Journal of the Royal Statistical Society Series A: Statistics in Society, 2025, **00**, 1–50
https://doi.org/10.1093/jrsssa/qnaf114

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series A
Statistics in Society
A

Original Article

Who increases emergency department use? New insights from the Oregon health insurance experiment

Augustine Denteh[1] and Helge Liebert[2]

[1]Department of Economics, Davidson College, 209 Ridge Rd, Box 5000, Davidson, NC 28035, USA
[2]Department of Economics, Swiss National Bank, Börsenstrasse 15, 8001 Zürich, Switzerland

Address for correspondence: Augustine Denteh, Department of Economics, Davidson College, 209 Ridge Rd, Box 5000, Davidson, NC 28035, USA. Email: audenteh@davidson.edu

Abstract

We provide new insights regarding the headline result that Medicaid increased emergency department (ED) use from the Oregon experiment. Using causal machine learning methods, we find meaningful heterogeneous impacts of Medicaid on ED use. The distribution of conditional treatment effects includes a wide range of negative and positive values, suggesting that the average effect masks substantial heterogeneity. A small group—about 14% of participants—in the right tail of the distribution drives the overall effect. We identify priority groups with economically significant increases in ED use based on demographics and previous utilization. Intensive margin effects are an important driver of increases in ED utilization.

Keywords: causal machine learning, ED use, effect heterogeneity, Medicaid, optimal policy

JEL codes: H75, I13, I38

1 Introduction

The finding that Medicaid increased emergency department (ED) utilization in the 2008 Oregon Health Insurance Experiment drew widespread national attention (Taubman et al., 2014). Economic theory predicts that health insurance coverage reduces the out-of-pocket cost of care, leading to increased ED use. However, Medicaid may reduce ED use if people substitute primary care services for ED visits (Sommers & Simon, 2017). Empirically, the positive effect of Medicaid on ED use in Oregon contradicts most quasi-experimental studies that find a reduction in ED utilization following health insurance expansions (Chen et al., 2011; Chou et al., 2020; Giannouchos et al., 2022; Miller, 2012; Sommers et al., 2016). Recently, Kowalski (2023) attempts to reconcile the contradictory findings in the Oregon experiment and the Massachusetts reform. Nonetheless, the effect of Medicaid on ED use, especially for nonurgent conditions, remains a crucial consideration in health insurance expansions because of the continued rise in ED visits, the declining number of EDs, and the effects of ED crowding on health (Moore & Liang, 2020; Sabbatini & Dugan, 2022; Tang et al., 2010; Woodworth, 2020).

This paper provides new insights into how Medicaid affects ED utilization when we go beyond the *average* effect. Using the Oregon experiment, we estimate the heterogeneous impacts of Medicaid on ED use and characterize those driving the positive average effects. In 2008, Oregon randomly assigned the opportunity to apply for spots in its Oregon Health Plan Standard (OHP Standard). Unlike OHP Plus, which serves Oregon's typical Medicaid population, OHP Standard was newly offered to a group of uninsured adults who were categorically ineligible for Medicaid under federal guidelines (Allen et al., 2013). On average, the Medicaid expansion

Received: April 30, 2024. **Revised:** March 30, 2025. **Accepted:** June 11, 2025
© The Royal Statistical Society 2025.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

increased utilization for most types of ED visits, including those not requiring immediate care, for up to 2 years (Finkelstein et al., 2016; Taubman et al., 2014). We do not revisit the average impacts of Medicaid, which are well-documented in this literature. Instead, we extend the literature by investigating how Medicaid's impact on ED use varies by covariates, which helps to isolate the risk factors that matter for increased utilization.

This paper makes three salient contributions. First, using causal machine learning methods, we nonparametrically estimate the heterogeneous treatment effects of Medicaid on ED utilization. Using generalized random forests (GRF), we estimate various types of heterogeneous effects, including conditional treatment effects predicted using each individual's covariate values and higher-level group average treatment effects (ATEs; Athey et al., 2019). Previous studies report a small number of subgroup effects for a subset of the Oregon experiment who completed a follow-up survey (Taubman et al., 2014). Subgroup analysis can miss interesting patterns of heterogeneity due to nonlinear interactions between multiple covariates (Angrist, 2004; Deaton, 2010 ; Heckman et al., 1997). In addition to the need to prespecify the covariates, subsample heterogeneity analysis is prone to multiple hypothesis testing problems (e.g. List et al., 2019; Young, 2019). Considering that Oregon expanded Medicaid to a previously uninsured population not typically covered by it, insights from examining its heterogeneous effects can help interpret its findings and inform policy-making.

Second, we document the risk factors for increasing ED use upon Medicaid coverage, opening the doorway for identifying priority groups for targeted policy interventions. Unlike Kowalski (2023), we do not attempt to reconcile Oregon's ED finding with the Massachusetts health insurance reform results. In particular, Kowalski attributes the discrepancy to different *local* ATEs (LATEs) arising from the same marginal treatment effect (MTE) function. In doing so, Kowalski estimates an MTE function for Oregon, assuming that treatment effects vary linearly with the fraction treated. The MTE function shows that treatment effects are positive for compliers but negative for never-takers. Kowalski (2023) then extrapolates the estimated MTE function to Massachusetts, finding that ED utilization is predicted to decrease for Massachusetts compliers comparable to a subset of Oregon never-takers. Relatedly, Marx (2024) shows that a different set of distribution assumptions within the MTE framework predicts a positive treatment effect for Oregon's never-takers. Nonetheless, a key contribution of Kowalski (2023) is showing that treatment effect heterogeneity is crucial for understanding Medicaid's impacts and transporting those results to new environments. Without making additional assumptions beyond those needed to recover the LATE, our paper complements the above studies by directly linking effect heterogeneity in Oregon to several observed characteristics in a data-driven manner. Doing so sheds light on the drivers of Medicaid's average impact on ED use, providing new insights into the complex relationship between insurance coverage and ED use in low-income populations.

Third, we discuss the potential to leverage the heterogeneous effects of Medicaid to propose decision rules to identify enrollees for targeted educational/outreach interventions. Reducing 'unnecessary' ED use (i.e. ED use for nonemergent conditions) is a high policy-making priority in Medicaid. This is partly because one-third of ED visits are preventable and it costs up to five times less to treat the same health problems in a doctor's office (Adams, 2013; Giannouchos et al., 2022). As such, policymakers are experimenting with various policies to curb unnecessary ED use among Medicaid recipients. For instance, as of 2020, about 14 states use blunt tools such as higher copayments to dissuade nonemergency ED use. Medicaid's heterogeneous effects may be essential for exploring targeted policies (e.g. outreach programmes) for reducing nonemergency ED use, but policymakers might need guidance on how to use such evidence. Therefore, we illustrate one way of using heterogeneous treatment effects to estimate decision rules based on policy learning algorithms (Athey & Wager, 2021; Kitagawa & Tetenov, 2018).

Several broad findings emerge from our study. First, we find substantial heterogeneity in the impact of Medicaid coverage on ED visits. The conditional treatment effects predicted for each individual span a wide range of negative and positive changes in ED utilization. In addition, the significant increases in overall ED use are concentrated in the right tail of the distribution. Specifically, for ED use on the extensive margin, a small group of people—about 14% of recipients—with significant increases in ED visits appear to drive the positive average impact of Medicaid. We find similar patterns of heterogeneity for ED use on the intensive margin. These findings persist for various types of ED visits. In some cases, we find that average null effects hide important countervailing forces—reductions in ED

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

use by some people offset increases in utilization by others. For instance, Taubman et al. (2014) find no average Medicaid coverage effects for conditions classified as 'emergent, not preventable' (i.e. illnesses that require immediate medical care and could not have been prevented). In this case, the conditional treatment effects show a large share of both significant increases and decreases in ED use, culminating in the average null effect.

Second, when we aggregate the individual-level conditional effects to higher levels, the heterogeneous effects of Medicaid yield insights into who drives the average impact. The main risk factors for increased ED use are gender, age, participation in other safety net programmes, and past ED use for conditions not requiring emergency care. We identify four subgroups estimated with statistically significant increases in ED use of at least twice the magnitude of the average effect. These groups are men, prior SNAP participants, younger adults under 50 years, and those with prelottery ED use classified as primary care treatable. We do not find significant group-level effects defined by other prelottery ED use variables or information collected during the lottery sign-up. Overall, the results suggest that the increases in ED use occur primarily for newly insured people with a history of using the ED for conditions that do not require it.

Finally, we illustrate one way to use the heterogeneous effects to estimate decision rules predicting those likely to increase nonemergent ED use. Based on a small set of covariates, the policy learning algorithm identifies two groups of people at risk of increased ED use. The first group consists of people with minimal prior outpatient ED use but who did so for nonemergent conditions. The second group is those with higher previous outpatient ED visits who rarely use the ED for conditions requiring inpatient care. Policymakers can use those assignment rules to prioritize outreach to new recipients at risk of unnecessary ED use. We caution that our policy learning exercise is illustrative, and policymakers should tailor policy-relevant decision rules to meet their objectives subject to legal and other practical constraints.

The rest of the paper is organized as follows. Section 2 describes the context of Oregon's health insurance experiment and the data. Section 3 presents the causal machine learning framework. Section 4 presents and discusses the results. Section 5 concludes. Additional results are collected in the attached appendices.

2 The Oregon health insurance experiment

2.1 Background

We provide a brief overview of the essential features of the Oregon Health Insurance Experiment. Previous studies contain detailed institutional information (Allen et al., 2013; Baicker et al., 2013; Finkelstein et al., 2012, 2016; Taubman et al., 2014). Oregon split its Medicaid programme into two due to a budget shortfall in 2003. The first programme (OHP Plus) served categorically eligible people under federal rules—low-income income children, pregnant women, blind or disabled people, and Temporary Assistance to Needy Families (TANF) recipients. The second programme (OHP Standard) served low-income and nondisabled adults not categorically eligible for the OHP Plus programme. Oregon's philosophy was to provide more people with coverage for fewer services rather than limiting participation. Thus, OHP Standard covered fewer services and came with higher premium and cost-sharing requirements, resulting in a decline in enrolment (Allen et al., 2013). By 2007, the resulting attrition in OHP Standard participation led to an accumulated budgetary surplus. Then, Oregon decided to expand its OHP Standard programme by offering about 10,000 spots using a lottery to solicit or invite applications.

Only the lottery winners could apply for Medicaid and the eligible applicants received coverage. The main eligibility criteria were as follows—being 19–64 years of age; being an Oregon resident who is not otherwise eligible for public insurance; being a US citizen or legal immigrant; being uninsured for the previous six months, and having income below the federal poverty level with assets not exceeding $2,000. Of the 89,824 individuals who signed up for the lottery, 35,169 won, and about 30% of the winners eventually enrolled in Medicaid.

When interpreting the Oregon experiment results, the low-income, uninsured population served by OHP Standard should be considered. Finkelstein et al. (2012) elaborates the difficulties with extrapolating Oregon's finding to other Medicaid expansions. The population served by OHP Standard is neither representative of the typical Medicaid population served by OHP Plus nor the low-income uninsured population in the United States. For example, compared to the low-

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

income US population, OHP Standard's target population has more whites (84%), fewer Blacks (2%), is older, and reports being in worse health (Allen et al., 2010). Also, people voluntarily signed up for the lottery. These issues do not threaten the internal validity of the Oregon experiment. However, the inherent sample selection on observed (and unobserved dimensions) suggests a crucial role for effect heterogeneity in understanding its findings.

2.2 Data

We use the publicly available data files from Taubman et al. (2014). The sample contains ED visit information for 24,646 individuals from 2007 to 2009. The prelottery period spans 1 January 2007 through 9 March 2008. The study period is 18 months from the earliest notification date on 10 March 2008 through 30 September 2009. Medicaid coverage is constructed from state administrative records and defined as any receipt during the study period.

We analyse fourteen ED visit measures. These variables measure overall ED use and three categories of ED visits. The first category groups ED use by hospital admission, including outpatient and inpatient ED visits. The second category groups ED visits by the time of occurrence and consists of those occurring during on-time (7 a.m. to 8 p.m. on Monday to Friday) and off-time hours (nights and weekends).

The final category groups ED visits by whether they required immediate care or not. We adopt the classification of ED visits based on the primary ICD-9 diagnosis code (Billings et al., 2000; Taubman et al., 2014). While we refer the reader to Taubman et al. (2014) for a detailed description, the algorithm assigns to each ED visit a probability that it is one of four types. The first type—emergent, nonpreventable—includes unpreventable illnesses requiring immediate ED care (e.g. heart attacks and nonspecific chest pain). The second type—emergent, preventable—includes ED visits that require immediate ED care but are avoidable (e.g. asthma attacks and urinary tract infections). The third type—primary care treatable—includes ED visits requiring immediate care but not through the ED (e.g. sprains, strains, and abdominal pain). The final type—nonemergent—contains ED visits that do not require immediate care (e.g. headaches and back problems). Each visit is assigned a probability of being in all four categories. The number of visits for each type is then obtained by summing the assigned probabilities across all visits (Taubman et al., 2014). We do not analyse unclassified visits that are not assigned to any of the above four categories.

We analyse both the binary and continuous versions of these types of ED visits. While Taubman et al. (2014) only analyses the continuous versions, we create their binary counterparts as follows. Since the number of ED visits for each type is computed as a sum of probabilities, we conservatively classify the individual as having only one type of visit for people with one ED visit, which is set equal to the visit type with the highest probability. For those with two ED visits, we classify the individual as having at most two types of visits, that is, visit types with the top two highest probabilities. We classify the individual as having at most three visit types for those with three ED visits, with the types being those with the top three highest probabilities. Finally, we classify individuals with four or more ED visits as having all visit types with nonzero assigned probabilities.

All of our analyses include the size of the individual's household to ensure unbiased estimation of treatment effects. Given our central focus on heterogeneous effects along observed dimensions, we include two additional groups of baseline variables which have not previously being used to study effect heterogeneity in the Oregon experiment studies—lottery list variables and household characteristics measured before the lottery. Henceforth, we refer to these additional variables as *heterogeneity variables*. We focus on these predetermined variables in exploring heterogeneity to ensure that they are not endogenous to Medicaid receipt.

The first group of heterogeneity variables comprises of eight lottery list variables constructed from the lottery sign-up sheet—age; sex; indicators for whether English is the preferred language for receiving materials; whether the individuals signed themselves up for the lottery; whether an individual provided a phone number on the sign-up form; whether the individuals listed their address as a P.O. Box; whether the individual signed up on the first day of the lottery, and the median household income in the applicant's zip code from the 2,000 decennial census. Compared with Taubman et al. (2014), we are only missing the last variable (the median household income in the applicant's zip code from the 2,000 decennial census) because it is not publicly available, but this does not affect the internal validity of the estimated treatment effects.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

The second group of heterogeneity variables includes prelottery ED use measures, participation in the Supplemental Nutrition and Assistance Program (SNAP), and receiving TANF in the prelottery period. We include twenty prelottery ED measures covering overall ED use, on-/off-time use, ED visits resulting in hospital admission, visits for each type described above, and visits for specific types of injuries and health conditions. We include this expanded list of prelottery ED visit information to capture historical emergency care demand. The prelottery SNAP/TANF participation variables, which proxy for experience with the social safety net system, come from state administrative data. Those variables include indicators of programme receipt and the total household benefit amounts received between 1 January 2007 and the individual's notification date.

In using the heterogeneity variables to explore variation in treatment effects, we would like to minimize differences between our analytic sample and Taubman et al.'s (2014) sample ($N = 24{,}646$). Across all our ED use outcomes and the heterogeneity variables, only 58 observations have nonmissing information. If we were to impose this restriction across all outcomes, we would have a sample of 24,588 observations, which is practically the same as Taubman et al.'s (2014) sample. To be more conservative and retain most of those 58 observations, we create analytic samples for each ED use outcome reported in Table 1. For each ED use outcome, we create a sample of individuals with nonmissing information on the outcome in question and all the heterogeneity variables. For instance, doing so for the 'any overall visit' outcome leads to a sample of 24,613 individuals (hereinafter the baseline sample), where we only lose 33 observations relative to Taubman et al.'s (2014) sample. We then repeat this exercise separately for all the remaining ED use outcomes. The analytic samples for other outcomes are slightly smaller, but the differences are negligible. Table C.4 reports the analytic sample sizes for all ED use outcomes, with the smallest sample being 24,588 individuals. Our results are unaffected by conditioning on a uniform sample of nonmissing observations across all outcomes (i.e. using the smallest sample of 24,588 individuals).

Table 1 presents the summary statistics for our baseline sample. The sample is 55% female with an average age of 40 years. About 54% of the sample received SNAP with an average SNAP benefit amount of 1,332 in the prelottery period. TANF receipt is much lower at 2% of the sample with an average benefit of 96. The sample averaged 0.77 ED visits in the prelottery period, with the total ED facility charges being 895 on average.

3 Methods

Our analysis of the impacts of Medicaid on ED use is twofold. First, we use GRF to estimate the heterogeneous effects of Medicaid on ED utilization (Athey et al., 2019). In the second part of our analysis, we combine the estimated heterogeneous effects with policy learning algorithms to estimate decision rules to achieve well-defined objectives (Athey & Wager, 2021; Kitagawa & Tetenov, 2018).

3.1 Notation, definition, and identification

In this section, we define our main treatment effect parameters in the potential outcomes framework, state assumptions, and discuss the identification of these parameters. Suppose our sample consists of the following measurements on each individual, (X, D, Y, Z), where $X \in \mathcal{X}$ is a set of observed (predetermined) characteristics, $D \in \{0, 1\}$ is a binary treatment variable, $Y \in \mathbb{R}$ is the outcome, and $Z \in \{0, 1\}$ is a binary instrumental variable (IV).

Let $Y(1)$ and $Y(0)$ denote the potential outcomes that would have been observed if treatment was externally set to $D = 1$ and $D = 0$, respectively. Further, let $D(1)$ and $D(0)$ be the potential treatment that would have been observed if the instrument takes on $Z = 1$ and $Z = 0$, respectively. Generally, we let $Y(Z, D)$ be the counterfactual outcome given the treatment D and the instrument Z, although the IV assumptions will rule out direct effects of Z on Y below. In this paper, Y represents various ED visit measures described in Section 2.2. D denotes Medicaid receipt, which is endogenous because the decision to receive Medicaid may be correlated with unobserved characteristics affecting insurance choice and health care utilization. The instrument, Z, indicates winning the randomized lottery, which gave the household the opportunity to apply for Medicaid.

We distinguish between two types of covariates, $X = (X_1, X_2)$, where X_1 refers to measured confounders between D and Y, and X_2 refers to all the heterogeneity variables that do not confound

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Table 1. Descriptive statistics

Variable	Mean	Min	p25	p50	p75	Max
Lottery list and baseline characteristics						
Gave phone number	87%					
English as preferred language	86%					
Female	55%					
Provided P.O. box address	3%					
Signed up self for lottery	90%					
Prelottery SNAP recipient	54%					
Prelottery TANF recipient	2%					
Week of lottery sign-up	58%	0	0	1.0	3.00	5
Age (years)	39.60	20	29	39.0	50.00	63
Prelottery SNAP benefit amount ($)	1331.94	0	0	522.5	2205.00	20745
Prelottery TANF benefit amount ($)	96.16	0	0	0.0	0.00	16031
Baseline prerandomization ED use						
Number of overall visits	0.77	0	0	0.0	1.00	17
Number of inpatient visits	0.09	0	0	0.0	0.00	6
Number of outpatient visits	0.69	0	0	0.0	1.00	16
Number of on-hours visits	0.45	0	0	0.0	0.00	13
Number of off-hours visits	0.33	0	0	0.0	0.00	10
Number of emergent, nonpreventable visits	0.16	0	0	0.0	0.00	9
Number of emergent, preventable visits	0.06	0	0	0.0	0.00	6
Number of primary care treatable visits	0.27	0	0	0.0	0.10	12
Number of nonemergent visits	0.16	0	0	0.0	0.00	12
Number ambulatory-care-sensitive visits	0.05	0	0	0.0	0.00	5
Number of visits (chronic conditions)	0.14	0	0	0.0	0.00	9
Number of visits (injury)	0.17	0	0	0.0	0.00	6
Number of visits (skin conditions)	0.05	0	0	0.0	0.00	5
Number of visits (abdominal pain)	0.04	0	0	0.0	0.00	5
Number of visits (back pain)	0.03	0	0	0.0	0.00	5
Number of visits (chest pain)	0.02	0	0	0.0	0.00	3
Number of visits (headache)	0.02	0	0	0.0	0.00	4
Number of visits (mood disorders)	0.02	0	0	0.0	0.00	5
Number of visits (psychiatric conditions)	0.06	0	0	0.0	0.00	6
Sum of total ED charges	894.85	0	0	0.0	542.47	42315
ED use outcomes						
Any overall visit	34%					
Any inpatient visit	7%					
Any outpatient visit	32%					
Any emergent, nonpreventable visit	16%					
Any emergent, preventable visit	10%					
Any primary care treatable visit	22%					
Any nonemergent visit	16%					
Number of overall visits	1.00	0	0	0.0	1.00	22
Number of inpatient visits	0.11	0	0	0.0	0.00	7

(*continued*)

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Table 1. Continued

Variable	Mean	Min	p25	p50	p75	Max
Number of outpatient visits	0.89	0	0	0.0	1.00	21
Number of emergent, nonpreventable visits	0.21	0	0	0.0	0.00	14
Number of emergent, preventable visits	0.07	0	0	0.0	0.00	8
Number of primary care treatable visits	0.35	0	0	0.0	0.24	16
Number of nonemergent visits	0.20	0	0	0.0	0.00	13

Note. This table presents descriptive statistics based on the Oregon Health Insurance Experiment data. The sample consists of 24,646 individuals in the Taubman et al. (2014) sample. The prerandomization SNAP and TANF benefits are total household benefits received from January 2007 through the individual's lottery notification date. The prerandomization ED use measures are with respect to utilization between 1 January 2007 and 9 March 2008. SNAP = Supplemental Nutrition and Assistance Program; TANF = Temporary Assistance to Needy Families; ED = emergency department.

the relationship between D and Y. Although Oregon randomly selected people from a lottery list, everyone in the lottery winner's household is eligible to apply, making Medicaid receipt correlated with household size (Finkelstein et al., 2012; Taubman et al., 2014). As such, conditioning on household size, the only element of X_1, is required to identify Medicaid's treatment effects.

