c \in \mathbb{R}^{|S_{\mathrm{TVA}}|}$, where $\sigma^2(c) := \sigma^2\|c\|^2$.

The second term can be handled by showing, for any $c \in \mathbb{R}^{|S_{\mathrm{TVA}}|}$, that we have $\mathbf{1}\{\mathcal{E}^c\}\cdot\sqrt{n}c'(Z'_{S_{\mathrm{TVA}}}Z_{S_{\mathrm{TVA}}})^{-1}Z'_{S_{\mathrm{TVA}}}\boldsymbol{\epsilon} = o_p(1)$. We already showed that $\sqrt{n}c'(Z'_{S_{\mathrm{TVA}}}Z_{S_{\mathrm{TVA}}})^{-1}Z'_{S_{\mathrm{TVA}}}\boldsymbol{\epsilon}$ is asymptotically normal and so $O_p(1)$. Since $\mathbf{1}\{\mathcal{E}^c\}$ is $o_p(1)$, the whole term is $o_p(1)$.

Regarding the third one: $\mathbf{1}\{\mathcal{E}^c\}\cdot\sqrt{n}c'(Z'_{\widehat{S}_{\mathrm{TVA}}}Z_{\widehat{S}_{\mathrm{TVA}}})^{-1}Z'_{\widehat{S}_{\mathrm{TVA}}}\boldsymbol{\epsilon}$. So here we study $\mathbf{1}\{\mathcal{E}^c\}\cdot\sqrt{n}c'(Z'_{\widehat{S}_{\mathrm{TVA}}}Z_{\widehat{S}_{\mathrm{TVA}}})^{-1}Z'_{\widehat{S}_{\mathrm{TVA}}}\boldsymbol{\epsilon}$, where $c \in \mathbb{R}^{|\widehat{S}_{\mathrm{TVA}}|}$, and show that it is $o_p(1)$. The point is that $(Z'_{\widehat{S}_{\mathrm{TVA}}}Z_{\widehat{S}_{\mathrm{TVA}}})^{-1}Z'_{\widehat{S}_{\mathrm{TVA}}}\boldsymbol{\epsilon}$, which potentially inherits omitted variable bias by including incorrect regressors, is nevertheless uniformly bounded in n and K. In fact, we can bound the $\|\cdot\|_\infty$ norm which implies bounding the size of the inner product with any aforementioned c. In detail, first note that $\|(Z'_{\widehat{S}_{\mathrm{TVA}}}Z_{\widehat{S}_{\mathrm{TVA}}})^{-1}Z'_{\widehat{S}_{\mathrm{TVA}}}\boldsymbol{\epsilon}\|_\infty \le \|Z'_{\widehat{S}_{\mathrm{TVA}}}\boldsymbol{\epsilon}\|_\infty$ because $(Z'_{\widehat{S}_{\mathrm{TVA}}}Z_{\widehat{S}_{\mathrm{TVA}}})^{-1}$ is a positive definite block diagonal matrix with every entry < 1. Second, $\|Z'_{\widehat{S}_{\mathrm{TVA}}}\boldsymbol{\epsilon}\|_\infty < Kn\boldsymbol{\epsilon}$

since $Z_{\widehat{S}_{\text{TVA}}}$ is a binary matrix. Since ϵ is $O_p(1)$, $\|(Z'_{\widehat{S}_{\text{TVA}}} Z_{\widehat{S}_{\text{TVA}}})^{-1} Z'_{\widehat{S}_{\text{TVA}}} \epsilon\|_\infty$ is uniformly $O_p(Kn)$ over all misspecifications $\widehat{S}_{\text{TVA}}$.

Thus, $\mathbf{1}\{\mathcal{E}^c\} \cdot \sqrt{n}c'(Z'_{\widehat{S}_{\text{TVA}}} Z_{\widehat{S}_{\text{TVA}}})^{-1} Z'_{\widehat{S}_{\text{TVA}}} \epsilon$ is uniformly bounded in probability by $\mathrm{P}(\mathcal{E}^c) \times O_p(Kn^{\frac{3}{2}})$. But since $\mathrm{P}(\mathcal{E}^c) \leq 2K \exp(-\frac{n(\lambda/K)^2}{2\sigma^2}) = 2\exp(\log(K) - \frac{n(\lambda/K)^2}{2\sigma^2})$, recalling the proof of Proposition 2.2,

$$O_p\big(Kn^{\frac{3}{2}}\big)\mathrm{P}\big(\mathcal{E}^c\big) = O_p\big(e^{(\frac{3}{2}\log(n)+2\log(K)-\frac{n(\lambda/K)^2}{2\sigma^2})}\big) = O_p\big(e^{(\frac{3}{2}\log(n)+2\gamma\log(n)-\frac{n^{1-2\gamma}\lambda^2}{2\sigma^2})}\big) = o_p(1).$$

