NBER WORKING PAPER SERIES

EFFECTS OF LOTTERY INCENTIVES FOR INFLUENZA VACCINATION: EVIDENCE FROM A LARGE-SCALE RANDOMIZED TRIAL AND CAUSAL FOREST ANALYSIS

Kelsey Moran
Gail M. Rosenbaum
Amir Goren
Michelle N. Meyer
Christopher F. Chabris
Joseph J. Doyle Jr.

Working Paper 35537
http://www.nber.org/papers/w35537

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2026

Kelsey Moran and Gail M. Rosenbaum are co-primary authors on this work. This research was supported by the NBER Roybal Center for Behavior Change in Health (NIA P30 AG034532). We thank Dr. John Pellegrino who suggested the lottery incentive idea. We also thank Katy Milkman, Richard Thaler, and seminar participants at MIT Sloan, Harvard Kennedy School, and Harvard Health Care Policy for helpful comments. We thank Carroll Walter for project management support, Dr. Stanley Martin for his support of vaccine outreach at Geisinger, the Geisinger Business Intelligence and Advanced Analytics team for assistance extracting data, and the Geisinger Marketing team for help implementing the messages. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2026 by Kelsey Moran, Gail M. Rosenbaum, Amir Goren, Michelle N. Meyer, Christopher F. Chabris, and Joseph J. Doyle Jr. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Effects of Lottery Incentives for Influenza Vaccination: Evidence from a Large-Scale Randomized Trial and Causal Forest Analysis

Kelsey Moran, Gail M. Rosenbaum, Amir Goren, Michelle N. Meyer, Christopher F. Chabris, and Joseph J. Doyle Jr.

NBER Working Paper No. 35537
July 2026
JEL No. I12

ABSTRACT

Financial incentives may encourage vaccination, but their cost-effectiveness depends on targeting patients who are most responsive. We conducted a randomized controlled trial with 57,579 adult patients in an integrated health system during the 2021–2022 influenza season. Patients with an upcoming appointment were randomized to a no-contact control, a simple reminder, a $1 cash incentive, or a scratch-off lottery ticket incentive worth up to $5,000. Any active intervention increased vaccination at the scheduled appointment by 2 percentage points (an 8% increase over the control mean) and by 1.5 percentage points over the full influenza season. The three active arms did not differ significantly, indicating that small financial incentives provided no additional benefit beyond a reminder. Using electronic health records and area-level characteristics, we used causal forests to estimate conditional average treatment effects. Predicted responsiveness varied substantially: patients in the top quartile were 3–4 times more likely to respond to incentives than those in the bottom quartile, whereas reminder effects were relatively homogeneous. These findings suggest that low-cost reminders can increase influenza vaccination, while machine-learning methods can help identify patients most responsive to small financial incentives.

Kelsey Moran
University of Miami
Miami Herbert Business School
Department of Health Management and Policy
kcmoran@miami.edu

Gail M. Rosenbaum
Geisinger
grosenbaum@geisinger.edu

Amir Goren
Geisinger
agoren@geisinger.edu

Michelle N. Meyer
Geisinger
michellenmeyer@gmail.com

Christopher F. Chabris
Geisinger
chabris@gmail.com

Joseph J. Doyle Jr.
Massachusetts Institute of Technology
MIT Sloan School of Management and NBER
jjdoyle@mit.edu

A randomized controlled trials registry entry is available at https://clinicaltrials.gov/study/NCT05012163

Influenza is both common and dangerous, resulting in approximately 400,000 deaths worldwide each year (Cozza et al., 2021). Despite the proven effectiveness of influenza vaccines in reducing morbidity and mortality, vaccination rates remain suboptimal. Although everyone over the age of 6 months without contraindications is recommended to receive an annual vaccination by the Centers for Disease Control and Prevention (Grohskopf et al., 2025), only about one-half of individuals receive one (Centers for Disease Control and Prevention, 2024). Previous research has identified several approaches to improve vaccination uptake, including educational campaigns, reminders, default scheduling, alerting health risks, and financial incentives (Lau et al., 2012; Moran et al., 1996; Corace et al., 2016; Campos-Mercade et al., 2024; Rosenbaum et al., ming). Timely reminders can increase vaccination rates by 2–3 percentage points, or roughly a 10% increase compared with mean uptake (Klüver et al., 2021; Milkman et al., 2022). In both influenza and COVID-19 vaccination campaigns, reminders are thought to work primarily by reducing forgetfulness and procrastination and by making the opportunity to vaccinate more salient at the moment when action is feasible.

Financial incentives represent a complementary tool. They have been used to encourage a range of healthy behaviors such as smoking cessation and weight loss (Glanz et al., 2026), but evidence on their effectiveness for vaccination uptake is mixed (Michaud et al., 2022; Campos-Mercade et al., 2023, 2024; Jacobson et al., 2022; Chang et al., 2026). Offers of small, sure payments may increase the salience of the message to get vaccinated; they may also compensate for immediate hassles and time costs, particularly when individuals are close to indifferent about vaccinating. Meanwhile, offers of financial incentives in the form of lotteries draw on insights from behavioral science: because people tend to overweight small probabilities and are attracted to skewed payoffs, a chance for a large prize may have greater motivational impact than a small guaranteed payment (Kahneman and Tversky, 1979). Indeed, during the COVID-19 pandemic, several U.S. states used lottery-based incentives for vaccination (Thirumurthy et al., 2022). Heterogeneous responses to different incentive designs may help reconcile the mixed evidence on financial incentives for vaccination.

A better understanding of heterogeneity in treatment effects across individuals is therefore crucial for policy design and cost effectiveness. From a policy perspective, health systems and governments face fixed budgets for incentive programs and must decide which types of nudges to deploy and their targets. Offering incentives uniformly can be wasteful if many recipients would have vaccinated anyway. Financial incentives, including lotteries, may appeal to some patients more than others depending on their preferences and familiarity with them. Detecting reliable patterns of heterogeneity is challenging, however. Conventional subgroup analyses based on a handful of prespecified characteristics can have low statistical power and are vulnerable to multiple-comparisons concerns, especially in the presence of complex interactions among predictors.

In this study, we conducted a large randomized controlled trial among 57,579 adult patients at Geisinger, an integrated health system in Pennsylvania. Each day, eligible patients with a primary care or vaccine-administering specialty appointment scheduled 3 days later were randomized to one of four arms: no contact; a reminder that influenza vaccination was available at the visit; an offer of $1 conditional on vaccination at that visit; or an offer of a Pennsylvania scratch-off lottery ticket worth up to $5,000 conditional on vaccination at that visit. Patients in the three active arms received text messages

3 days and 1 day before the appointment. This design allows us to test whether small financial incentives increase vaccination beyond reminders alone and whether a lottery incentive performs differently from a small sure payment.

Our setting is also well suited to explore heterogeneity in treatment effects using machine-learning methods. The trial includes a large sample size and a rich set of attributes, including individual-level information from electronic health records and contextual data at the county and neighborhood levels. We use causal forests to estimate conditional average treatment effects, allowing treatment effects to vary flexibly across combinations of demographic characteristics, prior health-care utilization, baseline vaccination behavior, clinical risk factors, and local socioeconomic conditions. The causal forest methodology exploits potential nonlinearities and interactions in treatment effectiveness while using sample-splitting and subsampling to control overfitting and to provide valid inferences about heterogeneity (Wager and Athey, 2018; Athey et al., 2019; Athey and Wager, 2019).

This paper contributes to the existing literature in three ways. First, we provide new evidence on how reminders and small financial incentives affect influenza-shot uptake in the context of an integrated health system, complementing prior work focused on pharmacy or community settings. Second, we apply the causal forest methodology to uncover treatment-effect heterogeneity across a rich set of covariates and to identify subgroups of patients for whom financial incentives, and lotteries in particular, are most effective. Third, we evaluate how attribute selection and data granularity affect causal forest performance, comparing models based on continuous versus discretized covariates and county versus ZIP-code measures. By advancing our understanding of which nudges work best for which patients, and of how to reliably learn such patterns from routine administrative data, we aim to inform future public health strategies that seek to improve vaccination uptake in a cost-effective and scalable way.

Results

Participants were 58% female, 94% white, 4% Black, with a mean age of 51 years, and a risk score that measures the probability of influenza complications was 5%. About half of the sample received an influenza vaccination in the prior season, and most of the upcoming appointments targeted by the intervention were for family medicine rather than specialty appointments. Two county characteristics of interest are population, averaging 150,000, and Pennsylvania state lottery sales per person in the county during 2019, averaging close to $400. Baseline characteristics were balanced across the trial arms (*SI Appendix*, Table S1).

Fig. 1 reports the influenza vaccination rates across the four arms. Compared with the passive control group, the messaging campaign increased influenza vaccination uptake at the scheduled appointment by 2.4 percentage points for the simple reminder, 1.6 percentage points for the small cash incentive, and 1.9 percentage points for the scratch-off lottery ticket incentive. All three differences are significantly different from the passive control group ($P < 0.001$ for the reminder and scratch-off lottery, and $P = 0.002$ for the cash incentive). Meanwhile, effects for the reminder and each of the financial incentives are not statistically significantly different ($P = 0.38$ for the scratch-off lottery; $P = 0.12$ for the small cash incentive). When examining the effect of the campaign over the entire influenza season, 48% of the passive

control group were vaccinated, while 49% of the active treatment arms were vaccinated. The effects are again statistically significant ($P < 0.01$ for the reminder and scratch-off, $P = 0.052$ for the small cash incentive). Taken together with the increase in uptake at the initial appointment, the results imply that the active messaging interventions increased influenza vaccination by approximately 1 percentage point and accelerated the timing of vaccination.

SI Appendix, Tables S2–S4 show that these results are robust to prespecified checks, including exclusion of study participants who cohabitate with others in the study, logit estimation, restricting to those who received both pre-appointment messages, the use of demographic controls, and assuming homoskedastic errors when calculating standard errors. In addition, the checks imply that the increase in vaccination occurred at the scheduled appointment rather than before or afterwards. The nudges increased completion of the scheduled appointment by 1.35 percentage points ($P < 0.01$) over the control-group mean of 69% (*SI Appendix*, Table S2).

The average treatment effects may mask heterogeneity in treatment effects. Causal forest estimation provides predicted treatment effectiveness for each individual based on their combination of attributes used in the model. The strength of the relationship between the predicted and actual treatment effects measures how well the causal forest estimates reveal heterogeneity in effectiveness. Fig. 2 plots the actual treatment effects against the predicted treatment effects across twenty equally sized bins. Panel A reports results for any active treatment versus the passive control group. Examining the horizontal spread of the data, the predicted treatment effects are nearly all positive, and they range up to a 6 percentage-point effect in the top vigintile. Similar ranges in the predicted treatment effects are found for each of the treatment arms as well. The slopes of the bin scatter plots show that the causal forest estimates reveal heterogeneity. For any treatment, the slope of 0.81 means that those with a 10% higher out-of-sample predicted treatment effect have an 8.1% higher actual treatment effect, on average. Panels B–D show that the heterogeneity is found for the financial incentives rather than for the simple reminder: there is relatively little relationship between the predicted and actual treatment effects for the reminder (slope = 0.17), while there is a more positive slope for cash (0.41) and for the scratch-off lottery ticket (0.54). This is consistent with reminders being valuable across many types of patients, while small financial incentives may be effective for only a subset of patients.

Table 1 shows that those in the top quartile of predicted effectiveness according to the causal forest estimates are 3.8 percentage points more likely to be vaccinated compared with the passive control group when they are sent any message or offered either of the small financial incentives. In contrast, those in the bottom quartile of predicted effectiveness show much smaller treatment effects: 0.3 percentage points for any message, and smaller and statistically insignificant treatment effects for the two financial incentives. For the simple reminder, the treatment effects are similar regardless of the predicted-effectiveness quartile, reflecting the lack of heterogeneity for this treatment. Similarly, standard checks on model fit for the causal-forest estimation reveal that the forest's treatment-effect predictions are well-calibrated out of sample (*SI Appendix*, Table S5), while statistically significant heterogeneity across the top and bottom quartiles is detected for any treatment ($P = 0.004$) and the scratch-off lottery ticket ($P = 0.007$), but not for the simple reminder ($P = 0.938$) or the small cash incentive ($P = 0.114$).

