Estimating population average treatment effects from experiments with noncompliance*

Kellie Ottoboni† Jason Poulos‡

December 8, 2018

Abstract

This paper improves on the transportability of clinical trial results to a population by extending a method of estimating population average treatment effects in settings with noncompliance. We identify the complier-average causal effect for a target population with few additional assumptions. Simulations show the compliance-adjusted estimator performs better than the unadjusted estimator when compliance is relatively low and can be predicted by observed covariates. We apply the proposed estimator to measure the effect of Medicaid coverage on health care use for a target population of adults who may benefit from expansions to the Medicaid program, using data from a large-scale health insurance experiment in which a small subset of those randomly selected to receive Medicaid benefits actually enrolled.

*The authors thank Jon McAuliffe and Jas Sekhon and his research group at UC Berkeley for valuable comments. Poulos acknowledges support from the National Science Foundation Graduate Research Fellowship [grant number DGE 1106400]. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) supercomputer Stampede2 at the Texas Advanced Computing Center (TACC) through allocation SES180010.

†Department of Statistics, University of California, Berkeley. email: kellieotto@berkeley.edu.

‡*Corresponding author:* Department of Political Science, University of California, Berkeley. email: poulos@berkeley.edu.

1 Introduction

Randomized control trials (RCTs) are the gold standard for estimating the causal effect of a treatment. An RCT may give an unbiased estimate of the Sample Average Treatment Effect (SATE), but external validity is an issue when the individuals in the RCT are unrepresentative of the actual population of interest. For example, the participants in an RCT in which individuals volunteer to sign up for health insurance may be in poorer health at baseline than the overall population. External validity is particularly relevant to policymakers who want to know how the treatment effect would generalize to the broader population.

This paper extends the literature on extrapolating clinical trial results to populations to deal with noncompliance. Previous approaches to the problem of extrapolating RCT results to a population (Imai, King, and Stuart 2008; Stuart et al. 2011; Hartman et al. 2015) are designed for settings where there is full compliance with treatment. This paper contributes to the literature by defining the assumptions required to identify complier–average causal effects for the target population and proposing a procedure to recover this estimand.

Hartman et al. (2015) propose a method of reweighting the responses of individuals in an RCT according to the distribution of covariates in the target population in order to estimate the population average treatment effect on the treated (PATT). Under a series of assumptions, the PATT is identified from the RCT outcomes. We extend the method to estimate the complier–average causal effects for the target population from RCT data with noncompliance. Noncompliance, which occurs when individuals who are assigned to the treatment group do not comply with the treatment, is a prevalent issue in RCTs. It dilutes the estimated effect of treatment assignment, and the resulting intention–to–treat (ITT) estimate is biased towards zero.

The proposed estimator involves estimating the expectation of the response of compliers in the RCT sample, conditional on their covariates, where the expectation is taken over the distribution of population covariates. Note that our estimation strategy differs from reweighting methods that use propensity scores to adjust the RCT data (Stuart et al. 2011).

In this context, the propensity score model predicts participation in the RCT, given pretreatment covariates common to both the RCT and population data. Individuals in the RCT and population are then weighted according to the inverse of the estimated propensity score. We use an ensemble of algorithms to predict the response surface for RCT compliers and use the predicted values from the response surface model to estimate the potential outcomes of population members who received treatment, given their covariates.

When estimating the average causal effect from an RCT, researchers typically divide the ITT estimate by the compliance rate under the identifying assumptions outlined in Angrist, Imbens, and Rubin (1996). When extrapolating RCT results to a population, one might simply weight the PATT estimate by the population compliance rate in order to yield a population average effect of treatment on treated compliers.[1] However, the compliance rate is likely to differ between the sample and population, as well as across subgroups. We propose an alternative approach of actually identifying the likely compliers in the control group. By explicitly modeling compliance, our approach allows researchers to decompose population estimates by covariate group and also predict which population members are likely to comply with treatment. Both of these features are useful for policymakers in evaluating the efficacy of policy interventions for subgroups of interest in a population.

We apply the proposed estimator to measure the effect of Medicaid coverage on health care use for a target population of adults who may benefit from government-backed expansions to the Medicaid program. We draw RCT data from a large-scale health insurance experiment, in which only 30% of those randomly selected to receive Medicaid benefits actually enrolled. We find substantial differences between sample and population estimates in terms of race, education, and health status subgroups.

The paper proceeds as follows: Section 2 presents the proposed estimator, necessary assumptions for its identifiability, and an estimation procedure; Section 4 reports the estimator's performance in simulations; Section 5 uses the estimator to identify the effect

1. A similar approach is used by Imai, Tingley, and Yamamoto 2013 for estimating average complier indirect effects.

of extending Medicaid coverage to the low–income adult population in the U.S; Section 6 discusses the results.

2 Estimator

We are interested in using the outcomes from an RCT to estimate the average treatment effect on the treated for a target population. Treatment in the population is not assigned at random, but rather may depend on other variables, confounding the effect of treatment on the outcome of interest. RCTs are needed to isolate the effect of treatment. However, strict exclusion criteria for RCTs often result in a sample whose distribution of covariates differs substantially from the target population.

Ideally, we would take the results of an RCT and reweight the sample such that the reweighted covariates match the those in the population. In practice, one rarely knows the true covariate distribution in the target population. Instead, we consider data from a nonrandomized, observational study in which participants are representative of the target population. Our proposed estimator combines RCT and observational data to overcome these issues.

2.1 Assumptions

Let Y_{ist} be the potential outcome for individual i in group s, where $s = 0$ is the population and $s = 1$ is the RCT, and let t be the treatment received. For simplicity, we assume there are only two possible treatments: $t = 1$ for the active treatment and $t = 0$ for the control or no treatment. Let S_i denote the sample assignment, T_i denote the treatment assigned, and D_i denote treatment received. Let W_i be individual i's observable pretreatment covariates that are related to the sample selection mechanism for membership in the RCT, treatment assignment in the population, and complier status. Let C_i be an indicator for individual i's compliance to treatment, which is only observable in the RCT. Treatment is assigned at

random by the investigator in the RCT, so we observe both D_i and T_i when $S_i = 1$. For compliers in the RCT, $D_i = T_i$.

In the population, we suppose that treatment is made available to individuals based on their covariates W_i. Individuals with $T_i = 0$ do not receive treatment, while those with $T_i = 1$ may decide whether or not to accept treatment. For individuals in the population, we only observe D_i — not T_i. We frame the Assumptions 3 and 4 defined below in terms of C_i and T_i in order to distinguish among the population controls noncompliers who should have received treatment (i.e., individuals with $T_i = 1$ and $D_i = 0$) from noncompliers assigned to control (i.e., individuals with $T_i = 1$ and $D_i = 0$). Differentiating population controls is important for deriving the estimator for τ_{PATT}.

Assumptions 1, 3, 4, and 5 are made by Hartman et al. (2015) to identify PATT from an RCT:

Assumption 1. *Consistency under parallel studies:*

$$Y_{i0t} = Y_{i1t}, \qquad \forall\, i\,,\, t = \{0, 1\}.$$

Assumption 1 requires that each individual i has the same response to treatments, whether i is in the RCT or not. Compliance status C_i is not a factor in this assumption because we assume that compliance is independent of sample and treatment assignment for all individuals with covariates W_i.

Assumption 2. *Conditional independence of compliance and assignment:*

$$C_i \perp\!\!\!\perp S_i, T_i \mid W_i, \qquad 0 < \mathbb{P}(C_i = 1 \mid W_i) < 1.$$

Together, Assumptions 1 and 2 ensure that potential outcomes do not differ based on sample assignment or receipt of treatment.

Assumption 3. *Strong ignorability of sample assignment for treated:*

$$(Y_{i01}, Y_{i11}) \perp\!\!\!\perp S_i \mid (W_i, T_i = 1, C_i = 1), \qquad 0 < \mathbb{P}(S_i = 1 \mid W_i, T_i = 1, C_i = 1) < 1.$$

Assumption 3 ensures the potential outcomes for treatment are independent of sample assignment for individuals with the same covariates W_i and assignment to treatment.[2] We make a similar assumption for the potential outcomes under control:

Assumption 4. *Strong ignorability of sample assignment for controls:*

$$(Y_{i00}, Y_{i10}) \perp\!\!\!\perp S_i \mid (W_i, T_i = 1, C_i = 1), \qquad 0 < \mathbb{P}(S_i = 1 \mid W_i, T_i = 1, C_i = 1) < 1.$$

RCT study designs that apply restrictive exclusion criteria may increase the likelihood that there are unobserved differences between the RCT and target population, which would violate the strong ignorability assumptions.[3]

Interference undermines the framework because it creates more than two potential outcomes per participant, depending on the treatment assignment of other participants (Rubin 1990). We therefore assume no interference between units:

Assumption 5. *The potential outcomes Y_{ist} do not depend on T_j, $\forall j \neq i$.*

Figure 1 shows Assumptions 3, 4, and 2 in a directed acyclic graph. Treatment assignment T_i may only depend on C_i through W_i, and the potential outcomes (Y_{is0}, Y_{is1}) may only depend on S_i through W_i. From the internal validity standpoint, the role of W_i is critical: if any relevant observed covariates are not controlled, then there is a backdoor pathway from T_i back to W_i and into Y_{ist}. We use the same W_i across all identifying assumptions, which implicitly assumes that the observable covariates that determine sample selection

2. Throughout, we assume individuals are sampled randomly from an infinite population.

3. Note that Assumptions 3 and 3 also imply strong ignorability of sample assignment for treated and control non-compliers since we assume in that compliance is also independent of sample and treatment assignment, conditional on W_i (Assumption 2). However, we are interested only on modeling the response surfaces for compliers.

also determine population treatment assignment and complier status. This choice reflects a modeling assumption of our estimation procedure described in Section 3.