The heterogeneity variables vector X_2, discussed in Section 2.2, is not required for unbiased estimation of the treatment effects and we include them solely to explore effect heterogeneity. As shown in Taubman et al. (2014), on average, the variables in X_2 do not differ statistically between lottery winners and losers. As such, the inclusion of X_2 in X does not affect the validity of the identification assumptions below.

We now summarize the two broad types of treatment effect parameters in this paper. First, we estimate the *unconditional* intent-to-treat (ITT) effect of winning the lottery by comparing lottery winners and losers. The ITT represents the ATE of expanding the opportunity to apply for Medicaid, but not necessarily the effect of coverage. If there is no ambiguity, we would refer to the ITT estimates as ATE, denoted by $\tau_{\text{ATE}} = \mathbb{E}[Y(1) - Y(0)]$. To explore effect heterogeneity, we also estimate *conditional* ATE (CATE) estimates for subsets of the population defined by X. Broadly, CATE captures the heterogeneous effects of winning the lottery with respect to covariates, denoted by $\tau(x)_{\text{CATE}} = \mathbb{E}[Y(1) - Y(0) \mid X]$. We use CATE to capture the treatment effect in the smallest covariate partition defined by an individual's covariate values. We use Group ATE (GATE), denoted as $\tau(G)_{\text{GATE}} = \mathbb{E}[Y(1) - Y(0) \mid G]$, to represent ATEs for any subgroup G defined by low-dimensional functions of X (Lechner et al., 2022). We later use the ITT estimates for policy learning exercises exploring alternative decision rules, which may help to target programme recipients with educational interventions. Since the lottery Z is randomized, standard unconfoundedness assumptions identify the ITT treatment parameters and we omit a detailed discussion of their identification for brevity (Imbens & Rubin, 2015).

Second, beyond the ITT analysis, we estimate the effect of Medicaid coverage on ED utilization. IV methods are popularly used to estimate treatment effects when the treatment is not randomly assigned, but the researcher has access to a valid instrument. Under a new set of assumptions, Imbens and Angrist (1994) introduced the LATE, which is the ATE for compliers—those who got Medicaid because they won the lottery and would otherwise have remained uninsured. In the LATE framework, the existence of an instrument creates four latent groups: always-takers ($D(1) = D(0) = 1$), never-takers ($D(1) = D(0) = 0$), compliers ($D(1) = 1$ and $D(0) = 0$ or $D(1) > D(0)$), and defiers ($D(1) = 0$ and $D(0) = 1$ or $D(1) < D(0)$) (Angrist et al., 1996; Imbens & Angrist, 1994). In the absence of covariates, the celebrated LATE theorem in Imbens and Angrist (1994) states that the IV estimand identifies the unconditional LATE, defined as $\tau_{\text{LATE}} = \mathbb{E}[Y(1) - Y(0) \mid D(1) > D(0)]$. We also estimate the *conditional* version of LATE for subsets of the population defined by X, which we refer to as the Conditional LATE (CLATE), $\tau(x)_{\text{CLATE}} = \mathbb{E}[Y(1) - Y(0) \mid X, D(1) > D(0)]$. For compliers, CLATE captures the heterogeneous effects of Medicaid with respect to covariates. We use CLATE to capture the LATE in the smallest covariate partition defined by an individual's covariate values and Group

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

LATE (GLATE), denoted by $\tau(G)_{\text{GLATE}} = \mathbb{E}[Y(1) - Y(0) \mid G, D(1) > D(0)]$, to represent LATEs for a subgroup G defined by low-dimensional functions of X.

Following Abadie (2003), we now state assumptions and formally discuss the identification of LATE parameters using IV techniques.

Assumption 1 (Conditional Independence of the Instrumental Variable).

$$(Y(0, 0), Y(0, 1), Y(1, 0), Y(1, 1), D(0), D(1)) \perp Z|X.$$

Assumption 2 (Exclusion Restriction of the Instrumental Variable).

$$P[Y(1, D = d) = Y(0, D = d)|X] = 1 \quad \text{for} \quad d \in \{0, 1\}.$$

Assumption 3 (Relevance of the Instrumental Variable).

$$0 < P[Z = 1|X] < 1 \quad \text{and} \quad P[D(1) = 1|X] \neq P[D(0) = 1|X].$$

Assumption 4 (Monotonicity/No-Defiers).

$$P[D(1) \geq D(0)|X] = 1.$$

Assumption 1 states that, conditional on X, the instrument is essentially randomly assigned. By design, Assumption 1 is satisfied in our study because Z represents winning the lottery, which is random after conditioning on only household size in X. Assumption 2 states that the instrument has no direct effect on potential outcomes, which simplifies the potential outcomes to $Y(1) = Y(0, 1) = Y(1, 1)$ and $Y(0) = Y(0, 0) = Y(1, 0)$. Assumption 2 is plausible in our study because winning the lottery arguably only affects ED use through Medicaid coverage. Assumption 3 states that Z and D are correlated, conditional on X; this assumption has been verified in previous studies on the Oregon experiment (e.g. Finkelstein et al., 2012). Assumption 4 rules out defiers in the population, stating that if a lottery loser will enroll in Medicaid, they will also get Medicaid if they were to win the lottery. Assumption 4 is reasonable, especially given the fact that the lottery was based on a sign-up list. Given that Assumptions 1–4 are conditional versions of those in Imbens and Angrist (1994), the following lemma states that the conditional IV estimand, denoted by $\tau(x)_{IV}$, identifies the conditional ATE for compliers—CLATE.

Lemma 1 Under Assumptions 1–4,

$$\begin{aligned} \tau(x)_{IV} &= \frac{Cov(Y, Z \mid X = x)}{Cov(D, Z \mid X = x)} = \frac{\mathbb{E}[Y \mid X = x, Z = 1] - \mathbb{E}[Y \mid X = x, Z = 0]}{\mathbb{E}[D \mid X = x, Z = 1] - \mathbb{E}[D \mid X = x, Z = 0]} \\ &= \mathbb{E}[Y(1) - Y(0) \mid X, D(1) > D(0)] = \tau(x)_{\text{CLATE}}. \end{aligned}$$

We provide a proof of Lemma 1 in Appendix A.

3.2 Estimation and implementation

Following Imbens and Angrist (1994) and Angrist et al. (1996) , several papers have examined the identification of the LATE in the presence of covariates (Abadie, 2003; Angrist et al., 2000; Angrist & Imbens, 1995; Das, 2005; Hirano et al., 2000; Sloczyński et al., 2025; Yau & Little, 2001). In the presence of covariates, Frölich (2007) studies the nonparametric IV identification of the LATE, proposing nonparametric estimators that make no functional form assumptions about the outcome and treatment models. Abadie (2003) considers nonparametric models of the outcome and treatment with covariates, proposing parametric and semiparametric estimators of the LATE.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Despite the above nonparametric approaches, linear two-stage least squares (2SLS) specifications are typically used to estimate LATE in the presence of covariates. In the case of discrete covariates, where X contains only binary covariates, one such specification is the fully saturated/interacted specification of Angrist et al. (1996) , where the first-stage and reduced-form regressions uses a vector of instruments that fully interacts Z with X. Recent work shows that the Angrist et al. (1996) specification works because it satisfies the condition $\mathbb{E}[Z \mid X = x] = \mathbb{E}[ZX'][\mathbb{E}[XX']]^{-1}x$ (Blandhol et al., 2022; Kolesár, 2013). Linear IV specifications that do not satisfy this condition do not recover the LATE. Moreover, the above condition automatically holds when Z is randomly assigned, which is the case for the Oregon experiment (Blandhol et al., 2022).

Beyond linear 2SLS, partially linear IV models can be used to estimate the LATE/CLATE without additional assumptions beyond Assumptions 1–4 (Abadie, 2003; Blandhol et al., 2022). Partially linear models have a long history in statistics and econometrics [e.g. Robinson (1988)] and can be written as

$$Y = \beta(X) + \alpha(X)D + \varepsilon, \tag{1}$$

where D is a potentially endogenous treatment, Z is an instrument, and $\mathbb{E}[\varepsilon|X, Z] = 0$. The coefficient on D is given by unknown functions of X, $\alpha(X)$. We do not use equation (1) as a structural model of potential outcomes because doing so would restrict unobserved effect heterogeneity (Frölich, 2007). As such, although the partially linear IV model in equation (1) is additively separable in the error term, using it for estimating the LATE/CLATE does not restrict (observed or unobserved) effect heterogeneity because equation (1) makes no additional assumptions beyond those needed for the (conditional) IV estimand to identify the (conditional) LATE.

Importantly, Abadie (2003) proves that the function $\alpha(X)$ identifies the CLATE, $\tau(X)_{\text{CLATE}}$, defined in Lemma 1, allowing us to use equation (1) to conduct inference on the CLATE.

Lemma 2 (Abadie (2003) Proposition 5.3). Under Assumptions 1–4, the functions $\alpha(X)$ and $\beta(X)$ in equation (1) are equal to

$$\alpha(X) = \frac{\mathbb{E}[Y \mid X, Z = 1] - \mathbb{E}[Y \mid X, Z = 0]}{\mathbb{E}[D \mid X, Z = 1] - \mathbb{E}[D \mid X, Z = 0]} = \mathbb{E}[Y(1) - Y(0) \mid X, D(1) > D(0)], \tag{2}$$

$$\begin{aligned}\beta(X) &= \frac{\mathbb{E}[Y \mid X, Z = 0]\mathbb{E}[D \mid X, Z = 1] - \mathbb{E}[Y \mid X, Z = 1]\mathbb{E}[D \mid X, Z = 0]}{\mathbb{E}[D \mid X, Z = 1] - \mathbb{E}[D \mid X, Z = 0]} \\ &= \mathbb{E}[Y(0)|X] + [\mathbb{E}[Y(1) - Y(0) \mid X, D(0) = D(1) = 1] \\ &\quad - \mathbb{E}[Y(1) - Y(0) \mid X, D(1) > D(0)]] \times P(D(0) = D(1) = 1|X).\end{aligned} \tag{3}$$

3.2.1 *Estimation of conditional effects: CATE and CLATE*

To conduct inference on the CATE and CLATE, our estimation task reduces to nonparametrically estimating $\alpha(X)$ in equation (1), which we do using GRF (Athey et al., 2019). Due to our focus on IV forest estimation, we only briefly review CATE estimation using causal forests and refer the reader to previous studies for a more detailed exposition on CATE estimation using forests and related approaches (Caron et al., 2022; Knaus et al., 2021, 2022; Strittmatter, 2023).

To discuss how causal forest estimation works, consider the case of a completely randomized experiment, where we are interested in estimating the CATE for prespecified covariates. Wager and Athey (2018) proposed *causal forest* for this scenario. The basic building block of causal forests is a causal tree (Athey & Imbens, 2016). Causal trees recursively split the sample into small leaves, $L(x)$, for any given point x defined by the vector of covariates. For any leaf, we can then estimate the ATE as the difference in means between treatment and control units. The causal forest is then an average of several causal trees. Instead of viewing forests as an average of trees, Athey et al. (2019) propose GRF as an adaptive nearest-neighbour estimator for any quantity identified

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

by local moment conditions. In addition to the fact that GRF can be used to estimate causal forests for randomized experiments, Athey et al. (2019) specifically apply their GRF architecture to estimate $\alpha(X)$ in equation (1), which equals $\tau(X)_{\text{CLATE}}$ in Lemma 1.

For brevity, we outline the instrumental variable GRF estimation of CLATE, noting that the details for estimating CATE are similar when we set $D = Z$. To motivate the IV forest estimation, we note that $\alpha(X)$ and $\beta(x)$ are identified by the local (conditional) estimating equations:

$$\mathbb{E}[\psi_{\alpha(x),\,\beta(x)}(O_i) \mid X_i = x] = 0 \quad \text{for all } x \in \mathcal{X}, \tag{4}$$

where $O_i = (Y_i, D_i, Z_i)$ and $\psi_{\alpha(x),\,\beta(x)}(O_i) = (Y_i - \beta(x) - \alpha(x)D_i)(1\ Z_i')'$. Based on equation (4), the GRF estimates, $(\widehat{\alpha}(x),\ \widehat{\beta}(x))$, are given by

$$(\widehat{\alpha}(x),\ \widehat{\beta}(x)) \in \underset{\alpha,\,\beta}{\operatorname{argmin}} \left\| \sum_{i=1}^{n} \omega_i(x)\psi_{\alpha(x),\beta(x)}(O_i) \right\|_2, \tag{5}$$

or alternatively, $(\widehat{\alpha}(x),\ \widehat{\beta}(x))$ solve

$$\sum_{i=1}^{n} \omega_i(x)\psi_{\hat{\alpha}(x),\,\hat{\beta}(x)}(O_i) = 0, \tag{6}$$

where the GRF weight $\omega_i(x)$ captures the relevance of observation i for estimating $(\alpha(X), \beta(X))$ at x; that is, how often observation i appears in the same leaf as x across all fitted trees.

It is instructive to outline the algorithmic GRF estimation steps in the R package `grf` from CRAN, which we use in this paper. First, perform *local centring* by regressing out the effect of X_i from Y_i, D_i, and Z_i (Nie & Wager, 2021; Robinson, 1988). Second, conceptually, the centred outcomes are used to perform a forest-based weighted conditional 2SLS estimation, that is, solve equation (5). Mechanically, one of the key innovations in Athey and Wager (2019) is to propose forest-based weights, $\omega_i(x)$, by averaging neighbourhoods produced by different trees. Athey et al. (2019) develops an algorithm for growing GRF that produces the weights $\omega_i(x)$, shows that the resulting GRF estimates are consistent and asymptotically normal, and provides an estimator for the asymptotic variance.

The trees for the forests are grown based on two principles—recursive partitioning on subsamples of the training data and the 'honesty' principle, where a subset of the subsample is used to grow the tree and another subset is used to make predictions in the leaves of the tree. The recursive partitioning is done to maximize heterogeneity in the target $\alpha(x)$. We summarize the process for obtaining the forest-based weights in Appendix B.

We note some key implementation details and choices in this paper. We grow all forests using $B = 100{,}000$ trees. In steps (a) and (b) of Appendix B, we randomly sample half the data to build each tree and divide the data in equal parts, respectively. For each randomly drawn subsample, we must also randomly select some covariates to build each tree. We use the default rule-of-thumb to determine the number of variables considered at each split point. That is, if p is the number of covariates, then the number of variables used in each split is the minimum of $(\sqrt{p} + 20)$ and p.

For inference, the variance of $\widehat{\alpha}(X)$ is constructed using the bootstrap of little bags, which amounts to training trees in small groups and comparing $\widehat{\alpha}(X)$ predictions within and across those groups (Athey et al., 2019; Sexton & Laake, 2009).

To quantify which covariates are most important for driving treatment effect heterogeneity in growing the forest, we report a variable importance ranking of all covariates used in growing the forest. This variable importance measure is based on the number of times the variable is used for splitting, weighted by the tree depth of the split.

We provide two remarks on our choice of GRF for examining heterogeneous effects. First, in the case of experimental data, which is relevant for our ITT analysis, previous studies have examined the performance of competing machine learning methods for CATE estimation. Based on their

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

empirical Monte Carlo evidence, Knaus et al. (2021) find that four methods, including the GRF we use in this paper, perform relatively well across all data generation processes.

Second, for the IV analysis, one may use other nonparametric methods (e.g. Frölich, 2007) to estimate the CLATE. However, a substantive comparison between competing methods for CLATE estimation using IV methods in observational data is outside the scope of this paper. In addition, as discussed above, the interpretation of the linear IV estimand as LATE/CLATE in the presence of covariates is not straightforward, except in a few instances of discrete covariates (Blandhol et al., 2022; Kolesár, 2013). Despite these challenges with the comparability of CLATE estimators, we compare the overall forest-based LATE and GLATE estimates with their linear IV counterparts in Section 4.2.

3.2.2 Estimation of average effects: ATE/GATE and LATE/GLATE

In addition to the nonparametric estimates of $\alpha(X)$, we can use doubly robust methods to average $\widehat{\alpha}(X)$ across the distribution of X to get an overall ATE/LATE estimate or average effects for subgroups defined by low-dimensional functions of X (GATE/GLATE). This aggregation is based on augmented inverse propensity weighting (AIPW) and available in the R package `grf` (Athey & Wager, 2019; Tibshirani et al., 2018).

We discuss the doubly robust estimation of average effects for the ITT and IV analysis in turn. First, for the ITT analysis, the estimated ATE is obtained as $\widehat{\tau}_{\text{ATE}} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\Gamma}_i$, where $\widehat{\Gamma}_i$ are the doubly robust scores constructed as

$$
\begin{aligned}
\widehat{\Gamma}_i &= [\widehat{\mu}_1(x_i) - \widehat{\mu}_0(x_i)] + \left[\frac{D_i(Y_i - \widehat{\mu}_1(x_i))}{\hat{e}(x_i)}\right] - \left[\frac{(1-D_i)(Y_i - \widehat{\mu}_1(x_i))}{1-\hat{e}(x_i)}\right] \\
&= \widehat{\alpha}(x_i) + \frac{D_i - \hat{e}(x_i)}{\hat{e}(x_i)(1-\hat{e}(x_i))}[Y_i - \widehat{\mu}_d(x)] \\
&= \widehat{\alpha}(x_i) + \frac{D_i - \hat{e}(x_i)}{\hat{e}(x_i)(1-\hat{e}(x_i))}[Y_i - \widehat{\mu}(x) - (D_i - \hat{e}(x_i))\widehat{\alpha}(x_i)],
\end{aligned} \tag{7}
$$

where $\widehat{\mu}_d(x)$, $\widehat{\mu}(x)$, and $\widehat{e}(x)$ are estimates of $\mu_d(x) = \mathbb{E}[Y \mid D = d, X = x]$, $\mu(x) = \mathbb{E}[Y \mid X = x]$, and $e(x) = P[D = 1 \mid X = x]$. Also, $\widehat{\alpha}(x) = \widehat{\mu}_1(x) - \widehat{\mu}_0(x)$, which is the estimated CATE from equation (1) assuming D is unconfounded conditional on X. The first line in equation (7) reproduces the doubly robust scores from Chernozhukov et al. (2018), and the last equality follows from the Robinson (1988) transformation discussed in Nie and Wager (2021).

If G denotes membership in a group defined by a low-dimensional function of X (e.g. any subgroups in Table 4), then we estimate the GATE as $\widehat{\tau}(G)_{\text{GATE}} = \frac{1}{n_G}\sum_{i\in G}\widehat{\Gamma}_i$. For both ATE and GATE, the estimated variances are computed as $\frac{n}{n-1}\sum_{i=1}^{n}(\widehat{\Gamma}_i - \widehat{\tau}_{\text{ATE/GATE}})^2$.

Similarly, for the IV analysis, the estimated average CLATE (i.e. the estimated LATE) is obtained as $\widehat{\tau}_{\text{LATE}} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\lambda}_i$, where $\widehat{\lambda}_i$ are the doubly robust scores constructed as

$$
\widehat{\lambda}_i = \widehat{\alpha}(x_i) + \frac{1}{\widehat{\Delta}(x_i)}\frac{Z_i - \hat{z}(x_i)}{\hat{z}(x_i)(1-\hat{z}(x_i))}[Y_i - \widehat{\mu}(x) - (D_i - \hat{e}(x_i))\widehat{\alpha}(x_i)], \tag{8}
$$

where $\widehat{\alpha}(x)$ is the estimated CLATE from equation (1) using the instrument Z, $\widehat{\Delta}(x_i)$ is the estimated conditional ATE of Z on D, $P[D = 1 \mid Z = 1, X = x] - P[D = 0 \mid Z = 1, X = x]$ (i.e. the compliance score), $\hat{z}(x_i)$ is an estimate of $P[Z = 1 \mid X = x]$, with $\widehat{\mu}(x)$ and $\hat{e}(x_i)$ being defined as in equation (7) (Athey & Wager, 2021). Finally, for subgroups defined by G, we estimate the GLATE as $\widehat{\tau}(G)_{\text{GLATE}} = \frac{1}{n_G}\sum_{i\in G}\widehat{\lambda}_i$, with its variance estimated as discussed above.

3.3 From conditional treatment effect estimation to policy learning

Estimating heterogeneous treatment effects opens the doorway for performing policy learning exercises to understand alternative treatment assignment regimes (decision rules) and target programme recipients with educational interventions. We use the CATE estimates from the ITT analysis for the policy learning exercises because Oregon's lottery of the opportunity to apply

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

for Medicaid is a randomized encouragement design (Hirano et al., 2000), which aligns with the practical problem of allocating Medicaid spots in Oregon's expansion of OHP Standard to a new population. Moreover, as evidenced by the results reported in Appendix D, the CATE estimates qualitatively summarize the Medicaid coverage effects presented in Section 4.

We perform two policy learning exercises, drawing heavily on the literature on statistical decision rules that uses the distribution of treatment effects to design treatment assignment rules to maximize mean social welfare, typically defined as the mean posttreatment outcome (Athey & Wager, 2021; Dehejia, 2005; Hirano & Porter, 2009; Kitagawa & Tetenov, 2018; Manski, 2004; Stoye, 2009, 2012). In addition, some recent papers have attempted to distinguish between the goals of heterogeneous effects estimation and causal decision-making, providing useful avenues for future research (Fernández-Loría & Provost, 2022).