Fig. 3 describes how patient characteristics vary with the predicted treatment effects for the scratch-

off lottery ticket, with similar figures for the other treatments in the *SI Appendix* Fig. S1. Each estimate in the figure is the difference in the mean of the characteristic for those in the top versus the bottom quartile of predicted effectiveness. Continuous characteristics were standardized to have a mean of zero and SD of one to aid comparisons with binary characteristics. Those in the top quartile of predicted effectiveness are older by 0.7 SD compared with those in the bottom quartile. Those who received an influenza vaccination in the prior year are also predicted to have larger treatment effects, with the average 40 percentage points higher in the top quartile versus the bottom quartile. Meanwhile, those in the top quartile of effectiveness averaged fewer encounters in the prior year, by 0.3 SD, and they lived in counties with smaller populations, by 0.2 SD. The scratch-off lottery offer appeared more effective in counties with less schooling and lower median incomes. In addition, counties with higher lottery sales were less likely to be among the most responsive.

The causal forest models reveal heterogeneity that would be difficult to find with a series of subgroup analyses. *SI Appendix*, Table S6 considers the 20 characteristics that were found to be most disparate when comparing the top and bottom quartiles of predicted effectiveness. Even without adjustment for multiple hypothesis testing, only two of the 20 characteristics exhibit statistically significant differences in treatment effects of the lottery offer across subgroups defined by these attributes: receipt of an influenza vaccination in the prior year and marital status. This shows that the characteristics by themselves are substantially less predictive than the combination of patient and area characteristics that the causal forest is designed to exploit. Prespecified heterogeneity analyses for any treatment show that those who received an influenza vaccination in the prior year are more likely to be affected by a nudge, but effects do not vary systematically by age group, insurer, ZIP code income level, sex, type of healthcare provider, previous COVID vaccination or infection, or county lottery ticket sales. (*SI Appendix*, Table S7).

We also explored how choices over the types of attributes included in the causal forest estimation affect model performance. Allowing the model to select from continuous variables provides better fit when estimating heterogeneity compared with discretized measures, demonstrating that the forests perform better when machine learning is used to categorize continuous measures. Indeed, when continuous measures are discretized into quartile indicators, the models no longer reveal any heterogeneity (*SI Appendix*, Fig. S2). The models perform worse when using attributes measured at the ZIP-code level rather than at the county level (*SI Appendix*, Fig. S3). Adding indicators for the county or ZIP code of residence, or adding ZIP-code attributes in addition to the county measures, results in substantially worse performance (*SI Appendix*, Fig. S4).

In addition to the targeted influenza vaccination, we also considered secondary outcomes, namely influenza diagnosis and complications, COVID-19 vaccination and diagnosis, and influenza vaccination in the following year when financial incentives were not provided. *SI Appendix*, Table S8 shows that the simple reminder group was less likely to be diagnosed with influenza, by 0.26 percentage points (SE = 0.10), which is large compared with the control-group mean of 0.92 percentage points. We do not find changes in complications from influenza, however. We also do not find that offering small financial incentives changed the likelihood of these secondary outcomes. In particular, when we examine influenza vaccination in the following influenza season (2022–2023), it appears that offering incentives one year did not discourage vaccination when such incentives were no longer present.

Discussion

In this large-scale randomized trial, simple text-message reminders and small financial incentives significantly increased influenza vaccination uptake. Relative to a passive control group, all three active messaging interventions increased vaccination at the scheduled appointment by approximately 2 percentage points and seasonal vaccination uptake by about 1 percentage point. Although modest in absolute terms, these effects are meaningful from a public health perspective because they were achieved through highly scalable, fully automated interventions at very low cost. The average cost of the reminder intervention was approximately $1.50 per participant, reflecting messaging costs alone, whereas the average cost of the incentive interventions was approximately $2.00 per participant, including incentive, shipping, and messaging costs. By comparison, estimates of the societal benefit of an additional influenza vaccination—including vaccination costs, reductions in mortality through contagion externalities, and gains in workforce productivity—range from roughly $100 to $300 per vaccination (White, 2021; Preaud et al., 2014).

Reminders likely operate by reducing procrastination and forgetfulness—bridging the gap between intention and action—while the lottery incentive may additionally motivate those who overweight low-probability payoffs. The similarity in average effects across arms suggests that, for most patients, a well-timed reminder is sufficient to prompt follow-through when vaccination is convenient, and that small financial incentives do not substantially increase uptake on top of reminders. This pattern is consistent with a view in which many patients already intend to vaccinate and primarily benefit from a prompt at the right time and place, while only a subset of patients are sufficiently sensitive to financial incentives for them to generate additional uptake beyond reminders.

Using the large sample combined with a wide array of attributes, causal forests revealed little heterogeneity in response to simple reminder messages. In contrast, the financial incentives produced greater variation in effect sizes. The top quartile of predicted responders to the incentives experienced effects that were roughly 2.5–3 times larger than the rest of the sample. A plausible explanation is that most patients are accustomed to reminders from their health system and respond in fairly uniform ways regardless of their individual characteristics, whereas a scratch-off lottery ticket or small cash incentive may appeal more selectively. Meanwhile, nudges appear to be more effective for those who are generally more likely to get an influenza vaccination: older individuals, patients with an influenza vaccination the previous year, and those with fewer recent encounters but higher clinical risk scores. This pattern is consistent with nudges being most effective when they help individuals follow through on existing intentions, as they are unlikely to change underlying preferences about vaccination.

The setting also allows us to learn more about the usefulness of causal forest methods in applied healthcare contexts. In particular, the method benefited from county-level measures rather than ZIP-code area measures, and adding indicators for each county or ZIP code of residence resulted in poorer performance. These findings underscore that more granular or more numerous covariates do not necessarily improve heterogeneity analyses: when additional variables are noisy or weakly related to treatment effects, they can dilute signal and degrade performance. Comparisons between models using continuous variables and those using discretized versions further demonstrate that allowing the algorithm to discover nonlinearities

in continuous covariates can yield richer insights than relying on ad hoc categorization.

This study has several limitations. First, the findings are most directly applicable to patients in our study population: primarily white adults living in northeastern and central Pennsylvania with established primary-care relationships within an integrated health system. Response to reminders and incentives may differ in more diverse populations, among individuals without a regular source of care, or in health systems with different communication norms. Second, we observe vaccination at Geisinger and at major pharmacy chains, but our data likely miss some vaccinations at smaller pharmacies, workplace clinics, other medical practices, and community events. Third, while causal forests help mitigate concerns about multiple testing and model overfitting, some detected patterns may not replicate in other settings, and operational use of targeting rules would require ongoing validation as patient populations and influenza seasons change.

Despite these caveats, our results provide actionable policy implications. Health systems and clinics can deploy low-cost reminder messages to reliably increase timely vaccination, leveraging existing scheduling and messaging infrastructure. Financial incentives appear less cost effective on average but may still be valuable when targeted to patients predicted to be particularly responsive. In practice, health systems could use routinely collected electronic health record data to generate causal-forest-based scores each season and offer incentives only to those above a chosen threshold, while monitoring equity implications. More broadly, the combination of scalable nudges with data-driven targeting holds promise for improving preventive care strategies beyond influenza vaccination. As health systems increasingly integrate behavioral science and machine learning into routine operations, understanding not only whether nudges work on average but also for whom they work best will be critical for designing cost-effective and equitable interventions.

Materials and Methods

Study design. The study was conducted as a randomized controlled trial at Geisinger, a large health system located in northeastern and central Pennsylvania, during the fall of 2021. Eligible participants were patients who had a primary care provider of record at Geisinger and a Pennsylvania mailing address, were at least 18 years old, and had not opted out of receiving messages from Geisinger. Each day, we identified all patients with an appointment scheduled 3 days later with a primary care physician or vaccine-administering specialist. Appointments could be scheduled between September 4, 2021 and December 31, 2021. Participants were randomized into four arms (Coppock et al., 2019): a no-contact control (N = 14,398); a reminder treatment notifying patients they could get an influenza vaccination at their appointment (N = 14,393); a cash-incentive treatment notifying patients they would receive $1 mailed to their home if vaccinated (N = 14,394); and a lottery incentive notifying patients they would receive a Pennsylvania scratch-off lottery ticket worth up to $5,000 if vaccinated ($N$ = 14,394). In the three treatment arms, two text messages were sent: one 3 days and one 1 day before the appointment. In the two incentive arms, the first message announced that vaccination would be rewarded and the second specified whether the reward was cash or a scratch-off lottery ticket. Further experimental design details, outcome measurement, complete SMS text, and information about distribution of the incentives

are provided in the *SI Appendix*.

Data and outcomes. The data are sourced from electronic health records and include baseline demographics (age, race, ethnicity, sex, marital status), prior utilization (number of encounters in the prior year and receipt of the influenza vaccination in the prior year), tenure as a Geisinger patient, smoking status, body mass index, indicators for 16 common health conditions, payer indicators, and month of the appointment. In addition, Geisinger used a wider set of data to calculate a risk score that predicts the likelihood of influenza complications (Wolk et al., 2022; Rosenbaum et al., ming). We also incorporated U.S. Census measures of income, education, and unemployment at the county and ZIP-code levels, as well as county-level state lottery participation. The preregistered primary outcome was whether the patient received an influenza vaccine at the appointment, 3 days after enrollment. Vaccination data were extracted from the Geisinger electronic health record, Geisinger Health Plan medical and pharmacy claims, and vendors that share vaccine data with Geisinger, including major pharmacy chains. If a patient had vaccine data from multiple sources available, the data pull prioritized data from the Geisinger electronic health record over other data sources. Secondary outcomes included vaccination receipt at other time horizons, regardless of location, COVID-19 vaccination, influenza diagnosis and complications, and influenza vaccination the following year. As an exploratory outcome that was not preregistered, we also examined attendance at the scheduled appointment.

Statistical power and study oversight. The study was powered at 80% to detect a 2-percentage-point increase in vaccination rates at a two-sided $\alpha = 0.05$ significance level. In practice, more patients entered the study than were forecast, and given the mean of the control group, the setting provides *post hoc* power to detect a difference of 1.45 percentage points when comparing any two arms. The Geisinger Institutional Review Board approved the study and granted waivers of consent and HIPAA authorization. The trial was registered at ClinicalTrials.gov (identifier NCT05012163).

Statistical analysis. We tested differences in vaccination rates using ordinary least-squares regression with heteroskedasticity-robust standard errors. For subject i, we estimate

$$Y_i = \alpha + \beta_1 \mathbf{1}(\text{Reminder})_i + \beta_2 \mathbf{1}(\text{Cash})_i + \beta_3 \mathbf{1}(\text{Lottery})_i + \varepsilon_i. \tag{1}$$

To examine heterogeneous treatment effects, we employed the generalized random forest, or causal forest, methodology (Wager and Athey, 2018; Athey et al., 2019), which grows an ensemble of honest regression trees that split the sample to separate individuals with differing treatment effects. This approach yields estimates of the conditional average treatment effect, the treatment effect for individuals with a given combination of covariates. Honesty—using separate subsamples to choose the splits and to estimate effects within them—together with subsampling guards against overfitting and supports valid inference about heterogeneity. Further details on the causal forest estimation, validation, and robustness checks are provided in the *SI Appendix*.

For the quartile-based heterogeneity analysis, we estimate

$$Y_i = \alpha + \sum_{q=1}^{3} \beta_q \mathbf{1}(Q_i = q) + \sum_{q=1}^{4} \gamma_q \mathbf{1}(\text{Treated})_i \times \mathbf{1}(Q_i = q) + \varepsilon_i, \tag{2}$$

where $Q_i \in \{1, 2, 3, 4\}$ denotes the quartile of the causal-forest predicted treatment effect for individual i, with $Q_i = 1$ denoting the top quartile and $Q_i = 4$ the bottom quartile. The bottom quartile is the omitted category for the main quartile indicators, and the γ_q coefficients give the treatment effect within quartile q. The model is again estimated using ordinary least-squares regression with heteroskedasticity-robust standard errors.

References

Athey, S., J. Tibshirani, and S. Wager (2019). Generalized random forests. *The Annals of Statistics 47*(2), 1148–1178.

Athey, S. and S. Wager (2019). Estimating treatment effects with causal forests: An application. *Observational Studies 5*(2), 37–51.

Campos-Mercade, P., A. N. Meier, S. Meier, D. G. Pope, F. H. Schneider, and E. Wengström (2024). Incentives to vaccinate. Working Paper 32899, National Bureau of Economic Research.

Campos-Mercade, P., A. N. Meier, D. Pope, and F. H. Schneider (2023). Motivating vaccination with financial incentives. *Trends in Cognitive Sciences 27*(12), 1099–1101.