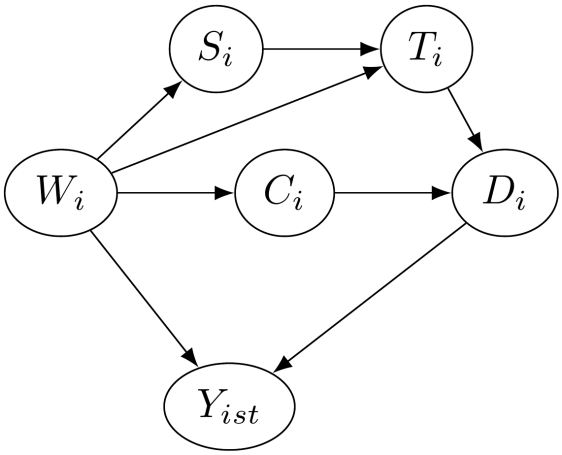

Figure 1: Causal diagram indicating the conditional independence assumptions needed to estimate the PATT-C.

We additionally include Assumptions 6 and 7, which are made by Angrist, Imbens, and Rubin (1996) to ensure identifiability. The former assumption ensures that crossover is only possible from treatment to control:

Assumption 6. *No defiers:*

$$T_i \geq D_i, \qquad \forall\, i\,,\, t = \{0, 1\}.$$

Assumption 7 ensures treatment assignment affects the response only through the treatment received. In particular, the treatment effect may only be nonzero for compliers.

Assumption 7. *Exclusion restriction: For noncompliers,*

$$Y_{i11} = Y_{i10},, \qquad \forall\, i.$$

2.2 PATT and PATT-C

The estimand of interest is the average treatment effect on those in the population who receive treatment:

$$\tau_{\text{PATT}} = \mathbb{E}\left(Y_{i01} - Y_{i00} \mid S_i = 0, D_i = 1\right), \tag{1}$$

It includes individuals who actually receive the treatment, but does not include those who are eligible for treatment and do not accept it (i.e., noncompliers). The following theorem, which we modify from Hartman et al. (2015) to account for noncompliance, relates the treatment effect in the population to the treatment effect in the RCT. The result enables us to estimate the Population Average Treatment Effect on Treated Compliers (PATT-C):

Theorem 1. *Under Assumptions 1 – 7,*

$$\tau_{PATT\text{-}C} = \mathbb{E}_{01}\left[\mathbb{E}\left(Y_{i11} \mid S_i = 1, D_i = 1, W_i\right)\right] - \mathbb{E}_{01}\left[\mathbb{E}\left(Y_{i10} \mid S_i = 1, T_i = 0, C_i = 1, W_i\right)\right] \tag{2}$$

where $\mathbb{E}_{01}\left[\mathbb{E}(\cdot \mid \ldots, W_i)\right]$ *denotes the expectation with respect to the distribution of* W_i *in the treated individuals in the target population.*

Proof. We separate the expectation linearly into two terms and consider each individually.

$$\begin{aligned}
\mathbb{E}\left(Y_{i01} \mid S_i = 0, D_i = 1\right) &= \mathbb{E}\left(Y_{i11} \mid S_i = 0, D_i = 1\right) && \text{by Assumption 1} \\
&= \mathbb{E}\left(Y_{i11} \mid S_i = 0, T_i = 1, C_i = 1\right) && \text{by Assumption 6} \\
&= \mathbb{E}_{01}\left[\mathbb{E}\left(Y_{i11} \mid S_i = 0, T_i = 1, C_i = 1, W_i\right)\right] \\
&= \mathbb{E}_{01}\left[\mathbb{E}\left(Y_{i11} \mid S_i = 1, T_i = 1, C_i = 1, W_i\right)\right] && \text{by Assumption 3} \\
&= \mathbb{E}_{01}\left[\mathbb{E}\left(Y_{i11} \mid S_i = 1, D_i = 1, W_i\right)\right]
\end{aligned}$$

Intuitively, conditioning on W_i makes sample selection ignorable under Assumption 3. This is the critical connector between the third and fourth lines of the first expectation derivation.

$$
\begin{aligned}
\mathbb{E}\left(Y_{i00} \mid S_i = 0, D_i = 1\right) &= \mathbb{E}\left(Y_{i10} \mid S_i = 0, D_i = 1\right) && \text{by Assumption 1} \\
&= \mathbb{E}\left(Y_{i10} \mid S_i = 0, T_i = 1, C_i = 1\right) && \text{by Assumption 6} \\
&= \mathbb{E}_{01}\left[\mathbb{E}\left(Y_{i10} \mid S_i = 1, T_i = 1, C_i = 1, W_i\right)\right] && \text{by Assumption 4} \\
&= \mathbb{E}_{01}\left[\mathbb{E}\left(Y_{i10} \mid S_i = 1, T_i = 0, C_i = 1, W_i\right)\right]
\end{aligned}
$$

The last line follows because of the randomization carried out in the RCT, which guarantees $Y_{i10} \perp\!\!\!\perp T_i \mid (W_i, S_i = 1)$. Finally, the result follows by plugging these two expressions into Eq. (1). □

3 Estimation procedure

There are two challenges in turning Theorem 1 into an estimator of $\tau_{\text{PATT-C}}$ in practice. First, we must estimate the inner expectation over potential outcomes of compliers in the RCT. In our empirical example, we use an ensemble of algorithms (Laan, Polley, and Hubbard 2007) to estimate the response surface for compliers in the RCT, given their covariates. Thus, the first term in the expression for $\tau_{\text{PATT-C}}$ is estimated by the weighted average of points on the response surface, evaluated for each treated population member's potential outcome under treatment. The second term is estimated by the weighted average of points on the response surface, evaluated for each treated population member's potential outcome under control.

The second challenge is that we cannot observe which individuals are included in the estimation of the second term. In the RCT control group, C_i is unobservable, as they always

receive no treatment ($D_i = 0$). We must estimate the second term of Eq. (2) by predicting who in the control group would be a complier, had they been assigned to treatment. An alternative approach, is to simply weight the PATT estimate by the population compliance rate in order to yield a population average effect of treatment on treated compliers. However, the compliance rate is likely to differ between the sample and population, as well as across subgroups. Explicitly modeling compliance allows us to decompose PATT-C estimates by subgroup according to covariates common to both RCT and observational datasets.

The procedure for estimating $\tau_{\text{PATT-C}}$ using Theorem 1 is as follows:

S.1 Using the group assigned to treatment in the RCT ($S_i = 1, T_i = 1$), train a model (or an ensemble of models) to predict the probability of compliance as a function of covariates W_i.

S.2 Using the model from S.1, predict who in the RCT assigned to control *would have* complied to treatment had they been assigned to the treatment group. Complier status C_i assumes the value 1 for individuals with a probability greater than 50%, and is otherwise 0.

S.3 For the observed compliers assigned to treatment and predicted compliers assigned to control, train a model to predict the response using W_i and T_i,[4] which gives $\mathbb{E}(Y_{i1t} \mid S_i = 1, C_i = 1, T_i = t, W_i)$ for $t \in \{0, 1\}$.

S.4 For all individuals who received treatment in the population ($S_i = 0, D_i = 1$), estimate their potential outcomes Y_{i10} and Y_{i11} using the model from S.3. The mean counterfactual Y_{i11} minus the mean counterfactual Y_{i10} is the estimate of $\tau_{\text{PATT-C}}$.

Assumptions 3 and 4 are particularly important for estimating $\tau_{\text{PATT-C}}$: the success of the proposed estimator hinges on the assumption that the response surface is the same for compliers in the RCT and target population. If this does not hold, then the potential

4. Note that assigned and observed treatment are the same for these individuals.

outcomes Y_{i10} and Y_{i11} for target population individuals cannot be estimated using the model from S.3.[5]

3.1 Modeling assumptions

In addition to the identification assumptions, we require additional modeling assumptions for the estimation procedure. We assume that the W_i that determine sample selection also determine population treatment assignment and complier status. As pointed out in Section 2.1, we also require that W_i is complete because if any relevant elements of W_i are not controlled, then there is a backdoor pathway from T_i back to W_i and into Y_{ist}. Lastly, we assume that the compliance model is accurate in predicting compliance in the training sample of RCT participants assigned to treatment and also generalizable to RCT participants assigned to control (S.1 and S.2). Section 3.2 below describes our method of evaluating the generalizability of the compliance model.

3.2 Ensemble method

In the empirical application, we use the weighted ensemble method described in Laan, Polley, and Hubbard (2007) for S.1 and S.3 of the estimation procedure. This ensemble method combines algorithms with a convex combination of weights based on minimizing cross-validated error. It is shown to control for overfitting and outperforms single algorithms selected by cross-validation (Polley and Van Der Laan 2010).

We choose a variety of candidate algorithms to construct the ensemble, with a preference towards algorithms that tend to outperform in supervised classification tasks. We also have a preference for algorithms that have a built-in variable selection property. The idea is that we input the same W_i and each candidate algorithm selects the most important covariates for predicting compliance status or potential outcomes. We select four types of candidate algorithms: gradient boosting algorithms (**friedman2000additive**; Friedman 2001, 2002);

5. Section 5.3 discusses whether the strong ignorability assumptions are plausible in the empirical application.

regularized linear models (i.e., Lasso or ridge regression) (Tibshirani et al. 2012); and random forests (Breiman 2001). Lasso regression is an attractive because it tends to shrink all but one of the coefficients of correlated covariates to zero.

4 Simulations

We conduct a simulation study to compare the performance of the PATT and PATT-C estimators. For comparison, we also estimate the Sample Average Treatment Effect on the Treated Compliers (SATT-C) by fitting the response curve for compliers in the RCT — which we do in S.3 for estimating $\tau_{\text{PATT-C}}$ — and then taking the mean difference in potential outcomes for treated compliers in the sample:

$$\tau_{\text{SATT-C}} = \mathbb{E}(Y_{i11} \mid S_i = 1, C_i = 1, T_i = 1, W_i) - \mathbb{E}(Y_{i10} \mid S_i = 1, C_i = 1, T_i = 1, W_i). \quad (3)$$

Our simulation is designed so that the effect of treatment is heterogeneous and depends on covariates which are different in the RCT and target population. The design satisfies the conditional independence assumptions in Figure 1.