3.3.1 Who should be invited to apply for Medicaid?

Relative to the default random lottery assignment, the first exercise is a theoretical one that estimates an alternative decision rule for selecting whom to solicit Medicaid applications from to achieve a well-defined objective. This alternative decision rule is necessarily infeasible because state Medicaid administrators cannot use it in practice to solicit applications. However, we estimate the alternative decision rule to contrast those selected under it and the observed Medicaid recipients under the random assignment scheme in the data. The question we seek to answer is: how should a social planner optimally choose people to treat (i.e. solicit Medicaid applications) to minimize the posttreatment outcome defined as the number of nonemergent ED visits. Such visits do not require immediate care and are most likely deemed unnecessary (i.e. not requiring service in the ED) by state administrators.

Formally, our goal is to estimate a decision rule (policy), $\pi \in \Pi$, that maps an individual's observed covariates to a binary decision of treatment assignment. That is, $\pi : X_i \in \mathcal{X} \rightarrow \{0, 1\}$, where $\pi(X_i) = 1$ means the policy invites individual i to apply for Medicaid while those with $\pi(X_i) = 0$ will not be invited to apply for Medicaid. Technically, we are interested in learning policies that minimize regret relative to the best $\pi \in \Pi$, $R(\pi)$, defined as

$$R(\pi) = \underset{\pi' \in \Pi}{\operatorname{argmax}} \{V(\pi') : \pi' \in \Pi\} - V(\pi), \tag{9}$$

where $V(\pi) = \mathbb{E}[Y_i(\pi(X_i)) - Y_i(0)]$ is the average effect of implementing a policy relative to not treating anyone (Athey & Wager, 2021).

Kitagawa and Tetenov (2018) shows how to use inverse propensity weighting to estimate the policy $\hat{\pi}$ from (9). Recently, Athey and Wager (2021) propose writing the regret in equation (9) in terms of the CATE and estimating the decision rule by solving

$$\hat{\pi} = \underset{\pi \in \Pi}{\operatorname{argmax}} \left\{ \frac{1}{n} \sum_{i=1}^{n} (2\pi(X_i) - 1)\widehat{\Gamma}_i : \pi \in \Pi \right\}, \tag{10}$$

where $\widehat{\Gamma}_i$ are the AIPW scores from (7). As noted in Athey and Wager (2021), solving (10) is equivalent to solving a weighted classification problem of the form

$$\hat{\pi} = \underset{\pi \in \Pi}{\operatorname{argmax}} \left\{ \frac{1}{n} \sum_{i=1}^{n} |\widehat{\Gamma}_i| \operatorname{sign}(\widehat{\Gamma}_i)(2\pi(X_i) - 1) : \pi \in \Pi \right\}, \tag{11}$$

where the goal is to train a classifier $\pi(\cdot)$ to classify the sign of the AIPW score using the absolute value of the same score as sample weights (Zhao et al., 2012).

Using the AIPW scores $\widehat{\Gamma}_i$ from the R package `grf`, we use the R package `lpSolve` for integer programming to solve (11). This approach lets us incorporate capacity constraints, allowing us to restrict the selected sample to equal the number of lottery winners in the Oregon experiment. We then compare the population selected based on the optimal linear programme to those who won the Oregon lottery.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

3.3.2 Decision rules for targeting educational interventions

The second exercise is more practical and aims to identify enrollees that programme administrators deem at risk of unnecessary ED utilization. As mentioned above, states are constantly exploring policy options to curb the nonemergent use of EDs by Medicaid recipients. However, the decision rules estimated in the previous subsection are not suitable for decision-making because they represent a complex mapping from the set of covariates to a binary invite/do-not-invite decision. To address the goal of developing a simple decision rule to identifying recipients most likely to increase unnecessary ED visits, we use Athey & Wager's (2021) policy learning algorithm. Using few variables, this approach uses shallow decision trees to estimate the optimal decision rule. Since this approach restricts the complexity of the estimated rules, programme administrators can use them to target enrollees (Medicaid recipients) for specific educational interventions. Given that the goal of this exercise is to help target recipients for interventions, we do not impose the capacity constraints in estimating the shallow decision trees.

We choose which covariates to include in the decision rule estimation. While we use the complete set of covariates to estimate the CATE, we limit the variables used to estimate the decision rules to nine prelottery ED utilization variables. Specifically, we include the overall number of prelottery ED visits, the number of inpatient ED visits, the number of outpatient ED visits, the number of on-hour ED visits, and the number of off-hours ED visits. The remaining variables indicate the type of prelottery ED utilization—indicators of any prelottery emergent ED use, any prelottery primary care treatable ED use, any prelottery nonemergent ED use, and the prelottery sum of total ED facility charges. Finally, we use the R package `policytree` (Sverdrup et al., 2020) to estimate the decision rules as shallow decision trees of depth 2 because the algorithm obtains the decision rule by exact (exhaustive) tree search, which makes the required time for estimating more complex decision rules exponential in tree depth.

4 Results

4.1 Conditional LATEs of Medicaid on ED use

We first present the heterogeneous impacts of Medicaid using GRF as described in Section 3. Our heuristic approach for assessing effect heterogeneity combines economic/practical and statistical significance, where the former includes examining aggregated group effects for various subgroups and the latter tests whether the conditional effects are different from zero at the 10% level. However, while statistical significance is a helpful heuristic approach to visualizing conditional effects, they provide suggestive evidence of effect heterogeneity because these tests do not adequately account for uncertainty in the overall ATE estimates.

In this paper, we mainly focus on the extensive margin results of receiving Medicaid on binary indicators of ED use. We briefly refer to additional intensive margin results of Medicaid on continuous ED visit measures in Appendix C. Also, we provide the ITT heterogeneous effects of winning the lottery on ED utilization in Appendix D.

Figure 1 displays the conditional LATE estimates of Medicaid on the probability of any visit, with the darker blue shade indicating statistical significance at the 10% level. The subplots in Figure 1 display results for any overall visits (panel a), any outpatient visits (panel b), and any inpatient visits (panel c). These results come from $\tau(X)$ estimates at the smallest covariate partition, with each point X denoting each individual's covariate values. Table 2 complements the graphical results by reporting selected quantiles of the empirical distribution of the CLATE estimates for all outcomes. Table 2 also reports the average effect in the first column (discussed momentarily in Section 4.2) and the share of the heterogeneous effects that are nonpositive in the last column.

We find substantial and meaningful heterogeneous effects of Medicaid across all ED visit outcomes. Specifically, for any ED visits, the heterogeneous effects of Medicaid span a wide range of negative and positive values. Although the average effect of Medicaid turns to be positive, about one-third of the CLATE estimates are negative, with the 25th percentile being a reduction of 3.1% points. Moreover, there is considerable variation in the magnitude of the CLATE estimates when we focus on the positive heterogeneous effects. For instance, Table 2 shows that the median heterogeneous effect is a 7% points increase, which is larger than the mean effect of 4.5% points.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

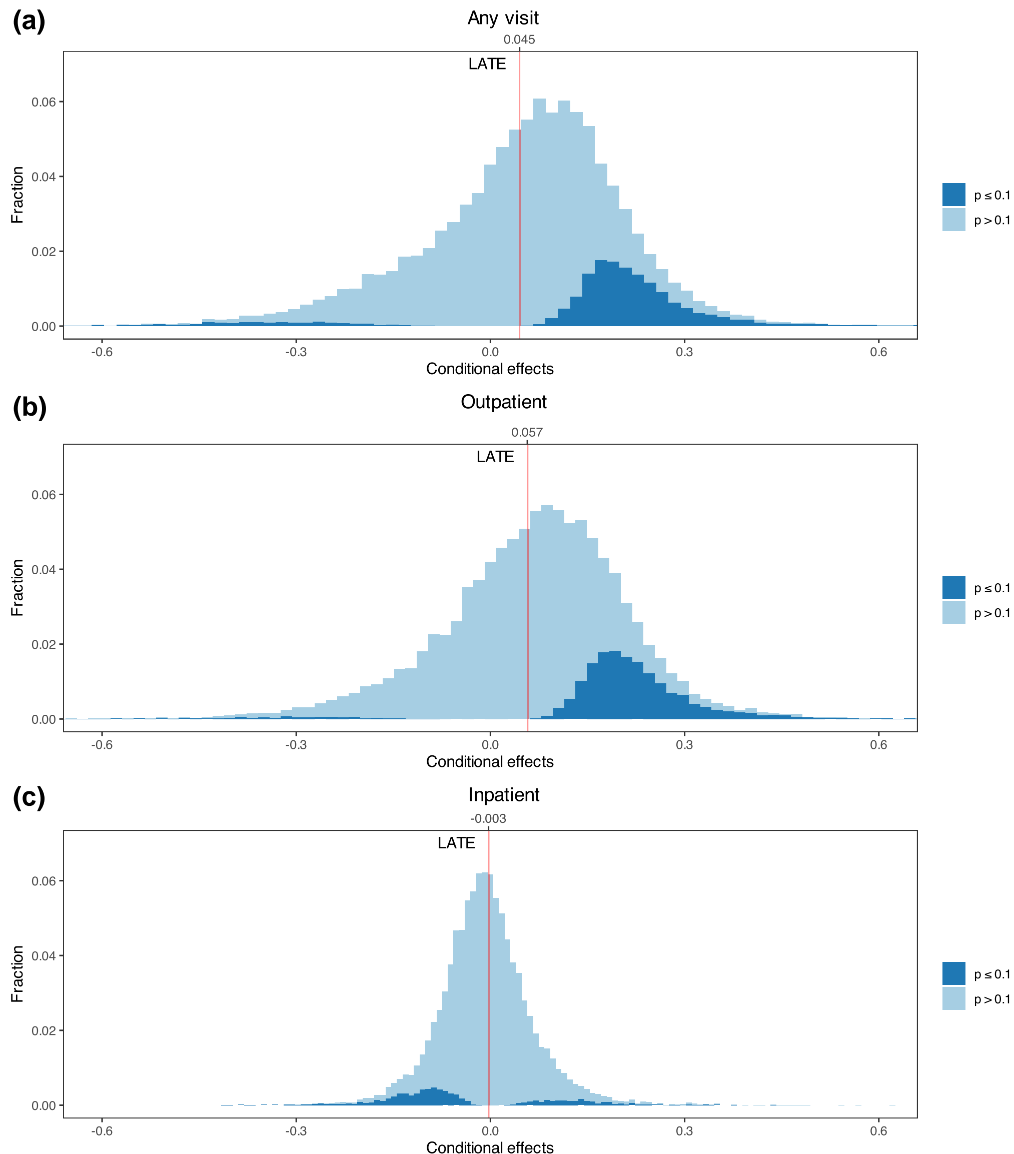

Figure 1. Conditional local average treatment effects (LATEs) of Medicaid on any emergency department (ED) visit. *Note.* This figure plots the conditional treatment effects of Medicaid on any overall ED visit (a), any outpatient ED visits (b), and any inpatient ED use (c) based on generalized random forests (GRF). The darker shade denotes statistical significance at the 10% level. The red vertical line indicates the LATE. The baseline sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery ED utilization and Supplemental Nutrition and Assistance Program (SNAP)/Temporary Assistance to Needy Families (TANF) receipt. The estimates displayed exclude less than half a percentile at the top and bottom of the distribution, resulting in the axes corresponding approximately to the percentile range [0.5%, 99.5%]. Bin size is chosen according to the Freedman–Diaconis rule.

In addition, the significant increases in ED use at the 10% level are limited to about 14.2% of the sample (with 8% being significant at the 5% level). Interestingly, there is no overlap between the significant CLATE estimates and the overall mean effect, further suggesting that the average effect masks heterogeneity. To put this nonoverlapping finding in context, at the 10% level, the smallest CLATE estimate among those who significantly increase ED usage is 4.9% points, about 10%

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Table 2. Empirical quantiles of the distribution of conditional LATEs of Medicaid on ED use

Variable	LATE	Min	25%	50%	75%	Max	Share nonpos.
Extensive margin							
Any overall visit	0.045	−0.815	−0.031	0.070	0.151	0.894	0.31
Any inpatient visit	−0.003	−0.415	−0.049	−0.010	0.031	0.619	0.57
Any outpatient visit	0.057	−0.879	−0.013	0.078	0.160	0.800	0.28
Any emergent, nonpreventable visit	−0.007	−0.622	−0.075	0.003	0.072	0.583	0.49
Any emergent, preventable visit	0.035	−0.313	−0.014	0.030	0.077	0.532	0.32
Any primary care treatable visit	0.048	−0.813	−0.011	0.067	0.141	0.785	0.28
Any nonemergent visit	0.073	−0.607	−0.004	0.060	0.132	0.604	0.27
Intensive margin							
Number of overall visits	0.354	−1.503	0.006	0.273	0.567	5.939	0.24
Number of inpatient visits	−0.012	−0.941	−0.074	−0.013	0.046	1.072	0.56
Number of outpatient visits	0.360	−1.199	0.052	0.291	0.571	5.828	0.20
Number of emergent, nonpreventable visits	0.028	−0.590	−0.066	0.020	0.109	3.025	0.44
Number of emergent, preventable visits	0.038	−1.533	−0.015	0.026	0.077	0.724	0.34
Number of primary care treatable visits	0.161	−0.550	0.018	0.131	0.255	3.689	0.22
Number of nonemergent visits	0.107	−0.663	−0.008	0.061	0.143	2.606	0.28

Note. This table reports selected quantiles of the conditional treatment effects of Medicaid on ED use based on GRF. The first column reports the average effect. The final column reports the share of the conditional effects that are nonpositive. The baseline sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery ED utilization and SNAP/TANF receipt. For overall and outpatient visits, one person with sparse features leading to an effect size below −1 was excluded. ED = emergency department; LATE = local average treatment effect; GRF = generalized random forests; SNAP = Supplemental Nutrition and Assistance Program; TANF = Temporary Assistance to Needy Families.

larger than the mean effect. While most of the significant CLATE estimates are positive, we find that about 2% of all conditional effects are statistically significant reductions in ED use.

An important distinction when examining ED visits is whether they are for outpatient or inpatient care since the latter signifies visits for conditions severe enough to warrant hospitalization. Regarding average effects, Taubman et al. (2014) find a significant effect for outpatient visits and null effects for inpatient ED use. Both types of ED visits exhibit substantial effect heterogeneity, but there are some crucial differences. A comparison of panels (a) and (b) in Figure 1 reveals that Medicaid's impacts on ED utilization are concentrated among outpatient visits. Moreover, the pattern of the distribution of conditional effect estimates for outpatient visits is similar to any overall ED visit, with negative and positive values. For instance, the 25th percentile for outpatient visits is a reduction of 1.3% points. However, the median conditional effect estimate of 7.8% points is larger than the mean effect of 5.7% points. Again, we find that the significant increases in outpatient ED use are concentrated in the right tail of the distribution. About 16% of all estimated outpatient effects are positive and significant, compared to 0.8% which are negative and significant.

For inpatient ED use, the distribution of the conditional treatment effects is relatively tighter than those for any ED or outpatient ED use. Notably, the distribution of the conditional effect estimates shows that the null effects of Medicaid on inpatient ED use mask an important heterogeneity. Figure 1c shows sizable proportions of significant negative and positive effects, with a slightly bigger mass for the negative conditional effects. Thus, Medicaid coverage appears to have significantly reduced inpatient ED use for a reasonable share of participants. In contrast to outpatient visits, significant and negative effects exceed significant and positive effects (6% vs. 2%). The

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

significant positive and negative heterogeneous effects appear to have counterbalanced and produced a null effect on average.

For the intensive margin, we obtain the same pattern of results for the number of ED visits. The corresponding graphs for the number of visits are shown in Figure C.1 with Table 2 showing selected quantiles of the estimated conditional effects distribution. Again, we find meaningful heterogeneity with conditional treatment effects ranging from a reduction of 1.50 to an increase of 5.94 ED visits. Unlike the extensive margin, the median of 0.27 is slightly smaller than the mean effect of 0.35 ED visits. However, we still find a concentration of significant increases in the number of ED visits in the distribution's right tail. On the intensive margin, we also see the similarities and differences mentioned above for the outpatient and inpatient care. First, the results for outpatient ED visits closely mirror the overall number of ED visits. The outpatient conditional effect estimates span a decrease of 1.20 to an increase of 5.83 ED visits, with the median and mean effects being 0.29 and 0.36, respectively. Second, the inpatient CLATE estimates are more tightly distributed. The distribution has substantially more mass to the left, outweighing the effect of high-use individuals in the right tail, which results in a null mean effect.

We turn our attention to the heterogeneous results by type of ED visit presented in Figure 2 for the extensive margin. Previous studies have highlighted the importance of looking beyond overall ED visits and categorizing them by medical urgency (Garthwaite et al., 2019; Giannouchos et al., 2022; Taubman et al., 2014). Consistent with the overall ED visit results, conditional treatment effects of Medicaid for all types of ED visits exhibit substantial heterogeneity. All ED visit types show a wide range of CLATE estimates, including negative and positive values. Moreover, the share of significant CLATE estimates differs across the ED visit types. For any emergent, nonpreventable visits, Figure 2a shows negative and positive significant effects of sizable proportions. These two opposing forces are consistent with the null average effect in Taubman et al. (2014). This finding accords with the large share of negative conditional effect estimates discussed above for inpatient ED care. Across all visit types, we find the largest share of significant reductions in ED uses for emergent, nonpreventable visits. A similar pattern holds for the number of emergent, nonpreventable visits, with the corresponding graph in Figure C.2.

Figure 2 shows some differences in the heterogeneous impacts for the remaining three types of ED visits. First, we find a larger share of positive and significant conditional effects for primary care treatable and nonemergent visits than emergent visits (i.e. nonpreventable and preventable) on both extensive and intensive margins. The two largest increases on the extensive and intensive margins occur for primary care treatable and nonemergent visits. However, the increase in primary care treatable visits is more pronounced on the intensive margin, whereas the increase in nonemergent visits is larger on the extensive margin.

In summarizing, the pattern of heterogeneous effects we uncover offers a more nuanced interpretation of Medicaid's effect on ED utilization that is not discernible by focusing on the mean effect alone. For instance, the large fraction of significant increases for primary care treatable and nonemergent visits suggests that gaining Medicaid coverage increased ED use in these categories for most people, not just on average. In contrast, we observe a sizable amount of negative conditional effects for the emergent, nonpreventable category. While the average effect is not statistically significant, this finding suggests that some people reduced ED usage by not seeking treatment for conditions that may require ED treatment.

It is also instructive to highlight the most important covariates driving these heterogeneous effects. Put differently, what are the most used individual characteristics for tree splitting in growing the causal forests? Table 1 lists the covariates utilized in growing the causal forests. Figure 3 displays the variable importance scores for the top 20 characteristics. We present the corresponding variable importance plot for the number of ED visits and the scores for all variables in Appendix C. For the extensive margin, Figure 3 shows that the causal forest splits the most on prelottery SNAP benefits followed by age, the sum of total ED facility charges, the number of prelottery emergent but nonpreventable ED visits, and the number of prelottery primary care treatable ED visits.

4.2 Aggregate effects of Medicaid on ED use

In this section, we present doubly robust estimates of the *average* Medicaid effect (i.e. the LATE). Although not the main focus of our analysis, a computationally cheap by-product of estimating

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

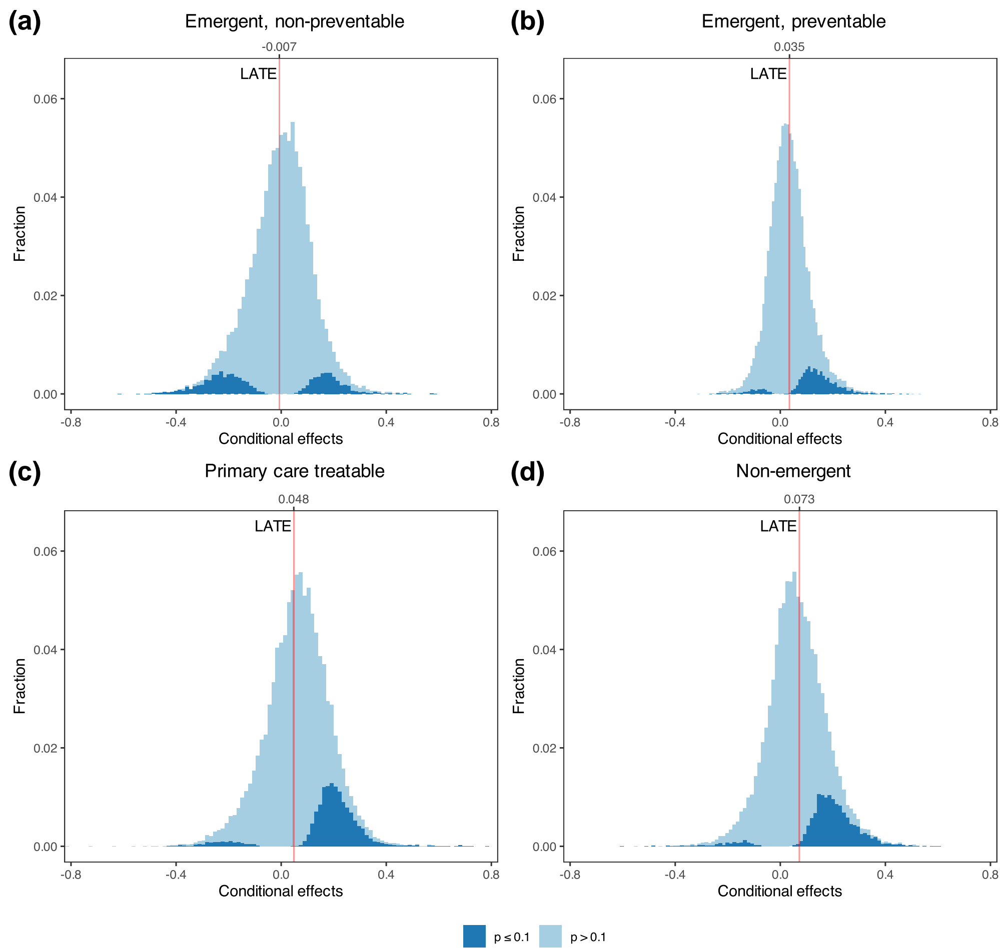

Figure 2. Conditional local average treatment effects (LATEs) of Medicaid on any emergency department (ED) visit by type of condition. *Note*. This figure plots the conditional treatment effects of Medicaid by type of ED visit based on generalized random forests (GRF) for any emergent, nonpreventable visit (a), any emergent, preventable visit (b), any primary care treatable visit (c), and any nonemergent visit (d). Measures of the type of ED visit are based on Billings et al.'s (2000) algorithm described in Taubman et al. (2014). We use these measures to construct binary indicators of ED visits by type of condition as described in the main text. The darker shade denotes statistical significance at the 10% level. The red vertical line indicates the LATE. The baseline sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery ED utilization and Supplemental Nutrition and Assistance Program (SNAP)/Temporary Assistance to Needy Families (TANF) receipt. The estimates displayed exclude less than half a percentile at the top and bottom of the distribution, resulting in the axes corresponding approximately to the percentile range [0.5%, 99.5%]. Bin size is chosen according to the Freedman–Diaconis rule.

heterogeneous effects is that we can aggregate them to produce the average effects (Section 3.2.2). Given that the standard linear IV estimator also recovers the LATE, we mainly discuss the forest-based LATE as a replication of Taubman et al. (2014) and show that the qualitative average findings are unchanged.

