Subpopulation Analysis in Causal Inference: A Healthcare Case Study

Georgios Mavroudeas
Department of Computer Science
Rensselaer Polytechnic Institute
Troy, NY, USA
mavrog2@rpi.edu

Nafis Neehal
Department of Computer Science
Rensselaer Polytechnic Institute
Troy, NY, USA
neehan@rpi.edu

Jason Kuruzovich
Lally School of Management
Rensselaer Polytechnic Institute
Troy, NY, USA
kuruzj@rpi.edu

Kristin P. Bennett
Department of Mathematical Sciences
Rensselaer Polytechnic Institute
Troy, NY, USA
bennek@rpi.edu

Malik Magdon-Ismail
Department of Computer Science
Rensselaer Polytechnic Institute
Troy, NY, USA
magdon@cs.rpi.edu

2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 978-1-6654-6819-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/BIBM55620.2022.9995265

***Abstract*— Treatment interventions are usually targeted to improve a specific outcome on a selected group of patients who are eligible to receive the treatment. The success of such treatments is determined by the post-intervention treatment effect on the population under consideration. There are cases when the treatment group contains multiple categories of eligible populations, with various effects, especially when the study's criteria are loosely defined. In such studies (non-targeted trials) non-eligible subjects may be treated, producing heterogeneous treatment effects within the treated group. Inferring the effectiveness of the treatment under this scenario is difficult since the average treatment effect on the treated is a combination of multiple effect levels. This can bias the resulting conclusion of the causal studies. We propose an end-to-end framework based on matching and unsupervised clustering for extracting population sub-groups based on their effect levels. We demonstrate our methods on a real-world healthcare application, highlighting the value of subpopulation analysis for recovering multiple effect groups.**

***Index Terms*—Causal Inference, Observational Studies, Subgroup Analysis**

I. Introduction

Randomized controlled trials (RCT) are the gold standard for determining the effect of medical interventions, referred to as the average treatment effect (ATE) for a sample population in which a random selection of the population receives the treatment. Many studies further look to estimate heterogeneous treatment effects (HTE), or the effects for different subgroups (e.g., the treatment effects by race or gender). When a targeted RCT is not possible, causal inference techniques (e.g., coarsened exact matching or propensity score matching) create a synthetic control group for treatment effect estimation. Increasingly machine learning is used to enhance existing approaches to causal inference, particularly in contexts in which the traditional approaches fail [1].

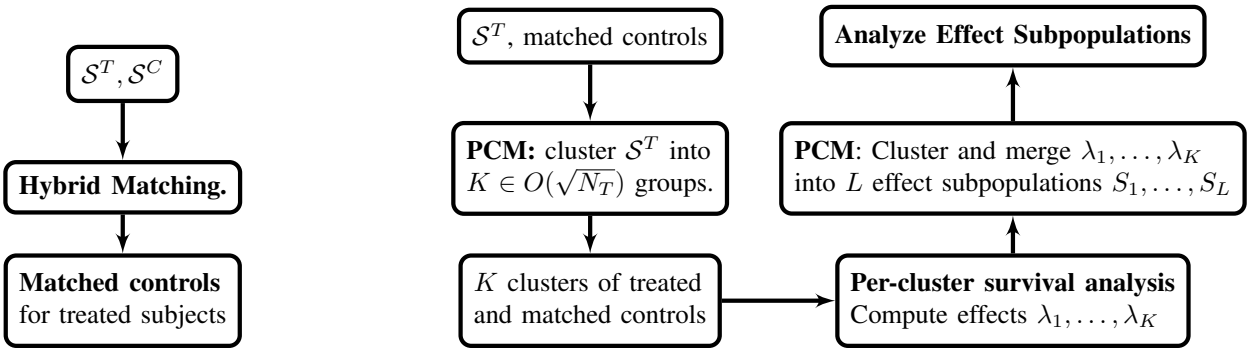
This paper develops a machine learning-based causal estimation procedure for contexts in which HTE are present in complex subpopulations, i.e., they are intricate functions of the patients' observed features. Imagine a population flooded with "healthy" individuals who do not respond to the treatment but are not easily separable from the "sick" individuals who do respond. The calculated ATE for this heterogeneous sample population will underestimate the effectiveness of the treatment and fail to correctly identify the value of the treatment on the sick. Further, standard subgroup analysis techniques to identify HTE (e.g., studying the effects varying by race or gender) may fail to identify the subpopulation in which the treatment is effective. To uncover effects in these kinds of non-targeted trials new methods are needed [2].