4.1 Simulation design

In the simulation, RCT eligibility, complier status, and treatment assignment in the population depend on multivariate normal covariates $(W_i^1, W_i^2, W_i^3, W_i^4)$ with mean $(0.5, 1, -1, -1)$ and covariances $\text{Cov}(W_i^1, W_i^2) = \text{Cov}(W_i^1, W_i^4) = \text{Cov}(W_i^2, W_i^4) = \text{Cov}(W_i^3, W_i^4) = 1$ and $\text{Cov}(W_i^1, W_i^3) = \text{Cov}(W_i^2, W_i^3) = 0.5$. The first three covariates are observed by the researcher and W_i^4 is unobserved.

The equation for selection into the RCT is

$$S_i = \mathbb{I}(e_2 + g_1 W_i^1 + g_2 W_i^2 + g_3 W_i^3 + e_4 W_i^4 + R > 0),$$

where R is standard normal. The parameter e_2 varies the fraction of the population eligible for the RCT and e_4 varies the degree of confounding with sample selection. We set the constants g_1, g_2, and g_3 to be 0.5, 0.25, and 0.75, respectively.

Complier status is determined by

$$C_i = \mathbb{I}(e_3 + h_2 W_i^2 + h_3 W_i^3 + e_5 W_i^4 + Q > 0),$$

where Q is standard normal, e_3 varies the fraction of compliers in the population, and e_5 varies the degree of confounding with treatment assignment. We set the constants h_2 and h_3 to 0.5.

For individuals in the population ($S_i = 0$), treatment is assigned by

$$T_i = \mathbb{I}(e_1 + f_1 W_i^1 + f_2 W_i^2 + e_6 W_i^4 + V > 0),$$

where V is standard normal. Varying e_1 changes the fraction eligible for treatment in the population and e_6 varies the degree of confounding with sample selection. We set the constants f_1 and f_2 to 0.25 and 0.75, respectively. For individuals in the RCT ($S_1 = 1$), treatment assignment is a sample from a Bernoulli distribution with probability $p = 0.5$. We set treatment received D_i according to T_i and C_i: $D_i = T_i$ if $C_i = 1$ and $D_i = 0$ if $C_i = 0$.

Finally, the response is determined by

$$Y_{ist} = a + bD + c_1 W_i^1 + c_2 W_i^2 + dU.$$

We assume that the treatment effect b is heterogeneous depending on W_i^1: $b = 1$ if $W_i^1 > 0.75$ and $b = -1$ if $W_i^1 \leq 0.75$. We set a, c_1, and d to 1 and c_2 to 2. U is standard

normal and $U, V, R, Q, (W_i^1, W_i^2, W_i^3, W_i^4)$ are mutually independent.

We generate a population of 30,000 individuals and randomly sample 5,000. Those among the 5,000 who are eligible for the RCT ($S_i = 1$) are selected. Similarly, we sample 5,000 individuals from the population and select those who are not eligible for the RCT ($S_i = 0$): these are our observational study participants. This set-up mimics the reality that a population census is usually impossible.

We set each individual's treatment received D_i according to their treatment assignment and complier status and observe their responses Y_{ist}. In this design, the way that we've set S_i, T_i, D_i, C_i, and Y_{ist} ensures that Assumptions 1 – 7 hold.

In the assigned-treatment RCT group ($S_i = 1, T_i = 1$), we train a gradient boosting algorithm on the covariates to predict who in the control group ($S_i = 1, T_i = 0$) has $C_i = 1$, which is unobservable. These individuals *would have* complied had they been assigned to the treatment group. For this group of observed compliers to treatment and predicted compliers from the control group of the RCT, we estimate the response surface using gradient boosting with features (W_i^1, W_i^2, W_i^3) and D_i. The PATT-C is estimated according to the estimation procedure outlined above.

4.2 Simulation results

We vary each of the parameters $e_1, e_2,, e_3$, e_4, e_5, and e_6 along a grid of $\{-2, -1, 0, 1, 2\}$ in order to generate different combinations of rates of compliance, treatment eligibility, RCT eligibility in the population, and confounding. For each possible combination of the six parameters, we run the simulation ten times and compute the average root mean squared error (RMSE) of PATT-C, PATT, and the SATT-C. All other parameters are held constant. The PATT and PATT-C estimates are obtained by estimating the response surface on all individuals in the RCT and applying S.4 of our estimation procedure to the nonrandomized trial individuals. SATT-C is estimated using the predicted values from the response model trained in S.3.

Figure 2 shows the relationship between the percent of compliers in the whole population, the percentage of people in the whole population eligible to participate in the RCT, and the RMSE of the PATT and PATT-C estimators. The pattern of performance is qualitatively similar: both perform best when compliance is high and perform badly when compliance is low. Figure 3 shows that for a fixed rate of compliance, the RMSE of both estimators increases with the fraction of the total population eligible to participate in the RCT. When less than 70% of the RCT is compliers, PATT-C tends to have lower RMSE than PATT.

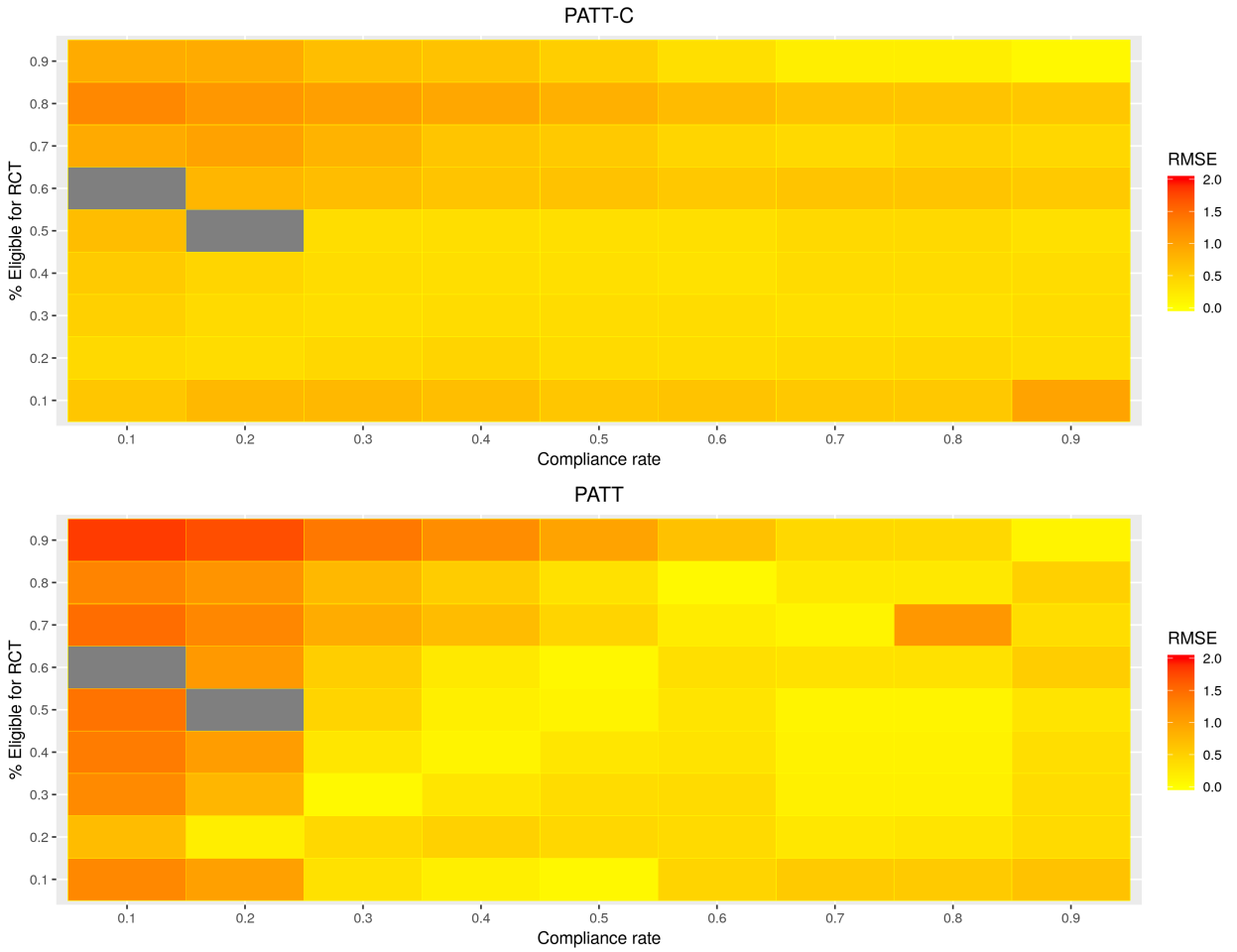

Figure 2: Simulated RMSE, binned by compliance rate and percent eligible for the RCT. Darker tiles correspond to higher RMSE.

Figure 4 compares the RMSE of PATT and PATT-C with the SATT-C at varying levels of compliance in the total population. When compliance is low, both PATT and PATT-C