Table 3 presents the nonparametric and linear IV estimates of Medicaid's effect on all measures of ED utilization. The linear IV estimates confirm that we closely replicate Taubman et al.'s (2014) main findings, with minor differences in magnitude plausibly due to rounding and the top-coding in the public-use Oregon data files. The fact that we replicate their findings is not surprising given the negligible difference between our samples. Therefore, we are confident that our linear IV

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

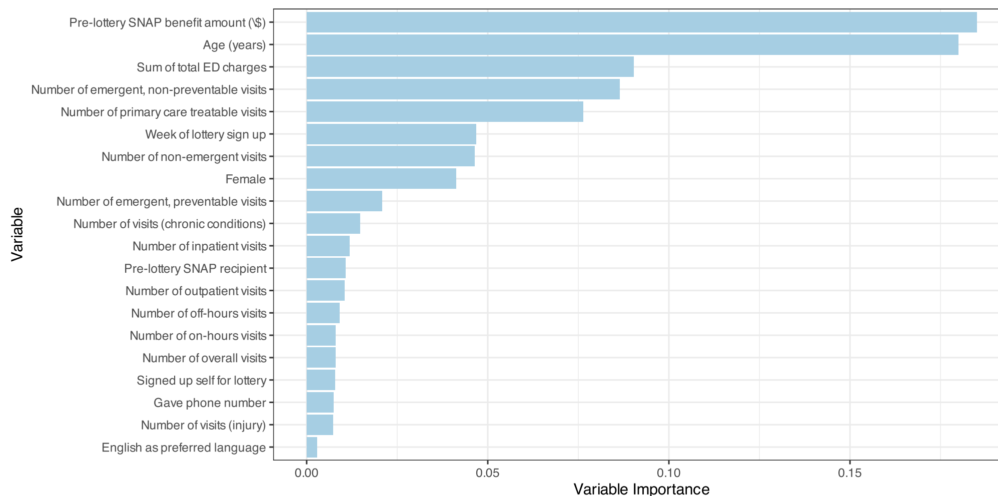

Figure 3. Variable importance scores in growing the causal forest (Any Visit). *Note.* This figure shows the variable importance scores of the top 20 characteristics in growing the generalized random forests (GRF) used to estimate the conditional treatment effects of Medicaid for any overall emergency department (ED) visit. The variable importance measure is a simple weighted sum of the proportion of times a variable is used in a tree splitting step at each depth in growing the forest. The scores roughly capture how important a variable is for driving treatment effect heterogeneity. The baseline sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery ED utilization and Supplemental Nutrition and Assistance Program (SNAP)/Temporary Assistance to Needy Families (TANF) receipt.

results are practically the same as those in Taubman et al. (2014) and serve as a useful baseline for comparing our nonparametric results.

Except for some differences in magnitudes in a few cases of the binary outcomes, the nonparametric LATE estimates are qualitatively the same as those based on the linear IV method. On the extensive margin, the nonparametric method appears slightly inefficient (higher standard errors), but there is no difference in significance between the estimates from the two methods. On the intensive margin, the nonparametric estimates more closely align with the linear method. Overall, our nonparametric estimates of the average effect comport with previous studies that Medicaid increased ED use while allowing us to explore heterogeneity at a more granular level.

4.3 Risk factors associated with changes in ED use

This section investigates the risk factors associated with significant changes in ED usage by examining treatment effects for selected subgroups. We summarize treatment effects for selected subgroups by aggregating the individual-level conditional treatment effects discussed above. Table 4 presents group LATEs from the causal forest and subgroup analysis based on the linear IV method for comparison.

There are three notable findings. First, using the causal forests, we find significant GATE estimates for four subgroups—men, prior SNAP participants, young and middle-aged adults below age 50, and individuals with any previous ED use classified as primary care treatable. These groups exhibit treatment effects between 8% and 10% points, at least twice the average effect. The forests do not reveal statistically significant group effects for other prelottery ED utilization or lottery list variables.

Figure 4 contrasts the distributions for the four main groups identified by the causal forest. The graphs highlight the differences in the empirical effect distribution that translate into the average group effects. Figure 4a shows a noticeable difference in the empirical distribution of the conditional treatment effects by prelottery SNAP receipt. The distribution of the CLATE estimates for SNAP recipients is narrower and exhibits a denser mass in the positive effects region. This

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Table 3. LATE estimates of Medicaid coverage on ED use

	GRF estimates			Linear estimates (Taubman et al. (2014) Replication)		
Variable	LATE	SE	*p*-value	LATE	SE	*p*-value
Extensive margin						
Any overall visit	0.045	0.028	0.111	0.069	0.024	0.004
Any inpatient visit	−0.003	0.015	0.855	−0.011	0.013	0.424
Any outpatient visit	0.057	0.028	0.037	0.082	0.024	0.001
Any emergent, nonpreventable visit	−0.007	0.021	0.743	0.007	0.019	0.731
Any emergent, preventable visit	0.035	0.016	0.029	0.039	0.015	0.009
Any primary care treatable visit	0.048	0.024	0.040	0.069	0.021	0.001
Any nonemergent visit	0.073	0.021	0.000	0.064	0.019	0.001
Intensive margin						
Number of overall visits	0.354	0.114	0.002	0.375	0.106	0.000
Number of inpatient visits	−0.012	0.028	0.676	−0.017	0.026	0.516
Number of outpatient visits	0.360	0.106	0.001	0.387	0.098	0.000
Number of emergent, nonpreventable visits	0.028	0.034	0.414	0.040	0.032	0.223
Number of emergent, preventable visits	0.038	0.018	0.039	0.036	0.017	0.034
Number of primary care treatable visits	0.161	0.049	0.001	0.171	0.045	0.000
Number of nonemergent visits	0.107	0.036	0.003	0.105	0.033	0.002

Note. This table reports the estimates of Medicaid coverage on ED use based on GRF and a linear IV model. The sample consists of 24,613 individuals in Taubman et al. (2014) sample with nonmissing information on prelottery ED utilization and SNAP/TANF receipt. ED = emergency department; GRF = generalized random forests; LATE = local average treatment effect; SNAP = Supplemental Nutrition and Assistance Program; TANF = Temporary Assistance to Needy Families.

difference in the distribution of the heterogeneous effects reflects in the aggregated group average effect. Previous SNAP recipients are estimated to increase ED use by 9% points, while there is no effect for SNAP nonrecipients.

Interestingly, Figure 4b shows that the CLATE distribution by prelottery primary care treatable ED use exhibits a similar pattern to prior SNAP receipt, but with a bigger group effect for those with any prior visit of 13% points. Finally, Figure 4c,d display the heterogeneous effects for age (age <50 vs. age ≥ 50 years) and gender (male vs. female). In contrast to the previous groups, the distributions for partitioning these groups are similar in shape but shifted to the right for the younger age group and men.

Second, the only subgroups with sizable negative effects are older adults (aged 50 and above) and those with any prelottery emergent, preventable ED visits. The effect for older adults is a 6% point reduction in the probability of using the ED. However, the estimate is insignificant (p-value = 0.24), possibly due to a lack of power. Further analyses in this direction suggest that those who reduce inpatient visits are mostly older people—individuals in the left tail of Figures 1c and 2a are older.

Third, we find important differences between the GLATE estimates based on the linear model and causal forests. While the linear method also uncovers heterogeneity based on the four main groups identified by the forest, the former yields a slightly bigger estimate for men (12% points), a slightly smaller estimate for prior SNAP recipients (8% points), a somewhat bigger estimate for people in the younger age group (10% points), and an almost identical group effect for those with prior primary care treatable ED use (12% points).

Contrary to the causal forest, the linear method estimates significant treatment effects for all binary indicators of prior ED use and those who did not receive TANF in the prelottery period. The same finding holds for some lottery list variables (e.g. households who provide a phone

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Table 4. Group LATEs of Medicaid on the propensity of ED use

Group	GRF estimates			Linear estimates			
	GLATE	SE	*p*-value	GLATE	SE	*p*-value	N%
Female:	0.00	0.04	0.97	0.03	0.04	0.34	0.55
Male:	0.10	0.04	0.01	0.12	0.03	0.00	0.45
Gave phone number: No	0.00	0.08	0.99	0.03	0.08	0.66	0.13
Gave phone number: Yes	0.05	0.03	0.11	0.08	0.03	0.00	0.87
English as preferred language: No	0.01	0.07	0.88	0.02	0.06	0.73	0.14
English as preferred language: Yes	0.05	0.03	0.12	0.08	0.03	0.01	0.86
First week sign-up: No	0.04	0.04	0.31	0.09	0.03	0.01	0.62
First week sign-up: Yes	0.06	0.04	0.20	0.06	0.04	0.16	0.38
Prelottery SNAP recipient: No	−0.01	0.05	0.79	0.02	0.05	0.62	0.46
Prelottery SNAP recipient: Yes	0.09	0.03	0.01	0.08	0.03	0.01	0.54
Prelottery TANF recipient: No	0.04	0.03	0.17	0.07	0.03	0.00	0.98
Prelottery TANF recipient: Yes	0.05	0.18	0.79	0.08	0.37	0.82	0.02
Age ≥50: No	0.08	0.03	0.02	0.10	0.03	0.00	0.75
Age ≥50: Yes	−0.06	0.05	0.24	0.01	0.04	0.77	0.25
Two+ household members on lottery list: No	0.03	0.03	0.42	0.05	0.03	0.06	0.80
Two+ household members on lottery list: Yes	0.09	0.06	0.15	0.19	0.07	0.00	0.20
Any prelottery ED visit No	0.03	0.03	0.34	0.07	0.03	0.02	0.69
Any prelottery ED visit: Yes	0.06	0.05	0.24	0.07	0.04	0.08	0.31
Any prelottery on-hours ED visit: No	0.03	0.03	0.35	0.07	0.03	0.01	0.77
Any prelottery on-hours ED visit: Yes	0.07	0.05	0.18	0.08	0.05	0.12	0.23
Any prelottery off-hours ED visit: No	0.04	0.03	0.21	0.06	0.03	0.03	0.81
Any prelottery off-hours ED visit: Yes	0.06	0.06	0.29	0.07	0.05	0.16	0.19
Any prelottery emergent, nonpreventable ED visit: No	0.04	0.03	0.19	0.07	0.03	0.01	0.87
Any prelottery emergent, nonpreventable ED visit: Yes	0.06	0.07	0.40	0.07	0.06	0.28	0.13
Any prelottery emergent, preventable ED visit: No	0.05	0.03	0.10	0.09	0.03	0.00	0.92
Any prelottery emergent, preventable ED visit: Yes	−0.05	0.08	0.55	0.02	0.07	0.83	0.08
Any prelottery primary care treatable ED visit: No	0.02	0.03	0.59	0.06	0.03	0.02	0.81
Any prelottery primary care treatable ED visit: Yes	0.13	0.06	0.03	0.12	0.05	0.03	0.19
Any prelottery nonemergent ED visit: No	0.03	0.03	0.34	0.07	0.03	0.01	0.86
Any prelottery nonemergent ED visit: Yes	0.08	0.07	0.21	0.09	0.06	0.17	0.14

Note. This table reports the group LATEs of Medicaid based on GRF. The baseline sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery ED utilization and SNAP/TANF receipt. ED = emergency department; GLATE = group local average treatment effect; GRF = generalized random forests; SNAP = Supplemental Nutrition and Assistance Program; TANF = Temporary Assistance to Needy Families.

number on lottery sign-up, households who requested English language materials, and those who did not sign-up in the first week). A caveat to this comparison is that the linear IV GLATE estimates were obtained by subgroup analysis. As such, these comparisons should be interpreted with caution.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

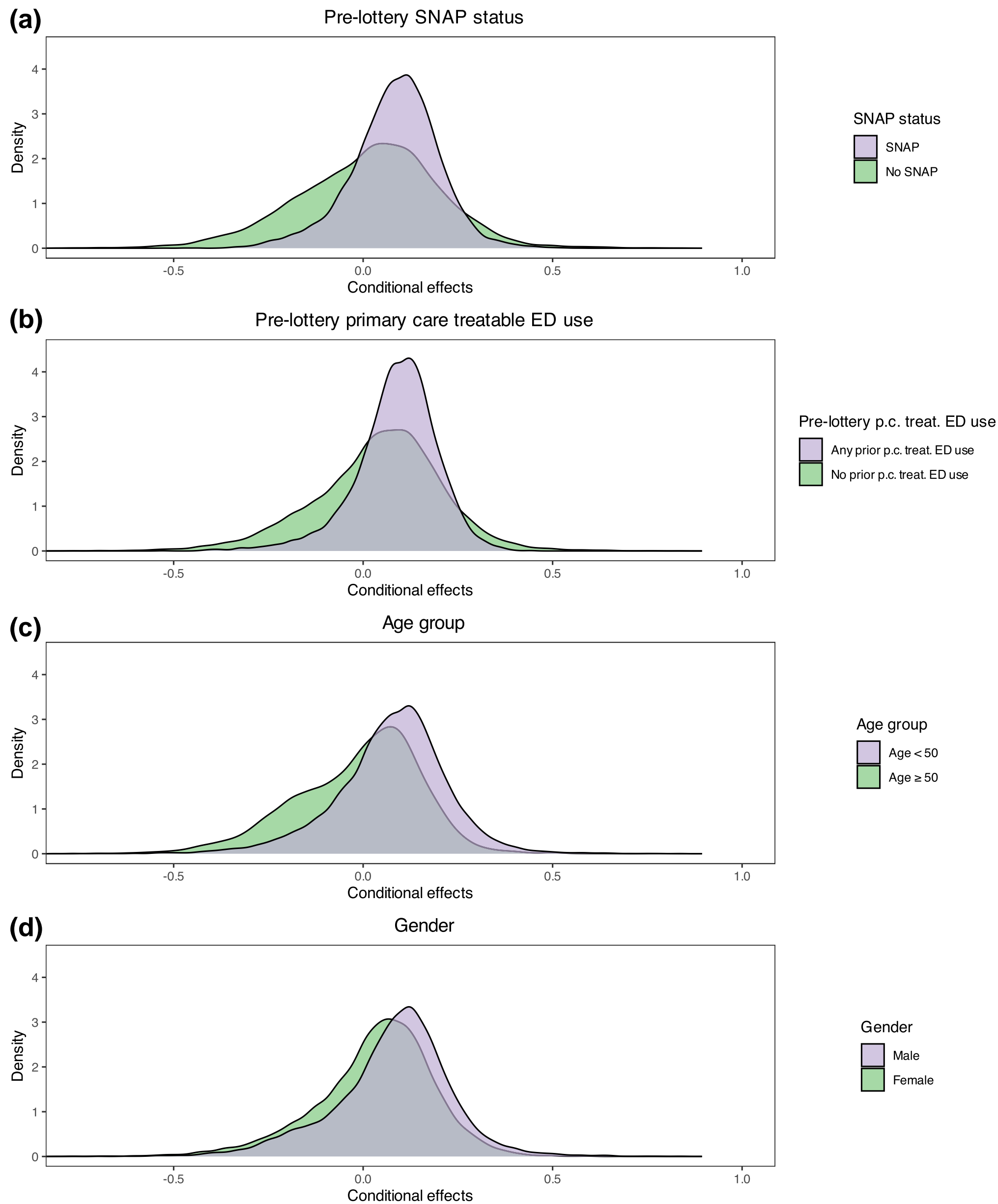

Figure 4. Conditional local average treatment effects (LATEs) of Medicaid by selected groups (Any visit). *Note*. This figure plots the conditional treatment effects of Medicaid on any overall emergency department (ED) visit for the four major groups identified with substantial group average effects—prelottery SNAP receipt (a), prelottery primary care treatable ED visit (b), age group (c), and gender (panel d). The baseline sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery ED utilization and Supplemental Nutrition and Assistance Program (SNAP)/Temporary Assistance to Needy Families (TANF) receipt.

Finally, we divide the sample based on the conditional effects to contrast the characteristics of those who increase and decrease usage ex-post. Table 5 presents the average characteristics of individuals identified to increase and reduce ED utilization (extensive margin) and their standardized mean differences. The corresponding summary of characteristics for the intensive margin of ED use are available in Appendix C. In almost all cases, we find that those who increase and

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Table 5. Characteristics of individuals who increased and decreased ED use

Variable	Increased ED use (1)	Decreased ED use (2)	Difference (std.) (2)–(1)
Lottery list characteristics			
Age (years)	38.76	41.48	−0.22
Gave phone number	0.87	0.87	0.00
English as preferred language	0.87	0.85	0.06
Female	0.51	0.63	−0.24
Week of lottery sign-up	1.57	1.61	−0.02
Provided P.O. box address	0.02	0.03	−0.06
Signed up self for lottery	0.89	0.92	−0.10
Prelottery SNAP recipient	0.62	0.35	0.56
Prelottery SNAP benefit amount ($)	1607.33	713.38	0.52
Prelottery TANF recipient	0.03	0.01	0.14
Prelottery TANF benefit amount ($)	111.08	63.09	0.07
Prelottery ED usage			
Number of overall visits	0.88	0.49	0.21
Number of inpatient visits	0.09	0.07	0.05
Number of outpatient visits	0.78	0.42	0.22
Number of on-hours visits	0.51	0.30	0.18
Number of off-hours visits	0.37	0.20	0.19
Number of emergent, nonpreventable visits	0.18	0.09	0.19
Number of emergent, preventable visits	0.07	0.04	0.11
Number of primary care treatable visits	0.30	0.17	0.19
Number of nonemergent visits	0.18	0.10	0.15
Number ambulatory-care-sensitive visits	0.05	0.04	0.03
Number of visits (chronic conditions)	0.14	0.11	0.05
Number of visits (injury)	0.20	0.09	0.21
Number of visits (skin conditions)	0.05	0.03	0.07
Number of visits (abdominal pain)	0.04	0.02	0.08
Number of visits (back pain)	0.04	0.02	0.08
Number of visits (chest pain)	0.02	0.02	0.00
Number of visits (headache)	0.03	0.02	0.05
Number of visits (mood disorders)	0.02	0.02	0.00
Number of visits (psychiatric conditions)	0.06	0.05	0.03
Sum of total ED charges	1013.28	584.41	0.17
N	17,011	7,602	24,613

Note. This table reports the means of individual characteristics and prerandomization ED use for those estimated to increase and decrease ED use upon receiving Medicaid coverage based on the GRF conditional LATE estimates. The sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery ED utilization and SNAP/TANF receipt. ED = emergency department; SNAP = Supplemental Nutrition and Assistance Program; TANF = Temporary Assistance to Needy Families; GRF = generalized random forests.

decrease are different on observed dimensions. To highlight a few key differences, those estimated to increase the likelihood of any ED use are almost 3 years younger (39.8 vs. 41.5 years), less likely to be female (0.51 vs. 0.63), more likely to have previously received SNAP (0.62 vs. 0.35) with

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

higher total household benefits averaging (\$1,607 vs. \$717), and more likely to have been a previous TANF recipient (0.03 vs. 0.01) with higher benefits (\$111 vs. \$63). In terms of prior ED visits, those estimated to increase use have a higher number of visits across all visits. The largest mean differences are for the overall number of ED visits and outpatient ED visits.

4.4 Results of policy learning exercises

This section presents the results of the policy learning exercises described in Section 3.3. Until now, we have focused on the effects of Medicaid coverage (i.e. the conditional LATE). However, the policy learning exercises are based on the ITT effects of winning the lottery, which are reported in Appendix D. The heterogeneous ITT effects follow a similar pattern as the IV coverage effects of Medicaid but the former is more natural for the policy learning exercises because it is the lottery and not the actual coverage that is random.

4.4.1 Comparing the observed lottery to an alternative assignment scheme

In this theoretical exercise, our objective is to estimate an alternative allocation scheme that prioritizes those least likely to increase unnecessary ED use, defined as the number of nonemergent ED visits. Doing so allows us to describe how the sample selected by such an optimal allocation scheme would compare to the sample of observed lottery winners. Again, while this exercise is infeasible from the point of policy-making, we use it to illustrate the value of the heterogeneous effects and verify that, given a well-defined objective, a random assignment scheme may be inferior to one based on the empirical distribution of treatment effects.

