Proposal: Extrapolating Sample Average Treatment Effects to the Population

Kellie Ottoboni, Jason Poulos

April 2, 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, the location where a RCT is conducted may result in a distribution of participant ethnicities that does not reflect the makeup of the population (perhaps come up with a better example; how sick patients are to participate in an RCT for example). This is particularly relevant to policy-maker who want to know how a treatment would affect the members of their constituency, rather than just an idealized sample of them. 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, this 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.

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 will make the same assumptions, and further assume that the potential outcomes for non-compliers are the same for treatment and control assignment. It does not make sense to talk about compliers and non-compliers in the population, as there is no imposed treatment assignment. We additionally assume that there are no "defiers," individuals in the RCT who never accept the group to which they are assigned.

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 a nonparametric method (e.g. ensemble tree methods) to estimate the response curve or surface. Then, we will use the model to estimate population members' response given their covariates. These estimates will be used to estimate the PATT.

3   Data

We will apply our method to study how Medicaid subscription affects use of the emergency department and other health outcomes. Medicaid is a federally-funded program, so understanding how its subscribers benefit from it is informative for public policy.

We draw RCT data from the Oregon Health Insurance Experiment [3, 2]. In 2008, approximately 90,000 low-income individuals participated in a lottery to receive Medicaid benefits.[1] Treatment occurred at the household

---

[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.

level: participants selected by the lottery won the opportunity for themselves and any household member to apply for Medicaid. 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.[2]

Health outcomes were assessed using surveys every six months. Emergency room records were obtained from 12 hospitals serving the participants and used to compare frequency of emergency room visits. The authors of the study found that in their sample, Medicaid coverage increased emergency room use by 40% relative to the control group.

The target population of interest for national policy-makers is the entire United States. We have data on the target population from the National Hospital Ambulatory Medical Care Survey (NHAMCS) [4]. The data comes from a random sample of visits to emergency and outpatient departments observed around the United States. It includes information on the reason for the visit and the type of insurance that the individual had.

4 Anticipated Results and Implications

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] 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.

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

---

[2]About half of the returned applications were deemed ineligible, primarily due to failure to demonstrate income below the FPL.