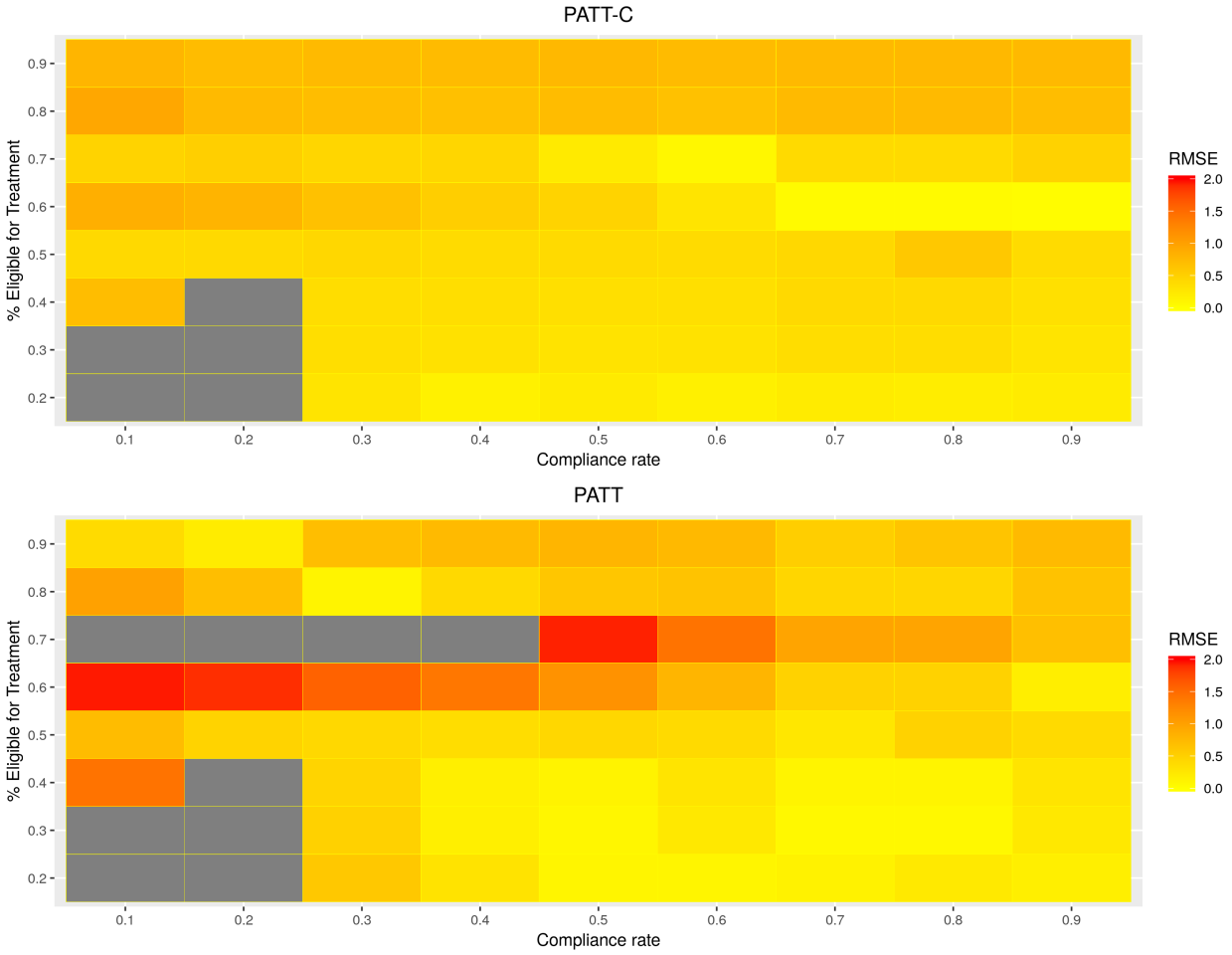

Figure 3: Simulated RMSE, binned by compliance rate and percent eligible for treatment.

estimators have a large RMSE. The adjusted estimator has a large variance because the subset of the RCT who are considered compliers is relatively small, whereas the unadjusted estimator has bias from the large amount of crossover. For low levels of compliance, the SATT-C actually tends to estimate the average causal effects for the target population more closely than either PATT or PATT-C. On the other hand, when the compliance rate is high, both PATT and PATT-C estimators have low RMSE, with the adjusted estimator performing slightly better. Meanwhile, the median RMSE for the SATT-C stays relatively constant regardless of compliance rate , and the estimator is unbiased for the SATT-C rather than the PATT-C. The SATT-C estimate accounts for noncompliance, but is unable to account for differences in pretreatment covariates between the RCT sample and target population. This shows that reweighting or response surface methods are needed to extrapolate RCT results to a wider population.

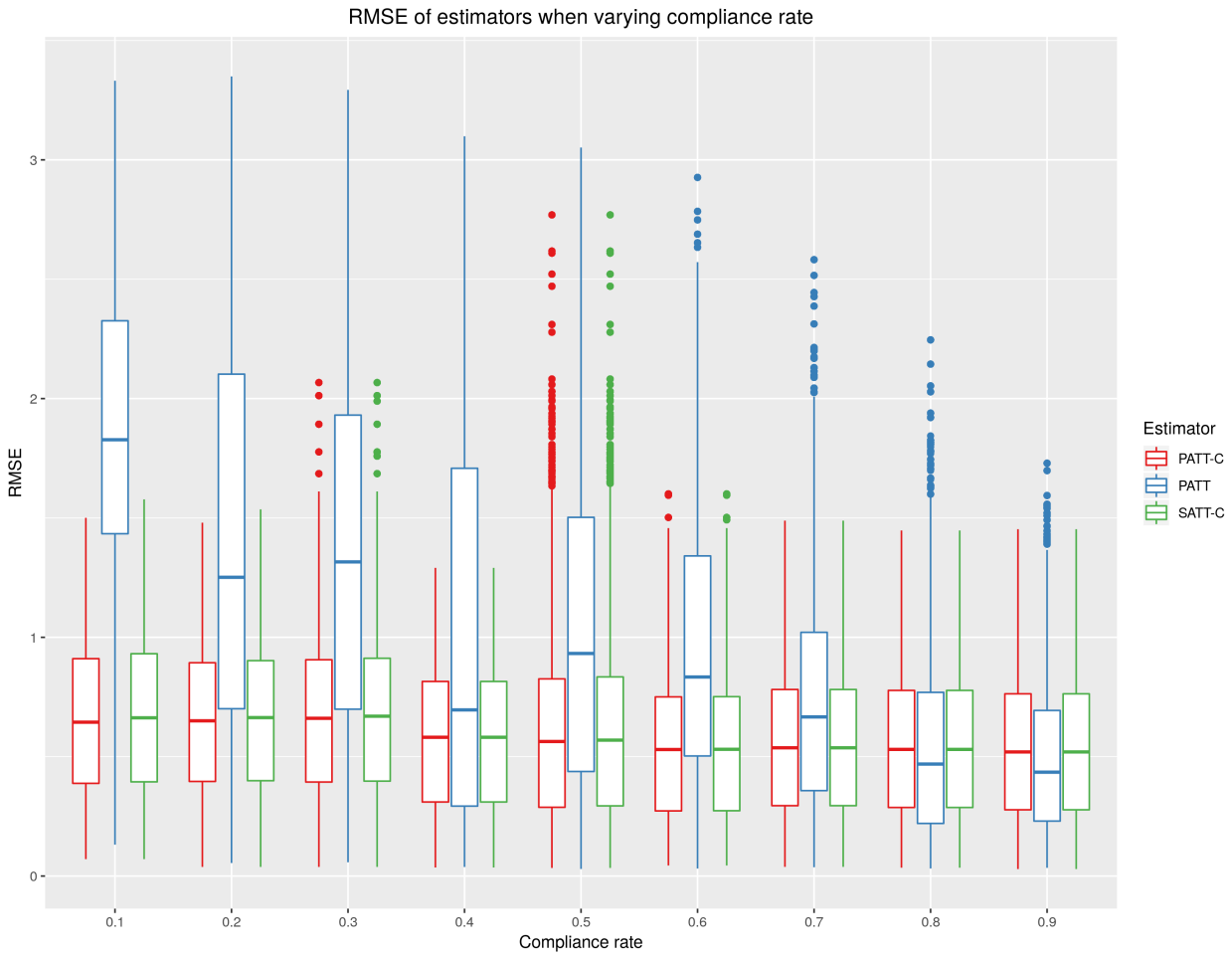

Figure 4: Simulated RMSE of PATT-C, PATT, and SATT-C, according to compliance rates in the total population.

5 Application: Medicaid and health care use

We apply the proposed estimator to measure the effect of Medicaid coverage on health care use for a target population of adults who may benefit from expansions to the Medicaid program. In particular, we examine the population of nonelderly adults in the U.S. with household incomes at or below 138% of the Federal Poverty Level (FPL) — which amounts to $32,913 for a four–person household in 2014 — who may be eligible for Medicaid following the Affordable Care Act (ACA) expansion.

5.1 RCT sample

We draw RCT data from the Oregon Health Insurance Experiment (OHIE) (Finkelstein et al. 2012; Katherine Baicker et al. 2013; Baicker et al. 2014; Taubman et al. 2014). In 2008, approximately 90,000 uninsured low-income adults participated in the OHIE to receive Medicaid benefits.[6] Treatment occurred at the household level: participants selected by the lottery won the opportunity for themselves and any household member to apply for Medicaid. Within a sample of 74,922 individuals representing 66,385 households, 29,834 participants were selected by the lottery; the remaining 45,008 participants served as controls in the experiment. Participants in selected households received benefits if they returned an enrollment application within 45 days of receipt. Among participants in selected households, about 60% mailed back applications and only 30% successfully enrolled.[7]

The response data originate from a mail survey that was administered to participants over July and August 2009 ($n = 23,741$ survey respondents). We use the same definition of insurance coverage as Finkelstein et al. (2012) to form our measure of compliance, which is a binary variable indicating whether the participant was enrolled in any Medicaid program

6. Eligible participants include Oregon residents (US citizens or legal immigrants) aged 19 to 64 not otherwise eligible for public insurance, who have been without insurance for six months, and have income below the FPL and assets below $2,000.

7. About half of the returned applications were deemed ineligible, primarily due to failure to demonstrate income below the FPL. Enrolled participants were required to recertify their eligibility status every six months.

between the notification date and 30 September 2009. The OHIE data include pretreatment covariates for gender, age, race, ethnicity, health status, education, and household income. Table 1 presents summary statistics on the $n = 10,619$ RCT participants in our sample who were on Medicaid during the study period.

Our outcomes of interest are binary variables for any emergency room (ER) and outpatient visits in the past 12 months. ER use is an important outcome because it is the main delivery system through which the the uninsured receive health care. The uninsured could potentially receive higher quality and less affordable healthcare through outpatient visits. An important question for policymakers is whether Medicaid expansions will decrease ER utilization by the previously uninsured.

Subsequent research calls in to question the external validity of the OHIE, which resulted in the counterintuitive finding that Medicaid increased ER use among RCT participants (Finkelstein et al. 2012; Taubman et al. 2014). For example, quasi-experimental studies on the impact of the 2006 Massachusetts health reform — which served as a model for the ACA — show that ER use decreased or remained constant following the reform (Miller 2012; Kolstad and Kowalski 2012). A challenge to the external validity of the OHIE is that it's exclusion criteria was likely more restrictive than government health insurance expansions.