As expected in Table 6, the selected sample based on the optimal allocation scheme has substantially lower ex-post ED utilization for nonemergent causes (0.18 fewer visits). Other ED use measures following treatment are also substantially lower. The same finding holds for prelottery ED use, suggesting that ED use patterns are complex and treatment effects are driven by a population that uses the ED for most of their health care needs (urgent or nonurgent).

4.4.2 Identifying people at risk of increased nonemergent ED use via decision rules

A more practical way to use the estimated heterogeneous effects for policy-making is to design simple decision rules identifying people at greater risk of increasing nonemergent ED use among Medicaid enrollees. Here, the administrator may still use a lottery to solicit Medicaid applications but target a subset of new enrollees with an outreach or educational intervention. For instance, policymakers may want to identify some enrollees who may be required to establish primary care upon receiving Medicaid coverage. Of course, a straightforward decision rule might be to target the intervention based on their predicted conditional treatment effects, such as those with $\hat{\tau}(X) \geq 0$. However, Athey and Imbens (2019) points out that such decision rules are not always optimal in the sense of minimizing the loss from not using the ideal assignment rule or policy.

Thus, we estimate the depth-2 decision trees to identify those most likely to increase nonemergent ED visits using only a subset of prelottery ED use variables (Athey & Wager, 2021). Figure 5 presents the estimated decision rule. The first node of the decision tree partitions the sample based on a threshold of two prior outpatient ED visits. On the one hand, for people with two or fewer outpatient ED visits, the decision rule selects them if they had no previous ED use classified as emergent. On the other hand, for those with more than two previous outpatient ED visits (potentially excessive ED use), those with at most one prior ED visit resulting in hospitalization are selected.

We mention two caveats on the interpretation of the shallow decision tree rules. First, we do not use a hold-out sample to evaluate the estimated rules in this paper, which makes it difficult to access the gains from targeting. Second, we emphasize that our estimated decision rule may be impractical, especially if the previous ED utilization variables are not easily accessible. Policymakers could also specify a different objective function to align with their goals or accommodate additional constraints.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Table 6. Individuals selected by lottery and the optimal linear programme

Variable	Selected by lottery (1)	Selected by optimal linear programme (2)	Difference (std.) (2)–(1)
Lottery list and baseline characteristics			
Age (years)	39.50	40.16	0.05
Gave phone number	0.88	0.87	−0.03
English as preferred language	0.85	0.79	−0.16
Female	0.54	0.53	−0.02
Week of lottery sign-up	1.55	1.66	0.07
Provided P.O. box address	0.03	0.03	0.00
Signed up self for lottery	0.85	0.85	0.00
Prelottery SNAP recipient	0.53	0.43	−0.20
Prelottery SNAP benefit amount ($)	1376.53	1120.72	−0.13
Prelottery TANF recipient	0.02	0.02	0.00
Prelottery TANF benefit amount ($)	100.09	72.82	−0.04
Number of overall visits	0.71	0.44	−0.18
Prerandomization ED use			
Number of inpatient visits	0.08	0.05	−0.09
Number of outpatient visits	0.63	0.39	−0.17
Number of on-hours visits	0.41	0.25	−0.16
Number of off-hours visits	0.30	0.19	−0.15
Number of emergent, nonpreventable visits	0.15	0.09	−0.15
Number of emergent, preventable visits	0.05	0.03	−0.08
Number of primary care treatable visits	0.25	0.15	−0.17
Number of nonemergent visits	0.14	0.08	−0.14
Number ambulatory-care-sensitive visits	0.05	0.03	−0.08
Number of visits (chronic conditions)	0.12	0.08	−0.08
Number of visits (injury)	0.17	0.10	−0.14
Number of visits (skin conditions)	0.04	0.03	−0.04
Number of visits (abdominal pain)	0.03	0.02	−0.05
Number of visits (back pain)	0.03	0.02	−0.05
Number of visits (chest pain)	0.02	0.01	−0.06
Number of visits (headache)	0.02	0.01	−0.06
Number of visits (mood disorders)	0.02	0.01	−0.05
Number of visits (psychiatric conditions)	0.05	0.03	−0.06
Sum of total ED charges	808.36	504.40	−0.14
Number of overall visits	0.98	0.43	−0.30
ED use outcomes			
Number of inpatient visits	0.10	0.06	−0.10
Number of outpatient visits	0.87	0.36	−0.30
Number of emergent, nonpreventable visits	0.20	0.10	−0.20

(*continued*)

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Table 6. Continued

Variable	Selected by lottery (1)	Selected by optimal linear programme (2)	Difference (std.) (2)–(1)
Number of emergent, preventable visits	0.07	0.04	−0.10
Number of primary care treatable visits	0.35	0.17	−0.24
Number of nonemergent visits	0.20	0.02	−0.37
N	9,607	9,607	

Note. This table reports the means of individual characteristics, prerandomization ED use and posttreatment outcomes for those selected by the lottery and selected by the optimal treatment assignment solution based on the optimal linear programme, which minimizes nonemergent ED use under the restriction that the same number of people are treated as within the lottery. The sample consists of 24,605 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery ED utilization and SNAP/TANF receipt. SNAP = Supplemental Nutrition and Assistance Program; TANF = Temporary Assistance to Needy Families; ED = emergency department.

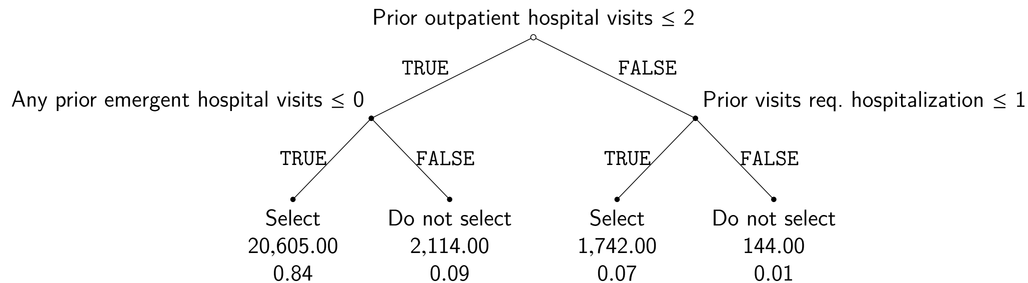

Figure 5. Estimated decision rule for targeting based on minimizing nonemergent ED use. *Note.* This figure shows the optimal policy assignment rule based on the doubly robust scores of the conditional average treatment effect (CATE) of being randomly selected as a lottery winner in the Oregon Health Insurance Experiment using the efficient policy learning framework of Athey and Wager (2021).

5 Conclusion

This paper estimates the heterogeneous impacts of Medicaid coverage on ED visits using records from 12 hospitals in Portland, Oregon, matched to the Oregon Health Insurance Experiment. The result that Medicaid increased ED utilization in the 2008 Oregon Health Experiment made headlines because of the hypothesis that insurance coverage should make it easier for recipients to access primary care and reduce the need to use the ED. However, this is an empirical question, given that insurance coverage also reduces out-of-pocket healthcare spending. We provide new insights into the ED results of the Oregon experiment when we go beyond the average impacts of Medicaid coverage. We do so by estimating the conditional treatment effects of Medicaid using nonparametric machine learning methods. We then leverage the heterogeneous effects to estimate decision rules to illustrate their usefulness for policymakers in similar settings.

We find substantial treatment effect heterogeneity in the impacts of Medicaid on ED utilization. The conditional treatment effects of Medicaid on ED use indicate positive and negative effects, suggesting a more nuanced interpretation of Medicaid's average impacts. We also find that coverage effects for different types of ED use exhibit meaningful heterogeneity. Due to the variation in the conditional effect distribution, the ATE appears to be driven by sections of the sample in the right tail of the distribution. A small proportion of high-use individuals drive the positive (average) Medicaid effect. Despite the positive average effects, the predicted treatment effects for many individuals are negative.

Our findings suggest that the average effect sometimes masks countervailing forces. In some instances, although many people increase ED usage due to obtaining coverage, a sizable proportion of people either decrease usage or do not respond. We find that reductions in ED use are

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

pronounced for inpatient visits and emergent, nonpreventable conditions. The overall null effects for these outcomes hide substantial opposing effects of obtaining coverage.

We also find substantial treatment effect heterogeneity on the intensive margin. Overall, the results suggest that increased ED usage is likely driven by intensive margin effects rather than extensive margin effects. Individuals who already utilize the ED further increase their usage after receiving Medicaid.

We also characterize the groups of people mainly driving the positive average Medicaid effects in the right tail of the distribution of the heterogeneous effects. Those with positive heterogeneous effects are predominantly younger, more likely to be men, more likely to receive SNAP and TANF in the prelottery period, and more likely to have higher baseline ED use. In particular, we identify four groups estimated to have statistically significant increases in ED use that are at least twice as large as the magnitude of the average effect. These groups are men, prior SNAP participants, younger adults under 50 years old, and people with prelottery ED use classified as primary care treatable.

Finally, we illustrate one way to use the estimated heterogeneous effects to estimate decision rules that may help to identify and target enrollees likely to increase nonemergent ED utilization. For people with minimal previous outpatient ED use, the decision rule identifies those with no previous emergent use as likely to increase nonemergent ED use. For those with higher previous outpatient ED use, the decision rule selects those whose ED visit resulted in hospitalization only once (i.e. those who have rarely presented with severe health problems justifying inpatient treatment). While our decision rules are illustrative, they highlight the potential for using statistical decision rules to guide policymakers to achieve context-specific objectives.

Acknowledgments

We thank Amy Finkelstein, Anirban Basu, Paul Goldsmith-Pinkham, Tymon Słoczyński, Timothy Layton, Jacob Wallace, Sherri Rose, Jason Fletcher, Carlos Lamarche, Mary Olson, Makayla Lavender, Prithvijit Mukherjee, Samuel Asare, Nicholas A. Wright, Anthony Strittmatter, Philip Marx, and seminar/conference participants at Tulane University, University of Washington, University of Kentucky, Institute for Research on Poverty (University of Wisconsin-Madison), Swiss Health Economics Workshop, and The American Society of Health Economists (ASHEcon) for helpful comments.

Conflicts of interest: The authors have no conflicts of interest to declare.

Data availability

This paper uses publicly available data from The Oregon Health Insurance Experiment, which can be accessed at www.nber.org/oregon. The authors' analysis code and data are attached as supporting materials.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series A*.

Appendix A. Proof of Lemma 1

To obtain the desired result, we begin by considering the expectation of $\mathbb{E}[Y_i \mid X_i = x, Z_i]$ at $Z_i = 1$ and $Z_i = 0$, given by

$$\begin{aligned}\mathbb{E}[Y_i \mid X_i = x, Z_i = 1] - \mathbb{E}[Y_i \mid X_i = x, Z_i = 0] &= \mathbb{E}[D_i(1)Y_i(1) + (1 - D_i(1))Y_i(0) \mid X_i = x, Z_i = 1] \\ &\quad - \mathbb{E}[D_i(0)Y_i(1) + (1 - D_i(0))Y_i(0) \mid X_i = x, Z_i = 0] \\ &= \mathbb{E}[(D_i(1) - D_i(0))(Y_i(1) - Y_i(0)) \mid X_i = x],\end{aligned}$$

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

where the first equality follows by Assumptions 1 and 2 and the second equality follows after some algebra. By Assumption 4, $D_i(1) - D_i(0) = \mathbb{I}\{D_i(1) > D_i(0)\} \in \{0, 1\}$, so we have that

$$\begin{aligned}\mathbb{E}[Y_i \mid X_i = x, Z_i = 1] - \mathbb{E}[Y_i \mid X_i = x, Z_i = 0] &= \mathbb{E}[(Y_i(1) - Y_i(0)) \mid D_i(1) > D_i(0), X_i = x] \\ &\quad \times P[D_i(1) > D_i(0) \mid X] \\ &= \mathbb{E}[(Y_i(1) - Y_i(0)) \mid D_i(1) > D_i(0), X_i = x] \\ &\quad \times \mathbb{E}[D_i(1) - D_i(0) \mid X_i = x] \\ &= \mathbb{E}[(Y_i(1) - Y_i(0)) \mid D_i(1) > D_i(0), X_i = x] \\ &\quad \times \mathbb{E}[D_i \mid Z_i = 1, X_i = x] - \mathbb{E}[D_i \mid Z_i = 0, X_i = x],\end{aligned}$$

where the last equality follows from Assumptions 1 and 3. After some rearrangement,

$$\begin{aligned}\mathbb{E}[(Y_i(1) - Y_i(0)) \mid D_i(1) > D_i(0), X_i = x] &= \frac{\mathbb{E}[Y_i \mid X_i = x, Z_i = 1] - \mathbb{E}[Y_i \mid X_i = x, Z_i = 0]}{\mathbb{E}[D_i \mid Z_i = 1, X_i = x] - \mathbb{E}[D_i \mid Z_i = 0, X_i = x]} \\ &= \frac{Cov(Y_i, Z_i \mid X_i = x)}{Cov(D, Z \mid X_i = x)},\end{aligned}$$

proving Lemma 1. The identification results for the group LATE and LATE follows immediately by an application of the Law of Iterated Expectations.

Appendix B. Estimating the forest-based weights using R package `grf`

Grow a set of B trees indexed by $b = 1, \ldots, B$ and use the following steps for each set of trees:

(a) From the training data, draw a random subsample of a chosen size without replacement.
(b) Randomly divide the selected subsample from step (a) in two equal nonoverlapping parts, labelled $\mathcal{I}_1$ and $\mathcal{I}_2$.
(c) Using subsample $\mathcal{I}_1$, greedily partition the covariate space to maximize the squared difference in subgroup treatment effects, which is an abstract version of the target criterion. Due to computational considerations, Athey et al. (2019) shows that this process formally reduces to splitting on a pseudo-outcome to optimize a gradient-based approximation to the target criterion. Let P denote a parent node, which we want to split into two child nodes, C_1 and C_2.
   (i) For each parent node, use all of its data to compute $\tilde{\alpha}_P$ and $\tilde{\beta}_P$ and an estimate of the gradient of the expectation of the ψ-function, denoted by A_p. $\tilde{\alpha}_P$ and $\tilde{\beta}_P$ are computed by solving equation (6) in the parent node once. A_p is computed as

$$A_p = \frac{1}{|\{i : X_i \in P\}|} \sum_{\{i : X_i \in P\}} \nabla \psi_{\hat{\alpha}(x), \hat{\beta}(x)}(O_i).$$

   (ii) Compute pseudo-outcomes as

$$\rho_i = \xi' A_p^{-1} \psi_{\hat{\alpha}_P(x), \hat{\beta}_P(x)}(O_i),$$

   where ξ picks out the α coordinate in $(\alpha(x), \beta(x))$.
   (iii) Using the pseudo-outcome ρ_i in a standard classification and regression tree (CART), split P into two child nodes C_1 and C_2 to maximize the criterion

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

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

(iv) After splitting into the two children, relabel the observations and repeat steps (c)(i)–(c)(iii) to output a tree, $\mathcal{T}$

(d) For tree $\mathcal{T}$, create a neighbourhood by collecting all observations in subsample $\mathcal{I}_2$ that fall in the same leaf as x, denoted by leaf $L_b(x)$, and compute

$$\omega_{bi}(x) = \frac{\mathbb{I}(\{X_i \in L_b(x)\})}{|L_b(x)|}.$$

(e) Repeat steps (a)–(d) through the total number of sets of trees, B and compute the GRF weight $\omega_i(x)$ as how often i falls in $L_b(x)$:

$$\omega_i(x) = \frac{1}{B}\sum_{1}^{B} \omega_{bi}(x).$$

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Appendix C. Supplementary Medicaid coverage effects

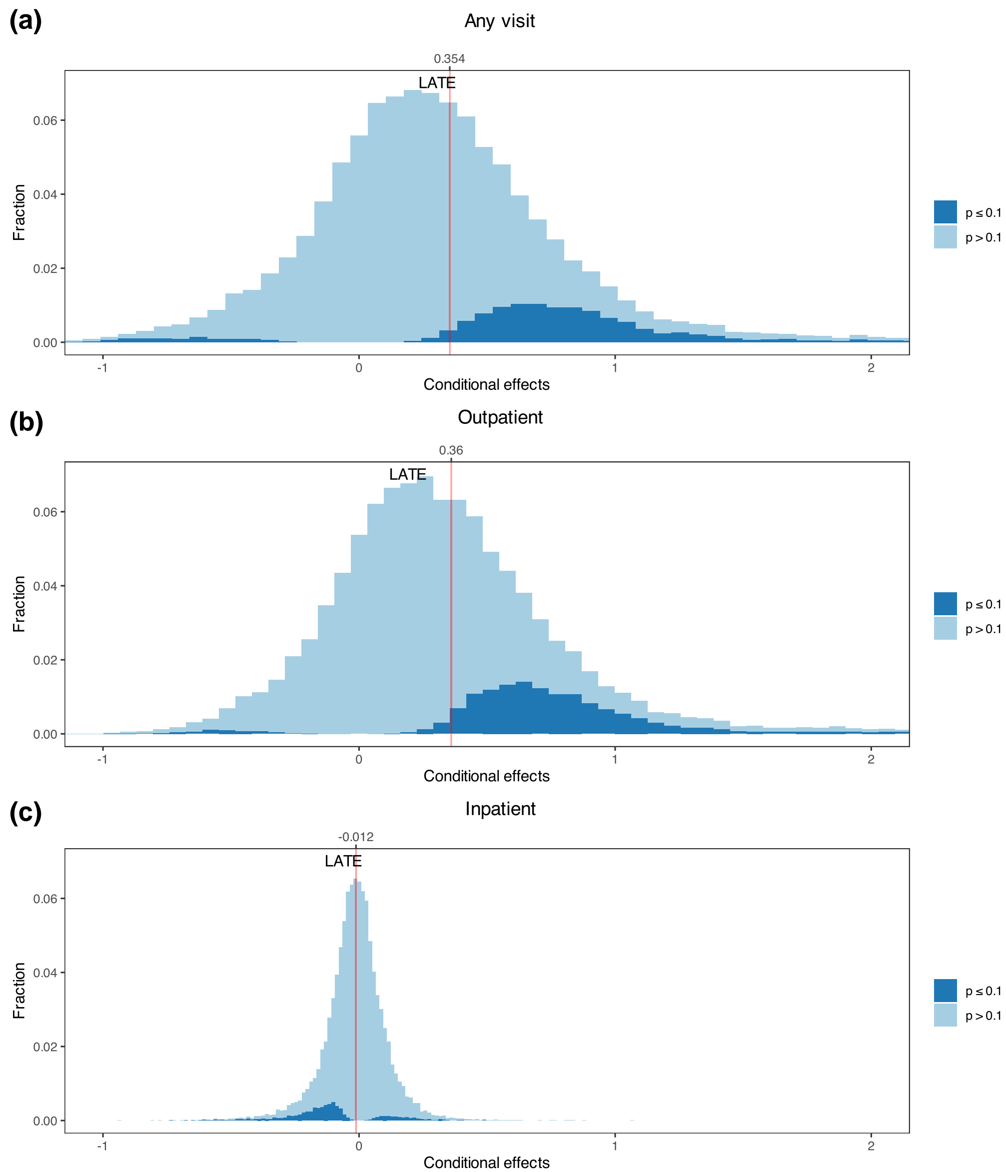

Figure C1. Conditional local average treatment effects of Medicaid on the number of emergency department (ED) visits. *Note*. This figure plots the conditional treatment effects of Medicaid on the number of overall ED visit (a), the number of outpatient ED visits (b), and the number of inpatient ED use (c) based on generalized random forests. The darker shade denotes statistical significance at the 10% level. The red vertical line indicates the local average treatment effect. The baseline sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and Supplemental Nutrition and Assistance Program (SNAP)/Temporary Assistance to Needy Families (TANF) receipt. The estimates displayed exclude less than half a percentile at the top and bottom of the distribution, resulting in the axes corresponding approximately to the percentile range [0.5%, 99.5%]. Bin size is chosen according to the Freedman–Diaconis rule.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