We use the non-parametric approach in [3] to uncover complex sub-populations with heterogeneous effects within the treated group. This is a general case of subgroup analysis. While we demonstrate the value of our work in a case study, the applicability stretches across healthcare and other domains.

II. Problem Setup

We use the Rubin-Neyman potential outcomes framework [4]. The problem setup is similar to [3]. A subject (patient) is a tuple $s = (x, c, t, y)$ sampled from a distribution D, where $x \in \mathbf{R}^d$ is a feature vector such as [age, weight], c indicates the effect-level subpopulation to which the subject belongs, $t \in \{0, 1\}$ indicates the subjects treatment cohort (1 for treated), and y is the observed outcome. The observed outcome is one of the two potential outcomes, v if treated or $\bar{v}$ if not treated. The probability $\mathbb{P}[t = 1 \mid x]$ is the propensity to treat function. Under assumptions of ignorability, overlap of support, stable treatment and non-interference (see [4]), the features to identify counterfactual controls for estimating effect. The effect-level c is central to the scope of our work. Mathematically, c is a hidden effect modifier, an unknown and possibly complex function of x. The ground-truth effect-level c should not be confused with the eligibility criteria of a trial. The effect-level c dichotomizes the feature space into subpopulations with different effects. The eligibility criteria for the trial are simply a part of the propensity to treat protocol.

Authorized licensed use limited to: University of Florida. Downloaded on March 24,2023 at 16:01:57 UTC from IEEE Xplore. Restrictions apply.



(a) Stage 1: Matched controls.

(b) Stage 2: Survival analysis and **PCM**.

Fig. 1: Overview of workflow. (a) Stage 1: Matching to get counterfactuals. (b) Stage 2: **PCM** to recover effect-subpopulations.

In practice, one tries to align the eligibility criteria with one of the effect-levels. We address how to estimate the effect levels when there is a misalignment of eligibility criteria and multiple effect-levels in the trial. A trial samples n subjects, $s_1, \ldots, s_n$. If subject i is treated, $t_i = 1$ and the observed outcome $y_i = v_i$, otherwise $t_i = 0$, and the observed outcome is $\bar{v}_i$. The main challenge we address is that the treated group contains subjects from multiple effect-level subpopulations.

State-of-the-art causal inference packages target accurate counterfactual estimation to compute ATT [5; 6]. These algorithms cannot handle multiple effect-level subpopulations in the treated group. We use the framework in [3] which provably extracts these subpopulations with different effect-levels as the trial's size increases. We demonstrate on a pre-diabetic intervention where the outcome of interest is time to inpatient hospitalizations and emergency visits. We use established methods to estimate counterfactual outcomes, e.g. [5; 6]. Extracting the sub-population effect levels is a special case of heterogeneous treatment effects [7; 8]. We compare with methods that directly extract the heterogeneous treatment effect from the individual effects. In our case study, direct causal effect analysis of the treated population as a whole leads to misleading conclusions, while our approach extracts three meaningful sub-populations with very different effects. These three sub-populations can be intuitively explained based on a deeper analysis of the subjects.