5.2 Observational data

We acquire data on the target population from the National Health Interview Study (NHIS) for years 2008 to 2017 (National Center for Health Statistics).[8] We restrict our sample to respondents with income below 138% of the FPL and who are uninsured or on Medicaid ($n = 11,129$). We select covariates on respondent characteristics that match the OHIE pretreatment covariates. The outcomes of interest from NHIS are variables on ER and outpatient visits in the past 12 months. We use a recoded variable that indicates whether respondents are on Medicaid as an analogue to the OHIE compliance measure.

8. A possible limitation of this application is that it ignores the complex sampling techniques of the NHIS sample design such as differential sampling, which is discussed in detail in Parsons et al. (2014).

5.3 Verifying assumptions

In order for $\tau_{\text{PATT-C}}$ to be identified, Assumptions 1 – 7 must be met. Assumption 1 ensures that potential outcomes for participants in the target population (i.e., respondents in the NHIS sample) would be identical to their outcomes in the RCT if they had been randomly assigned their observed treatment. Medicaid coverage for uninsured individuals was applied in the same manner in the RCT as it is in the population. Differences in potential outcomes due to sample selection might arise, however, if there are differences in the mail surveys used to elicit health care use responses between the RCT and the nonrandomized study.

We cannot directly test Assumptions 3 and 4, which state that potential outcomes for treatment and control are independent of sample assignment for individuals with the same covariates and assignment to treatment. The assumptions are only met if every possible confounder associated with the response and the sample assignment is accounted for. In our estimation of the response surface, we use all demographic, socioeconomic, and pre-existing health condition data that were common in the OHIE and NHIS data. Potentially important unobserved confounders include the number of hospital and outpatient visits in the previous year, proximity to health services, and enrollment in other federal programs.

Table 1 compares RCT participants selected for Medicaid with NRT participants on Medicaid. Compared to the RCT treated, the target population "treated" group is predominantly female, younger, more racially and ethnically diverse, less educated, and live in higher income households. Diagnoses of diabetes, asthma, high blood pressure, and heart disease are more common among the population on Medicaid then the RCT treated.

Table 1: Pretreatment covariates and responses for OHIE and NHIS respondents on Medicaid.

	OHIE control $n = 5,426$		OHIE treated $n = 5,193$		NHIS treated $n = 6,261$	
Covariate	**n**	%	**n**	%	**n**	%
Sex:						
Female	3,124	57.6	2,920	56.2	4,288	68.5
Age:						
19-49	1,546	28.5	1,367	26.3	4,324	69.1
50-64	3,880	71.5	3,826	73.7	1,937	30.9
Race:						
White	4,746	87.5	4,393	84.6	3,902	62.3
Black	243	4.5	197	3.8	1723	27.5
Hispanic	326	6.0	476	9.2	1570	25.1
Health status:						
Diabetes	557	10.3	539	10.4	866	13.8
Asthma	1,030	19.0	887	17.1	1,272	20.3
High blood pressure	1,649	30.4	1,418	27.3	2,166	34.6
Heart condition	168	3.1	141	2.7	529	8.4
Education:						
Less than high school	1,062	19.6	950	18.3	1,942	31.0
High school diploma or GED	3,093	57.0	2,775	53.4	2,076	33.2
Voc. training / 2-year degree	959	17.7	1,031	19.9	1,810	28.9
4-year college degree or more	312	5.8	437	8.4	433	6.9
Income:						
$< \$10$k	5,425	100	3,204	61.7	2,588	41.3
\$10k-\$25k	1	0	1,616	31.1	3,098	49.5
$> \$25$k	0	0	373	7.2	575	9.2
Response						
Any ER visit	1,377	25.4	1,301	25.1	1,659	26.5
Any outpatient visit	3,265	60.2	3,081	59.3	4,093	65.4

Strong ignorability assumptions may also be violated due to the fact that the OHIE applied a more stringent exclusion criteria compared to the NHIS sample. While the RCT and population sample both screened for individuals below the FPL, only the RCT required those enrolled to recertify their eligibility status every six months.

A violation of the assumption of no-interference (Assumption 5) biases the estimate of $\tau_{\text{PATT-C}}$ if, for instance, treated participants' Medicaid coverage makes control participants more likely to visit the ER. Interference is less likely in this experimental set–up because treatment occurs at the household level. Assumption 2 is violated if assignment to treatment influences the compliance status of individuals with the same covariates. Our compliance ensemble can accurately classify compliance status for 77% of treated RCT participants with only the covariates — and not treatment assignment — as model inputs.[9] This gives evidence in favor of the conditional independence assumption.

The exclusion restriction (Assumption 7) ensures treatment assignment affects the response only through enrollment in Medicaid. It is reasonable that a person's enrollment in Medicaid, not just their eligibility to enroll, would affect their hospital use. For private health insurance one might argue that eligibility may be be negatively correlated with hospital use, as people with pre-existing conditions are less often eligible yet go to the hospital more frequently. This should not be the case with a federally funded program such as Medicaid.

5.3.1 Sensitivity to no defiers assumption

Angrist, Imbens, and Rubin (1996) show that the bias due to violations of Assumption 6 is equivalent to the difference of average causal effects of treatment received for compliers and defiers, multiplied by the relative proportion of defiers, $\mathbb{P}(i \text{ is a defier})/(\mathbb{P}(i \text{ is a complier}]) - \mathbb{P}(i \text{ is a defier}))$.

Table A1 reports the distribution of participants in the OHIE by status of treatment assignment and treatment received. Assumption 6 does not hold due to the presence of

9. The compliance ensemble is evaluated in terms of 10–fold cross–validated MSE. The distribution of MSE for the ensemble and its candidate algorithms are provided in Table A3.

defiers; i.e., participants who were assigned to control and enrolled in Medicaid during the study period. About 6.7% of the RCT sample were assigned to control but were enrolled in Medicaid ($T_i < D_i$) and 65.5% of the sample complied with treatment assignment ($D_i = T_i$), which results in a bias multiplier of 0.11. Suppose that the difference of average causal effects of Medicaid received on ER use for compliers and defiers is 1.2%.[10] The resulting bias is only 0.1%, which would not meaningfully alter the interpretation of the SATT-C or PATT-C reported in Section 5.4.

5.4 Empirical results

We compare PATT-C and SATT estimates for ER and outpatient use. We obtain estimates for the overall group of participants and subgroups according to sex, age, race, health status, education, and household income. Subgroup treatment effects are estimated by taking differences across response surfaces for a given covariate subgroup, and response surfaces are estimated with the ensemble mean predictions. We use treatment assignment, number of household members, and the subgroup covariates as features in the response models. We generate 95% confidence intervals for these estimates using 1,000 bootstrap samples. The entire RCT sample size is $n = 23,205$ and the population sample is $n = 17,145$. There are $n = 10,619$ RCT compliers and $n = 6,261$ individuals in the target population are covered by Medicaid.

Table A2 compares PATT-C estimates of the effect of Medicaid coverage on ER and outpatient usage to SATT-C and compliance-adjusted SATE estimates. The estimated effect of Medicaid on the likelihood of visiting the ER is practically zero for both compliers in the population and sample. In comparison, the estimated treatment effect on RCT compliers is 1.2% [-1.1%, 3.8%]. For population and sample compliers, the effect of Medicaid coverage on the likelihood of outpatient use is also near zero. In comparison, the SATE estimate on outpatient use is -2% [-4.9%, 0.6%].

10. This quantity is the difference between the adjusted and unadjusted SATT on ER use reported in Section 5.4.

Figures 5 and 6 examine conditional treatment effect estimates on ER and outpatient use, respectively. There exists substantial heterogeniety in treatment effect estimates in our PATT-C estimator compared to the PATT and SATT-C estimators. For instance, Medicaid coverage decreases the probability of ER use for white and hispanic population compliers and increases ER use for black population compliers. Conversely, Medicaid coverage decreases outpatient use for black population compliers, while white and hispanic population compliers experience an increase in outpatient use due to Medicaid.