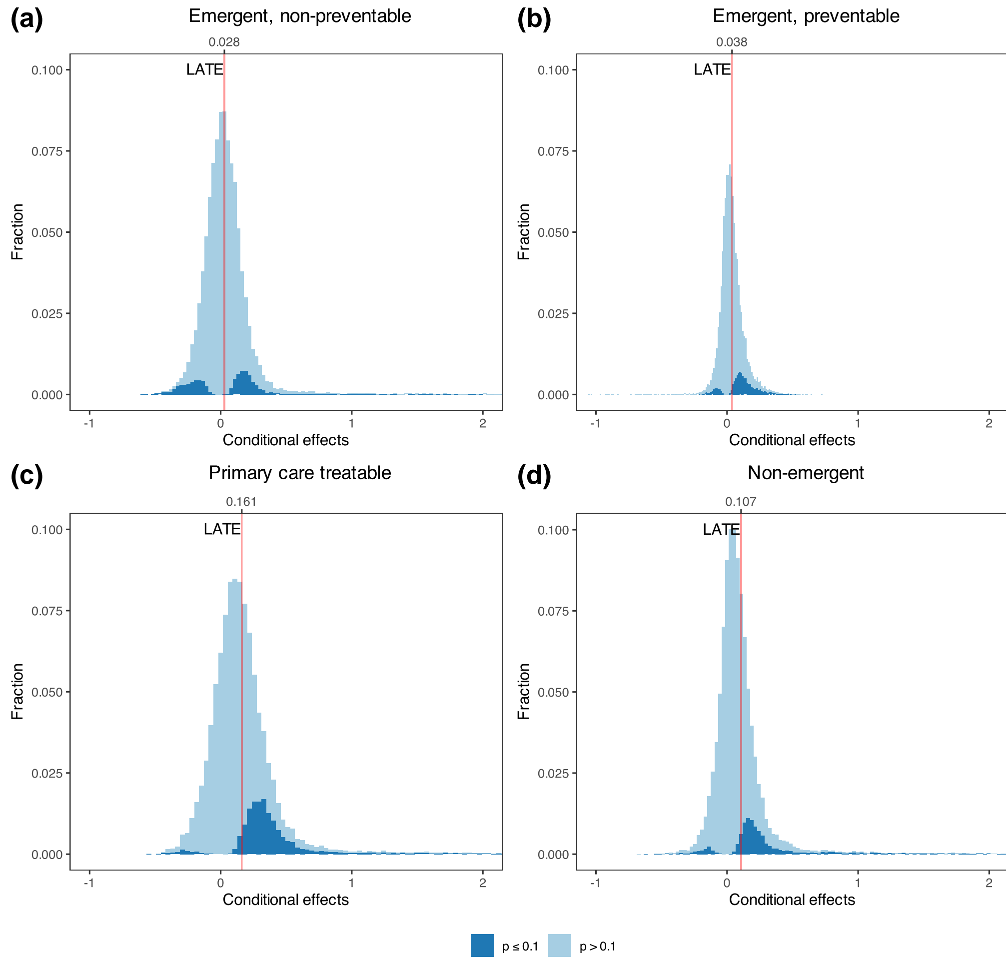

Figure C2. Conditional local average treatment effects of Medicaid on the number of emergency department (ED) visits by type of condition. *Note.* This figure plots the conditional treatment effects of Medicaid by type of ED visit based on generalized random forests for the number of emergent, nonpreventable visits (a), the number of emergent, preventable visits (b), the number of primary care treatable visits (c), and the number of nonemergent visits (d). Measures of the type of ED visit are based on Billings et al.'s (2000) algorithm described in Taubman et al. (2014). The number of visits of each type is then obtained by summing the probabilities across all visits for an individual. The darker shade denotes statistical significance at the 10% level. The red vertical line indicates the local average treatment effect. The baseline sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and Supplemental Nutrition and Assistance Program (SNAP)/Temporary Assistance to Needy Families (TANF) receipt. The estimates displayed exclude less than half a percentile at the top and bottom of the distribution, resulting in the axes corresponding approximately to the percentile range [0.5%, 99.5%]. Bin size is chosen according to the Freedman–Diaconis rule.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

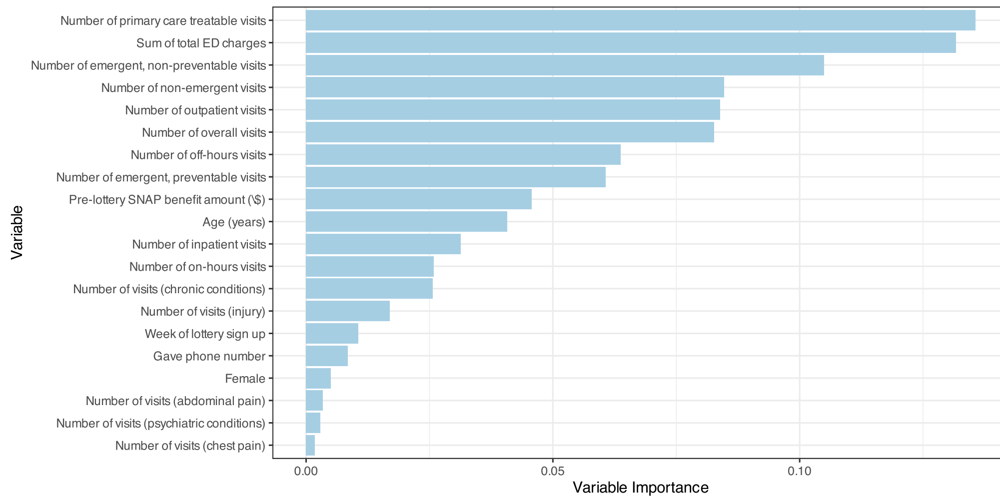

Figure C3. Variable importance scores in growing causal forest (number of visits). *Note*. This figure shows the variable importance scores of the top 20 characteristics in growing the generalized random forests used to estimate the conditional treatment effects of Medicaid for the number of overall ED visit. The variable importance measure is a simple weighted sum of the proportion of times a variable is used in a tree splitting step at each depth in growing the forest. The scores roughly capture how important a variable is for driving treatment effect heterogeneity. The baseline sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and Supplemental Nutrition and Assistance Program (SNAP)/ Temporary Assistance to Needy Families (TANF) receipt.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Table C1. Variable importance for all variables in growing causal forest (overall ED use)

Any visit		Number of visits	
Variable	Importance	Variable	Importance
Prelottery SNAP benefit amount ($)	0.18	Number of primary care treatable visits	0.14
Age (years)	0.18	Sum of total ED charges	0.13
Sum of total ED charges	0.09	Number of emergent, nonpreventable visits	0.10
Number of emergent, nonpreventable visits	0.09	Number of nonemergent visits	0.08
Number of primary care treatable visits	0.08	Number of outpatient visits	0.08
Week of lottery sign up	0.05	Number of overall visits	0.08
Number of nonemergent visits	0.05	Number of off-hours visits	0.06
Female	0.04	Number of emergent, preventable visits	0.06
Number of emergent, preventable visits	0.02	Prelottery SNAP benefit amount ($)	0.05
Number of visits (chronic conditions)	0.01	Age (years)	0.04
Number of inpatient visits	0.01	Number of inpatient visits	0.03
Prelottery SNAP recipient	0.01	Number of on-hours visits	0.03
Number of outpatient visits	0.01	Number of visits (chronic conditions)	0.03
Number of off-hours visits	0.01	Number of visits (injury)	0.02
Number of on-hours visits	0.01	Week of lottery sign up	0.01
Number of overall visits	0.01	Gave phone number	0.01
Signed up self for lottery	0.01	Female	0.01
Gave phone number	0.01	Number of visits (abdominal pain)	0.00
Number of visits (injury)	0.01	Number of visits (psychiatric conditions)	0.00
English as preferred language	0.00	Number of visits (chest pain)	0.00
Prelottery TANF benefit amount ($)	0.00	Prelottery TANF benefit amount ($)	0.00
Number of visits (psychiatric conditions)	0.00	Number of visits (back pain)	0.00
Number of visits (mood disorders)	0.00	Number of visits (skin conditions)	0.00
Number ambulatory-care-sensitive visits	0.00	Prelottery SNAP recipient	0.00
Number of visits (skin conditions)	0.00	Number ambulatory-care-sensitive visits	0.00
Number of visits (abdominal pain)	0.00	Number of visits (headache)	0.00
Number of visits (back pain)	0.00	Number of visits (mood disorders)	0.00
Prelottery TANF recipient	0.00	Signed up self for lottery	0.00
Number of visits (chest pain)	0.00	Prelottery TANF recipient	0.00
Number of visits (headache)	0.00	English as preferred language	0.00
Provided P.O. box address	0.00	Provided P.O. box address	0.00

Note. This table shows the top variable importance scores of all characteristics for growing the generalized random forests used to estimate the ITE of Medicaid coverage for overall ED visits. The variable importance measure is a simple weighted sum of the proportion of times a variable is used in a splitting step at each depth in growing the causal forest, thus, capturing how important a variable is for driving treatment effect heterogeneity. The sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and SNAP/TANF receipt. ED = emergency department; SNAP = Supplemental Nutrition and Assistance Program; TANF = Temporary Assistance to Needy Families.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Table C2. Group local average treatment effects of Medicaid on the number of ED visits

	GRF estimates			Linear estimates			
Group	GLATE	SE	*p*-value	GLATE	SE	*p*-value	*N*%
Female:	0.27	0.16	.08	0.23	0.18	.21	0.55
Male:	0.43	0.17	.01	0.46	0.18	.01	0.45
Gave phone number: No	−0.06	0.33	.86	−0.26	0.38	.50	0.13
Gave phone number: Yes	0.42	0.12	.00	0.42	0.14	.00	0.87
English as preferred language: No	0.14	0.17	.41	0.05	0.18	.78	0.14
English as preferred language: Yes	0.39	0.13	.00	0.36	0.14	.01	0.86
First week sign-up: No	0.38	0.14	.01	0.48	0.16	.00	0.62
First week sign-up: Yes	0.29	0.19	.12	0.13	0.21	.54	0.38
Prelottery SNAP recipient: No	0.21	0.15	.16	0.17	0.18	.34	0.46
Prelottery SNAP recipient: Yes	0.45	0.17	.01	0.34	0.17	.04	0.54
Prelottery TANF recipient: No	0.34	0.11	.00	0.32	0.13	.01	0.98
Prelottery TANF recipient: Yes	0.93	1.21	.44	1.46	2.44	.55	0.02
Age ≥ 50: No	0.43	0.14	.00	0.38	0.16	.02	0.75
Age ≥ 50: Yes	0.05	0.20	.81	0.25	0.21	.22	0.25
Two+ household members on lottery list: No	0.32	0.14	.02	0.27	0.15	.07	0.80
Two+ household members on lottery list: Yes	0.52	0.16	.00	0.74	0.23	.00	0.20
Any prelottery ED visit No	0.22	0.09	.01	0.25	0.08	.00	0.69
Any prelottery ED visit: Yes	0.67	0.32	.03	0.45	0.31	.15	0.31
Any prelottery on-hours ED visit: No	0.19	0.09	.04	0.26	0.09	.00	0.77
Any prelottery on-hours ED visit: Yes	0.87	0.41	.03	0.67	0.42	.11	0.23
Any prelottery off-hours ED visit: No	0.22	0.09	.02	0.21	0.09	.03	0.81
Any prelottery off-hours ED visit: Yes	0.95	0.45	.04	0.56	0.45	.21	0.19
Any prelottery emergent, nonpreventable ED visit: No	0.22	0.09	.02	0.28	0.09	.00	0.87
Any prelottery emergent, nonpreventable ED visit: Yes	1.15	0.63	.07	0.69	0.63	.27	0.13
Any prelottery emergent, preventable ED visit: No	0.25	0.09	.01	0.32	0.09	.00	0.92
Any prelottery emergent, preventable ED visit: Yes	1.50	0.96	.12	1.31	0.90	.15	0.08
Any prelottery primary care treatable ED visit: No	0.18	0.09	.04	0.26	0.09	.00	0.81
Any prelottery primary care treatable ED visit: Yes	1.03	0.46	.03	0.72	0.47	.13	0.19
Any prelottery nonemergent ED visit: No	0.27	0.09	.00	0.34	0.09	.00	0.86
Any prelottery nonemergent ED visit: Yes	0.82	0.61	.18	0.26	0.63	.68	0.14

Note. This table reports the group local average treatment effects of Medicaid based on generalized random forests and the linear IV method. The baseline sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and SNAP/TANF receipt. ED = emergency department; GLATE = group local average treatment effect; GRF = generalized random forests; SNAP = Supplemental Nutrition and Assistance Program; TANF = Temporary Assistance to Needy Families.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Table C3. Characteristics of individuals who increased and decreased ED use (number of visits)

Variable	Increased ED use (1)	Decreased ED use (2)	Difference (std.) (2)–(1)
Lottery list characteristics			
Age (years)	38.54	42.88	−0.35
Gave phone number	0.86	0.89	−0.09
English as preferred language	0.87	0.84	0.09
Female	0.53	0.58	−0.10
Week of lottery sign up	1.59	1.56	0.02
Provided P.O. box address	0.02	0.03	−0.06
Signed up self for lottery	0.89	0.92	−0.10
Prelottery SNAP recipient	0.59	0.36	0.47
Prelottery SNAP benefit amount ($)	1497.29	817.99	0.38
Prelottery TANF recipient	0.02	0.02	0.00
Prelottery TANF benefit amount ($)	96.79	94.84	0.00
Prelottery ED usage			
Number of overall visits	0.90	0.30	0.39
Number of inpatient visits	0.10	0.06	0.11
Number of outpatient visits	0.80	0.25	0.40
Number of on-hours visits	0.52	0.19	0.33
Number of off-hours visits	0.38	0.11	0.36
Number of emergent, nonpreventable visits	0.18	0.06	0.30
Number of emergent, preventable visits	0.07	0.03	0.16
Number of primary care treatable visits	0.32	0.08	0.41
Number of nonemergent visits	0.18	0.07	0.23
Number ambulatory-care-sensitive visits	0.05	0.03	0.07
Number of visits (chronic conditions)	0.15	0.07	0.15
Number of visits (injury)	0.20	0.05	0.33
Number of visits (skin conditions)	0.06	0.01	0.19
Number of visits (abdominal pain)	0.04	0.01	0.14
Number of visits (back pain)	0.04	0.02	0.09
Number of visits (chest pain)	0.02	0.01	0.07
Number of visits (headache)	0.03	0.01	0.10
Number of visits (mood disorders)	0.03	0.01	0.10
Number of visits (psychiatric conditions)	0.07	0.03	0.12
Sum of total ED charges	1027.60	387.86	0.29
N	18581.00	6018.00	24,599

Note. This table reports the means of individual characteristics and prerandomization ED use for those estimated to increase and decrease ED use upon receiving Medicaid coverage based on the causal forest CATE estimates. ED use is measured as the number of ED visits. Panel A reports the means for the full sample while Panel B is limited to effects significant at the 10% level. The sample consists of 24,599 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and SNAP/TANF receipt. ED = emergency department; SNAP = Supplemental Nutrition and Assistance Program; TANF = Temporary Assistance to Needy Families; CATE = conditional average treatment effect.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Table C4. Analytic sample sizes for ED use outcomes

Variable	Sample size
ED use outcomes	
Any overall visit	24,613
Any inpatient visit	24,613
Any outpatient visit	24,613
Any emergent, nonpreventable visit	24,588
Any emergent, preventable visit	24,588
Any primary care treatable visit	24,588
Any nonemergent visit	24,588
Number of overall visits	24,599
Number of inpatient visits	24,613
Number of outpatient visits	24,599
Number of emergent, nonpreventable visits	24,605
Number of emergent, preventable visits	24,611
Number of primary care treatable visits	24,600
Number of nonemergent visits	24,605

Note. This table reports the sample sizes for our ED use outcome variables after restricting to nonmissing information across the heterogeneity variables. The original Taubman et al. (2014) sample based on the Oregon Health Insurance Experiment data consists of 24,646 individuals. ED = emergency department.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Appendix D. Supplementary intent-to-treat effects

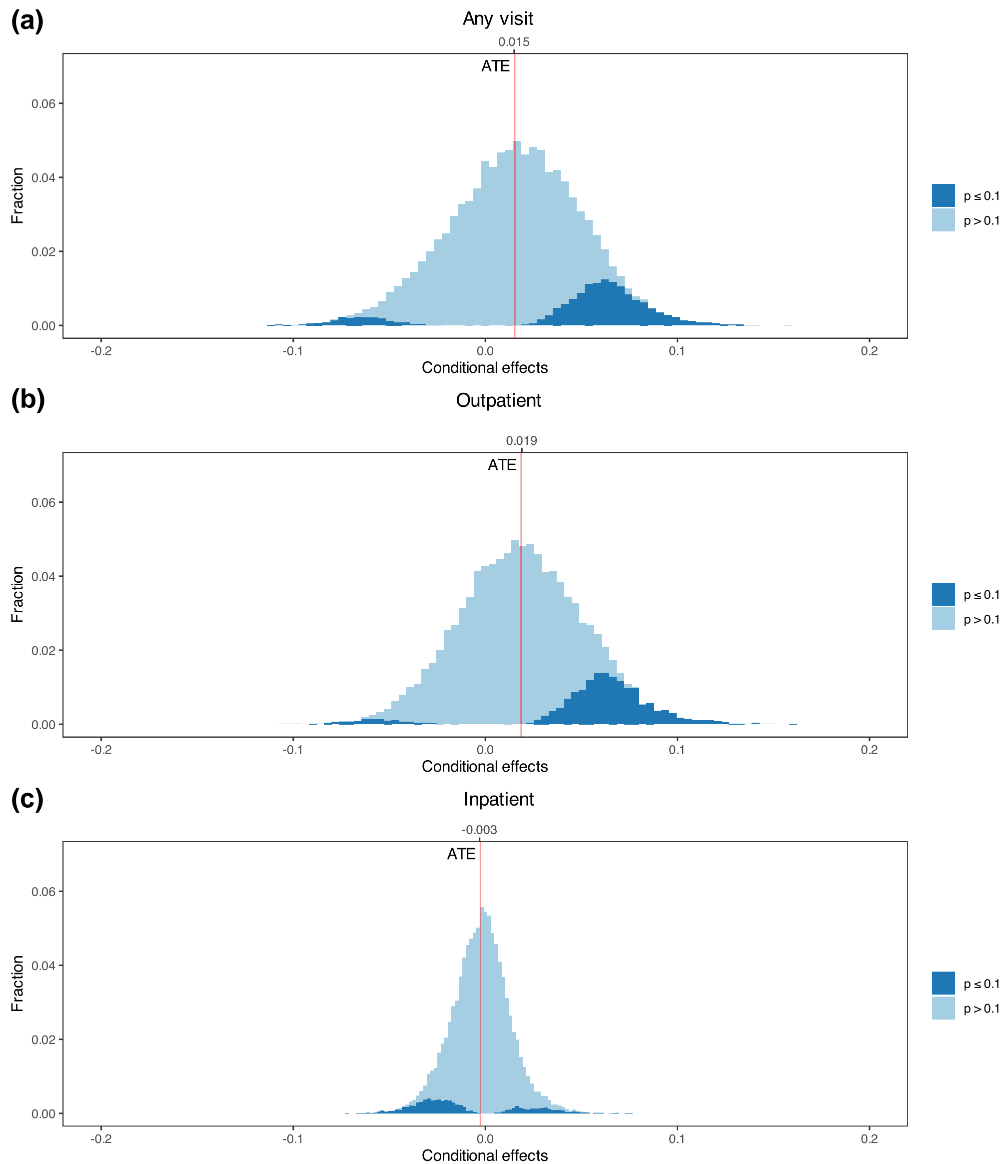

Figure D1. Conditional average treatment effects of winning the lottery on any ED use. *Note.* This figure plots the conditional treatment effects of winning the lottery (and being invited to apply for Medicaid) on any overall ED visit (a), any outpatient ED visits (b), and any inpatient ED use (c) based on generalized random forests. The darker shade denotes statistical significance at the 10% level. The red vertical line indicates the local average treatment effect. The baseline sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and Supplemental Nutrition and Assistance Program (SNAP)/ Temporary Assistance to Needy Families (TANF) receipt. The estimates displayed exclude less than half a percentile at the top and bottom of the distribution, resulting in the axes corresponding approximately to the percentile range [0.5%, 99.5%]. Bin size is chosen according to the Freedman–Diaconis rule.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

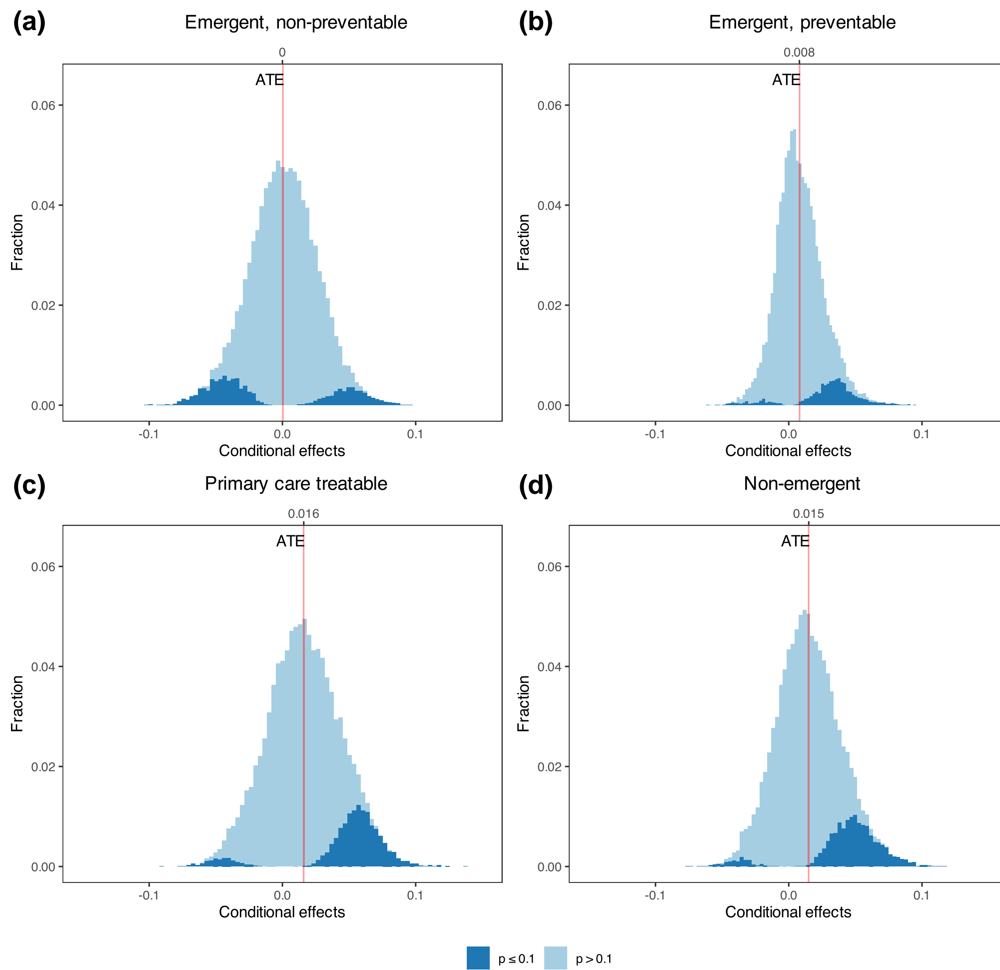

Figure D2. Conditional average treatment effects of winning the lottery on any emergency department (ED) use by type of condition. *Note.* This figure plots the conditional treatment effects of winning the lottery (and being invited to apply for Medicaid) by type of ED visit based on generalized random forests for any emergent, nonpreventable visit (a), any emergent, preventable visit (b), any primary care treatable visit (c), and any nonemergent visit (d). Measures of the type of ED visit are based on Billings et al.'s (2000) algorithm described in Taubman et al. (2014). We use these measures to construct binary indicators of ED visits by type of condition as described in the main text. The darker shade denotes statistical significance at the 10% level. The red vertical line indicates the local average treatment effect. The baseline sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and Supplemental Nutrition and Assistance Program (SNAP)/Temporary Assistance to Needy Families (TANF) receipt. The estimates displayed exclude less than half a percentile at the top and bottom of the distribution, resulting in the axes corresponding approximately to the percentile range [0.5%, 99.5%]. Bin size is chosen according to the Freedman–Diaconis rule.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Figure D3. Variable importance scores in growing causal forest (any visit). *Note.* This figure shows the estimates of variable importance for the top 20 characteristics used in growing the generalized random forests in estimating the conditional average treatment effect (CATE) of winning the lottery (and being invited to apply for Medicaid) for any overall emergency department (ED) visit. The variable importance measure is a simple weighted sum of the proportion of times a variable is used in a splitting step at each depth in growing the causal forest, thus, capturing how important a variable is for driving treatment effect heterogeneity. The sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and Supplemental Nutrition and Assistance Program (SNAP)/Temporary Assistance to Needy Families (TANF) receipt.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