III. General Approach

Denote by $\mathcal{S}^T$ the treated population and by $\mathcal{S}^C$ the untreated potential control subjects. The two main goals are (i) for each subject $\mathcal{S}^T$, estimate the counterfactual outcome $\tilde{u}$, (ii) uncover the hidden effect levels $c_1, .., c_L$, where L denotes the number of hidden effect levels, and assign each treated subject $s \in \mathcal{S}^T$ to its corresponding effect-level group c. For (i) we use a hybrid matching technique to estimate the counterfactual outcomes [9] that combines k-nearest-neighbors, exact matching and coarsened exact matching [10; 11]. The reason is some features in the health space should be exactly controlled for, like age and ER-visits, while others can be approximately matched, like blood pressure and weight.

After computing the counterfactual outcomes, each treated subject $s \in \mathcal{S}^T$ is now a tuple $s = (x, v, \tilde{v})$, where x is the feature vector, v the observed outcome, and $\tilde{v}$ the estimated counterfactual outcome. We now determine the number of subpopulation effect-levels L, and assign each treated subject $s_i \in \mathcal{S}^T$ to a level c_ℓ, using a provably accurate pre-cluster and merge algorithm, PCM, developed in [3].

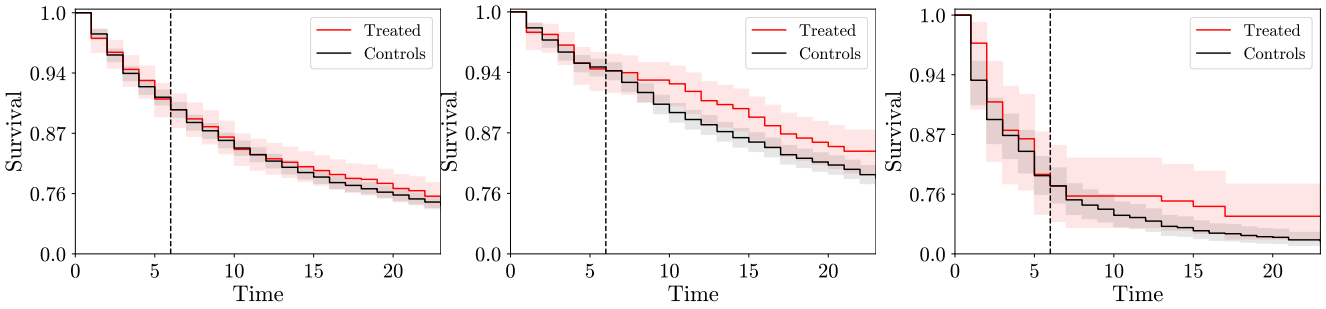
We briefly describe the algorithm and refer to [3] for the details. The input is the set of treated subjects $s_1, \ldots, s_{N_T}$, where $s_i = (x_i, v_i, \tilde{v}_i)$. There are four main steps: (i) Cluster using the features; (ii) Compute treatment effects within each cluster; (iii) Group clusters into effect levels; (iv) Estimate subpopulation effects and assign subjects to subpopulations. An overview of the case study workflow is given in Figure 1.

IV. Case Study

We examine the effectiveness of a health intervention program (HI) for pre-diabetics, using proprietary data from a local health-insurance collaborator. The data has 1,604 patients who enrolled in HI at some time between November 2017 and April 2021. HI is designed for pre-diabetics at risk of diabetes in the future. In addition to patients in the program, the treated group, we were provided with a database of about 350K patients, the control group. Patient features were: Age, Total Cost, Gender, Tobacco Use, Prior Blood Pressure, Diagnosis of several conditions (Obesity, Hypertension, Hypothyroid), Total number of Chornic Diseases, Events within last 2 months (Acute Care), Events within last 6 months (Acute Care, Inpatient Care, and Emergency Visits), and Line of Business. Our goal is to evaluate this program in terms of the positive impact that it had on the enrolled patients, measured by a survival analysis on the time it takes after enrollment in HI for a patient to use acute care (in-patient or ER usage).