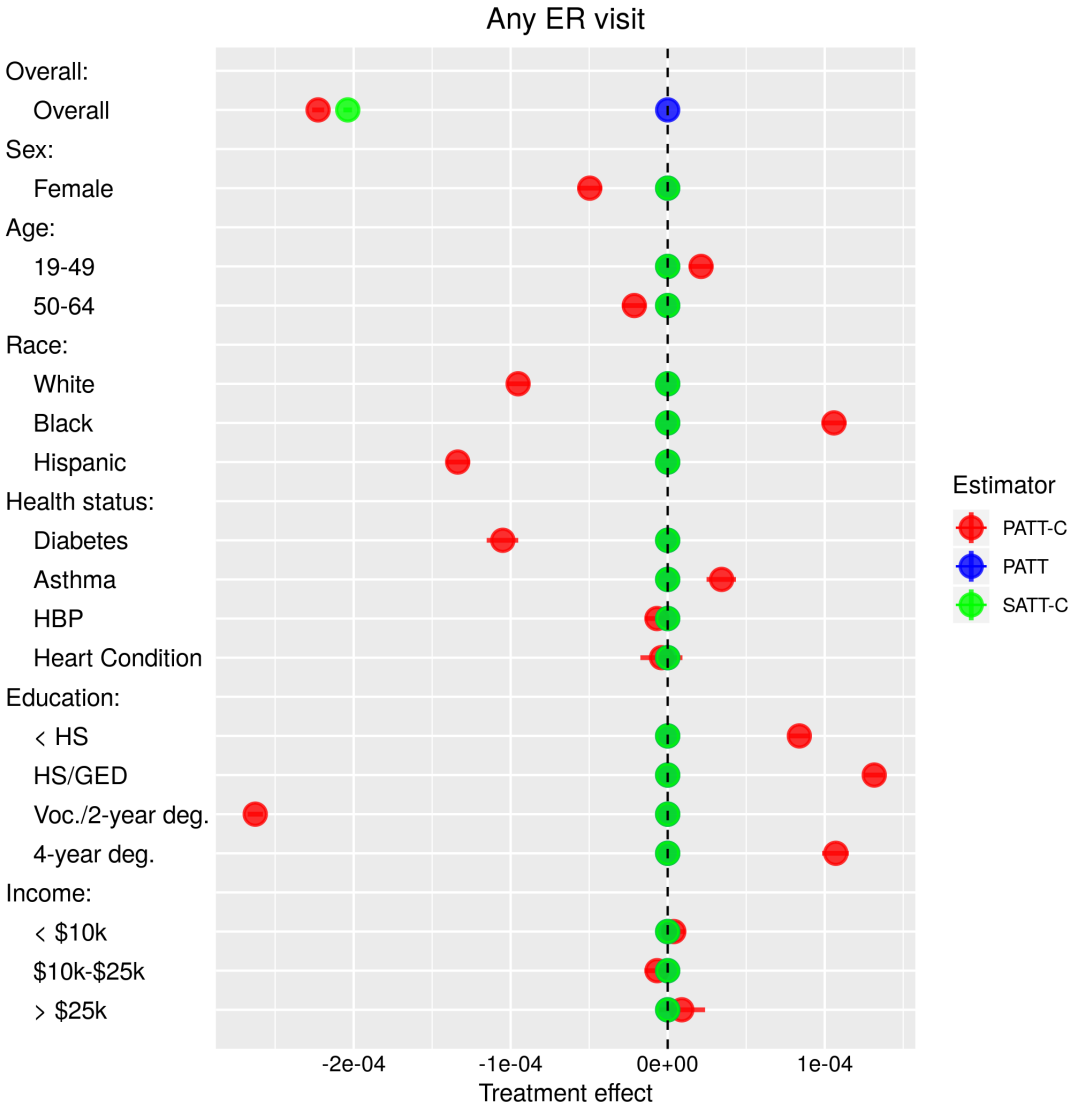

Figure 5: Comparison of estimators: any ER visit. Horizontal lines represent 95% bootstrap confidence intervals.

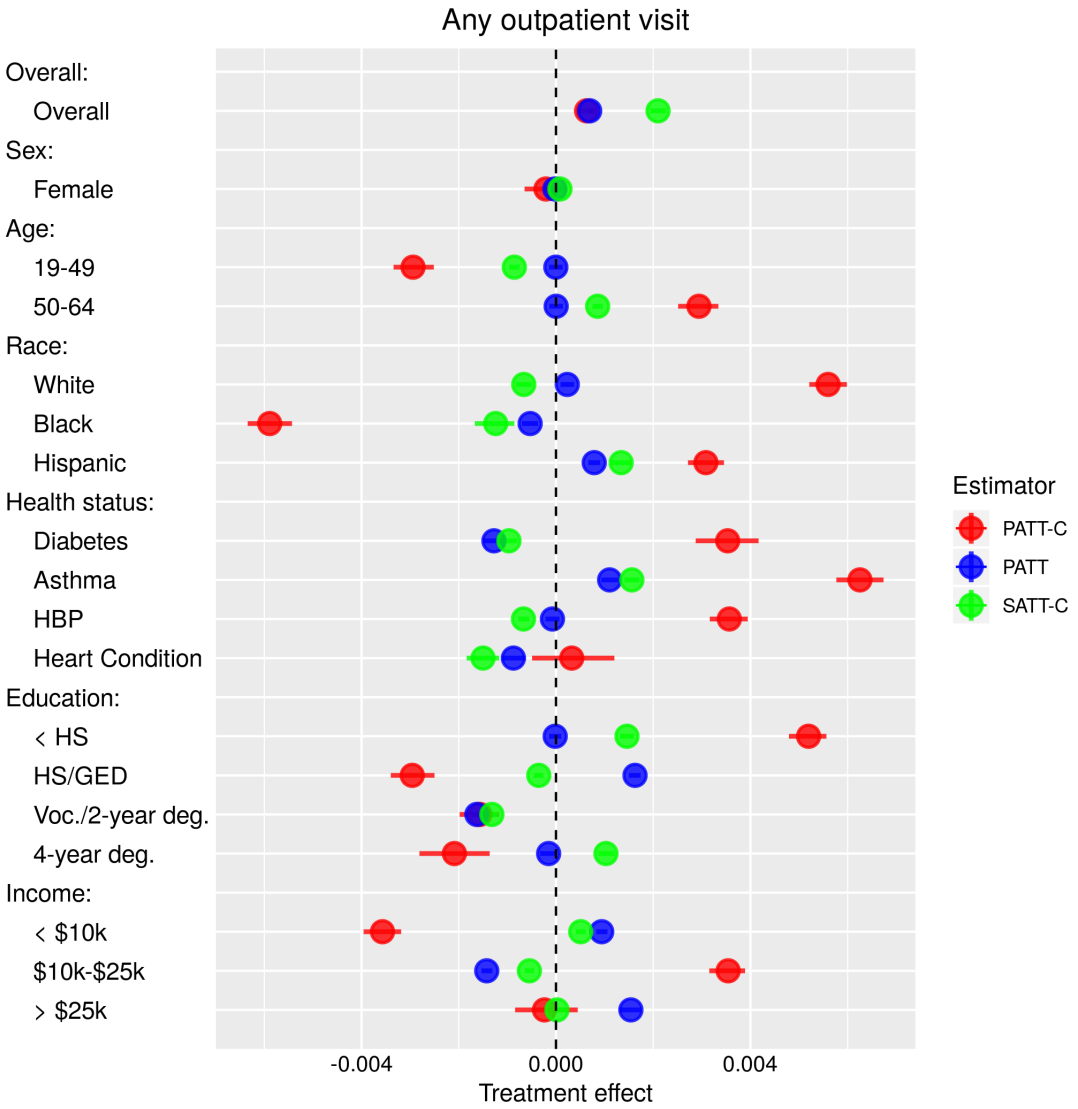

Figure 6: Comparison of estimators: any outpatient visit. Horizontal lines represent 95% bootstrap confidence intervals.

5.4.1 Comparison with previous results

In this section, we compare our population estimates with Finkelstein et al. (2012), who report population estimates of the effect of Medicaid coverage on *the number of* ER and out-patient visits using 2004–2009 NHIS data on adults aged 19–64 below 100 percent of the federal poverty line ($n = 15,528$). Finkelstein et al. (2012) estimates Medicaid coverage significantly increases the number of ER visits by 0.08 [0.05, 0.12] and increases the number

of outpatient visits by 1.45 [1.33, 1.57]. In contrast, our PATT-C estimates indicate that Medicaid coverage has little effect on the probability of ER and outpatient visits.

While our study is the first to our knowledge to estimate heterogeniety in treatment effects for the target population, Taubman et al. (2014) and Kowalski (2016) perform subgroup analyses on the RCT sample. Across subgroups in the sample, Taubman et al. (2014) estimates mostly positive effects of Medicaid on ER use. Similar to our PATT-C estimates, Taubman et al.'s [2014] subgroup analyses indicate that increases in ER use due to Medicaid are significantly larger for men and younger individuals in the sample. The authors similarly find that ndividuals with an education level of high school and those pre-lottery diagnoses experience larger increases in ER use due to Medicaid. Kowalski (2016) perform subgroup analyses on OHIE sample data and find larger increases in ER use as a result of Medicaid for men, English speakers, and individuals enrolled in a food stamp program prior to the lottery.

6 Discussion

The simulations presented in Section 4 shed light on the conditions under which the proposed estimator should work well. When the rate of crossover from treatment to control is low, the proposed estimator performs better than the estimator which doesn't adjust for noncompliance. Here, the unadjusted estimator gives the ITT effect, which tends to underestimate the average treatment effect on compliers. Of course, the simulation results depend on the particular way we parameterized the compliance, selection, treatment assignment, and response schedules. In particular, the strength of correlation between the covariates and compliance governs how well the estimator will perform, since S.1 of the estimation procedure is to predict who *would be* a complier in the RCT control group, had they been assigned to treatment. If it is difficult to predict compliance using the observed covariates, then the estimator will perform badly because of noise introduced by incorrectly treating

noncompliers as compliers. Further research should be done into ways to test how well the model of compliance works in the population.

The proposed estimator, adjusting for noncompliance, performs better in simulations than the unadjusted estimator when compliance is low and can be predicted by observed covariates (Figure 4). In the OHIE trial, only about 30% of those selected to receive Medicaid benefits actually enrolled. Our compliance ensemble accurately classified compliance status for 77% of treated RCT participants using only the pretreatment covariates as features. While we don't know how well the compliance ensemble predicts for the control group, the control group should be similar to the treatment group on pretreatment covariates because of the randomization done. The model's performance on the training set suggests that compliance is not purely random and depends on observed covariates. This gives evidence in favor of using the proposed estimator.

Further research might explore models to more accurately predict compliance in RCTs. Accurately predicting compliance is not only essential for yielding unbiased estimates of the average causal effects for target populations, it is also useful for researchers and policymakers to know which groups of individuals are unlikely to comply with treatment.

In our empirical application, the sample population differs in several dimensions from the target population of individuals who will be covered by other Medicaid expansions, such as the ACA expansion to cover all adults up to 138% of the FPL. For instance, the RCT participants are disproportionately white urban–dwellers (Taubman et al. 2014). The RCT participants volunteered for the study and therefore may be in poorer health compared to the target population. These differences in baseline covariates make reweighting or response surface methods necessary to extend the RCT results to the population.

Explicitly modeling compliance allows us to decompose SATT-C and PATT-C estimates by subgroup according to pretreatment covariates common to both RCT and observational datasets; e.g, demographic variables, pre–existing conditions, and insurance coverage. We find substantial differences between sample and population estimates in terms of race, ed-

ucation, and health status subgroups. This pattern is expected because RCT participants volunteered for the study and are predominately white and educated.

References

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91, no. 434 (June): 444–455.

Baicker, K, A Finkelstein, J Song, and S Taubman. 2014. "The Impact of Medicaid on Labor Market Activity and Program Participation: Evidence from the Oregon Health Insurance Experiment." *The American Economic Review* 104 (5): 322.

Baicker, Katherine, Sarah L Taubman, Heidi L Allen, Mira Bernstein, Jonathan H Gruber, Joseph P Newhouse, Eric C Schneider, Bill J Wright, Alan M Zaslavsky, and Amy N Finkelstein. 2013. "The Oregon experiment—effects of Medicaid on clinical outcomes." *New England Journal of Medicine* 368 (18): 1713–1722.