Centers for Disease Control and Prevention (2024, September). Flu vaccination coverage, United States, 2023–24 influenza season. https://www.cdc.gov/fluvaxview/coverage-by-season/2023-2024.html. Accessed: April 16, 2025.

Chang, T. Y., M. Jacobson, M. Shah, R. Pramanik, and S. B. Shah (2026). Impact of reminder messages, with and without financial incentives, on influenza vaccination: A randomized trial in a California health system. *Vaccine 75*, 128263.

Coppock, A., J. Cooper, and N. Fultz (2019). randomizr: Easy-to-Use Tools for Common Forms of Random Assignment and Sampling. *R package version 0.20.0*.

Corace, K. M., J. A. Srigley, D. P. Hargadon, D. Yu, T. K. MacDonald, L. R. Fabrigar, and G. E. Garber (2016). Using behavior change frameworks to improve healthcare worker influenza vaccination rates: A systematic review. *Vaccine 34*(28), 3235–3242.

Cozza, V., H. Campbell, H. H. Chang, A. D. Iuliano, J. Paget, N. N. Patel, R. C. Reiner, C. Troeger, C. Viboud, J. S. Bresee, et al. (2021). Global seasonal influenza mortality estimates: a comparison of 3 different approaches. *American Journal of Epidemiology 190*(5), 718–727.

Glanz, K., H. Thirumurthy, and N. A. D'Cruze (2026). Effectiveness of financial incentives for health behavior change: A narrative review and analysis. *Annual Review of Public Health 47*, 215–233.

Grohskopf, L. A., L. H. Blanton, J. M. Ferdinands, C. Reed, V. G. Dugan, and D. C. Daskalakis (2025). Prevention and control of seasonal influenza with vaccines: Recommendations of the Advisory Committee on Immunization Practices — United States, 2025–26 influenza season. *MMWR. Morbidity and Mortality Weekly Report 74*(32), 500–507.

Jacobson, M., T. Y. Chang, M. Shah, R. Pramanik, and S. B. Shah (2022). Can financial incentives and other nudges increase COVID-19 vaccinations among the vaccine hesitant? a randomized trial. *Vaccine 40*(43), 6235–6242.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–291.

Klüver, H., F. Hartmann, M. Humphreys, F. Geissler, and J. Giesecke (2021). Incentives can spur COVID-19 vaccination uptake. *Proceedings of the National Academy of Sciences 118*(36), e2109543118.

Lau, D., J. Hu, S. R. Majumdar, D. A. Storie, S. E. Rees, and J. A. Johnson (2012). Interventions to improve influenza and pneumococcal vaccination rates among community-dwelling adults: a systematic review and meta-analysis. *The Annals of Family Medicine 10*(6), 538–546.

Michaud, T. L., P. A. Estabrooks, W. You, J. Ern, D. Scoggins, K. Gonzales, K. M. King, H. Dai, and D. Su (2022). Effectiveness of incentives to improve the reach of health promotion programs- a systematic review and meta-analysis. *Preventive Medicine 162*, 107141.

Milkman, K. L., L. Gandhi, M. S. Patel, H. N. Graci, D. M. Gromet, H. Ho, J. S. Kay, T. W. Lee, J. Rothschild, J. E. Bogard, et al. (2022). A 680,000-person megastudy of nudges to encourage vaccination in pharmacies. *Proceedings of the National Academy of Sciences 119*(6), e2115126119.

Moran, W. P., K. Nelson, J. L. Wofford, R. Velez, and L. D. Case (1996). Increasing influenza immunization among high-risk patients: education or financial incentive? *The American Journal of Medicine 101*(6), 612–620.

Preaud, E., L. Durand, B. Macabeo, N. Farkas, B. Sloesen, A. Palache, F. Shupo, and S. I. Samson (2014). Annual public health and economic benefits of seasonal influenza vaccination: a European estimate. *BMC Public Health 14*(1), 813.

Rosenbaum, G., A. Goren, M. Shermohammed, D. M. Wolk, A. M. Tice, J. Doyle, M. N. Meyer, et al. (Forthcoming). Algorithm aversion or appreciation? three randomized field trials of personalized risk-communication nudges to encourage flu vaccination. *Nature Human Behaviour*.

Thirumurthy, H., K. L. Milkman, K. G. Volpp, A. M. Buttenheim, and D. G. Pope (2022). Association between statewide financial incentive programs and COVID-19 vaccination rates. *PLOS ONE 17*(3), e0263425.

Wager, S. and S. Athey (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association 113*(523), 1228–1242.

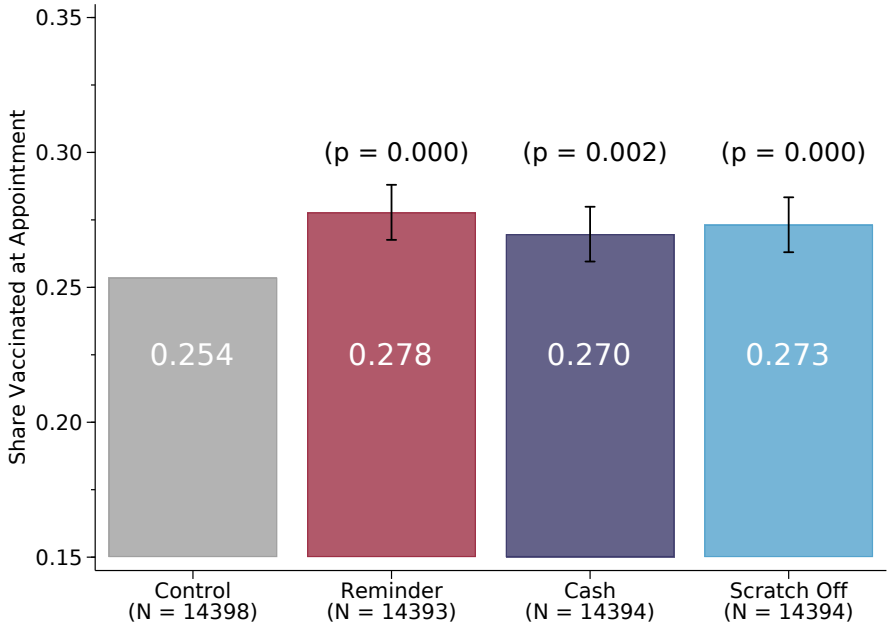
White, C. (2021). Measuring social and externality benefits of influenza vaccination. *Journal of Human Resources 56*(3), 749–785.

Wolk, D. M., A. Lanyado, A. M. Tice, M. Shermohammed, Y. Kinar, A. Goren, C. F. Chabris, M. N. Meyer, A. Shoshan, and V. Abedi (2022). Prediction of influenza complications: Development and validation of a machine learning prediction model to improve and expand the identification of vaccine-hesitant patients at risk of severe influenza complications. *Journal of Clinical Medicine 11*(15), 4342.

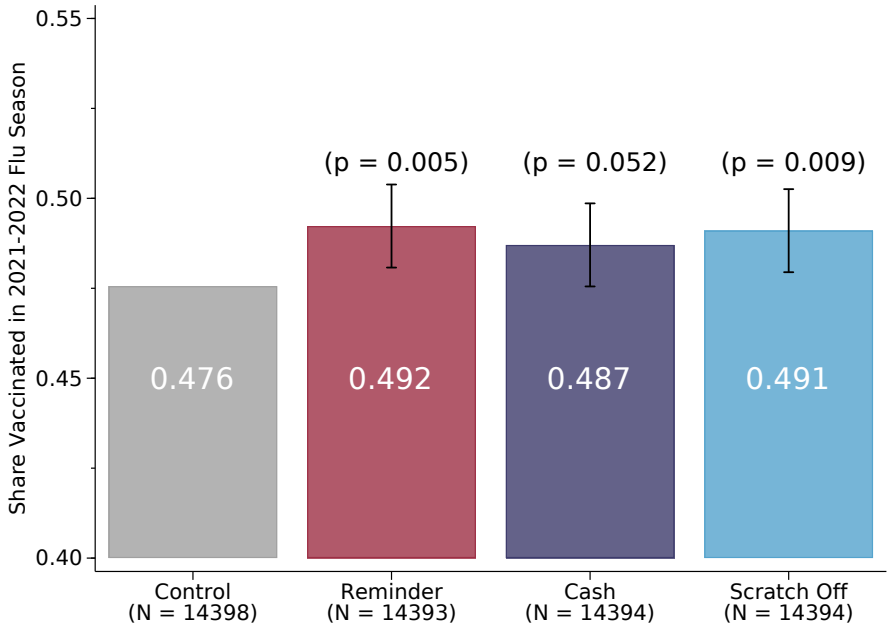
Figures and Tables

Figure 1: Average Treatment Effects

(a) Share that received the influenza vaccine at appointment



(b) Share that received the influenza vaccine in 2021–2022 influenza season

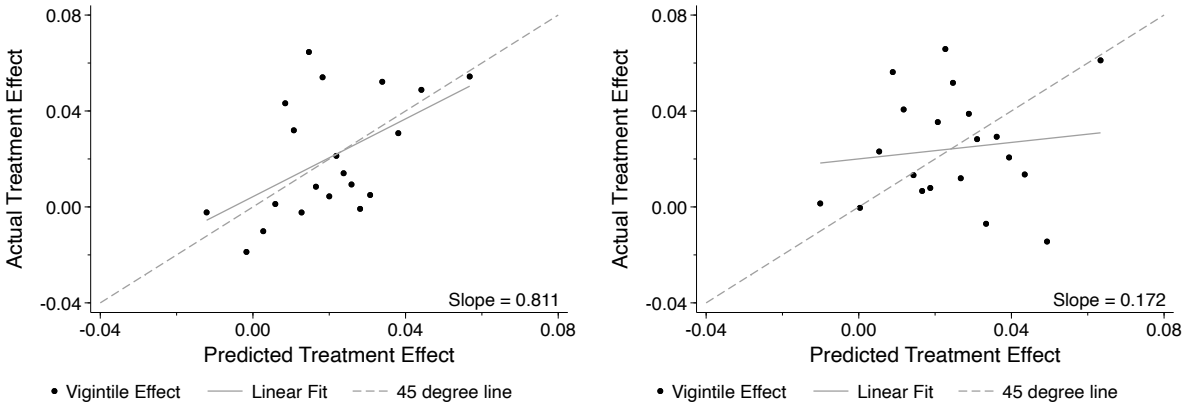


Notes: Figure shows the share of patients in the control arm and in each treatment arm that received an influenza vaccination at their appointment (Panel A) or during the full 2021–2022 influenza season (Panel B). Whiskers are 95% confidence intervals based on heteroskedasticity-robust standard errors from Eq. 1. P values relative to the control arm are shown above the bars. In Panel A, the P-value from testing the equivalence of the reminder and cash treatment effects is 0.124, that for reminder and scratch-off is 0.382, and that for cash and scratch-off is 0.508. The analogous P-values in Panel B are 0.373, 0.827, and 0.502, respectively.

