Proposal: Estimating population average treatment effects from experiments with noncompliance

Kellie Ottoboni, Jason Poulos

April 3, 2015

1 Introduction

Randomized control trials (RCTs) are the "gold standard" for estimating the causal effect of a treatment. However, external validity is often an issue when RCTs use samples of individuals who do not represent the population of interest. For example, RCTs in which participants volunteer to sign up for health insurance coverage may exhibit a sample population that is in poorer health than the target population. External validity is particularly relevant to policymakers who want to know how a sample average treatment effect (SATE) would generalize to the broader population. Hartman et al. propose a method of reweighting the responses of individuals in an RCT study according to the distribution of covariates in the target population in order to estimate the population average treatment effect on the treated (PATT). Under a series of assumptions, the PATT estimate is asymptotically unbiased [1].

A prevalent issue in RCTs is noncompliance. In one-way crossover from treatment to control, individuals who are randomly assigned to the treatment group refuse the treatment. The serves to dilute the treatment effect, and the resulting intention to treat estimate is biased towards 0. We propose to extend the method of Hartman et. al. to RCTs with one-way crossover from treatment to control in order to identify the complier–average causal effect for the target population. After deriving the estimator, we will apply the method to identify the effect of extending Medicare coverage to the low–income population in the U.S.

2 Statistical Analysis

We propose to modify the estimator of PATT from [1] by allowing for the possibility of one-way crossover from treatment to control in the RCT. We make the same assumptions needed for [1]; these include the stable unit treatment value assumption and strong ignorability. To ensure identifiability, we also assume monotonicity (no defiers) and exclusion restriction as in [2]. It does not make sense to talk about compliers and non-compliers in the population, as there is no imposed treatment assignment; compliance is observable only in the RCT sample.

The estimator of PATT involves the expectation of the response of individuals in the RCT sample, conditional on their covariates, where the expectation is taken over the distribution of population covariates. To estimate the conditional expectation of responses in the RCT, we plan to use an ensemble learning method to estimate the response surface. Then, we will use the response model to estimate population members' outcomes given their covariates. These estimates will be used to estimate the PATT.

3 Data

We will apply the proposed method to estimate the effect of extending Medicaid coverage on emergency–room use and other health outcomes. Medicaid is a federally-funded program, so understanding how the RCT estimates generalize to a broader population of individuals who will be covered by other Medicaid expansions is informative for public policy.

Oregon Health Insurance Experiment We draw RCT data from the Oregon Health Insurance Experiment [3, 4]. In 2008, approximately 90,000 uninsured low-income adults participated in a lottery to receive Medicaid benefits.[1] Treatment occurred at the household level: participants selected by the lottery won the opportunity for themselves and any household member to apply for Medicaid.[2] In total, about 35,000 participants (representing about 30,000 households) were selected by the lottery; the remaining participants were not able to apply for Medicaid and served as controls in the experiment. Participants in selected households received benefits if they returned an enrollment application within 45 days of receipt. Among participants in selected households, about 60% mailed back applications and only 30% successfully enrolled.[3] The RCT data includes demographic variables such as age, gender, ethnicity, pre-existing medical conditions, education, employment, income, and insurance coverage.

The Oregon Health Study obtained data on the number of emergency room visits for every RCT participant who resides near twelve hospitals in the Portland area ($N = 24,646$). The authors find Medicaid coverage increased emergency-room use over an 18–month period by 40% relative to the control group [4].

Observational data We have data on the target population from the National Hospital Ambulatory Medical Care Survey (NHAMCS) [5]. These data come from a random sample of visits to emergency and outpatient departments observed around the United States. It includes categorical demographic variables such as age range, gender, ethnicity, and insurance coverage. The medical record from an individual's visit to the emergency department contains information on the reason for the visit, pre-existing conditions, drugs prescribed, and diagnostic tests run.

4 Anticipated Results and Implications

The treatment effect of Medicaid on emergency–room use applies to uninsured adults with income below the FPL who express interest in health insurance coverage. The sample population differs in several dimensions from the target population of individuals who will be covered by other Medicaid expansions, such as the Affordable Care Act expansion to cover all adults up to 138% of the FPL. For instance, the RCT participants are disproportionately white urban–dwellers [4]. The RCT participants volunteered for the study and therefore may be in poorer health compared to the population.

The proposed method allows us to decompose both SATE and PATT estimates by subgroup according to covariates common to both RCT and observational datasets (e.g, demographic variables, pre–existing conditions, and insurance coverage). Since the RCT participants are predominately white and volunteered for the study, we expect to find substantial differences between sample and population estimates in terms of ethnicity and pre–existing condition subgroups. Overall, we expect the PATT estimate to be lower than the SATE estimate due to the adverse selection problem in the RCT.

References

[1] Erin Hartman, Richard Grieve, Roland Ramsahai, and Jasjeet S. Sekhon. From sate to patt: Combining experimental with observational studies to estimate population treatment effects. *Journal of the Royal Statistical Society, Series A*, (to appear).

[2] Joshua D. Angrist, Guido W. Imbens, and Donald B. Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455, June 1996.

[3] Amy Finkelstein, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P Newhouse, Heidi Allen, Katherine Baicker, Oregon Health Study Group, et al. The oregon health insurance experiment: Evidence from the first year. *The quarterly journal of economics*, 127(3):1057, 2012.

[4] Sarah Taubman, Heidi Allen, Bill Wright, Katherine Baicker, Amy Finkelstein, and the Oregon Health Study Group. Medicaid increases emergency department use: Evidence from oregon's health insurance experiment. *Science*, 343(6168):263–268, January 2014.

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

[2] Since randomization is applied on the household level, we will need to account for the number of each participant's household members when specifying the probability of treatment assignment.

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

[5] National Center for Health Statistics. National hospital ambulatory medical care survey. http://www.cdc.gov/nchs/ahcd/ahcd_questionnaires.htm. April 2015.