Breiman, Leo. 2001. "Random forests." *Machine learning* 45 (1): 5–32.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P Newhouse, and Heidi Allen. 2012. "The Oregon Health Insurance Experiment: Evidence from the First Year." *The Quarterly Journal of Economics* 127 (3): 1057.

Friedman, Jerome H. 2001. "Greedy function approximation: A gradient boosting machine." *Annals of Statistics:* 1189–1232.

———. 2002. "Stochastic gradient boosting." *Computational Statistics & Data Analysis* 38 (4): 367–378.

Hartman, Erin, Richard Grieve, Roland Ramsahai, and Jasjeet S Sekhon. 2015. "From SATE to PATT: Combining experimental with observational studies to estimate population treatment effects." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 10:1111.

Imai, Kosuke, Gary King, and Elizabeth Stuart. 2008. "Misunderstandings Among Experimentalists and Observationalists about Causal Inference." *Journal of the Royal Statistical Society, Series A* 171, part 2:481–502.

Imai, Kosuke, Dustin Tingley, and Teppei Yamamoto. 2013. "Experimental designs for identifying causal mechanisms." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176 (1): 5–51.

Kolstad, Jonathan T, and Amanda E Kowalski. 2012. "The impact of health care reform on hospital and preventive care: evidence from Massachusetts." *Journal of Public Economics* 96 (11-12): 909–929.

Kowalski, Amanda E. 2016. *Doing More When You're Running LATE: Applying Marginal Treatment Effect Methods to Examine Treatment Effect Heterogeneity in Experiments.* Working Paper, Working Paper Series 22363. National Bureau of Economic Research, June. doi:10.3386/w22363. http://www.nber.org/papers/w22363.

Laan, Mark J van der, Eric C Polley, and Alan E Hubbard. 2007. "Super learner." *Statistical Applications in Genetics and Molecular Biology* 6 (1).

Miller, Sarah. 2012. "The effect of insurance on emergency room visits: an analysis of the 2006 Massachusetts health reform." *Journal of Public Economics* 96 (11-12): 893–908.

National Center for Health Statistics. *National Health Interview Survey, 2008–2013.* http://www.cdc.gov/nchs/nhis.htm. [Online; accessed 06-April-2015].

Parsons, Van L, Christopher L Moriarity, Kimball Jonas, Thomas F Moore, Karen E Davis, and Linda Tompkins. 2014. "Design and estimation for the National Health Interview Survey, 2006–2015." *National Center for Health Statistics. Vital and Health Statistics* 2 (165): 1–53.

Polley, Eric C, and Mark J Van Der Laan. 2010. *Super learner in prediction.* Working Paper, U.C. Berkeley Division of Biostatistics Working Paper Series 266. Division of Biostatistics, University of California, Berkeley. http://biostats.bepress.com/ucbbiostat/paper266.

Rubin, Donald B. 1990. "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies." *Statistical Science* 5 (4): 472–480.

Stuart, Elizabeth A, Stephen R Cole, Catherine P Bradshaw, and Philip J Leaf. 2011. "The use of propensity scores to assess the generalizability of results from randomized trials." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 174 (2): 369–386.

Taubman, Sarah, Heidi Allen, Bill Wright, Katherine Baicker, and Amy Finkelstein. 2014. "Medicaid Increases Emergency Department Use: Evidence from Oregon's Health Insurance Experiment." *Science* 343, no. 6168 (January): 263–268.

Tibshirani, Robert, Jacob Bien, Jerome Friedman, Trevor Hastie, Noah Simon, Jonathan Taylor, and Ryan J Tibshirani. 2012. "Strong rules for discarding predictors in lasso-type problems." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 74 (2): 245–266.

Appendix

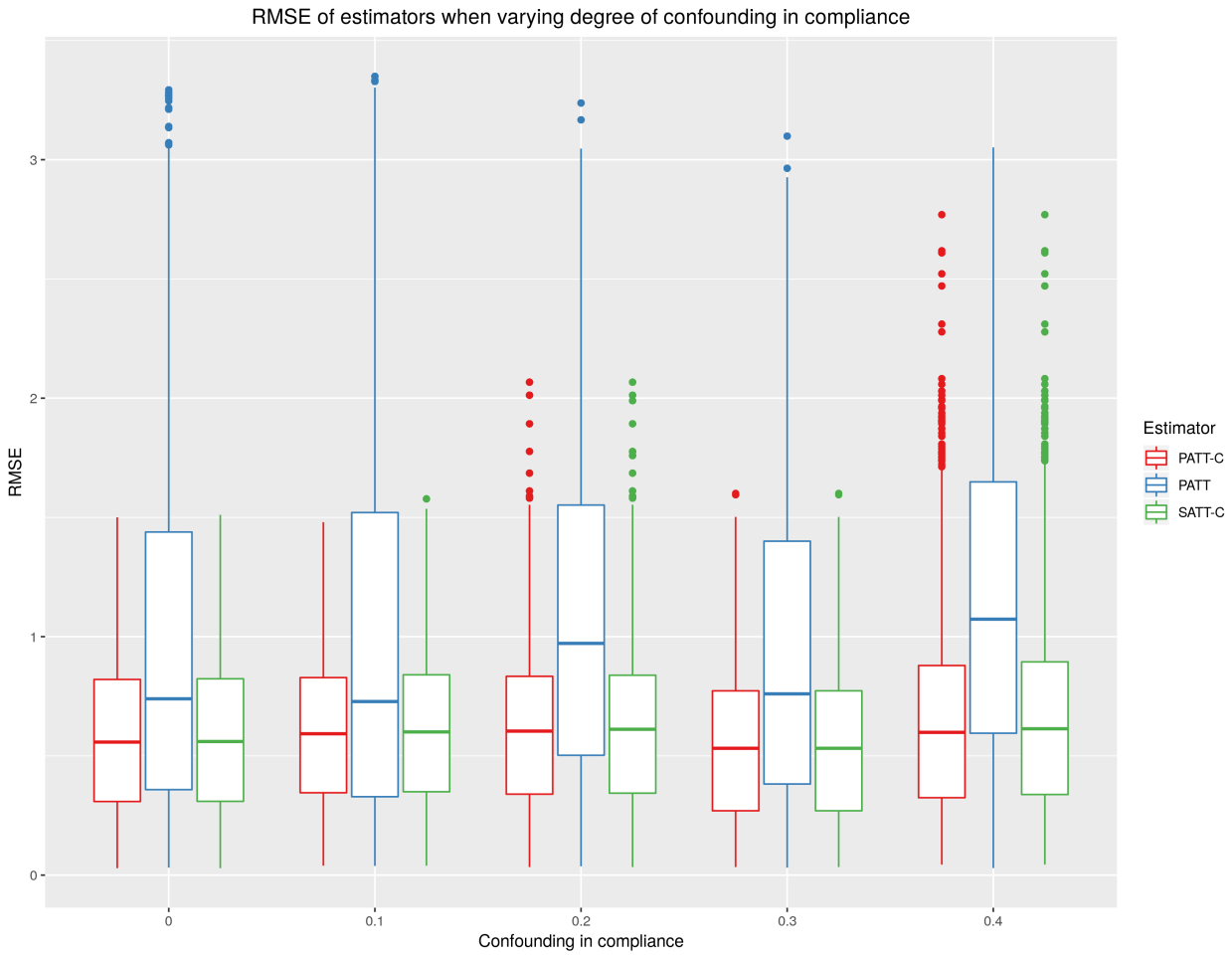

Figure A1: Simulated RMSE of PATT-C, PATT, and SATT-C, according to degree of confounding in compliance.

Table A1: Distribution of OHIE participants by status of treatment assignment (T_i) and treatment received (D_i).

	$D_i = 0$	$D_i = 1$	n
$T_i = 0$	10,010	1,556	11,566
$T_i = 1$	6,446	5,193	11,639
n	16,456	6,749	23,205