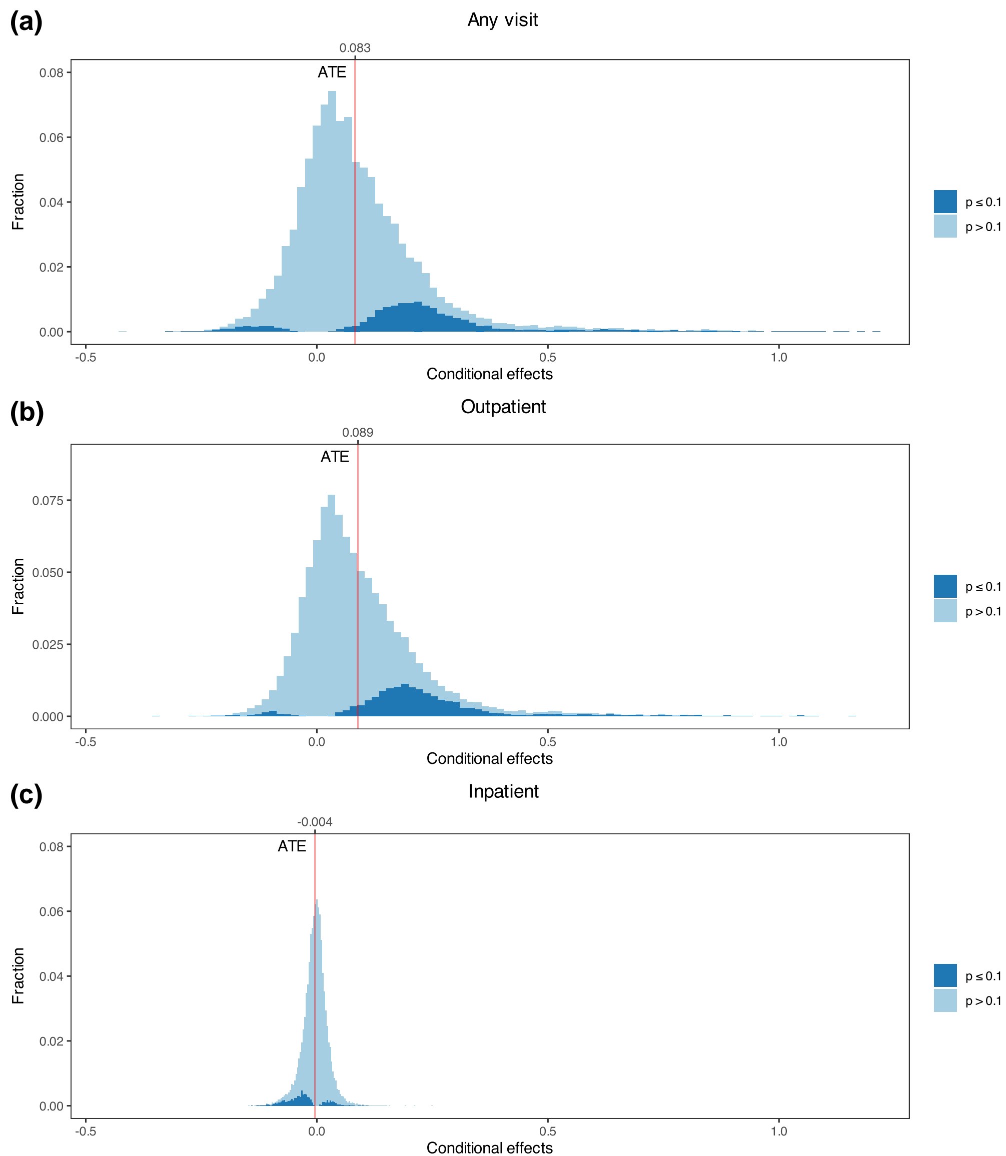

Figure D4. Conditional average treatment effects of winning the lottery on the number of emergency department (ED) visits. *Note.* This figure plots the conditional treatment effects of winning the lottery (and being invited to apply for Medicaid) on the number of overall ED visit (a), the number of outpatient ED visits (b), and the number of inpatient ED use (c) based on generalized random forests. The darker shade denotes statistical significance at the 10% level. The red vertical line indicates the local average treatment effect. The baseline sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and Supplemental Nutrition and Assistance Program (SNAP)/Temporary Assistance to Needy Families (TANF) receipt. The estimates displayed exclude less than half a percentile at the top and bottom of the distribution, resulting in the axes corresponding approximately to the percentile range [0.5%, 99.5%]. Bin size is chosen according to the Freedman–Diaconis rule.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

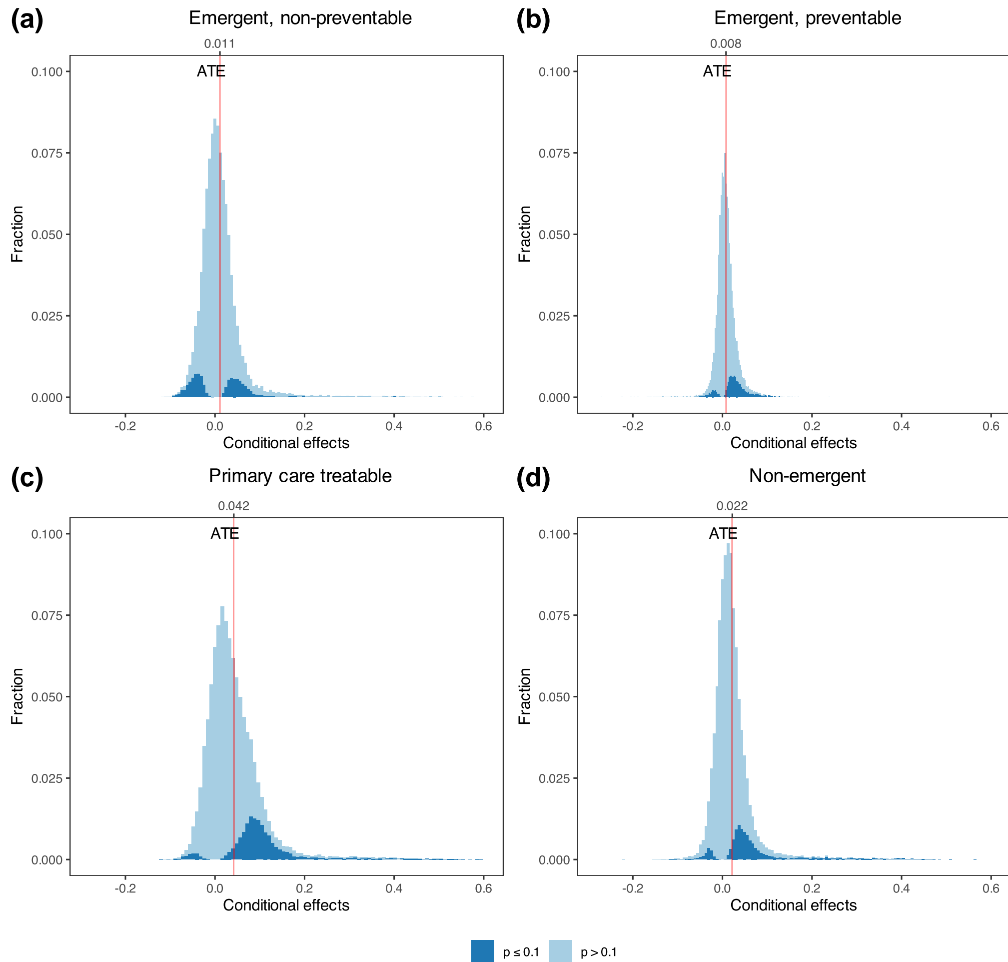

Figure D5. Conditional average treatment effects of winning the lottery on the number of emergency department (ED) visits by type of condition. *Note*. This figure plots the conditional treatment effects of winning the lottery (and being invited to apply for Medicaid) by type of ED visit based on generalized random forests for the number of emergent, nonpreventable visit (a), the number of emergent, preventable visit (b), the number of primary care treatable visit (c), and the number of nonemergent visit (d). Measures of the type of ED visit are based on Billings et al.'s (2000) algorithm described in Taubman et al. (2014). The darker shade denotes statistical significance at the 10% level. The red vertical line indicates the local average treatment effect. The baseline sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and Supplemental Nutrition and Assistance Program (SNAP)/Temporary Assistance to Needy Families (TANF) receipt. The estimates displayed exclude less than half a percentile at the top and bottom of the distribution, resulting in the axes corresponding approximately to the percentile range [0.5%, 99.5%]. Bin size is chosen according to the Freedman–Diaconis rule.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

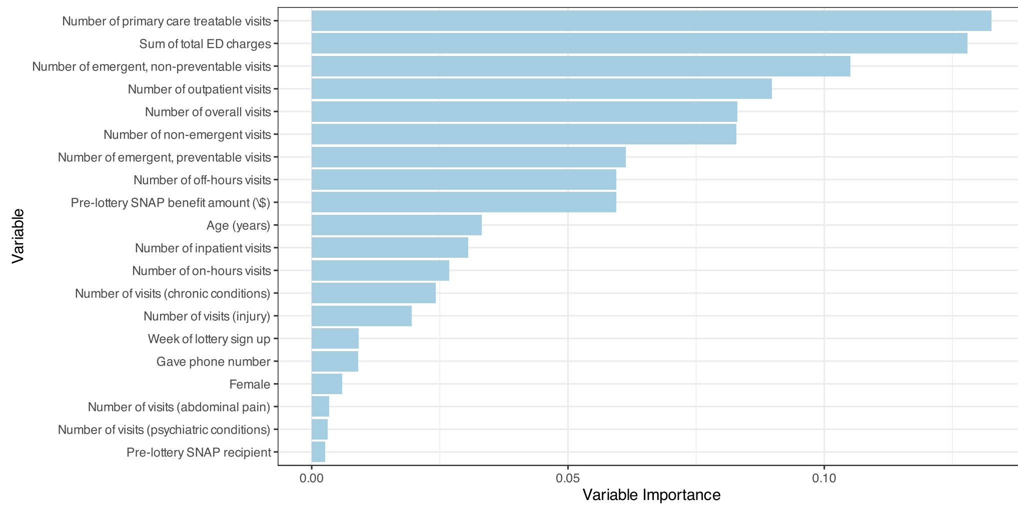

Figure D6. Variable importance scores in growing causal forest (number of visits) *Note.* This figure shows the estimates of variable importance for the top 20 characteristics used in growing the generalized random forests in estimating the conditional average treatment effect (CATE) of winning the lottery (and being invited to apply for Medicaid) for the number of overall emergency department (ED) visits. The variable importance measure is a simple weighted sum of the proportion of times a variable is used in a splitting step at each depth in growing the causal forest, thus, capturing how important a variable is for driving treatment effect heterogeneity. The sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and Supplemental Nutrition and Assistance Program (SNAP)/Temporary Assistance to Needy Families (TANF) receipt.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Table D1. Treatment effect estimates of winning the lottery on ED use

	GRF estimates			Linear estimates		
Variable	ATE	SE	*p*-value	ATE	SE	*p*-value
Extensive margin						
Any overall visit	0.015	0.006	.008	0.017	0.006	.004
Any inpatient visit	−0.003	0.003	.430	−0.003	0.003	.424
Any outpatient visit	0.019	0.006	.001	0.020	0.006	.001
Any emergent, nonpreventable visit	0.000	0.005	.940	0.002	0.005	.731
Any emergent, preventable visit	0.008	0.004	.026	0.010	0.004	.010
Any primary care treatable visit	0.016	0.005	.002	0.017	0.005	.001
Any nonemergent visit	0.015	0.005	.001	0.016	0.005	.001
Intensive margin						
Number of overall visits	0.083	0.027	.002	0.093	0.026	.000
Number of inpatient visits	−0.004	0.006	.519	−0.004	0.006	.516
Number of outpatient visits	0.089	0.024	.000	0.096	0.024	.000
Number of emergent, nonpreventable visits	0.011	0.008	.162	0.010	0.008	.224
Number of emergent, preventable visits	0.008	0.004	.056	0.009	0.004	.034
Number of primary care treatable visits	0.042	0.011	.000	0.042	0.011	.000
Number of nonemergent visits	0.022	0.008	.007	0.026	0.008	.002

Note. This table reports the estimates of winning the lottery on ED use based on generalized random forests and a linear model. The sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and SP/TANF receipt. ED = emergency department; ATE = average treatment effect; TANF = Temporary Assistance to Needy Families.

Table D2. Empirical quantiles of the distribution of individualized treatment effects of winning the lottery on ED use

Variable	ATE	Min	25%	50%	75%	Max
Extensive margin						
Any overall visit	0.015	−0.112	−0.007	0.017	0.039	0.159
Any inpatient visit	−0.003	−0.072	−0.012	−0.002	0.007	0.076
Any outpatient visit	0.019	−0.107	−0.003	0.018	0.041	0.159
Any emergent, nonpreventable visit	0.000	−0.103	−0.016	0.001	0.018	0.096
Any emergent, preventable visit	0.008	−0.061	−0.003	0.007	0.019	0.095
Any primary care treatable visit	0.016	−0.089	−0.003	0.016	0.035	0.137
Any nonemergent visit	0.015	−0.077	−0.001	0.014	0.031	0.116
Intensive margin						
Number of overall visits	0.083	−0.427	0.000	0.061	0.145	1.205
Number of inpatient visits	−0.004	−0.148	−0.018	−0.003	0.011	0.248
Number of outpatient visits	0.089	−0.347	0.009	0.065	0.146	1.154
Number of emergent, nonpreventable visits	0.011	−0.116	−0.015	0.004	0.026	0.576
Number of emergent, preventable visits	0.008	−0.270	−0.003	0.006	0.018	0.238
Number of primary care treatable visits	0.042	−0.118	0.003	0.031	0.067	0.686
Number of nonemergent visits	0.022	−0.218	−0.002	0.014	0.034	0.563

Note. This table reports selected quantiles of the individualized treatment effects of winning the lottery (and being invited to apply for Medicaid) on ED use based on generalized random forests. The first column reports the average effect (intent-to-treat effect). The baseline sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and SNAP/TANF receipt. ED = emergency department; ATE = average treatment effect; SNAP = Supplemental Nutrition and Assistance Program; TANF = Temporary Assistance to Needy Families.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Table D3. GATE estimates of winning the lottery on the propensity of ED use

	GRF estimates			Linear estimates			
Group	GATE	SE	*p*-value	GATE	SE	*p*-value	*N*%
ATE	0.02	0.01	.01	0.02	0.01	.00	100.00
Female:	0.00	0.01	.72	0.01	0.01	.34	0.55
Male:	0.03	0.01	.00	0.03	0.01	.00	0.45
Gave phone number: No	0.01	0.02	.54	0.01	0.02	.66	0.13
Gave phone number: Yes	0.02	0.01	.01	0.02	0.01	.00	0.87
English as preferred language: No	0.01	0.01	.59	0.00	0.01	.73	0.14
English as preferred language: Yes	0.02	0.01	.01	0.02	0.01	.01	0.86
First week sign-up: No	0.01	0.01	.06	0.02	0.01	.01	0.62
First week sign-up: Yes	0.02	0.01	.06	0.01	0.01	.16	0.38
Prelottery SNAP recipient: No	0.00	0.01	.94	0.00	0.01	.62	0.46
Prelottery SNAP recipient: Yes	0.03	0.01	.00	0.02	0.01	.01	0.54
Prelottery TANF recipient: No	0.02	0.01	.01	0.02	0.01	.00	0.98
Prelottery TANF recipient: Yes	−0.01	0.04	.78	0.01	0.04	.82	0.02
Age ≥ 50: No	0.02	0.01	.00	0.02	0.01	.00	0.75
Age ≥ 50: Yes	−0.01	0.01	.67	0.00	0.01	.77	0.25
Two+ household members on lottery list: No	0.01	0.01	.06	0.01	0.01	.06	0.80
Two+ household members on lottery list: Yes	0.03	0.01	.02	0.04	0.01	.00	0.20
Any prelottery ED visit No	0.01	0.01	.07	0.02	0.01	.02	0.69
Any prelottery ED visit: Yes	0.02	0.01	.05	0.02	0.01	.09	0.31
Any prelottery on-hours ED visit: No	0.01	0.01	.03	0.02	0.01	.01	0.77
Any prelottery on-hours ED visit: Yes	0.02	0.01	.11	0.02	0.01	.13	0.23
Any prelottery off-hours ED visit: No	0.01	0.01	.04	0.01	0.01	.03	0.81
Any prelottery off-hours ED visit: Yes	0.02	0.01	.08	0.02	0.01	.16	0.19
Any prelottery emergent, nonpreventable ED visit: No	0.01	0.01	.07	0.01	0.01	.03	0.79
Any prelottery emergent, nonpreventable ED visit: Yes	0.03	0.01	.02	0.03	0.01	.07	0.21
Any prelottery emergent, preventable ED visit: No	0.01	0.01	.02	0.02	0.01	.00	0.90
Any prelottery emergent, preventable ED visit: Yes	0.03	0.02	.15	0.03	0.02	.13	0.10
Any prelottery primary care treatable ED visit: No	0.01	0.01	.10	0.01	0.01	.03	0.74
Any prelottery primary care treatable ED visit: Yes	0.03	0.01	.02	0.03	0.01	.05	0.26
Any prelottery nonemergent ED visit: No	0.01	0.01	.04	0.02	0.01	.01	0.86
Any prelottery nonemergent ED visit: Yes	0.03	0.02	.07	0.02	0.02	.18	0.14

Note. This table reports the GATE estimates of winning the lottery based on generalized random forests and the linear method. The overall effect is reproduced in the first row. The sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and SNAP/TANF receipt. GATE = Group ATE; ED = emergency department; GRF = generalized random forests; ATE = average treatment effect; SNAP = Supplemental Nutrition and Assistance Program; TANF = Temporary Assistance to Needy Families.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Table D4. GATE estimates of winning the lottery on the number of ED visits

	GRF estimates			Linear estimates			
Group	GATE	SE	*p*-value	GATE	SE	*p*-value	%N
ATE	0.08	0.03	0.00	0.08	0.03	0.01	100.00
Female:	0.05	0.03	0.19	0.05	0.04	0.21	0.55
Male:	0.13	0.04	0.00	0.12	0.05	0.01	0.45
Gave phone number: No	−0.04	0.08	0.62	−0.06	0.09	0.50	0.13
Gave phone number: Yes	0.10	0.03	0.00	0.10	0.03	0.00	0.87
English as preferred language: No	0.03	0.03	0.43	0.01	0.04	0.78	0.14
English as preferred language: Yes	0.09	0.03	0.00	0.09	0.04	0.01	0.86
First week sign-up: No	0.09	0.03	0.00	0.11	0.04	0.00	0.62
First week sign-up: Yes	0.06	0.05	0.17	0.03	0.06	0.54	0.38
Prelottery SNAP recipient: No	0.04	0.03	0.17	0.03	0.03	0.34	0.46
Prelottery SNAP recipient: Yes	0.12	0.04	0.01	0.11	0.05	0.04	0.54
Prelottery TANF recipient: No	0.08	0.03	0.00	0.08	0.03	0.01	0.98
Prelottery TANF recipient: Yes	0.07	0.27	0.78	0.17	0.29	0.55	0.02
Age $\geq$ 50: No	0.10	0.03	0.00	0.09	0.04	0.02	0.75
Age $\geq$ 50: Yes	0.03	0.05	0.50	0.07	0.06	0.23	0.25
Two+ household members on lottery list: No	0.07	0.03	0.02	0.07	0.04	0.07	0.80
Two+ household members on lottery list: Yes	0.12	0.03	0.00	0.15	0.05	0.00	0.20
Any prelottery ED visit No	0.05	0.02	0.01	0.06	0.02	0.00	0.69
Any prelottery ED visit: Yes	0.15	0.07	0.04	0.12	0.09	0.15	0.31
Any prelottery on-hours ED visit: No	0.05	0.02	0.01	0.06	0.02	0.00	0.77
Any prelottery on-hours ED visit: Yes	0.19	0.09	0.04	0.17	0.11	0.12	0.23
Any prelottery off-hours ED visit: No	0.05	0.02	0.02	0.05	0.02	0.03	0.81
Any prelottery off-hours ED visit: Yes	0.22	0.11	0.04	0.16	0.12	0.21	0.19
Any prelottery emergent, nonpreventable ED visit: No	0.06	0.02	0.01	0.07	0.02	0.00	0.87
Any prelottery emergent, nonpreventable ED visit: Yes	0.23	0.14	0.11	0.18	0.17	0.28	0.13
Any prelottery emergent, preventable ED visit: No	0.06	0.02	0.00	0.08	0.02	0.00	0.92
Any prelottery emergent, preventable ED visit: Yes	0.30	0.22	0.17	0.36	0.25	0.15	0.08
Any prelottery primary care treatable ED visit: No	0.05	0.02	0.03	0.06	0.02	0.00	0.81
Any prelottery primary care treatable ED visit: Yes	0.24	0.11	0.03	0.19	0.13	0.13	0.19
Any prelottery nonemergent ED visit: No	0.07	0.02	0.00	0.08	0.02	0.00	0.86
Any prelottery nonemergent ED visit: Yes	0.14	0.14	0.30	0.07	0.16	0.68	0.14

