Economics Letters 167 (2018) 131–135

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Randomization bias in field trials to evaluate targeting methods

Eric Potash

Harris School of Public Policy, The University of Chicago, 1155 E 60th St., Chicago, IL 60637, United States

HIGHLIGHTS

- Methods for targeting limited resources to high-risk subpopulations are studied.
- An RCT is considered for measuring the difference in efficiency between methods.
- The RCT is shown to suffer from a form of randomization bias.
- A survey-based design is shown to be unbiased.
- An application to targeting lead hazard investigations is presented.

ARTICLE INFO

Article history:
Received 15 November 2017
Received in revised form 16 February 2018
Accepted 15 March 2018
Available online 29 March 2018

Keywords:
Targeting
Field experiments
Randomized controlled trials
Survey methods

ABSTRACT

This paper studies the evaluation of methods for targeting the allocation of limited resources to a high-risk subpopulation. We consider a randomized controlled trial to measure the difference in efficiency between two targeting methods and show that it is biased. An alternative, survey-based design is shown to be unbiased. Both designs are simulated for the evaluation of a policy to target lead hazard investigations using a predictive model. Based on our findings, we advised the Chicago Department of Public Health to use the survey design for their field trial. Our work anticipates further developments in economics that will be important as predictive modeling becomes an increasingly common policy tool.

1. Introduction

Policymakers may choose to target the allocation of scarce resources to a subpopulation according to risk or need. Rapid advances in predictive modeling in recent decades have the potential to make significant contributions to this age-old economic problem (Kleinberg et al., 2015). Some of the programs where predictive targeting is employed or has been proposed include: residential lead hazard investigations (Potash et al., 2015), restaurant hygiene inspections (Kang et al., 2013), and violence education (Chandler et al., 2011).

Of course, the impact of any targeting method should be evaluated. However, as we shall see, care must be taken in applying the existing economic field trial framework when different treatments (targeting methods) operate on different subsets of the population. We develop a framework for this analysis by drawing on the machine learning (Baeza-Yates et al., 1999) and targeted therapies (Mandrekar and Sargent, 2009) literatures.

Concretely, suppose we have a population of units (e.g. homes) $X = \{1, \ldots, N\}$ and the resources to perform k observations (e.g. investigations) of some binary outcome y (e.g. lead hazards).[1] Next suppose we have a targeting method S which selects a subset S_k of k units for observation.

We define the *precision* of S at k to be the proportion of positive outcomes among the targets S_k. When the goal of targeting is to observe positive outcomes, precision is a measure of efficiency (e.g. the proportion of home investigations finding lead hazards).

In this paper our task is to compare the precision at k of two different targeting methods S and T using k observations. Denoting the precisions of S and T by μ_{S_k} and μ_{T_k}, respectively, we wish to measure their difference

$$\delta := \mu_{S_k} - \mu_{T_k}. \quad (1.1)$$

When δ is positive, S is more efficient than T as a targeting method.

With k observations we can measure the precision of S or of T. But we would need up to[2] $2k$ observations to measure them both and so measure δ. Thus we estimate δ statistically.

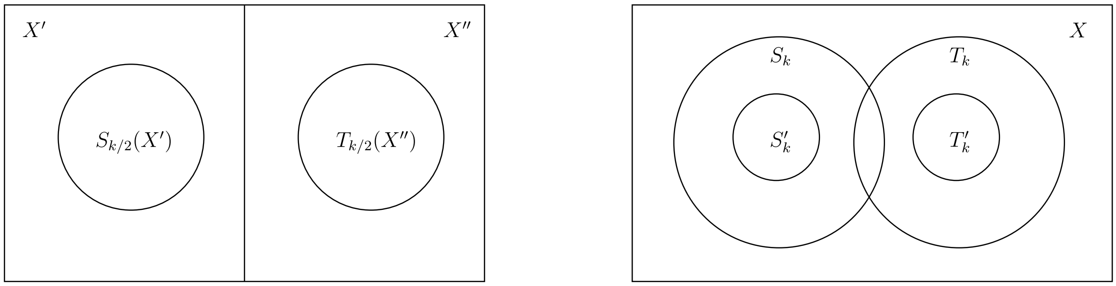
A natural design for a field trial to estimate δ is an RCT in which the population is randomly split in half and each targeting method

[1] We consider interventions in Section 4. Continuous outcomes may be accommodated but binary outcomes are more common.

[2] Depending on the size of the intersection $S_k \cap T_k$.

E-mail address: epotash@uchicago.edu.

https://doi.org/10.1016/j.econlet.2018.03.012
0165-1765/



(a) RCT. (b) Survey.

Fig. 1. In the RCT (a), the population is randomly split in halves X', X'' and each targeting method is applied to one half. In the survey (b), the targeting methods are applied on the population and randomly sampled after excluding their intersection.

is applied to one half. Then we observe the top $k/2$ units in each half, resulting in k total observations.