A patient is a time series of features, $x_i = (x_{i,1}, ..., x_{i,\tau_i})$. To evaluate HI, we use the survival probability $S(\tau) = \mathbb{P}[\text{event time} \geq \tau]$, where the event of interest is utilization of acute care. Computing $S(\tau)$ is challenging due to censoring, [12; 13]. Thus, we use the Kaplan-Meier curves, [14], a non-parametric technique that accounts for censoring in computing the survival probability. We use $S(\tau)$ to get the restricted

Authorized licensed use limited to: University of Florida. Downloaded on March 24,2023 at 16:01:57 UTC from IEEE Xplore. Restrictions apply.



(a) Sick, No Eff. (b) Healthy, Positive Eff. (c) Critical Very Positive Eff.

Fig. 2: Kaplan Meier's, survival curves for "Time to Acute Care." (a) "Sick" subpopulation with no effect ($p = 0.44$). (b) "Healthy" subpopulation with positive effect ($p = 0.01$). (c) "Critical" subpopulation with large positive effect ($p = 0.08$).

mean survival time (RMST) to 18 months [15]. To evaluate the treatment, we compare the RMST for treated patients with the RMST for matched controls. Since our matching process utilizes historical features, we restricted the analysis to treated subjects with at least six months of history prior to their registered date who also had at least five matching controls.

A. Results

First, we show the survival time to acute care for the full treated population (compared to controls).

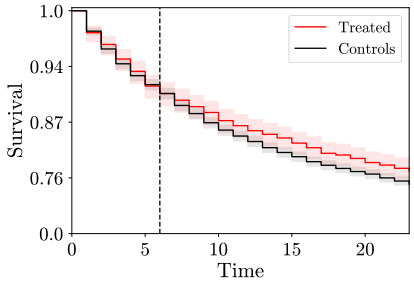
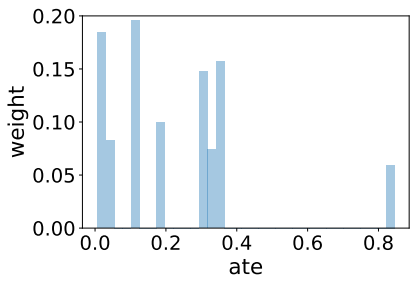


Fig. 3: Kaplan Meier curves, on the outcome: "Time to Acute Care" for all treated population.

The vertical line is the start of the intervention. There is a significant positive treatment effect with p-value 0.01. Note that the matching process produces near identical survival curves prior to HI, as it should.

Our main goal is to identify the different subpopulations in the treatment group with heterogeneous effects. We used our PCM algorithm with agglomerative clustering and 10 clusters to automatically uncover the subpopulations (10 clusters comes from the theory in [3] which suggests about $O(\sqrt{N_T})$ clusters). Our cluster ATEs (18 month RMST) are shown below.



Visual inspection of the **PCM**-cluster effects suggests three effect levels, [0,0.2], [0.3,0.4] and 0.85. Competing techniques for learning heterogeneous treatment effects based on decision trees [16] were not able to recover meaningful subpopulations.

We take a deep dive into the three subpopulations found by PCM after merging cluster ATEs into three effect levels. We call the three subpopulations "sick" (with zero effect), "healthy" (with positive effect) and "critical" (with very positive effect). These names derive from a feature analysis of the corresponding sub-populations. This feature analysis is shown in Table I. For an analysis of the whole population as opposed to the subpopulation analysis see [9]. Survival curves for these three subpopulations are shown in Figure 2. Competing methods for heterogeneous treatment effects (detailed comparison with these methods can be found in [3]).

The first **no effect** group has a mildly positive effect at 18 months, but this effect is insignificant with a p-value of 0.44: the survival curves of the treated and matched controls are statistically indistinguishable, Figure 2 (a)). The second group has a very significant positive effect with a p-value of 0.01. The third group a significant very positive effect with a p-value of 0.08, see Figures 2(b),(c). The most exciting finding is that while the whole treated population does display a positive effect, this effect is by no means uniform across the population. Indeed, the sick group does not appear to derive any benefit from the program when compared to matched controls. This sick group is about half the treated population. The majority of the positive effect is concentrated in a relatively healthy (perhaps health conscious) subpopulation, and a critical severely sick subpopulation which derives extreme benefit.