Figure 2: Causal Forest Predicted vs. Actual Treatment Effects

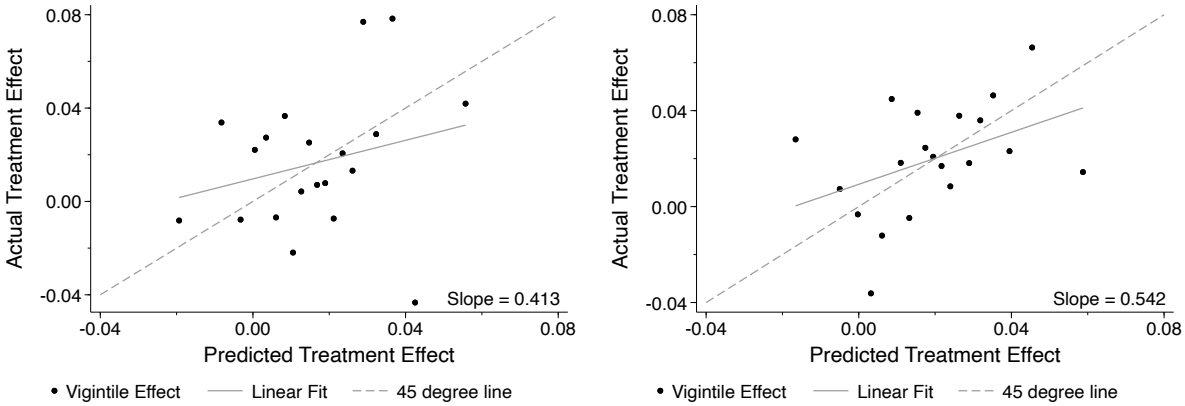
(a) Any treatment vs. control

(b) Reminder treatment vs. control



(c) Cash treatment vs. control

(d) Scratch-off treatment vs. control



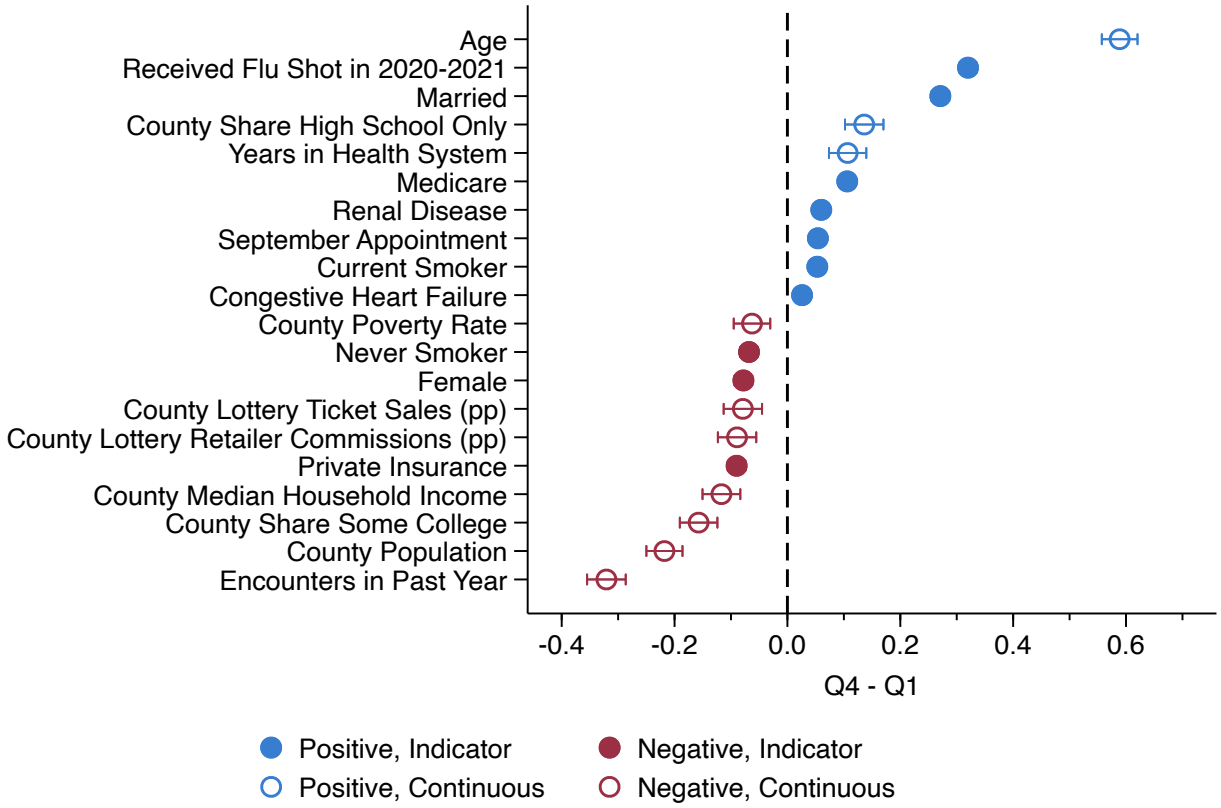
Notes: Figure plots the causal forest predicted treatment effects (x axis) and the actual treatment effects (y axis) for individuals binned into twenty vigintiles of the causal forest predicted treatment effects. All causal forests are run using basic controls and 100,000 trees. The outcome is whether the patient received an influenza vaccination at the appointment. Solid line shows the linear relationship between the actual and predicted effects. Regression-estimated slope of the solid line is included in the right-hand bottom corner. Dashed line is the 45-degree line, included for reference.

Table 1: Evidence of Heterogeneous Treatment Effects

	Shot at Appointment			
	(1) Any Treatment	(2) Reminder	(3) Cash	(4) Scratch-Off
Top Quartile of Predicted Treatment Effects	0.0383***	0.0225**	0.0375***	0.0376***
	(0.0092)	(0.0111)	(0.0112)	(0.0113)
Second Top Quartile of Predicted Treatment Effects	0.0095	0.0249**	0.0077	0.0201*
	(0.0083)	(0.0101)	(0.0101)	(0.0103)
Third Top Quartile of Predicted Treatment Effects	0.0311***	0.0259***	0.0078	0.0244**
	(0.0077)	(0.0100)	(0.0098)	(0.0097)
Bottom Quartile of Predicted Treatment Effects	0.0030	0.0237**	0.0138	-0.0024
	(0.0080)	(0.0102)	(0.0100)	(0.0098)
$H0$: Top Quartile equals Bottom Quartile, $p =$	0.004	0.938	0.114	0.007
N	57579	28791	28792	28792

Notes: Table shows an assessment of the causal forest predictions, where the causal forest is run for all three treatment arms jointly (Column 1), the reminder treatment arm only (Column 2), the cash treatment arm only (Column 3), and the scratch-off treatment arm only (Column 4). All causal forests are run using basic controls and 100,000 trees. Table shows treatment effect estimates for patients in each quartile of predicted treatment effects, following Eq. 2. P values from tests of whether quartiles have the same treatment effects are shown. Standard errors are heteroskedasticity-robust and displayed in parentheses. $^{*}P < 0.10$, $^{**}P < 0.05$, $^{***}P < 0.01$.

Figure 3: Causal Forest Heterogeneity for Scratch-off Treatment



Notes: Figure shows the difference between the mean of each characteristic for those in the top quartile versus the bottom quartile of predicted scratch-off lottery treatment effects stemming from the causal forest estimation. For continuous variables, differences are divided by the SDs of the control group to ease comparisons with binary characteristics. Whiskers show 95% confidence intervals. The figure shows only characteristics with the top 10 largest positive and negative difference magnitudes. Causal forest models are run using basic controls over 100,000 trees.

Supplemental Information

Effects of lottery incentives for influenza vaccination: Evidence from a large-scale randomized trial and causal forest analysis

Kelsey Moran[1,*], Gail Rosenbaum[2,*], Amir Goren[2], Michelle N. Meyer[2], Christopher F. Chabris[2], and Joseph J. Doyle, Jr.[3]

[1]University of Miami, corresponding author: kcmoran@miami.edu
[2]Geisinger
[3]MIT Sloan School of Management and NBER
[*]Co-first author.

Overview. This Supplemental Information provides additional detail underlying the trial reported in the main text. Section A documents the randomization procedure and the construction of the vaccination outcome; Section B reproduces the full text of the messages sent to patients in each arm; Section C describes the causal forest estimation and the procedures we use to validate it; and Section D collects the supplemental tables and figures referenced in the main text.

A Details of the Experimental Design and Measurement

This section provides further detail on the trial's implementation and measurement: how patients were randomized into arms and sent messages, how we constructed the vaccination outcome, and a note on two patients excluded after enrollment.

A.1 Randomization mechanisms

Each day, eligible patients with scheduled appointments in three days were evenly randomized into one of the four arms using the `complete_ra` function from the R package `randomizr` (Coppock et al., 2019). All patients who were sent messages three days pre-appointment (the full text of the messages for each arm is reproduced in Section B) were also sent messages one day pre-appointment, as long as they appeared on the list of eligible patients the day prior to their appointment. Patients may no longer have appeared on that list for several reasons: their appointment had been rescheduled so that they no longer had an appointment eligible for influenza vaccination on that date; Geisinger records had been updated between the patient's enrollment date and the day prior to their appointment to indicate that they were already vaccinated, rendering the patient ineligible for an influenza vaccination at their appointment; or they had opted out of text-message communications after enrollment. In cases where a patient canceled their appointment or rescheduled it for a different day, they were not sent a one-day pre-appointment message. In cases where a patient rescheduled their appointment to a different time on the same day, or where the patient had a different appointment eligible for influenza vaccination on the same day, the patient was still sent a one-day reminder message.

A.2 Vaccination measurement

Vaccine administration and timing data were extracted from the Geisinger electronic health record (EHR), which automatically imports data from the Pennsylvania state vaccine registry (including external pharmacies and other health systems) and also includes patient self-report of external vaccination; Geisinger Health Plan medical and pharmacy claims; and Geisinger vendors that report vaccine data. For each patient, if vaccine data were reported in the EHR, the record from that data source was prioritized over other sources in the vaccine data pull. Across all arms, over 90% of the vaccination records came from the EHR, and the share of records from EHR versus non-EHR sources did not differ by study arm. Within the EHR data, our dataset does not include the source of the vaccine record (i.e., whether the patient was vaccinated at Geisinger or the data were from the vaccine registry).

A.3 Additional note

Two additional patients beyond the 57,579 patients reported in the main text were inadvertently enrolled at age 17 (though they turned 18 in the three days between

enrollment and their target appointment for the study): one in the no-contact control arm and one in the reminder arm. These two patients were removed from the study and excluded from all analyses because they did not meet the age cutoff defined in our eligibility criteria.

B Message Text

This section reproduces the complete text of the messages described in Section A. For each arm, we present the messages sent three days and one day before the appointment, the automated replies to patient responses, and—for the two incentive arms—the informational materials and the mailings that accompanied the incentives.

B.1 Scratch-Off Lottery Arm Messages

3 days before the appointment
Initial message:

> Geisinger: It's flu season. [FirstName][Suffix], get a $1 PA Bee Lucky™ scratch-off ticket for receiving a flu shot at your upcoming appt:
>
> [Date], @[Time] with [DoctorName] @[CenterName]
>
> Text Y to learn how. Text FLUHELP for help. Text FLUSTOP to stop flu shot messages.

Reply to Y:

> If medical records confirm you received a flu shot at your appt, you will automatically be mailed a Pennsylvania $1 Bee Lucky™ or similar scratch-off ticket.
>
> You could win $5,000. [LINK TO INFORMATION PDF SHOWN ON NEXT PAGE]
>
> The ticket will be mailed to your address on file. You can verify or update your address at your appt. Text NO if you DON'T want the ticket.

Reply to NO:

> Thank you for your reply. You will not receive a ticket if you get a flu shot at your appointment.

1 day before the appointment
If they previously replied "Y" to get more information about the ticket and did not reply "NO" to indicate they do not want the ticket:

> Don't forget to get a flu shot at your appt tomorrow – if you do, you will be mailed a scratch-off ticket.

If they previously replied "Y" to get more information about the ticket, and replied "NO" to indicate they do not want the ticket:

> Don't forget to get a flu shot at your appt tomorrow.

If they did not reply to the original messages:

> Don't forget to get a flu shot at your appt tomorrow – if you do, you will receive a $1 PA Bee Lucky™ scratch-off ticket. Text Y to learn how.

Reply to Y – same as 3-day messages for this arm

B.2 Scratch-Off Lottery Arm - Informational PDF

The informational PDF shown on this page refers to patients learning about the incentive from an email. This is because the team had considered sending study messages via email but ultimately decided to limit the intervention to text messages.

Geisinger Flu Shot Scratch-Off Ticket
2021–2022 Flu Season

If medical records confirm you received a flu shot at the Geisinger appointment indicated in the text message or email, you will automatically be mailed a Pennsylvania $1 Bee Lucky™ scratch-off ticket, or a similar Pennsylvania $1 scratch-off ticket with a top prize of $5,000 (or the largest top prize available if no $5,000 tickets are available at the time tickets are sent).
The ticket will automatically be mailed to your address on file, unless you indicate you DON'T want a ticket. You can indicate you DON'T want a ticket in one of 2 ways:

- If you learned about the ticket via text, you can reply N to the text message to indicate you DON'T want the ticket. If you prefer, you can instead call (570) 214-6607 or email behavioralinsightsteam@geisinger.edu to indicate you DON'T want the ticket.
- If you learned about the ticket via email, you can call (570) 214-6607 or email behavioralinsightsteam@geisinger.edu to indicate you DON'T want the ticket.

You can verify or update your address at your appointment.
Tickets will be mailed no later than February 28, 2022.
Eligibility requirements:

- You must be a PA resident with a valid Pennsylvania mailing address–if the mail is undeliverable, your ticket will be forfeited
- You must be at least 18 years old
- You must get your flu shot at a particular appointment:
  - If you learned about the appointment via text, you must get a flu shot at the appointment indicated in the text message
  - If you learned about the appointment via email, you must get a flu shot at your next scheduled Geisinger appointment where a flu shot is available

For questions, call (570) 214-6607 or email behavioralinsightsteam@geisinger.edu.