So $\mathbf{1}\{\mathcal{E}^c\} \cdot \sqrt{n}c'(Z'_{\widehat{S}_{\text{TVA}}} Z_{\widehat{S}_{\text{TVA}}})^{-1} Z'_{\widehat{S}_{\text{TVA}}} \epsilon$ vanishes in probability and altogether,

$$\sqrt{n}c'\big(\hat{\eta}_{\widehat{S}_{\text{TVA}}} - \eta^0_{\widehat{S}_{\text{TVA}}}\big)/\sigma^2(c) \rightsquigarrow \mathcal{N}(0,1) + o_p(1) + o_p(1) \rightsquigarrow \mathcal{N}(0,1),$$

which completes the proof. *Q.E.D.*

REFERENCES

AKBARPOUR, MOHAMMAD, SURAJ MALLADI, AND AMIN SABERI (2017): "Diffusion, Seeding, and the Value of Network Information," Available at SSRN. https://ssrn.com/abstract=3062830. [1206]

ALATAS, VIVI, ARUN G. CHANDRASEKHAR, MARKUS MOBIUS, BENJAMIN A. OLKEN, AND CINDY PALADINES (2019): "When Celebrities Speak: A Nationwide Twitter Experiment Promoting Vaccination in Indonesia," Tech. rep, National Bureau of Economic Research. [1188]

ANDREWS, ISAIAH, TORU KITAGAWA, AND ADAM MCCLOSKEY (2021): "Inference on Winners," Tech. rep. [1185,1187,1196-1198,1202,1213,1214]

ARAL, SINAN, AND DYLAN WALKER (2012): "Creating Social Contagion Through Viral Product Design: A Randomized Trial of Peer Influence in Networks," *Management Science*. [1207]

BANERJEE, ABHIJIT, EMILY BREZA, ARUN G. CHANDRASEKHAR, AND BENJAMIN GOLUB (2018): "When Less Is More: Experimental Evidence on Information Delivery During India's Demonetization," Tech. rep, National Bureau of Economic Research. [1206]

BANERJEE, ABHIJIT, ARUN CHANDRASEKHAR, ESTHER DUFLO, SURESH DALPATH, JOHN FLORETTA, MATHEW JACKSON, HARINI KANNAN, ANNA SCHRIMPF, AND MAHESHWOR SHRESTHA (2021a): "Evaluating the Impact of Interventions to Improve Full Immunisation Rates in Haryana, India." [1204]

BANERJEE, ABHIJIT, ARUN CHANDRASEKHAR, ESTHER DUFLO, AND MATTHEW O. JACKSON (2013): "Diffusion of Microfinance," *Science*, 341 (6144). [1206,1207]

BANERJEE, ABHIJIT, ARUN G. CHANDRASEKHAR, SURESH DALPATH, ESTHER DUFLO, JOHN FLORETTA, MATTHEW O. JACKSON, HARINI KANNAN, FRANCINE N. LOZA, ANIRUDH SANKAR, ANNA SCHRIMPF, AND MAHESHWOR SHRESTHA (2021b): "Selecting the Most Effective Nudge: Evidence From a Large-Scale Experiment on Immunization," Working Paper 28726, National Bureau of Economic Research. [1184]

BANERJEE, ABHIJIT, ARUN G. CHANDRASEKHAR, ESTHER DUFLO, AND MATTHEW O. JACKSON (2019): "Using Gossips to Spread Information: Theory and Evidence From Two Randomized Controlled Trials," *The Review of Economic Studies*. [1188,1207,1209,1211]

BANERJEE, ABHIJIT VINAYAK, ABHIJIT VINAYAK, ESTHER DUFLO, RACHEL GLENNERSTER, AND DHRUVA KOTHARI (2010): "Improving Immunisation Coverage in Rural India: Clustered Randomised Controlled Evaluation of Immunisation Campaigns With and Without Incentives," *Bmj*, 340, c2220. [1188,1205]

BASSANI, DIEGO G., PAUL ARORA, KERRI WAZNY, MICHELLE F. GAFFEY, LINDSEY LENTERS, AND ZULFIQAR A. BHUTTA (2013): "Financial Incentives and Coverage of Child Health Interventions: A Systematic Review and Meta-Analysis," *BMC Public Health*, 13 (S3), S30. [1188,1205]