Based on Table I, we named our subpopulations sick, healthy and critical. The sick subpopulation is male dominated, has high cost and several comorbidities. The healthy subpopulation is female dominated, has low cost and fewer comorbidities. The critical subpopulation is extremely high cost, all smokers with high historical acute care usage. An interesting feature is line of business, which is 0 for medicaid and 1 otherwise. This critical group is over-represented by medicaid patients, which may suggest an inequity of care that is only visible when one looks at subpopulations.

Authorized licensed use limited to: University of Florida. Downloaded on March 24,2023 at 16:01:57 UTC from IEEE Xplore. Restrictions apply.

TABLE I: Feature breakdown of the subpopulations from **PCM**. The p-value quantifies how well matched the subpopulation is w.r.t. its controls, with respect to a given feature (high p-value means the controls match the subpopulation). The "*" means that in both treated and matched controls the feature was always 0.

		Subpopulations found by PCM		
	Treated Population	***Sick, No Effect***	***Healthy, Positive Effect***	***Critical, Positive Effect***
	mean (p-value), N=1364	***mean (p-value), N=767***	***mean (p-value), N=516***	***mean (p-value), N=81***
Age	50.77 (0.86)	51.52 (0.88)	50.08 (0.92)	48.06 (0.99)
Total Cost	705.78 (0.34)	798.16 (0.66)	462.72 (0.56)	1379.32 (0.19)
Gender	0.21(1.0)	0.35 (1.0)	0.02 (1.0)	0.16 (1.0)
Tobacco Use	0.06 (0.37)	0.0 (0.0)	0.0 (0.08)	1.0 (0.0)
Pressure	0.0 (0.4)	0.0 (0.4)	0.0 (*)	0.0 (*)
Obesity	0.5 (0.51)	0.74 (0.18)	0.13 (0.41)	0.6 (0.68)
Hypertension	0.34 (0.36)	0.38 (0.65)	0.26 (0.51)	0.46 (0.44)
Hypothyroid	0.1 (0.05)	0.18 (0.02)	0.0 (0.12)	0.04 (0.91)
Disease Count	2.9 (0.66)	3.48 (0.66)	1.74 (0.95)	4.79 (0.55)
Acute Care 2	0.04 (0.35)	0.04 (0.32)	0.02 (0.95)	0.12 (0.77)
Acute Care 6	0.11 (0.97)	0.12 (0.95)	0.06 (0.96)	0.3 (0.97)
Inpatient Care 6	0.02 (1.0)	0.03 (1.0)	0.0 (1.0)	0.07 (1.0)
Emergency Visits 6	0.09 (0.91)	0.09 (0.91)	0.06 (1.0)	0.23 (0.94)
Line of Bussiness	0.96 (1.0)	0.95 (1.0)	0.99 (1.0)	0.84 (1.0)

V. Conclusion

Our work extends causal analysis to non-targeted health interventions and clinical trials where the treated population can consist of subpopulations exhibiting different effects to the treatment. In the simplest case, there is an eligible population for whom we think treatment works, the positive-effect group, and the ineligible population who might experience a side-effect if treated. Using the pre-cluster and merge strategy in [3], we found three subpopulations with significantly different effects. The challenge we addressed was to untangle the different effect-levels that are co-mingled in the treated population of a non-targeted trial. We used a non-parametric pre-cluster and merge algorithm for untangling the effect-levels.