B.3 $1 Cash Incentive Arm Messages

3 days before the appointment
Initial message:

> Geisinger: It's flu season. [FirstName][Suffix], get $1 in cash for receiving a flu shot at your upcoming appt:
>
> [Date], @[Time] with [DoctorName] @[CenterName]
>
> Text Y to learn how. Text FLUHELP for help. Text FLUSTOP to stop flu shot messages.

Reply to Y:

> If medical records confirm you received a flu shot at your appt, you will automatically be mailed $1 in cash. [LINK TO INFORMATION PDF SHOWN ON NEXT PAGE]
>
> The cash will be mailed to your address on file. You can verify or update your address at your appt. Text NO if you DON'T want the $1.

Reply to NO:

> Thank you for your reply. You will NOT receive $1 if you get a flu shot at your appointment.

1 day before the appointment
If they previously replied "Y" to get more information about the $1 and did not reply "NO" to indicate they do not want the $1:

> Don't forget to get a flu shot at your appt tomorrow – if you do, you will be mailed $1 in cash.

If they previously replied "Y" to get more information about the $1, and replied "NO" to indicate they do not want the $1:

> Don't forget to get a flu shot at your appt tomorrow.

If they did not reply to the original messages:

> Don't forget to get a flu shot at your appt tomorrow – if you do, you will be mailed $1 in cash. Text Y to learn how.

Reply to Y – same as 3-day messages for this arm

B.4 $1 Cash Incentive Arm - Informational PDF

The informational PDF shown below refers to patients learning about the incentive from an email. This is because the team had considered sending study messages via email but ultimately decided to limit the intervention to text messages.

Geisinger Flu Shot $1 Cash
2021–2022 Flu Season

If medical records confirm you received a flu shot at the Geisinger appointment indicated in the text message or email, you will automatically be mailed $1 in cash.
The cash will automatically be mailed to your address on file, unless you indicate you DON'T want the $1. You can indicate you DON'T want the $1 in one of 2 ways:

- If you learned about the $1 via text, you can reply N to the text message to indicate you DON'T want the $1. If you prefer, you can instead call (570) 214-6607 or email behavioralinsightsteam@geisinger.edu to indicate you DON'T want the $1.
- If you learned about the $1 via email, you can call (570) 214-6607 or email behavioralinsightsteam@geisinger.edu to indicate you DON'T want the $1.

You can verify or update your address at your appointment.
Cash will be mailed no later than February 28, 2022.
Eligibility requirements:

- You must be a PA resident with a valid Pennsylvania mailing address–if the mail is undeliverable, your cash will be forfeited
- You must be at least 18 years old
- You must get your flu shot at a particular appointment:
  - If you learned about the appointment via text, you must get a flu shot at the appointment indicated in the text message
  - If you learned about the appointment via email, you must get a flu shot at your next scheduled Geisinger appointment where a flu shot is available

For questions, call (570) 214-6607 or email behavioralinsightsteam@geisinger.edu.

B.5 Simple Reminder (Active Control) Arm Messages

3 days before the appointment

Geisinger: It's flu season. [FirstName][Suffix], you can get a flu shot at your upcoming appt:

[Date], @[Time] with [DoctorName] @[CenterName]

Text FLUHELP for help. Text FLUSTOP to stop flu shot messages

1 day before the appointment

Don't forget to get a flu shot at your appt tomorrow.

B.6 Replies to FLUHELP and FLUSTOP for All Message Arms

Reply to FLUHELP

> This year, Geisinger is sending messages to patients who can get a flu shot at a scheduled appointment. Text FLUSTOP to stop receiving messages about flu shots

Reply to FLUSTOP

> Thank you for your message. You will no longer receive text messages about flu shots from this number. To re-enroll, text FLUSTART

B.7 Mailing Incentives for Scratch-Off and Cash Arms

In February 2022, patients in the Scratch-Off or Cash arms who received an influenza vaccination at Geisinger on their appointment date were sent their assigned incentive, along with the following message:

> Dear Geisinger Patient,
>
> Last year, we sent you a text message reminding you to get your annual flu shot. At that time, we offered to send you a [SCRATCH-OFF ARM: $1 PA scratch-off lottery ticket / CASH ARM: $1 bill] if you received a flu shot at the upcoming appointment indicated in the message. Your [SCRATCH-OFF ARM: lottery ticket / CASH ARM: $1 bill] is enclosed.
>
> Thank you for getting your flu shot!
>
> Sincerely,
>
> Your Geisinger Care Team

Additionally, to prompt patients to watch their mail for the incentives, patients who qualified for either incentive were sent the following via SMS text message:

> Thanks for getting your flu shot this year! Check your mail in the next week or two for an important envelope from Geisinger.

C Details of Causal Forest Analysis

We estimate conditional average treatment effects using the generalized random forest (GRF) of Athey et al. (2019), as implemented in the `grf` package in `R` (Wager and Athey, 2018; Athey and Wager, 2019). A causal forest is an adaptive nearest-neighbor method: it uses an ensemble of honest, randomized trees to learn, for each individual i, which other individuals are most similar in the way that matters for treatment-effect estimation, and then estimates that individual's effect from a locally weighted moment condition. We run the forest separately for each treatment arm against the pure control group and for any treatment against control. Throughout, $W_i \in \{0, 1\}$ denotes the treatment indicator, Y_i the outcome (vaccination at the appointment), and X_i the vector of patient- and area-level covariates described in the main text. Section C.1 details how the forest is estimated, and Section C.2 the procedures we use to validate its predictions.

C.1 Causal Forest Details

At a high level, the algorithm proceeds in four stages: (1) it strips out the baseline differences in the outcome and the treatment that the covariates can predict on their own, leaving the forest to focus on treatment-effect heterogeneity (*centering*); (2) it grows many randomized trees, each of which repeatedly splits the sample to separate patients whose treatment effects differ (*the splitting rule*); (3) each tree is grown "honestly" on a random subsample, using one set of patients to decide where to split and a separate set to estimate effects within the resulting groups (*honesty*); and (4) the trees are combined into a set of similarity weights, which are used to estimate each patient's treatment effect from the patients most similar to them (*prediction*). We describe each stage in turn.

Step 1: Centering. The algorithm first removes the part of the outcome and the treatment that the covariates can predict on their own—the "R-learner" transformation. We estimate the marginal outcome $\hat{m}(x) = \widehat{\mathbb{E}}[Y_i \mid X_i = x]$ and the propensity score $\hat{e}(x) = \widehat{\mathbb{E}}[W_i \mid X_i = x]$ using separate regression forests, and then work with the residuals $\tilde{Y}_i = Y_i - \hat{m}(X_i)$ and $\tilde{W}_i = W_i - \hat{e}(X_i)$. Intuitively, this lets the forest concentrate on explaining differences in how patients *respond* to treatment rather than differences in their baseline likelihood of vaccinating. The treatment effect is then identified by the residualized partially linear moment $\psi_\tau(O_i) = \tilde{W}_i\big(\tilde{Y}_i - \tau\,\tilde{W}_i\big)$, whose solution is a regression of $\tilde{Y}$ on $\tilde{W}$. Because treatment is randomly assigned in our trial, the propensity score is constant by design and the treatment-side residualization is nearly inert; the outcome centering, however, still removes baseline variation in vaccination propensity and improves precision.

Step 2: Growing each tree (the splitting rule). Each tree sorts patients into groups ("leaves") so that patients within a leaf have similar treatment effects while patients in different leaves differ. The ideal way to split a parent node P into two children C_1 and C_2 is the split that most separates their estimated effects,

$$\Delta(C_1, C_2) = \frac{n_{C_1}\, n_{C_2}}{n_P^2}\big(\hat{\tau}_{C_1} - \hat{\tau}_{C_2}\big)^2, \tag{S1}$$

where n_{C_j} is the number of patients in child j. Searching for this split directly would

require re-estimating the treatment effect for every candidate split of every covariate, which is computationally prohibitive. GRF instead uses a shortcut that closely approximates Eq. S1 in two steps. In the *labeling step*, the treatment effect is estimated *once* in the parent node, and each patient is assigned a pseudo-outcome that summarizes how much they push that estimate up or down,

$$\rho_i = -\xi^\top A_P^{-1}\, \psi_{\hat{\tau}_P, \hat{\nu}_P}(O_i), \qquad A_P = \frac{1}{n_P} \sum_{i \in P} \nabla \psi_{\hat{\tau}_P, \hat{\nu}_P}(O_i), \tag{S2}$$

where ξ selects the treatment-effect coordinate. In the *regression step*, the algorithm then makes an ordinary regression-tree split on these pseudo-outcomes, choosing the split that maximizes

$$\widetilde{\Delta}(C_1, C_2) = \sum_{j=1}^{2} \frac{1}{|\{i : X_i \in C_j\}|} \left(\sum_{i : X_i \in C_j} \rho_i \right)^2. \tag{S3}$$

Because the pseudo-outcomes are computed only once per node, finding the best split reduces to the fast cumulative-sum calculation used in any regression tree, recovering the heterogeneity criterion in Eq. S1 at a fraction of the cost. At each node the search is over a random subset of covariates (the number drawn from a Poisson distribution with mean `mtry`), and a split is allowed only if each child retains at least 5 treated and 5 control patients, so that a treatment effect remains estimable in every leaf.

Step 3: Honesty and subsampling. Each tree is grown on a random subsample of half of the data, drawn without replacement. Crucially, that subsample is itself split in half: the first half is used only to choose the splits (Eqs. S2–S3) and the disjoint second half only to estimate effects within the resulting leaves. This "honesty" matters because using the same patients both to decide where to split and to estimate effects inside each leaf would manufacture spurious heterogeneity—the tree would carve out groups that look extreme by chance and then report those same extremes as findings. Separating the two tasks removes this bias and underlies the validity of the forest's confidence intervals (Wager and Athey, 2018; Athey et al., 2019). We grow $B = 100{,}000$ trees in every specification, each on a subsample fraction of 0.5, with a minimum leaf size of 5.

Step 4: From the forest to a prediction. A single honest tree places the target patient x in one leaf and offers only a coarse, noisy effect from the few neighbors there. Averaging over the whole forest smooths this into a continuous measure of how similar each training patient i is to x, namely the share of trees in which the two fall in the same leaf,

$$\alpha_i(x) = \frac{1}{B} \sum_{b=1}^{B} \frac{\mathbf{1}\{X_i \in L_b(x)\}}{|L_b(x)|}, \tag{S4}$$

where $L_b(x)$ is the leaf containing x in tree b. The forest is therefore not used to average a separate effect from each tree; instead, the full ensemble defines a single set of weights $\alpha_i(x)$, and the predicted effect is estimated once by solving the weighted moment condition

$$\big(\hat{\tau}(x), \hat{\nu}(x)\big) \in \arg\min_{\tau, \nu} \left\| \sum_{i=1}^{n} \alpha_i(x)\, \psi_{\tau, \nu}(O_i) \right\|_2, \tag{S5}$$

where ν is a nuisance parameter. For our single binary treatment, this has the closed form of a locally weighted comparison of residualized outcomes,

$$\hat{\tau}(x) = \frac{\sum_i \alpha_i(x)\, \tilde{W}_i\, \tilde{Y}_i}{\sum_i \alpha_i(x)\, \tilde{W}_i^2}. \tag{S6}$$

In words, a patient's predicted treatment effect is a weighted comparison of treated and control outcomes among the patients the forest judges most similar to them, with the weights learned to track treatment-effect heterogeneity rather than mere covariate proximity. This adaptive weighting is what allows GRF to deliver valid pointwise inference on $\hat{\tau}(x)$ (Athey et al., 2019).

C.2 Causal Forest Evaluation

We assess the causal forest predictions in three complementary ways, following the approach of Davis and Heller (2017, 2020).

1. **Predicted versus actual treatment effects.** We sort individuals into twenty equally sized bins (vigintiles) by their out-of-bag predicted treatment effect, compute the actual treatment effect within each bin, and plot the two against one another (Figure 2 in the main text). If the predictions were perfect, the points would lie on the 45-degree line; if the predictions were pure noise, the relationship would be flat. The horizontal spread of the binned predictions therefore measures how much heterogeneity the forest detects, and the slope of the fitted line measures how well that predicted heterogeneity tracks the actual treatment effects.

