Subpopulation Analysis in Causal Inference: A Healthcare Case Study

Georgios Mavroudeas, Nafis Neehal,
Jason Kuruzovich, Kristin P. Bennett,
Malik Magdon-Ismail

Rensselaer Polytechnic Institute,
Troy, NY, USA.

Background

- RCTs - gold standard for treatment evaluation
- Problem arises in presence of heterogeneous treatment effects (HTE)
  - ATT becomes a combination of multiple effect levels
  - Biased treatment effect estimation

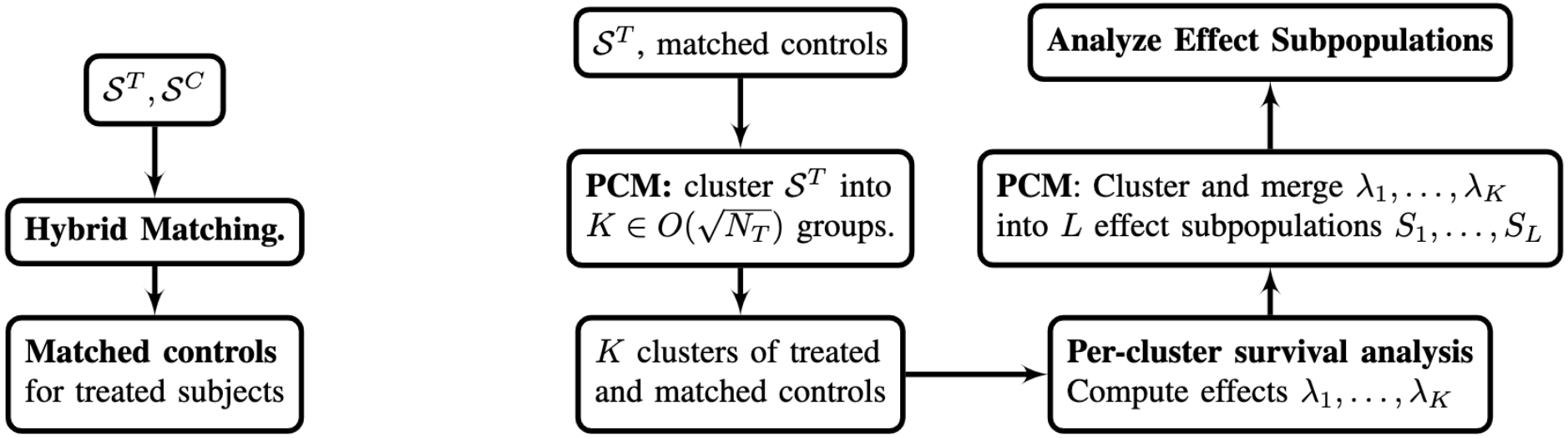
Our Contribution

- Machine Learning based Causal Estimation Procedure
  - Applicable for contexts in which HTE is present in complex subpopulations
- End-to-end framework based on matching and unsupervised learning
  - Non-parametric approach

Problem Setup

- A subject is a tuple s = (x,c,t,y)
  - Sampled from distribution D
  - x is the d-dimensional feature vector
  - c is the effect-level subpopulation to which the subject belongs
  - t is binary treatment assignment
  - Y is the observed outcome
- S^T and S^C are all treated and control population with N^T and N^C number of subjects
- We want to identify -
  - If there exist meaningful subpopulations with heterogenous treatment effects within a population
  - How this heterogeneity affects the population-level estimation of ATE

Methodology



(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.

Goals

- Let S^T and S^C be treated and untreated controls respectively
- Goal 1:
  - for each subject S^T estimate the counterfactual $\tilde{v}$
- Goal 2:
  - uncover the hidden effect levels C_1,….. ,C_L where L denotes the number of hidden effect levels
  - Assign each treated subject $s \in S^T$ to its corresponding effect-level group c

Stage 1: Counterfactual Estimation

- We use a hybrid matching technique described in [9]
  - To produce counterfactual outcomes ($\tilde{v}$) for each treated subject (S^T)
- Combines
  - K-nearest neighbors
  - Exact matching
  - Coarsened exact matching
- Reason:
  - 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.