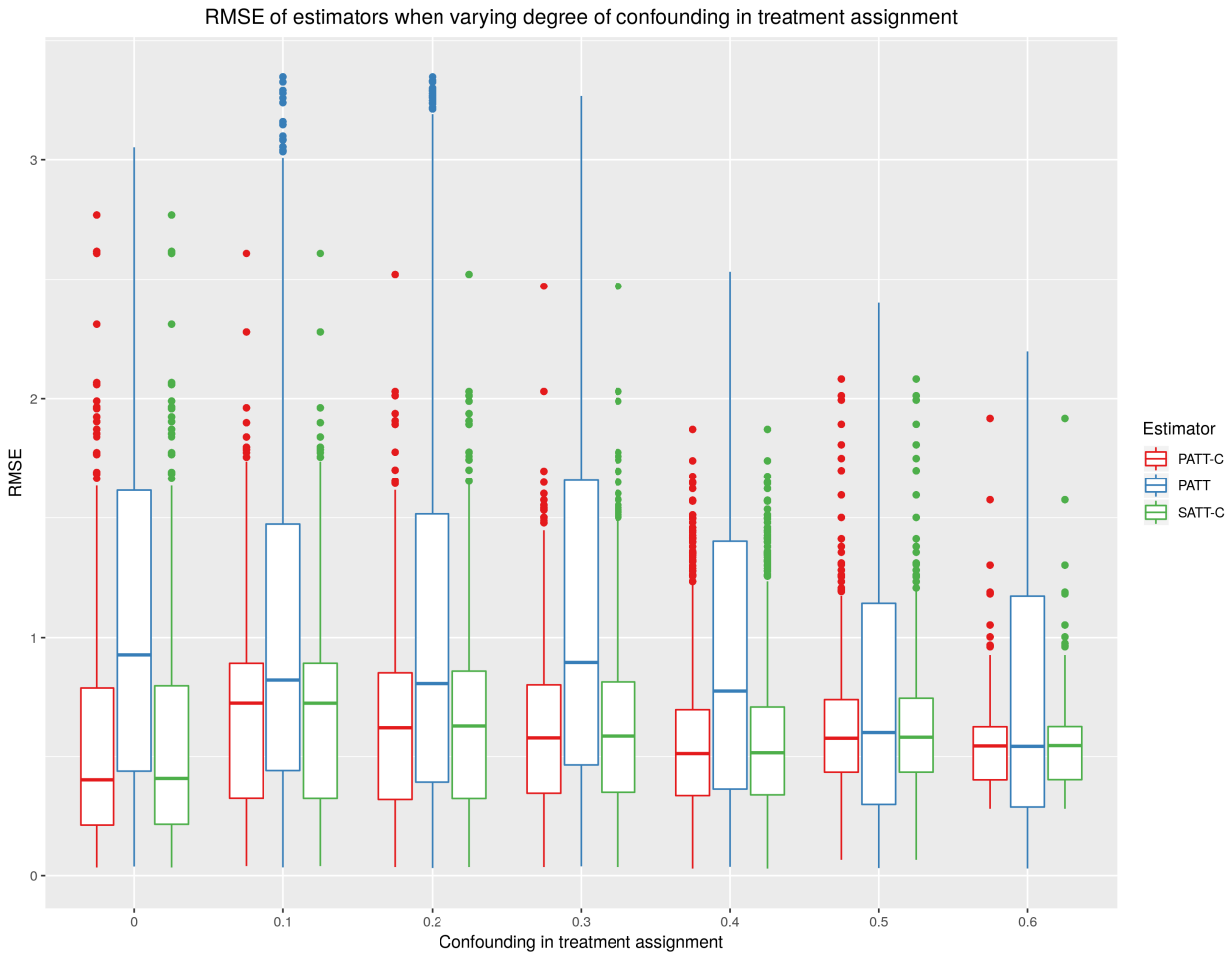

Figure A2: Simulated RMSE of PATT-C, PATT, and SATT-C, according to degree of confounding in treatment assignment.

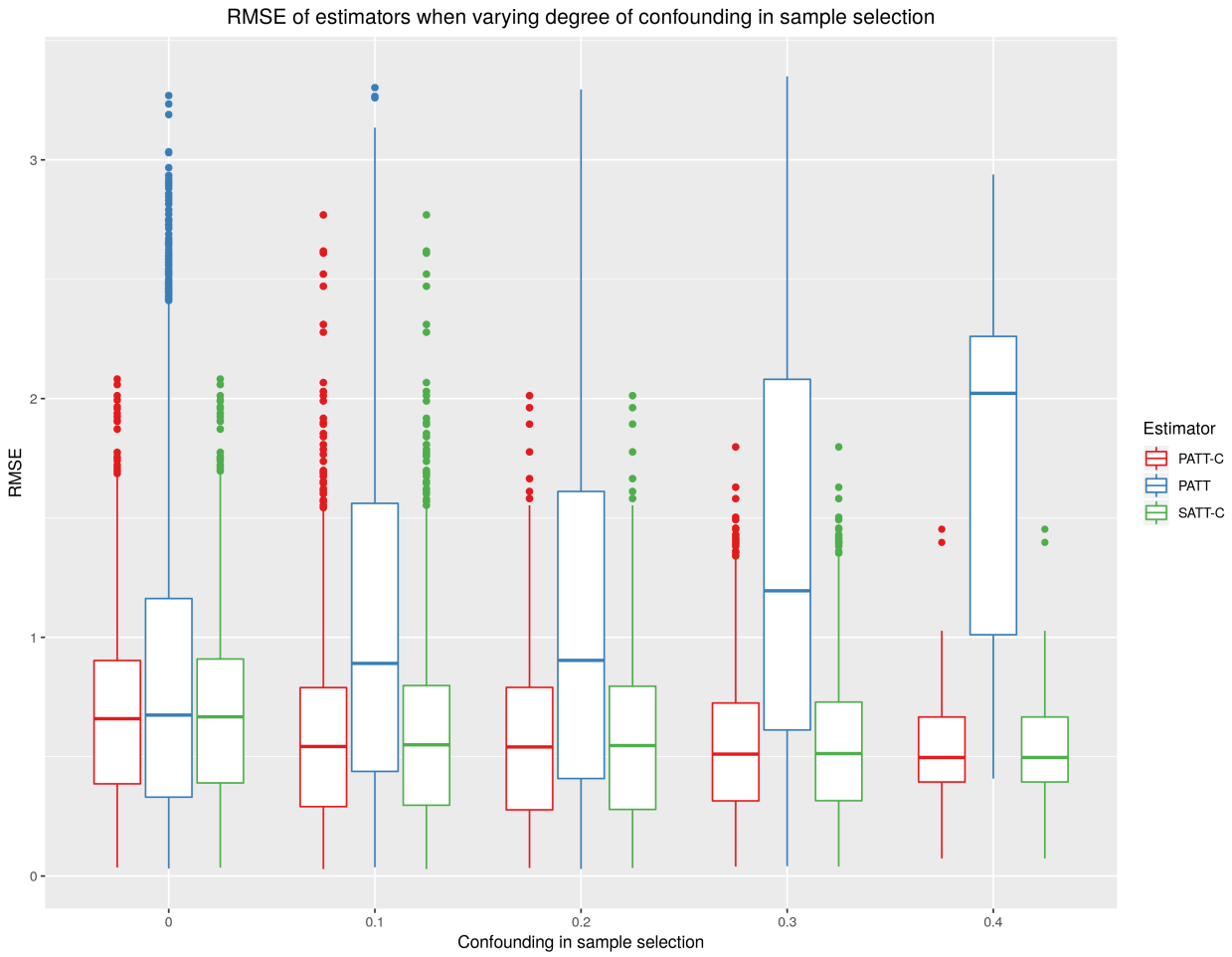

Figure A3: Simulated RMSE of PATT-C, PATT, and SATT-C, according to degree of confounding in sample selection.

Table A2: Comparison of compliance-adjusted estimates on OHIE and NHIS data.

Estimator \ Outcome	Any ER visit	Any outpatient visit
PATT-C	-0.0002 [-0.0002, -0.0002]	0.0006 [0.0004, 0.0008]
SATT-C	-0.0002 [-0.0002, -0.0002]	0.002 [0.002, 0.002]
SATE	0.12 [-0.01, 0.04]	-0.02 [-0.05, 0.01]

Notes: SATE is the ITT effect adjusted by the sample compliance rate in the RCT. Estimates in brackets represent 95% bootstrap confidence intervals.

Table A3: Distribution of MSE for compliance ensemble.

Algorithm	Mean	SE	Min.	Max.
Super learner (`SuperLearner`)	0.22	0.001	0.21	0.23
Lasso regression (`glmnet`)	0.22	0.001	0.21	0.23
Random forests (`randomForest`)	0.27	0.002	0.25	0.29
Ridge regression (`glmnet`)	0.22	0.001	0.21	0.23

Notes: MSE is 10-fold cross-validated error for super learner ensemble and candidate algorithms. R package used for implementing each algorithm in parentheses.

Table A4: Error and weights for candidate algorithms in response ensemble for RCT compliers.

Any ER visit		
Algorithm	MSE	Weight
Adaboost (`gbm`)	0.18	0
Extreme gradient boosting (`XGBoost`)	0.19	0
Gradient boosting, Bernoulli dist. (`gbm`)	0.18	0.55
Lasso regression (`glmnet`)	0.18	0
Logistic regression (`LogicReg`)	0.18	0
Random forests, $\#preds. = 1$ (`randomForest`)	0.25	0
Random forests, $\#preds. = 10$ (`randomForest`)	0.25	0
Regularized linear regression, $\alpha = 0.25$ (`glmnet`)	0.18	0.45
Regularized linear regression, $\alpha = 0.5$ (`glmnet`)	0.18	0
Any outpatient visit		
Algorithm	MSE	Weight
Adaboost (`gbm`)	0.24	0.52
Extreme gradient boosting (`XGBoost`)	0.24	0.06
Gradient boosting, Bernoulli dist. (`gbm`)	0.24	0
Lasso regression (`glmnet`)	0.24	0
Logistic regression (`LogicReg`)	0.24	0
Random forests, $\#preds. = 1$ (`randomForest`)	0.39	0.42
Random forests, $\#preds. = 10$ (`randomForest`)	0.33	0
Regularized linear regression, $\alpha = 0.25$ (`glmnet`)	0.24	0
Regularized linear regression, $\alpha = 0.5$ (`glmnet`)	0.24	0

Notes: cross-validated error and weights used for each algorithm in super learner ensemble. *MSE* is the ten-fold cross-validated mean squared error for each algorithm. *Weight* is the coefficient for the Super Learner, which is estimated using non-negative least squares based on the Lawson-Hanson algorithm. Weights are rounded so that they sum to 1. R package used for implementing each algorithm in parentheses. $\#preds.$ is the number of predictors randomly sampled as candidates in each decision tree in random forests algorithm. α is a regularization parameter.

Table A5: Error and weights for candidate algorithms in response ensemble for all RCT participants.

Any ER visit		
Algorithm	MSE	Weight
Adaboost (`gbm`)	0.18	0
Extreme gradient boosting (`XGBoost`)	0.19	0
Gradient boosting, Bernoulli dist. (`gbm`)	0.18	0
Lasso regression (`glmnet`)	0.18	0.18
Logistic regression (`LogicReg`)	0.18	0.68
Random forests, $\#preds. = 1$ (`randomForest`)	0.25	0
Random forests, $\#preds. = 10$ (`randomForest`)	0.24	0
Regularized linear regression, $\alpha = 0.25$ (`glmnet`)	0.18	0.14
Regularized linear regression, $\alpha = 0.5$ (`glmnet`)	0.18	0
Ridge regression (`glmnet`)	0.25	0
Any outpatient visit		
Algorithm	MSE	Weight
Adaboost (`gbm`)	0.23	0
Extreme gradient boosting (`XGBoost`)	0.24	0
Gradient boosting, Bernoulli dist. (`gbm`)	0.23	0
Lasso regression (`glmnet`)	0.23	0.74
Logistic regression (`LogicReg`)	0.23	0
Random forests, $\#preds. = 1$ (`randomForest`)	0.39	0.25
Random forests, $\#preds. = 10$ (`randomForest`)	0.32	0.01
Regularized linear regression, $\alpha = 0.25$ (`glmnet`)	0.23	0
Regularized linear regression, $\alpha = 0.5$ (`glmnet`)	0.23	0

See notes to Fig.A4.