2. **Calibration test.** We use the `test_calibration` routine in the `grf` package, which implements the best-linear-predictor calibration test of Chernozhukov et al. (2023) (*SI Appendix*, Table S5). It regresses the residualized outcome on two synthetic predictors built from the forest: a "mean forest prediction" term and a "differential forest prediction" term. A coefficient of one on the mean forest prediction indicates that the average prediction is correctly calibrated, while a positive and significant coefficient on the differential forest prediction indicates that the forest has captured genuine heterogeneity in treatment effects. This differential coefficient is the forest-based analog of the interaction test of Davis and Heller (2020), for which a significant estimate implies jointly that treatment effects are heterogeneous, that the covariates are associated with that heterogeneity, and that the forest predictions capture it.

3. **Quartile heterogeneity.** Finally, we test whether individuals with higher predicted treatment effects exhibit larger actual treatment effects. For patient i and for each treatment, we estimate

$$Y_i = \alpha + \sum_{q=1}^{3} \beta_q\, \mathbf{1}(Q_i = q) + \sum_{q=1}^{4} \gamma_q\, \mathbf{1}(\text{Treated})_i \times \mathbf{1}(Q_i = q) + \varepsilon_i, \tag{S7}$$

where $Q_i \in \{1, 2, 3, 4\}$ denotes the quartile of the causal-forest predicted treatment effect for individual i, with $Q_i = 1$ the top quartile and $Q_i = 4$ the bottom. The bottom quartile is the omitted category among the quartile level indicators, and the γ_q coefficients give the treatment effect within each quartile. The model is estimated by ordinary least squares with heteroskedasticity-robust standard errors. A rejection of equality across the γ_q provides a transparent, regression-based confirmation that the heterogeneity detected by the forest translates into different actual treatment effects across the predicted-effect distribution.

D Supplemental Exhibits

Table S1: Summary Statistics

	(1) Control	(2) Reminder	(3) Cash	(4) Scratch-Off
Panel A. Patient-Level Non-Health Characteristics				
Age	51.08	51.07	50.9	51.09
	(16.73)	(16.8)	(16.76)	(16.86)
Female	0.58	0.58	0.58	0.58
White	0.94	0.94	0.94	0.94
Black	0.04	0.04	0.04	0.04
Asian	0.01	0.01	0.01	0.01
Hispanic	0.05	0.05	0.05	0.04
Married	0.51	0.51	0.51	0.51
Medicare	0.16	0.16	0.16	0.16
Medicaid	0.03	0.03	0.03	0.03
Panel B. Patient-Level Health Characteristics				
BMI	32.39	33.54	32.67	32.44
	(52.93)	(101.69)	(58.95)	(64.91)
Medial Risk Score	0.05	0.05	0.05	0.05
	(0.05)	(0.05)	(0.05)	(0.05)
Encounters in Past Year	54.51	54.63	54.95	55.15
	(54.7)	(54.15)	(54.74)	(55.94)
Years in Health System	18.36	18.27	18.3	18.32
	(8.78)	(8.91)	(8.90)	(8.77)
Current Smoker	0.2	0.2	0.2	0.21
Former Smoker	0.29	0.29	0.29	0.28
Never Smoker	0.51	0.51	0.51	0.51
Myocardial Infarction	0.03	0.03	0.03	0.03
Congestive Heart Failure	0.04	0.04	0.04	0.04
Peripheral Vascular Disease	0.06	0.06	0.06	0.06
Cerebrovascular Disease	0.05	0.05	0.05	0.05
Dementia	0.01	0.01	0.01	0.01
Chronic Pulmonary Disease	0.19	0.19	0.19	0.19
Connective Tissue Disease	0.02	0.02	0.02	0.02
Peptic Ulcer Disease	0.01	0.01	0.01	0.01
Mild Liver Disease	0.07	0.07	0.07	0.07
Moderate/Severe Liver Disease	0	0	0	0
Diabetes without Complications	0.18	0.19	0.18	0.18
Diabetes with Complications	0.08	0.08	0.08	0.08
Paraplegia, Hemiplegia	0.01	0.01	0.01	0.01
Renal Disease	0.1	0.1	0.1	0.1
Cancer	0.06	0.06	0.06	0.06
Metastatic Carcinoma	0	0	0	0.01
Panel C. Patient-Level Prior Flu Shot Behavior				
Received Flu Shot in 2020–2021	0.49	0.49	0.48	0.49
Received Flu Shot Before Appt.	0.08	0.09	0.08	0.08
Panel D. Patient-Level Appointment Type				

Continued on next page

Table S1 – continued from previous page

	(1) Control	(2) Reminder	(3) Cash	(4) Scratch-Off
Dermatology	0	0	0	0
Family Medicine	0.93	0.93	0.93	0.93
Gynecology	0	0	0	0
Internal Medicine	0.06	0.06	0.07	0.07
Orthopedic	0	0	0	0
Pediatric	0.01	0.01	0	0.01
Rheumatology	0	0	0	0
Urology	0	0	0	0
Panel E. County-Level Characteristics (2019)				
Population (thousands)	150.99	151.18	150.86	150.55
	(108.05)	(108.74)	(107.36)	(106.95)
Lottery Ticket Sales (per person)	397.99	396.84	396.32	396.95
	(97.86)	(98.8)	(98.67)	(98.22)
Lottery Prizes Paid (per person)	281.78	281.16	280.65	281.42
	(67.8)	(68.38)	(68.48)	(67.83)
Lottery Retailer Commissions (per person)	23.04	22.97	22.93	22.97
	(5.82)	(5.87)	(5.86)	(5.83)
Poverty Rate	13.55	13.56	13.58	13.56
	(1.82)	(1.81)	(1.8)	(1.79)
Median Household Income	54953.95	54957.54	54926.86	54940.37
	(4764.48)	(4872.39)	(4800.74)	(4681.45)
Unemployment Rate	5.13	5.12	5.12	5.13
	(0.75)	(0.75)	(0.75)	(0.75)
Share with Less Than High School	10.02	10.02	10.02	10.03
	(2.3)	(2.33)	(2.3)	(2.31)
Share with High School Only	41.33	41.29	41.26	41.31
	(6.21)	(6.29)	(6.27)	(6.23)
Share with Some College	24.64	24.63	24.63	24.65
	(3.12)	(3.13)	(3.14)	(3.14)
Share College Graduates	24.01	24.06	24.09	24.01
	(7.94)	(8.07)	(8.04)	(7.94)
Panel F. ZIP-Code-Level Characteristics (2020)				
Median Household Income	56344.47	56553.91	56520.51	56609.83
	(12853.52)	(12891.91)	(12925.72)	(13030.94)
Share with Less Than High School	0.1	0.1	0.1	0.1
	(0.04)	(0.04)	(0.05)	(0.04)
Share with High School Only	0.42	0.41	0.41	0.41
	(0.09)	(0.09)	(0.09)	(0.09)
Share with Some College	0.15	0.15	0.15	0.16
	(0.03)	(0.04)	(0.04)	(0.04)
Share College Graduates	0.33	0.33	0.33	0.33
	(0.11)	(0.12)	(0.12)	(0.12)
N	14398	14393	14394	14394

Notes: Table shows means (standard deviations) of baseline patient-level, county-level, and ZIP-code-level characteristics for patients across control and treatment arms. Column (1) shows the control arm. Column (2) shows the Reminder arm. Column (3) shows the Cash arm. Column (4) shows the Scratch-Off arm.

D.1 Pre-registered Robustness Checks

Table S2: Average Treatment Effects Across Time Horizons and Appointment Completion

	(1) Vaccinated at Appointment	(2) Vaccinated in 2021-2022	(3) Vaccinated in 3 Days Before Appointment	(4) Vaccinated After Appointment	(5) Completed the Appointment
Panel A. Treatment groups separate					
Reminder	0.0242***	0.0167***	0.0008	-0.0084**	0.0093*
	(0.0052)	(0.0059)	(0.0010)	(0.0039)	(0.0054)
Cash	0.0161***	0.0115*	0.0005	-0.0039	0.0131**
	(0.0052)	(0.0059)	(0.0010)	(0.0039)	(0.0054)
Scratch-Off	0.0196***	0.0154***	-0.0003	-0.0027	0.0179***
	(0.0052)	(0.0059)	(0.0010)	(0.0039)	(0.0054)
Panel B. Treatment groups together					
Any Treatment	0.0200***	0.0145***	0.0003	-0.0050	0.0135***
	(0.0042)	(0.0048)	(0.0008)	(0.0032)	(0.0044)
Control Group Mean	0.2536	0.4756	0.0072	0.1293	0.6907
Observations	57579	57579	57579	57579	57579

Notes: Each column in each panel is a separate regression of the effect of treatment on the given outcome. Panel A shows the treatment effects separately by treatment arm, while Panel B pools the treatment arms together and shows the average treatment effect. Standard errors are heteroskedasticity-robust and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table S3: Average Treatment Effects – Unadjusted SEs

	(1) Vaccinated at Appointment	(2) Vaccinated in 2021-2022	(3) Vaccinated in 3 Days Before Appointment	(4) Vaccinated After Appointment	(5) Completed the Appointment
Panel A. Treatment groups separate					
Reminder	0.0242***	0.0167***	0.0008	-0.0084**	0.0093*
	(0.0052)	(0.0059)	(0.0010)	(0.0039)	(0.0054)
Cash	0.0161***	0.0115*	0.0005	-0.0039	0.0131**
	(0.0052)	(0.0059)	(0.0010)	(0.0039)	(0.0054)
Scratch-Off	0.0196***	0.0154***	-0.0003	-0.0027	0.0179***
	(0.0052)	(0.0059)	(0.0010)	(0.0039)	(0.0054)
Panel B. Treatment groups together					
Any Treatment	0.0200***	0.0145***	0.0003	-0.0050	0.0135***
	(0.0043)	(0.0048)	(0.0008)	(0.0032)	(0.0044)
Control Group Mean	0.2536	0.4756	0.0072	0.1293	0.6907
Observations	57579	57579	57579	57579	57579

Notes: As noted in the pre-registration, this table provides a robustness check on the calculation of the standard errors under homoskedasticity. Each column in each panel is a separate regression of the effect of treatment on the given outcome. Panel A shows the treatment effects separately by treatment arm, while Panel B pools the treatment arms together and shows the average treatment effect. Standard errors are displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table S4: Average Treatment Effects – Robustness

	Vaccinated at Appointment				
	(1)	(2)	(3)	(4)	(5)
Panel A. Treatment groups separate					
Reminder	0.0242***	0.0267***	0.1241***	0.0445***	0.0245***
	(0.0052)	(0.0057)	(0.0267)	(0.0059)	(0.0051)
Cash	0.0161***	0.0177***	0.0834***	0.0294***	0.0168***
	(0.0052)	(0.0057)	(0.0268)	(0.0058)	(0.0051)
Scratch-Off	0.0196***	0.0201***	0.1010***	0.0368***	0.0195***
	(0.0052)	(0.0057)	(0.0268)	(0.0058)	(0.0051)
Panel B. Treatment groups together					
Any Treatment	0.0200***	0.0215***	0.1029***	0.0368***	0.0203***
	(0.0042)	(0.0046)	(0.0220)	(0.0046)	(0.0041)
Drop Cohabitants		X			
Logit			X		
Keep Protocol Adherents				X	
Control for Age, Sex					X
Control Group Mean	0.2536	0.2504	0.2536	0.2713	0.2536
Observations	57579	47307	57579	46273	57579

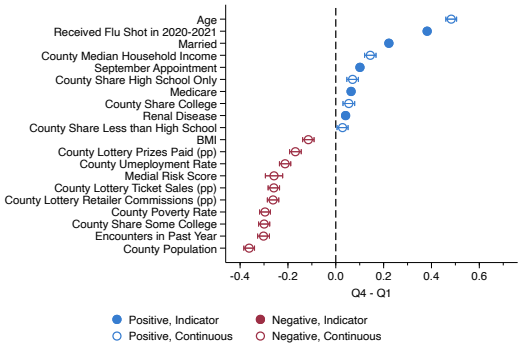
Notes: As noted in the pre-registration, this table provides a variety of robustness checks on estimation decisions. Each column in each panel is a separate regression of the effect of treatment on whether a participant received an influenza vaccination at the appointment. Panel A shows the treatment effects separately by treatment arm, while Panel B pools the treatment arms together and shows the average treatment effect. Column 1 displays the baseline specification, as in Table S2. Column 2 removes participants who reside at the same address as other participants. Column 3 runs logistic regressions instead of OLS. Column 4 restricts the sample to participants who were sent messages according to protocol. Column 5 adds controls for binned age, sex, and their interactions. Standard errors are heteroskedasticity-robust and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table S5: Assessing Causal Forest Predictions using "Test_Calibration" in the Generalized Random Forest Package