BEAMAN, LORI, ARIEL BENYISHAY, JEREMY MAGRUDER, AND AHMED MUSHFIQ MOBARAK (2018): "Can Network Theory-Based Targeting Increase Technology Adoption?" Tech. rep, National Bureau of Economic Research. [1207]

BERK, RICHARD, LAWRENCE BROWN, ANDREAS BUJA, KAI ZHANG, AND LINDA ZHAO (2013): "Valid Post-Selection Inference," *The Annals of Statistics*, 802–837. [1196]

BICKEL, PETER J., YA'ACOV RITOV, AND ALEXANDRE B. TSYBAKOV (2009): "Simultaneous Analysis of Lasso and Dantzig Selector," *The Annals of statistics*, 37 (4), 1705–1732. [1193]

BILL AND FOUNDATION MELINDA GATES (2020): *COVID-19: A Global Perspective: 2020 Goalkeepers Report*. [1215]

BLOCH, FRANCIS, MATTHEW O. JACKSON, AND PIETRO TEBALDI (2016): "Centrality Measures in Networks,", http://ssrn.com/abstract=2749124. [1206]

CHANDIR, SUBHASH, DANYA ARIF SIDDIQI, SARA ABDULLAH, ESTHER DUFLO, AAMIR JAVED KHAN, AND RACHEL GLENNERSTER (2022): "Small Mobile Conditional Cash Transfers (mCCTs) of Different Amounts, Schedules and Design to Improve Routine Childhood Immunization Coverage and Timeliness of Children Aged 0–23 Months in Pakistan: An Open Label Multi-Arm Randomized Controlled Trial," *EClinicalMedicine*, 50. [1205,1210]

CHERNOZHUKOV, VICTOR, MERT DEMIRER, ESTHER DUFLO, AND IVAN FERNANDEZ-VAL (2018): "Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments," Tech. rep, National Bureau of Economic Research. [1212]

DLHS (2013): "District Level Household and Facility Survey-4." [1204]

DOMEK, GRETCHEN J., INGRID L. CONTRERAS-ROLDAN, SEAN T. O'LEARY, SHEANA BULL, ANNA FURNISS, ALLISON KEMPE, AND EDWIN J. ASTURIAS (2016): "SMS Text Message Reminders to Improve Infant Vaccination Coverage in Guatemala: A Pilot Randomized Controlled Trial," *Vaccine*, 34 (21), 2437–2443. [1188]

GIBSON, DUSTIN G., BENARD OCHIENG, E. WANGECI KAGUCIA, JOYCE WERE, KYLA HAYFORD, LAWRENCE H. MOULTON, ORIN S. LEVINE, FRANK ODHIAMBO, KATHERINE L. O'BRIEN, AND DANIEL 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 (4), e428–e438. [1188,1205]

HADAD, VITOR, DAVID A. HIRSHBERG, RUOHAN ZHAN, STEFAN WAGER, AND SUSAN ATHEY (2021): "Confidence Intervals for Policy Evaluation in Adaptive Experiments," *Proceedings of the National Academy of Sciences*, 118 (15), e2014602118. [1185]

HAMMER, SCOTT M., KATHLEEN E. SQUIRES, MICHAEL D. HUGHES, JANET M. GRIMES, LISA M. DEMETER, JUDITH S. CURRIER, JOSEPH J. JR. ERON, JUDITH E. FEINBERG, HENRY H. JR. BALFOUR, LAWRENCE R. DEYTON et al. (1997): "A Controlled Trial of Two Nucleoside Analogues Plus Indinavir in Persons With Human Immunodeficiency Virus Infection and CD4 Cell Counts of 200 per Cubic Millimeter or Less," *New England Journal of Medicine*, 337 (11), 725–733. [1185]

HE, XUMING, AND QI-MAN SHAO (2000): "On Parameters of Increasing Dimensions," *Journal of multivariate analysis*, 73 (1), 120–135. [1196,1219]

HINZ, OLIVER, BERND SKIERA, CHRISTIAN BARROT, AND JAN U. BECKER (2011): "Seeding Strategies for Viral Marketing: An Empirical Comparison," *Journal of Marketing*, 75 (6), 55–71. [1206]

IYENGAR, RAGHURAM, CHRISTOPHE VAN DEN BULTE, AND THOMAS W. VALENTE (2010): "Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science*, 30 (2), 195–212. [1206]