Note. This table reports the GATE estimates of winning the lottery based on generalized random forests and the linear method. The overall effect is reproduced in the first row. The sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and SNAP/TANF receipt.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Table D5. Characteristics of individuals who increased and decreased ED use upon winning the lottery

Variable	Increased ED use (1)	Decreased ED use (2)	Difference (std.) (2)–(1)
Lottery list characteristics			
Age (years)	38.81	41.32	−0.20
Gave phone number	0.87	0.87	0.00
English as preferred language	0.87	0.84	0.09
Female	0.51	0.63	−0.24
Week of lottery sign up	1.57	1.61	−0.02
Provided P.O. box address	0.02	0.03	−0.06
Signed up self for lottery	0.89	0.92	−0.10
Prelottery SNAP recipient	0.62	0.35	0.56
Prelottery SNAP benefit amount ($)	1613.82	715.39	0.52
Prelottery TANF recipient	0.02	0.01	0.07
Prelottery TANF benefit amount ($)	110.21	65.85	0.07
Prelottery ED usage			
Number of overall visits	0.87	0.52	0.19
Number of inpatient visits	0.09	0.07	0.05
Number of outpatient visits	0.78	0.45	0.20
Number of on-hours visits	0.50	0.32	0.15
Number of off-hours visits	0.37	0.21	0.18
Number of emergent, nonpreventable visits	0.18	0.10	0.17
Number of emergent, preventable visits	0.07	0.05	0.07
Number of primary care treatable visits	0.30	0.18	0.17
Number of nonemergent visits	0.18	0.11	0.13
Number ambulatory-care-sensitive visits	0.05	0.04	0.03
Number of visits (chronic conditions)	0.14	0.12	0.03
Number of visits (injury)	0.20	0.10	0.19
Number of visits (skin conditions)	0.05	0.03	0.07
Number of visits (abdominal pain)	0.04	0.02	0.08
Number of visits (back pain)	0.04	0.02	0.08
Number of visits (chest pain)	0.02	0.02	0.00
Number of visits (headache)	0.03	0.02	0.05
Number of visits (mood disorders)	0.02	0.03	−0.04
Number of visits (psychiatric conditions)	0.06	0.05	0.03
Sum of total ED charges	1002.57	615.50	0.15
N	16,871	7,742	24,613

Note. This table reports the means of individual characteristics and prerandomization ED use for those estimated to increase and decrease ED use upon winning the lottery based on the causal forest CATE estimates. ED use is measured as the propensity to use the ED. Panel A reports the means for the full sample while Panel B is limited to individuals with effects significant at the 10% level. The sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and SNAP/TANF receipt.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Table D6. Characteristics of individuals who increased and decreased ED use upon winning the lottery (Number of visits)

Variable	Increased ED use (1)	Decreased ED use (2)	Difference (std.) (2)–(1)
Lottery list characteristics			
Age (years)	38.52	42.89	−0.35
Gave phone number	0.87	0.88	−0.03
English as preferred language	0.87	0.84	0.09
Female	0.53	0.59	−0.12
Week of lottery sign up	1.59	1.55	0.02
Provided P.O. box address	0.02	0.03	−0.06
Signed up self for lottery	0.89	0.92	−0.10
Prelottery SNAP recipient	0.59	0.36	0.47
Prelottery SNAP benefit amount ($)	1502.03	813.33	0.39
Prelottery TANF recipient	0.02	0.02	0.00
Prelottery TANF benefit amount ($)	95.68	98.23	−0.00
Prelottery ED usage			
Number of overall visits	0.90	0.32	0.38
Number of inpatient visits	0.10	0.06	0.11
Number of outpatient visits	0.80	0.26	0.38
Number of on-hours visits	0.52	0.21	0.30
Number of off-hours visits	0.38	0.12	0.35
Number of emergent, nonpreventable visits	0.18	0.07	0.27
Number of emergent, preventable visits	0.07	0.03	0.16
Number of primary care treatable visits	0.32	0.08	0.41
Number of nonemergent visits	0.18	0.08	0.20
Number ambulatory-care-sensitive visits	0.05	0.03	0.07
Number of visits (chronic conditions)	0.15	0.08	0.13
Number of visits (injury)	0.20	0.06	0.30
Number of visits (skin conditions)	0.06	0.01	0.19
Number of visits (abdominal pain)	0.04	0.01	0.14
Number of visits (back pain)	0.04	0.02	0.09
Number of visits (chest pain)	0.02	0.01	0.06
Number of visits (headache)	0.03	0.01	0.10
Number of visits (mood disorders)	0.03	0.02	0.05
Number of visits (psychiatric conditions)	0.07	0.03	0.12
Sum of total ED charges	1022.87	411.30	0.28
N	18494.00	6105.00	24,599

Note. This table reports the means of individual characteristics and prerandomization ED use for those estimated to increase and decrease ED use upon winning the lottery based on the causal forest CATE estimates. ED use is measured as the number of total visits. Panel A reports the means for the full sample while Panel B is limited to individuals with effects significant at the 10% level. The sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and SNAP/TANF receipt.

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Table D7. Variable importance for all variables in growing causal forest (overall ED use)

Any visit		Number of visits	
Variable	Importance	Variable	Importance
Prelottery SNAP benefit amount ($)	0.23	Number of primary care treatable visits	0.13
Age (years)	0.16	Sum of total ED charges	0.13
Sum of total ED charges	0.09	Number of emergent, nonpreventable visits	0.11
Number of emergent, nonpreventable visits	0.08	Number of outpatient visits	0.09
Number of primary care treatable visits	0.08	Number of overall visits	0.08
Female	0.04	Number of nonemergent visits	0.08
Number of nonemergent visits	0.04	Number of emergent, preventable visits	0.06
Week of lottery sign up	0.04	Number of off-hours visits	0.06
Prelottery SNAP recipient	0.02	Prelottery SNAP benefit amount ($)	0.06
Number of emergent, preventable visits	0.02	Age (years)	0.03
Number of visits (chronic conditions)	0.01	Number of inpatient visits	0.03
Number of inpatient visits	0.01	Number of on-hours visits	0.03
Number of outpatient visits	0.01	Number of visits (chronic conditions)	0.02
Number of off-hours visits	0.01	Number of visits (injury)	0.02
Number of overall visits	0.01	Week of lottery sign up	0.01
Number of on-hours visits	0.01	Gave phone number	0.01
Gave phone number	0.01	Female	0.01
Signed up self for lottery	0.01	Number of visits (abdominal pain)	0.00
Number of visits (injury)	0.01	Number of visits (psychiatric conditions)	0.00
Prelottery TANF benefit amount ($)	0.00	Prelottery SNAP recipient	0.00
English as preferred language	0.00	Number of visits (chest pain)	0.00
Number of visits (mood disorders)	0.00	Prelottery TANF benefit amount ($)	0.00
Number of visits (psychiatric conditions)	0.00	Number of visits (skin conditions)	0.00
Number ambulatory-care-sensitive visits	0.00	Number of visits (back pain)	0.00
Number of visits (abdominal pain)	0.00	Number of visits (headache)	0.00
Number of visits (skin conditions)	0.00	Number of visits (mood disorders)	0.00
Prelottery TANF recipient	0.00	Number ambulatory-care-sensitive visits	0.00
Number of visits (chest pain)	0.00	Signed up self for lottery	0.00
Number of visits (back pain)	0.00	Prelottery TANF recipient	0.00
Number of visits (headache)	0.00	English as preferred language	0.00
Provided P.O. box address	0.00	Provided P.O. box address	0.00

Note. This table shows the top variable importance scores of all characteristics for growing the generalized random forests used to estimate the ITE of winning the lottery for overall ED visits. The variable importance measure is a simple weighted sum of the proportion of times a variable is used in a splitting step at each depth in growing the causal forest, thus, capturing how important a variable is for driving treatment effect heterogeneity. The sample consists of 24,613 individuals in the Taubman et al. (2014) sample with nonmissing information on prelottery emergency department utilization and SNAP/TANF receipt.

References

Abadie A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics*, *113*(2), 231–263. https://doi.org/10.1016/S0304-4076(02)00201-4

Adams J. G. (2013). Emergency department overuse: Perceptions and solutions. *JAMA*, *309*(11), 1173–1174. https://doi.org/10.1001/jama.2013.2476

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Allen H., Baicker K., Finkelstein A., Taubman S., Wright B. J., & Group O. H. S. (2010). What the Oregon health study can tell us about expanding Medicaid. *Health Affairs*, *29*(8), 1498–1506. https://doi.org/10.1377/hlthaff.2010.0191

Allen H., Baicker K., Taubman S., Wright B., & Finkelstein A. (2013). The Oregon health insurance experiment: When limited policy resources provide research opportunities. *Journal of Health Politics, Policy and Law*, *38*(6), 1183–1192. https://doi.org/10.1215/03616878-2373244

Angrist J. D. (2004). Treatment effect heterogeneity in theory and practice. *The Economic Journal*, *114*(494), C52–C83. https://doi.org/10.1111/j.0013-0133.2003.00195.x

Angrist J. D., Graddy K., & Imbens G. W. (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *The Review of Economic Studies*, *67*(3), 499–527. https://doi.org/10.1111/roes.2000.67.issue-3

Angrist J. D., & Imbens G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, *90*(430), 431–442. https://doi.org/10.1080/01621459.1995.10476535

Angrist J. D., Imbens G. W., & Rubin D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455. https://doi.org/10.1080/01621459.1996.10476902

Athey S., & Imbens G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, *113*(27), 7353–7360. https://doi.org/10.1073/pnas.1510489113

Athey S., & Imbens G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, *11*(1), 685–725. https://doi.org/10.1146/economics.2019.11.issue-1

Athey S., Tibshirani J., & Wager S. (2019). Generalized random forests. *Annals of Statistics*, *47*(2), 1148–1178. https://doi.org/10.1214/18-AOS1709

Athey S., & Wager S. (2019). Estimating treatment effects with causal forests: An application. *Observational Studies*, *5*(2), 37–51. https://doi.org/10.1353/obs.2019.0001

Athey S., & Wager S. (2021). Policy learning with observational data. *Econometrica: Journal of the Econometric Society*, *89*(1), 133–161. https://doi.org/10.3982/ECTA15732

Baicker K., Taubman S. L., Allen H. L., Bernstein M., Gruber J. H., Newhouse J. P., Schneider E. C., Wright B. J., Zaslavsky A. M., & Finkelstein A. N. (2013). The Oregon experiment-effects of Medicaid on clinical outcomes. *New England Journal of Medicine*, *368*(18), 1713–1722. https://doi.org/10.1056/NEJMsa1212321

Billings J., Parikh N., & Mijanovich T. (2000). Emergency room use: The New York story. *The Commonwealth Fund*, *434*, 1–12. https://www.commonwealthfund.org/sites/default/files/documents/___media_files_publications_issue_brief_2000_nov_emergency_room_use__the_new_york_story_billings_nystory_pdf.pdf

Blandhol C., Bonney J., Mogstad M., & Torgovitsky A. (2022). *When is TSLS actually late?* (Technical Report). National Bureau of Economic Research.

Caron A., Baio G., & Manolopoulou I. (2022). Estimating individual treatment effects using non-parametric regression models: A review. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *185*(3), 1115–1149. https://doi.org/10.1111/rssa.12824

Chen C., Scheffler G., & Chandra A. (2011). Perspective: Massachusetts' health care reform and emergency department utilization. *New England Journal of Medicine*, *365*(12), e25. https://doi.org/10.1056/NEJMp1109273

Chernozhukov V., Chetverikov D., Demirer M., Duflo E., Hansen C., Newey W. K., & Robins J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, *21*(1), C1–C68. https://doi.org/10.1111/ectj.12097

Chou S.-C., Gondi S., Weiner S. G., Schuur J. D., & Sommers B. D. (2020). Medicaid expansion reduced emergency department visits by low-income adults due to barriers to outpatient care. *Medical Care*, *58*(6), 511–518. https://doi.org/10.1097/MLR.0000000000001305

Das M. (2005). Instrumental variables estimators of nonparametric models with discrete endogenous regressors. *Journal of Econometrics*, *124*(2), 335–361. https://doi.org/10.1016/j.jeconom.2004.02.001

Deaton A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*, *48*(2), 424–455. https://doi.org/10.1257/jel.48.2.424

Dehejia R. H. (2005). Program evaluation as a decision problem. *Journal of Econometrics*, *125*(1–2), 141–173. https://doi.org/10.1016/j.jeconom.2004.04.006

Fernández-Loría C., & Provost F. (2022). Causal decision making and causal effect estimation are not the same… and why it matters. *INFORMS Journal on Data Science*, *1*(1), 4–16. https://doi.org/10.1287/ijds.2021.0006

Finkelstein A., Taubman S., Wright B., Bernstein M., Gruber J., Newhouse J. P., Allen H., Baicker K., & Group O. H. S. (2012). The Oregon health insurance experiment: Evidence from the first year. *The Quarterly Journal of Economics*, *127*(3), 1057–1106. https://doi.org/10.1093/qje/qjs020

Finkelstein A. N., Taubman S. L., Allen H. L., Wright B. J., & Baicker K. (2016). Effect of Medicaid coverage on ed use—further evidence from Oregon's experiment. *New England Journal of Medicine*, *375*(16), 1505–1507. https://doi.org/10.1056/NEJMp1609533

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Frölich M. (2007). Nonparametric iv estimation of local average treatment effects with covariates. *Journal of Econometrics*, *139*(1), 35–75. https://doi.org/10.1016/j.jeconom.2006.06.004

Garthwaite C., Graves J. A., Gross T., Karaca Z., Marone V. R., & Notowidigdo M. J. (2019). *All Medicaid expansions are not created equal: The geography and targeting of the affordable care act* (Technical Report). National Bureau of Economic Research.

Giannouchos T. V., Ukert B., & Andrews C. (2022). Association of Medicaid expansion with emergency department visits by medical urgency. *JAMA Network Open*, *5*(6), e2216913–e2216913. https://doi.org/10.1001/jamanetworkopen.2022.16913

Heckman J. J., Smith J., & Clements N. (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies*, *64*(4), 487–535. https://doi.org/10.2307/2971729

Hirano K., Imbens G. W., Rubin D. B., & Zhou X.-H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics*, *1*(1), 69–88. https://doi.org/10.1093/biostatistics/1.1.69

Hirano K., & Porter J. R. (2009). Asymptotics for statistical treatment rules. *Econometrica: Journal of the Econometric Society*, *77*(5), 1683–1701. https://doi.org/10.3982/ECTA6630

Imbens G. W., & Angrist J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica: Journal of the Econometric Society*, *62*(2), 467–475. https://doi.org/10.2307/2951620

Imbens G. W., & Rubin D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Kitagawa T., & Tetenov A. (2018). Who should be treated? Empirical welfare maximization methods for treatment choice. *Econometrica: Journal of the Econometric Society*, *86*(2), 591–616. https://doi.org/10.3982/ECTA13288

Knaus M. C., Lechner M., & Strittmatter A. (2021). Machine learning estimation of heterogeneous causal effects: Empirical Monte Carlo evidence. *The Econometrics Journal*, *24*(1), 134–161. https://doi.org/10.1093/ectj/utaa014

Knaus M. C., Lechner M., & Strittmatter A. (2022). Heterogeneous employment effects of job search programs: A machine learning approach. *Journal of Human Resources*, *57*(2), 597–636. https://doi.org/10.3368/jhr.57.2.0718-9615R1

Kolesár M. (2013). *Estimation in an instrumental variables model with treatment effect heterogeneity* (Technical Report). Working Paper.

Kowalski A. E. (2023). Reconciling seemingly contradictory results from the Oregon health insurance experiment and the Massachusetts health reform. *Review of Economics and Statistics*, *105*(3), 1–45. https://doi.org/10.1162/rest_a_01069

Lechner M., Strittmatter A., & Knaus M. (2022). Heterogeneous employment effects of job search programs: A machine learning approach. *Journal of Human Resources*, *57*(2), 597–636. https://doi.org/10.3368/jhr.57.2.0718-9615R1

List J. A., Shaikh A. M., & Xu Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, *22*(4), 773–793. https://doi.org/10.1007/s10683-018-09597-5

Manski C. F. (2004). Statistical treatment rules for heterogeneous populations. *Econometrica: Journal of the Econometric Society*, *72*(4), 1221–1246. https://doi.org/10.1111/ecta.2004.72.issue-4

Marx P. (2024). Sharp bounds in the latent index selection model. *Journal of Econometrics*, *238*(2), 105561. https://doi.org/10.1016/j.jeconom.2023.105561

Miller S. (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. https://doi.org/10.1016/j.jpubeco.2012.07.004

Moore B. J., & Liang L. (2020). Costs of emergency department visits in the United States, 2017. In *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs [Internet]* Agency for Healthcare Research and Quality (US).

Nie X., & Wager S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, *108*(2), 299–319. https://doi.org/10.1093/biomet/asaa076

Robinson P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, *56*(4), 931–954. https://doi.org/10.2307/1912705

Sabbatini A. K., & Dugan J. (2022). Medicaid expansion and avoidable emergency department use—implications for us national and state government spending. *JAMA Network Open*, *5*(6), e2216917–e2216917. https://doi.org/10.1001/jamanetworkopen.2022.16917

Sexton J., & Laake P. (2009). Standard errors for bagged and random forest estimators. *Computational Statistics & Data Analysis*, *53*(3), 801–811. https://doi.org/10.1016/j.csda.2008.08.007

Słoczyński T. (2020). *When should we (not) interpret linear IV estimands as LATE?* [Technical Report]. Working Paper.

Słoczyński T., Uysal S. D., & Wooldridge J. M. (2025). Abadie's kappa and weighting estimators of the local average treatment effect. *Journal of Business & Economic Statistics*, *43*(1), 164–177. https://doi.org/10.1080/07350015.2024.2332763

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026

Sommers B. D., Blendon R. J., Orav E. J., & Epstein A. M. (2016). Changes in utilization and health among low-income adults after Medicaid expansion or expanded private insurance. *JAMA Internal Medicine*, *176*(10), 1501–1509. https://doi.org/10.1001/jamainternmed.2016.4419

Sommers B. D., & Simon K. (2017). Health insurance and emergency department use-a complex relationship. *The New England Journal of Medicine*, *376*(18), 1708–1711. https://doi.org/10.1056/NEJMp1614378

Stoye J. (2009). Minimax regret treatment choice with finite samples. *Journal of Econometrics*, *151*(1), 70–81. https://doi.org/10.1016/j.jeconom.2009.02.013

Stoye J. (2012). Minimax regret treatment choice with covariates or with limited validity of experiments. *Journal of Econometrics*, *166*(1), 138–156. https://doi.org/10.1016/j.jeconom.2011.06.012

Strittmatter A. (2023). What is the value added by using causal machine learning methods in a welfare experiment evaluation? *Labour Economics*, *84*(5), 102412. https://doi.org/10.1016/j.labeco.2023.102412

Sverdrup E., Kanodia A., Zhou Z., Athey S., & Wager S. (2020). policytree: Policy learning via doubly robust empirical welfare maximization over trees. *Journal of Open Source Software*, *5*(50), 2232. https://doi.org/10.21105/joss

Tang N., Stein J., Hsia R. Y., Maselli J. H., & Gonzales R. (2010). Trends and characteristics of us emergency department visits, 1997–2007. *JAMA*, *304*(6), 664–670. https://doi.org/10.1001/jama.2010.1112

Taubman S. L., Allen H. L., Wright B. J., Baicker K., & Finkelstein A. N. (2014). Medicaid increases emergency-department use: Evidence from Oregon's health insurance experiment. *Science*, *343*(6168), 263–268. https://doi.org/10.1126/science.1246183

Tibshirani J., Athey S., Friedberg R., Hadad V., Hirshberg D., Miner L., Sverdrup E., Wager S., Wright M., & Tibshirani M. J. (2018). grf: generalized random forests, comprehensive R archive network. https://doi.org/10.32614/CRAN.package.grf

Wager S., & Athey S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242. https://doi.org/10.1080/01621459.2017.1319839

Woodworth L. (2020). Swamped: Emergency department crowding and patient mortality. *Journal of Health Economics*, *70*(8), 102279. https://doi.org/10.1016/j.jhealeco.2019.102279

Yau L. H., & Little R. J. (2001). Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association*, *96*(456), 1232–1244. https://doi.org/10.1198/016214501753381887

Young A. (2019). Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results*. *The Quarterly Journal of Economics*, *134*(2), 557–598. https://doi.org/10.1093/qje/qjy029

Zhao Y., Zeng D., Rush A. J., & Kosorok M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, *107*(499), 1106–1118. https://doi.org/10.1080/01621459.2012.695674

Downloaded from https://academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnaf114/8232851 by guest on 22 May 2026