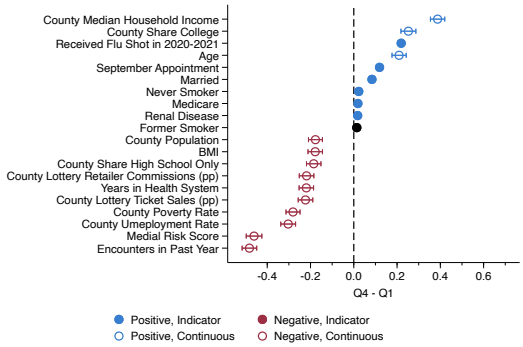
	Flu Shot at Appointment			
	(1)	(2)	(3)	(4)
	Any Treatment	Reminder	Cash	Scratch-Off
Mean Forest Prediction	0.9979***	1.0010***	1.0043***	0.9993***
	(0.1939)	(0.1971)	(0.2922)	(0.2470)
Differential Forest Prediction	0.6138***	0.1432	0.3640	0.5468**
	(0.2617)	(0.2958)	(0.2872)	(0.2841)

Notes: Table shows a test of the goodness of fit of the causal forest predictions using the "test_calibration" command within the grf R package, where the causal forest is run for all three treatment arms jointly (Column 1), the Reminder treatment arm only (Column 2), the Cash treatment arm only (Column 3), and the Scratch-Off treatment arm only (Column 4). All causal forests are run using basic controls and 100,000 trees. A coefficient of 1 for "Mean Forest Prediction" suggests the mean causal forest prediction is correctly calibrated, while a coefficient of 1 for "Differential Forest Prediction" suggests the forest's estimates of heterogeneity are well-calibrated; a positive and statistically significant "Differential Forest Prediction" coefficient indicates that the forest has captured genuine heterogeneity in treatment effects. Standard errors are heteroskedasticity-robust (HC3) and displayed in parentheses, and the reported p-values are one-sided. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

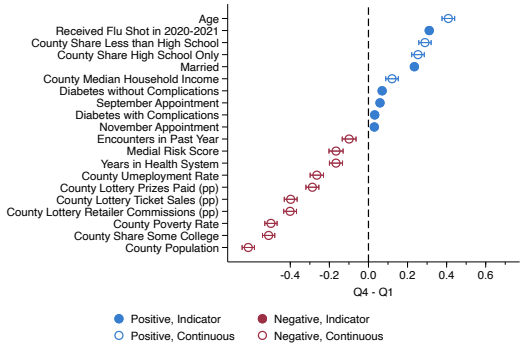
Figure S1: Causal Forest Heterogeneity for Any Treatment, Reminder, and Cash – Patient Characteristics for Top vs. Bottom Quartile of Predicted Effects



(a) Any Treatment



(b) Reminder



(c) Cash

Notes: Figure shows the difference between the mean of each characteristic for those in the top quartile versus the bottom quartile of predicted treatment effects stemming from the causal forest estimation. Panel A uses predicted treatment effects from any treatment, Panel B uses predicted treatment effects from the reminder treatment, and Panel C uses predicted treatment effects from the cash treatment. For continuous variables (empty circles), differences are divided by the standard deviations of the control group to ease comparisons with binary characteristics. Whiskers show 95% confidence intervals. The figure is limited to characteristics with the 10 largest positive and 10 largest negative difference magnitudes. Causal forest models are run using basic controls over 100,000 trees.

Table S6: OLS Heterogeneity for Scratch-Off Treatment – Top and Bottom 10 Characteristics Flagged by the Causal Forest

	Outcome = Shot at Appointment		
	(1) Interaction Coefficient	(2) Interaction SE	(3) Interaction Significance
Age	0.0100	0.0051	*
Received Flu Shot in 2020-2021	0.0274	0.0098	***
Married	0.0231	0.0104	**
County Share High School Only	0.0068	0.0053	
Years in Health System	0.0010	0.0052	
Medicare	0.0012	0.0125	
Renal Disease	0.0083	0.0185	
September Appointment	0.0061	0.0119	
Current Smoker	0.0166	0.0118	
Congestive Heart Failure	0.0374	0.0279	
County Poverty Rate	-0.0025	0.0053	
Never Smoker	-0.0105	0.0104	
Female	-0.0176	0.0106	*
County Lottery Ticket Sales (pp)	0.0017	0.0053	
County Lottery Retailer Commissions (pp)	0.0009	0.0053	
Private Insurance	0.0039	0.0120	
County Median Household Income	-0.0062	0.0053	
County Share Some College	0.0011	0.0052	
County Population	-0.0078	0.0052	
Encounters in Past Year	-0.0042	0.0054	

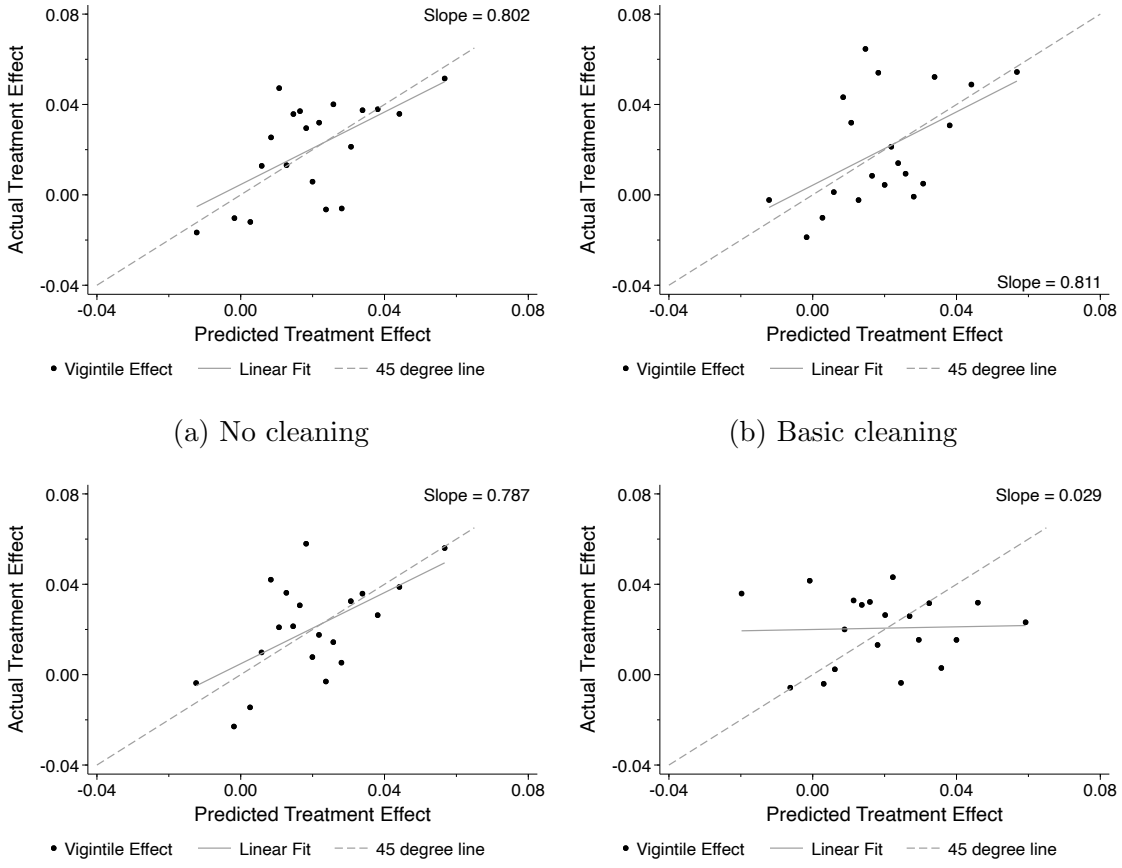
Notes: Each row is a separate regression of the effect of the scratch-off lottery treatment on whether the patient received an influenza vaccination at the appointment, where the treatment is interacted with the variable given in the row. Column 1 shows the estimated coefficient on the interaction term. Column 2 shows the estimated heteroskedasticity-robust standard error. Column 3 shows significance levels. Continuous variables are standardized to have a mean of 0 and a standard deviation of 1.

Table S7: Pre-Registered OLS Heterogeneity for Any Treatment

	Outcome = Shot at Appointment		
	(1) Interaction Coefficient	(2) Interaction SE	(3) Interaction Significance
Age: 25-34	0.0289	0.0171	*
Age: 35-44	0.0448	0.0166	***
Age: 45-54	0.0237	0.0166	
Age: 55-64	0.0461	0.0166	***
Age: 65+	0.0447	0.0169	***
Insurer: Medicaid	-0.0092	0.0216	
Insurer: Medicare	0.0005	0.0102	
Zipcode Income: 60-100% of median	-0.0165	0.0530	
Zipcode Income: 100-140% of median	-0.0034	0.0531	
Zipcode Income: >140% of median	-0.0235	0.0547	
Female	-0.0101	0.0086	
Primary Care Appointment	0.0124	0.0088	
Received Flu Shot in 2020-2021	0.0240	0.0080	***
Previously Received Covid Shot	0.0119	0.0078	
Previously Received Covid Diagnosis	-0.0067	0.0171	
County Lottery Ticket Sales (pp)	-0.0002	0.0043	

Notes: Each row is a separate regression of the effect of any treatment on whether the patient received an influenza vaccination at the appointment, where the treatment is interacted with the variable given in the row. Age, insurer, and ZIP-code-level income are binned categorical variables, so each of those rows comes from the same regression. Column 1 shows the estimated coefficient on the interaction term. Column 2 shows the estimated heteroskedasticity-robust standard error. Column 3 shows significance levels. Continuous variables are standardized to have a mean of 0 and a standard deviation of 1.

Figure S2: Causal Forest Heterogeneity – Add Noise, Part 1



(a) No cleaning

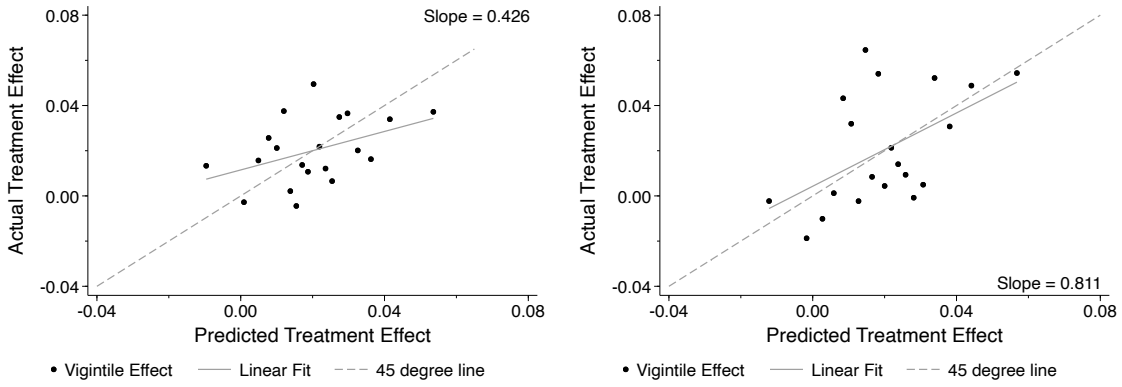
(b) Basic cleaning

(c) Winsorize

(d) Quartile bins

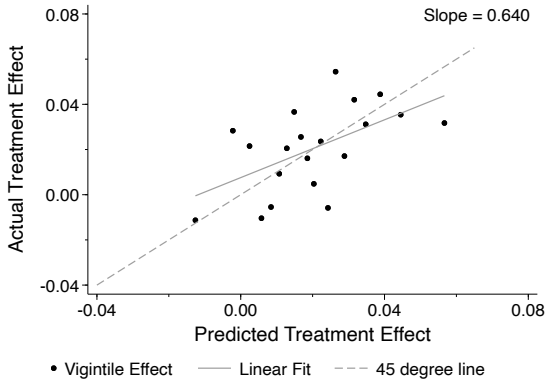
Notes: Figure plots the causal forest predicted treatment effects (x-axis) and the actual treatment effects (y-axis) for individuals binned into twenty vigintiles of the causal forest predicted treatment effects. All causal forests are run using basic controls, 100,000 trees, and all three treatment arms together. The outcome is whether the patient received an influenza vaccination at the appointment. Solid line shows the linear relationship between the actual and predicted effects. Regression-estimated slope of the solid line is included in the right-hand corner. Dashed line is the 45-degree line, included for reference. Panel A does not clean any of the individual-level continuous variables. Panel B replaces extreme BMI values with missing values. Panel C winsorizes the top 2.5% and bottom 2.5% of each variable. Panel D bins each variable into quartiles.