We used our algorithm in a health intervention case study at an insurance company. Estimating the causal effect of an intervention in a specific population is a critical aspect of understanding its value. If the program is deployed very widely across a mostly healthy population, the effects on average may appear quite small or be absent altogether. A similar small or absent effect may be observed if the application is deployed only to patients who are too sick to be influenced. Our work allows interventions to be deployed widely, yet we can robustly identify subpopulations with differing effects. Such an approach is essential if one is to best understand the benefits and side-effects of a treatment, and identify the subpopulations who would greatly benefit and/or the subpopulations who would be adversely affected. This line of algorithms can also help in identifying inequities between the subpopulations.

Acknowledgments

This material is based upon work supported by a grant from the Capital District Physicians' Health Plan, Inc (CDPHP).

References

[1] J. Pearl, "The seven tools of causal inference, with reflections on machine learning," Comm. of the ACM, vol. 62, no. 3, pp. 54–60, 2019.

[2] Y. Zhang, P. Schnell, C. Song, B. Huang, and B. Lu, "Subgroup causal effect identification and estimation via matching tree," Comp. Stat. & Data Analysis, vol. 159, p. 107188, 2021.

[3] G. Mavroudeas, M. Magdon-Ismail, J. Kuruzovich, and K. P. Bennett, "Untangling effect and side effect: Consistent causal inference in non-targeted trials (submitted under review)," May 2022.

[4] D. B. Rubin, "Causal inference using potential outcomes: Design, modeling, decisions," J. Am. Stat. Assoc., vol. 100, no. 469, pp. 322–331, 2005.

[5] K. Battocchi, E. Dillon, M. Hei, G. Lewis, P. Oka, M. Oprescu, and V. Syrgkanis, "EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation." https://github.com/microsoft/EconML, 2019. Version 0.x.

[6] A. Sharma and E. Kiciman, "Dowhy: An end-to-end library for causal inference," arXiv preprint arXiv:2011.04216, 2020.

[7] S. R. Künzel, J. S. Sekhon, P. J. Bickel, and B. Yu, "Metalearners for estimating heterogeneous treatment effects using machine learning," PNAS, vol. 116, no. 10, pp. 4156–4165, 2019.

[8] U. Shalit, F. D. Johansson, and D. Sontag, "Estimating individual treatment effect: generalization bounds and algorithms," in ICML, pp. 3076–3085, 2017.

[9] N. Neehal, G. Mavroudeas, M. Magdon-Ismail, J. Kuruzovich, and K. P. Bennett, "Hybrid matching methods for treatment program evaluation: A case study," in International Conference on Health Informatics and Medical Systems (to appear), August 2022.

[10] S. M. Iacus, G. King, and G. Porro, "Causal inference without balance checking: Coarsened exact matching," Political Analysis, vol. 20, no. 1, pp. 1–24, 2012.

[11] E. A. Stuart, "Matching methods for causal inference: A review and a look forward," Statistical Science, vol. 25, no. 1, p. 1, 2010.

[12] S. W. Lagakos, "General right censoring and its impact on the analysis of survival data," Biometrics, pp. 139–156, 1979.

[13] K.-M. Leung, R. M. Elashoff, and A. A. Afifi, "Censoring issues in survival analysis," Annual Review of Public Health, vol. 18, no. 1, pp. 83–104, 1997.

[14] B. Efron, "Logistic regression, survival analysis, and the Kaplan-Meier curve," J. Am. Stat. Assoc., vol. 83, no. 402, pp. 414–425, 1988.

[15] L. Zhao, B. Claggett, L. Tian, H. Uno, M. A. Pfeffer, S. D. Solomon, L. Trippa, and L. Wei, "On the restricted mean survival time curve in survival analysis," Biometrics, vol. 72, no. 1, pp. 215–221, 2016.

[16] S. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," IEEE Trans Sys., Man, Cyb., vol. 21, no. 3, pp. 660–674, 1991.

Authorized licensed use limited to: University of Florida. Downloaded on March 24,2023 at 16:01:57 UTC from IEEE Xplore. Restrictions apply.