There is an alternative design: consider S_k and T_k as (after discarding their intersection) disjoint subpopulations and observe $k/2$ random units from each. We think of this design as a survey because it randomly samples the two target sets in the population as opposed to applying the targeting methods to random halves of the population. See Fig. 1 for a graphical comparison of the two designs.

The remainder of the paper is organized as follows. After defining a framework in Section 2, we show in Section 3 that the RCT provides unrepresentative observations and we derive a formula for the bias. In Section 4, we show that the survey gives an unbiased estimate of δ and discuss implementation details. In Section 5, we apply the above to a field trial to evaluate targeting of residential lead hazard investigations using a predictive model and simulate sampling distributions for both designs.

The issue in the RCT stems from the interaction between finite populations, partitions, and order statistics. It is of particular interest as an example of the failure of random assignment to solve an estimation problem. In this sense it is an example of randomization bias (Heckman and Smith (1995), Sianesi (2017)) and adds to the collection of pitfalls that researchers should consider before selecting an RCT design (Deaton and Cartwright, 2016). Our work anticipates further developments in economics that will be important as predictive modeling becomes an increasingly common policy tool.

2. Framework

A targeting method S is a function which, given a set X' of units and a number j selects a subset $S_j(X')$ of size j. When $X' = X$, the full population, we use the shorthand $S_j := S_j(X)$. Let $Y_{S_j(X')}$ denote the outcome y restricted to the set $S_j(X')$ and $\bar{Y}_{S_j(X')}$ its mean. This is the proportion of units in S_j with positive outcome, i.e. the precision of S at resource level j. The population precision at j is then $\bar{Y}_{S_j}$ but we denote it by μ_{S_j} to reflect that it is a population (albeit a finite population) object.

Any model of y is also a targeting method. That is, suppose we have such a model which estimates for any unit x the probability.[3] $P(y|x)$. The corresponding targeting method would select, from any subset X', the units in X' with the j highest model probabilities.[4]

An expert S may not practically be able to rank all units. Instead, they may only be able to produce a list $S_j(X')$. However, we assume that the expert is rational in the sense that there is an underlying

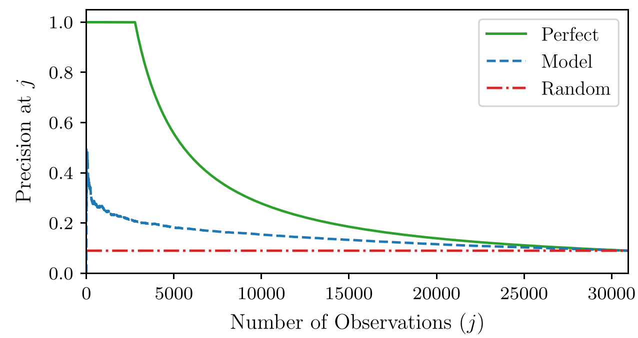


Fig. 2. Precision curves for targeting methods in Section 5.

ranking of all units X that is consistently applied to any subset X'. This implies that any $S_j(X')$ is ordered and we write

$$S_j(X') = (s_1(X'), s_2(X'), \ldots, s_j(X'))$$

to reference units by their rank. When $X' = X$, we use the shorthand $s_j := s_j(X)$.

Following the machine learning literature (Baeza-Yates et al., 1999), we define the *precision curve* of a targeting method S to be μ_{S_j} as a function of j. See Fig. 2. Note that when $k = N$ the entire population is selected, so precision at N of any targeting method is the proportion of positive outcomes in the population.

3. Randomized controlled trial design

A natural RCT to estimate δ using k observations is as follows (see Fig. 1(a)):

1. Randomly partition the population into disjoint halves: $X = X' \cup X''$ with $X' \cap X'' = \varnothing$.
2. Use S to select and observe the top $k/2$ units from X': $S_{k/2}(X')$.
3. Use T to select and observe the top $k/2$ units from X'': $T_{k/2}(X'')$.
4. Calculate

$$\hat{\delta}_{RCT} := \bar{Y}_{S_{k/2}(X')} - \bar{Y}_{T_{k/2}(X'')}.$$

Note we have assumed N and k are even so $N/2$ and $k/2$ are integers.

A hint of the problem with this design arises when carefully defining its terms. Since a traditional RCT applies the same treatment to all units in a treatment group, we must have that: there are just two "units", the subpopulations X' and X''; the "treatments" are $k/2$ selections and observations from each subpopulation; the "outcome" is the precision in the subpopulation, e.g. $\bar{Y}_{S_{k/2}(X')}$. The

[3] Or a score which is not necessarily a probability.

[4] Ties may be broken randomly. For simplicity, we do not explicitly consider stochastic targeting methods.