Figure S3: Causal Forest Heterogeneity – Add Noise, Part 2

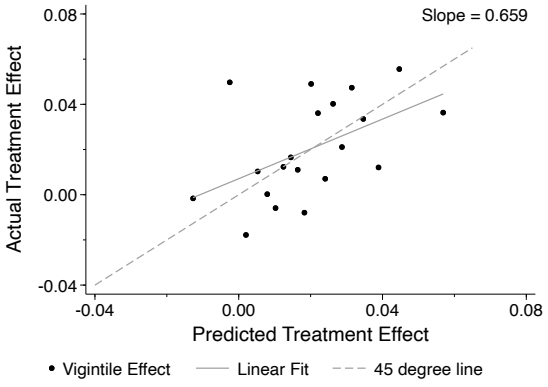


(a) ZIP-code-level

(b) County-level



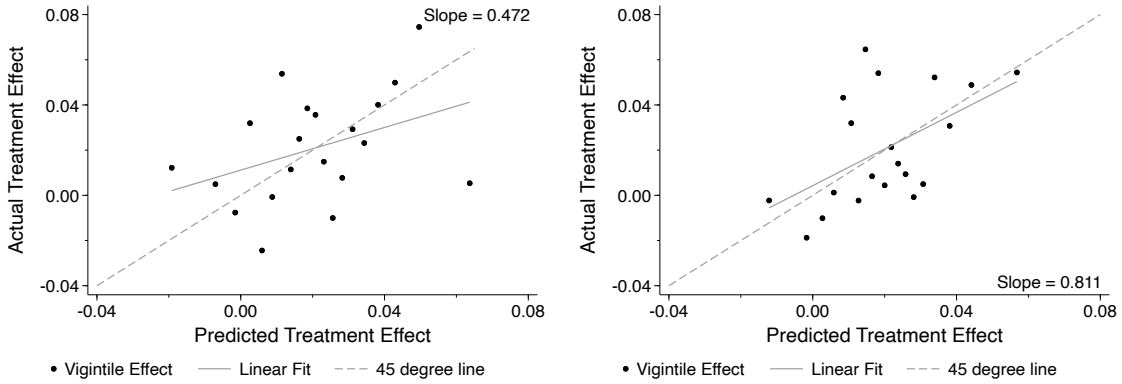
(c) CZ-level



(d) HRR-level

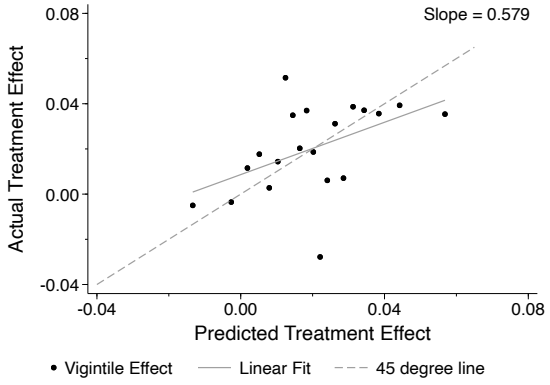
Notes: Figure plots the causal forest predicted treatment effects (x-axis) and the actual treatment effects (y-axis) for individuals binned into twenty vigintiles of the causal forest predicted treatment effects. All causal forests are run using basic controls, 100,000 trees, and all three treatment arms together. The outcome is whether the patient received an influenza vaccination at the appointment. Solid line shows the linear relationship between the actual and predicted effects. Regression-estimated slope of the solid line is included in the right-hand corner. Dashed line is the 45-degree line, included for reference. Panel A includes ZIP-code-level variables where possible and county-level variables otherwise. Panel B replaces all ZIP-code-level variables with county-level variables. Panel C replaces all county-level variables with commuting-zone-level variables. Panel D replaces all commuting-zone-level variables with health-referral-region-level variables.

Figure S4: Causal Forest Heterogeneity – Add Information

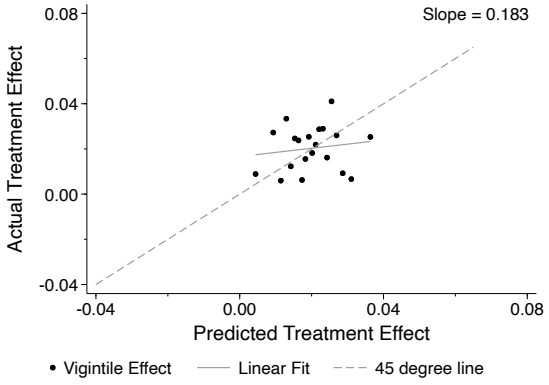


(a) Individual-level covariates only

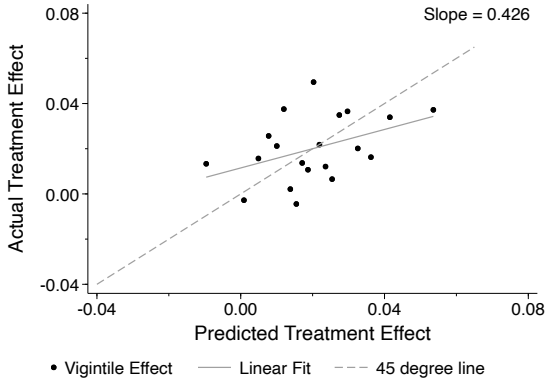
(b) Add county-level covariates



(c) Add county-level fixed effects



(d) Now ZIP-code-level fixed effects



(e) Now ZIP-code-level covariates

Notes: Figure plots the causal forest predicted treatment effects (x-axis) and the actual treatment effects (y-axis) for individuals binned into twenty vigintiles of the causal forest predicted treatment effects. All causal forests are run using basic controls, 100,000 trees, and all three treatment arms together. The outcome is whether the patient received an influenza vaccination at the appointment. Solid line shows the linear relationship between the actual and predicted effects. Regression-estimated slope of the solid line is included in the right-hand corner. Dashed line is the 45-degree line, included for reference. Panel A includes only individual-level covariates. Panel B adds county-level covariates. Panel C adds county-level fixed effects. Panel D replaces county-level fixed effects with ZIP-code-level fixed effects. Panel E is equivalent to Panel B but replaces county-level covariates with ZIP-code-level covariates where possible.

D.2 Pre-registered Secondary Outcomes

Table S8: Pre-Analysis Plan Secondary Outcomes

	(1) Flu Shot w/i 7 Days	(2) Flu Diagnosis (21-22)	(3) Flu Complications (21-22)	(4) Share HH Flu Shot w/i 28 Days	(5) Flu Shot (22-23)	(6) Covid Shot (21-22)	(7) Covid Diagnosis (21-22)
Panel A. Treatment groups separate							
Reminder	0.0230***	-0.0026**	0.0005	0.0026	0.0018	0.0006	-0.0017
	(0.0054)	(0.0010)	(0.0039)	(0.0052)	(0.0057)	(0.0054)	(0.0033)
Cash	0.0163***	-0.0001	-0.0008	-0.0038	-0.0011	0.0027	-0.0028
	(0.0053)	(0.0011)	(0.0039)	(0.0051)	(0.0057)	(0.0054)	(0.0033)
Scratch-Off	0.0152***	-0.0011	0.0045	-0.0015	0.0000	0.0011	-0.0048
	(0.0053)	(0.0011)	(0.0039)	(0.0052)	(0.0057)	(0.0054)	(0.0032)
Panel B. Treatment groups together							
Any Treatment	0.0182***	-0.0013	0.0014	-0.0009	0.0003	0.0014	-0.0031
	(0.0044)	(0.0009)	(0.0032)	(0.0042)	(0.0047)	(0.0044)	(0.0027)
Drop Cohabitants				X			
Control Group Mean	0.2819	0.0092	0.1245	0.1024	0.3748	0.2946	0.0843
Observations	57579	57579	57579	24491	57579	57579	57579

Notes: Each column in each panel is a separate regression of the effect of treatment on the given outcome. Panel A shows the treatment effects separately by treatment arm, while Panel B pools the treatment arms together and shows the average treatment effect. Column 4 removes participants who reside at the same address as other participants. Standard errors are heteroskedasticity-robust and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.3 Additional Estimates Not Referenced

Table S9: Evidence of Heterogeneous Treatment Effects – Top Quartile vs. Rest

	Shot at Appointment			
	(1) Any Treatment	(2) Reminder	(3) Cash	(4) Scratch-Off
Top Quartile of Predicted Treatment Effects	0.0383***	0.0225**	0.0375***	0.0376***
	(0.0092)	(0.0111)	(0.0112)	(0.0113)
Rest of Sample	0.0143***	0.0249***	0.0099*	0.0138**
	(0.0046)	(0.0058)	(0.0058)	(0.0058)
$H0$: subgroups equal, $p =$	0.020	0.852	0.029	0.060
N	57579	28791	28792	28792

Notes: Table shows an assessment of the causal forest predictions, where the causal forest is run for all three treatment arms jointly (Column 1), the Reminder treatment arm only (Column 2), the Cash treatment arm only (Column 3), and the Scratch-Off treatment arm only (Column 4). All causal forests are run using basic controls and 100,000 trees. Table shows treatment effect estimates for the patients in the top quartile of predicted treatment effects and the rest of the sample, following a modified version of Eq. S7. P-values from tests of whether these two groups have the same treatment effects are shown. Standard errors are heteroskedasticity-robust and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table S10: Means by Quartile of Predicted Treatment Effect for Scratch-Off

	(1) Q1 (Bottom)	(2) Q2	(3) Q3	(4) Q4 (Top)	(5) All
Age	46.65	49.00	52.18	56.49	51.08
Received Flu Shot in 2020-2021	0.39	0.38	0.46	0.71	0.49
Married	0.39	0.45	0.53	0.66	0.51
County Share High School Only	40.65	41.55	41.58	41.49	41.32
Years in Health System	18.28	17.95	17.90	19.22	18.34
Medicare	0.22	0.24	0.27	0.32	0.26
Renal Disease	0.08	0.09	0.10	0.14	0.10
September Appointment	0.25	0.28	0.29	0.30	0.28
Current Smoker	0.16	0.20	0.24	0.22	0.21
Congestive Heart Failure	0.03	0.04	0.04	0.05	0.04
County Poverty Rate	13.67	13.48	13.51	13.56	13.56
Never Smoker	0.57	0.50	0.46	0.50	0.51
Female	0.62	0.59	0.57	0.54	0.58
County Lottery Ticket Sales (pp)	397.32	401.08	401.91	389.59	397.47
County Lottery Retailer Commissions (pp)	23.02	23.23	23.27	22.50	23.00
Private Insurance	0.75	0.73	0.70	0.66	0.71
County Median Household Income	55303.07	54949.38	54790.41	54745.80	54947.16
County Share Some College	24.76	24.78	24.76	24.27	24.64
County Population	163751.62	150397.58	148741.92	140205.73	150770.19
Encounters in Past Year	66.05	54.16	50.70	48.51	54.83
Observations	7198	7198	7198	7198	28792

Notes: Columns report the mean of characteristics for individuals in the bottom quartile (Column 1), second quartile (Column 2), third quartile (Column 3), and top quartile (Column 4) of predicted scratch-off lottery treatment effects, as well as the mean across all individuals who were randomized into either the pure control group or the scratch-off treatment group.

References

Athey, S., J. Tibshirani, and S. Wager (2019). Generalized random forests. *The Annals of Statistics 47*(2), 1148–1178.

Athey, S. and S. Wager (2019). Estimating treatment effects with causal forests: An application. *Observational Studies 5*(2), 37–51.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernández-Val (2023). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india. Working Paper 24678, National Bureau of Economic Research.

Coppock, A., J. Cooper, and N. Fultz (2019). randomizr: Easy-to-Use Tools for Common Forms of Random Assignment and Sampling. *R package version 0.20.0*.

Davis, J. M. V. and S. B. Heller (2017). Using causal forests to predict treatment heterogeneity: An application to summer jobs. *American Economic Review 107*(5), 546–550.

Davis, J. M. V. and S. B. Heller (2020). Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs. *The Review of Economics and Statistics 102*(4), 664–677.

Wager, S. and S. Athey (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association 113*(523), 1228–1242.