JACKSON, MATTHEW O. (2008): "Average Distance, Diameter, and Clustering in Social Networks With Homophily," in *Proceedings of the Workshop in Internet and Network Economics (WINE 2008)*. Lecture Notes in Computer Science, ed. by Christos Papadimitriou and Shuzhong Zhang. Berlin Heidelberg: Springer-Verlag. arXiv:0810.2603v1. [1206]

——— (2017): "A Typology of Social Capital and Associated Network Measures,", http://ssrn.com/abstract=3073496. [1206]

JACKSON, MATTHEW O., AND LEEAT YARIV (2011): *Diffusion, Strategic Interaction, and Social Structure*. Handbook of Social Economics, ed. by Jess Benhabib, Alberto Bisin, AND Matthew O. Jackson. San Diego: North Holland. [1206]

JAVANMARD, ADEL, AND ANDREA MONTANARI (2013): "Model Selection for High-Dimensional Regression Under the Generalized Irrepresentability Condition," in *Advances in Neural Information Processing Systems*, 3012–3020. [1187,1196,1218]

——— (2014): "Confidence Intervals and Hypothesis Testing for High-Dimensional Regression," *The Journal of Machine Learning Research*, 15, 2869–2909. [1203]

——— (2018): "Debiasing the Lasso: Optimal Sample Size for Gaussian Designs," *The Annals of Statistics*, 46, 2593–2622. [1203]

JIA, JINZHU, AND KARL ROHE (2015): "Preconditioning the Lasso for Sign Consistency," *Electronic Journal of Statistics*, 9 (1), 1150–1172. [1187,1193,1194,1211,1216]

JOHRI, MIRA, MYRIAM CIELO PÉREZ, CATHERINE ARSENAULT, JITENDAR K. SHARMA, NITIKA PANT PAI, SMRITI PAHWA, AND MARIE-PIERRE 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. [1188]

KARING, ANNE (2018): *Social Signaling and Childhood Immunization: A Field Experiment in Sierra Leone*. Berkeley: University of California. [1188]

KASY, MAXIMILIAN, AND ANJA SAUTMANN (2021): "Adaptive Treatment Assignment in Experiments for Policy Choice," *Econometrica*, 89 (1), 113–132. [1185]

KATONA, ZSOLT, PETER PAL ZUBCSEK, AND MIKLOS SARVARY (2011): "Network Effects and Personal Influences: The Diffusion of an Online Social Network," *Journal of Marketing Research*, 48 (3), 425–443. [1206]

KATZ, ELIHU, AND PAUL F. LAZARSFELD (1955): *Personal Influence: The Part Played by People in the Flow of Mass Communication*. Glencoe, IL: Free Press. [1207]

KEMPE, DAVID, JON KLEINBERG, AND EVA TARDOS (2003): "Maximizing the Spread of Influence Through a Social Network," in *Proc. 9th Intl. Conf. on Knowledge Discovery and Data Mining*, 137–146. [1206]

KRACKHARDT, DAVID (1996): "Strucutral Leverage in Marketing," in *Networks in Marketing*, ed. by Dawn Iacobucci. Thousand Oaks: Sage, 50–59. [1206]

LEE, JASON D., DENNIS L. SUN, YUEKAI SUN, AND JONATHAN E. TAYLOR (2016): "Exact Post-Selection Inference, With Application to the Lasso," *The Annals of Statistics*, 44 (3), 907–927. [1196]

LEEB, HANNES, AND BENEDIKT M. PÖTSCHER (2005): "Model Selection and Inference: Facts and Fiction," *Econometric Theory*, 21 (1), 21–59. [1187,1196]

——— (2008): "Sparse Estimators and the Oracle Property, or the Return of Hodges' Estimator," *Journal of Econometrics*, 142, 201–211. [1196]

MCCLOSKEY, ADAM (2020): "Asymptotically Uniform Tests After Consistent Model Selection in the Linear Regression Model," *Journal of Business & Economic Statistics*, 38 (4), 810–825. [1197]

MCKENZIE, DAVID (2019): "Be Careful With Inference From 2-by-2 Experiments and Ther Cross-Cutting Designs," online, accessed October, 19, 2019. [1185]

MEKONNEN, ZELEKE ABEBAW KASSAHUN ALEMU GELAYE, MARTIN C. WERE KASSAHUN DESSIE GASHU, AND BINYAM CHAKILU TILAHUN (2019): "Effect of Mobile Text Message Reminders on Routine Childhood Vaccination: A Systematic Review and Meta-Analysis," *Systematic reviews*, 8 (1), 1–14. [1206]

MURALIDHARAN, KARTHIK, MAURICIO ROMERO, AND KASPAR WUTHRICH (2019): "Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments." [1185,1187,1215]

NFHS (2016): "National Family Health Survey-4 State Fact Sheet for Haryana." [1203]

NIE, LIZHEN, AND VERONIKA ROČKOVÁ (2022): "Bayesian Bootstrap Spike-and-Slab Lasso," *Journal of the American Statistical Association*, 1–35. Just-accepted. [1203]

OYO-ITA, ANGELA, CHARLES S. WIYSONGE, CHIOMA ORINGANJE, CHUKWUEMEKA E. NWACHUKWU, OLABISI ODUWOLE, AND MARTIN M. MEREMIKWU (2016): "Interventions for Improving Coverage of Childhood Immunisation in Low-and Middle-Income Countries," *Cochrane Database of Systematic Reviews*, 7. [1188]

OZAWA, SACHIKO, ANDREW MIRELMAN, MEGHAN L. STACK, DAMIAN G. WALKER, AND ORIN S. LEVINE (2012): "Cost-Effectiveness and Economic Benefits of Vaccines in Low-and Middle-Income Countries: A Systematic Review," *Vaccine*, 31 (1), 96–108. [1212]

REGAN, ANNETTE K., LAUREN BLOOMFIELD, IAN PETERS, AND PAUL V. EFFLER (2017): "Randomized Controlled Trial of Text Message Reminders for Increasing Influenza Vaccination," *The Annals of Family Medicine*, 15 (6), 507–514. [1188]

ROGERS, EVERETT M. (1995): *Diffusion of Innovations*. Free Press. [1206]

ROHE, KARL (2014): "A Note Relating Ridge Regression and ols p-Values to Preconditioned Sparse Penalized Regression," arXiv preprint. arXiv:1411.7405. [1187,1211]

TIBSHIRANI, RYAN J., JONATHAN TAYLOR, RICHARD LOCKHART, AND ROBERT TIBSHIRANI (2016): "Exact Post-Selection Inference for Sequential Regression Procedures," *Journal of the American Statistical Association*, 111 (514), 600–620. [1196]

UDDIN, MD JASIM, MD SHAMSUZZAMAN, LILY HORNG, ALAIN LABRIQUE, LAVANYA VASUDEVAN, KELSEY ZELLER, MRIDUL CHOWDHURY, CHARLES P. LARSON, DAVID BISHAI, AND NURUL 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 (2), 276–283. [1188]

UNICEF AND WHO (2019): "Progress and Challenges With Achieving Universal Immunization Coverage: 2018 Estimates of Immunization Coverage WHO." [1184]

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). https://math.stackexchange.com/users/1551/user1551. [1217]

VAN DE GEER, SARA (2019): "On the Asymptotic Variance of the Debiased Lasso," *Electronic Journal of Statistics*, 13 (2), 2970–3008. [1203]

WAKADHA, HOTENZIA, SUBHASH CHANDIR, ELIJAH VICTOR WERE, ALAN RUBIN, DAVID OBOR, ORIN S. LEVINE, DUSTIN G. GIBSON, FRANK ODHIAMBO, KAYLA F. LASERSON, AND DANIEL 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 (6), 987–993. [1188]

WHO (2019): *Situation Analysis of Immunization Expenditure*. [1184]
ZHAN, RUOHAN, ZHIMEI REN, SUSAN ATHEY, AND ZHENGYUAN ZHOU (2021): "Policy Learning With Adaptively Collected Data," arXiv preprint. arXiv:2105.02344. [1185]
ZHANG, CUN-HUI, AND STEPHANIE S. ZHANG (2014): "Confidence Intervals for Low Dimensional Parameters in High Dimensional Linear Models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76 (1), 217–242. [1202,1203]
ZHAO, PENG, AND BIN YU (2006): "On Model Selection Consistency of Lasso," *Journal of Machine learning research*, 7, 2541–2563. [1187,1193,1216]

Co-editors Aviv Nevo and Kate Ho handled this manuscript.

Manuscript received 10 May, 2021; final version accepted 31 May, 2024; available online 28 June, 2024.

The replication package for this paper is available at https://doi.org/10.5281/zenodo.11392582. The Journal checked the data and codes included in the package for their ability to reproduce the results in the paper and approved online appendices.