[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

Stage 2: Determine Effect-Levels

- Each treated $s \in S^T$ is now a tuple $s = (x, v, \tilde{v})$
  - x is the feature, v is the observed outcome, $\tilde{v}$ is the estimated counterfactual outcome
- Determine subpopulation effect-levels L
- Assign each treated subject $s \in S^T$ to a level c_l using a pre-cluster and merge (PCM) algorithm developed in [3]
  - Cluster using features x
  - Compute treatment effects within each cluster
  - Group clusters into effect levels using PCM
  - Assign subjects to subpopulations and estimate subpopulation effects

[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.

Case Study

- Effectiveness of health intervention (HI) program for pre-diabetics
  - Proprietary data from a local health insurance provider
  - 1604 patients enrolled between November 2017 and April 2021 – treated group
  - 350k patients in the control group
  - Features included demographics, lab results, prior health conditions, and history of events (Acute Care, Inpatient Care, and Emergency Visits) within the last 2 and 6 months
- The goal was to evaluate this program
  - Measured by survival analysis on the time it takes after enrollment in HI for a patient to use acute care (in-patient or ER usage)

Results and Discussion (Whole Treated Population)

- Compare Restricted Mean Survival Time (RMST) from Kaplan Meier Curves
  - Between treated and matched controls
- Dotted vertical line is the start of the intervention
- Significant positive treatment effect with p-value 0.01
- Matching process produces near identical survival curves prior to HI, as it should

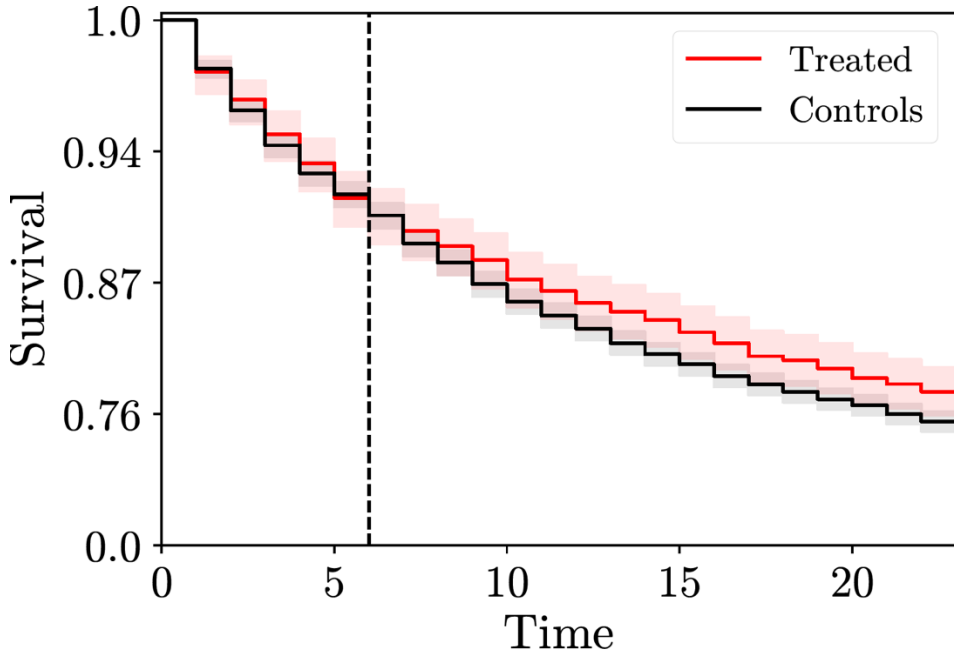


Fig: Kaplan Meier curves, on the outcome: “Time to Acute Care” for all treated population.

Results and Discussion (ATE Clusters)

- PCM Algorithm with agglomerative clustering and 10 clusters
- Visual inspection suggests three effect levels of [0, 0.2], [0.3, 0.4], and 0.85
- Competing techniques for learning HTE based on decision trees were not able to recover this
- Merging clusters results in three final effect-levels
  - Sick (with zero effect)
  - Healthy (with positive effect)
  - Critical (with very positive effect)

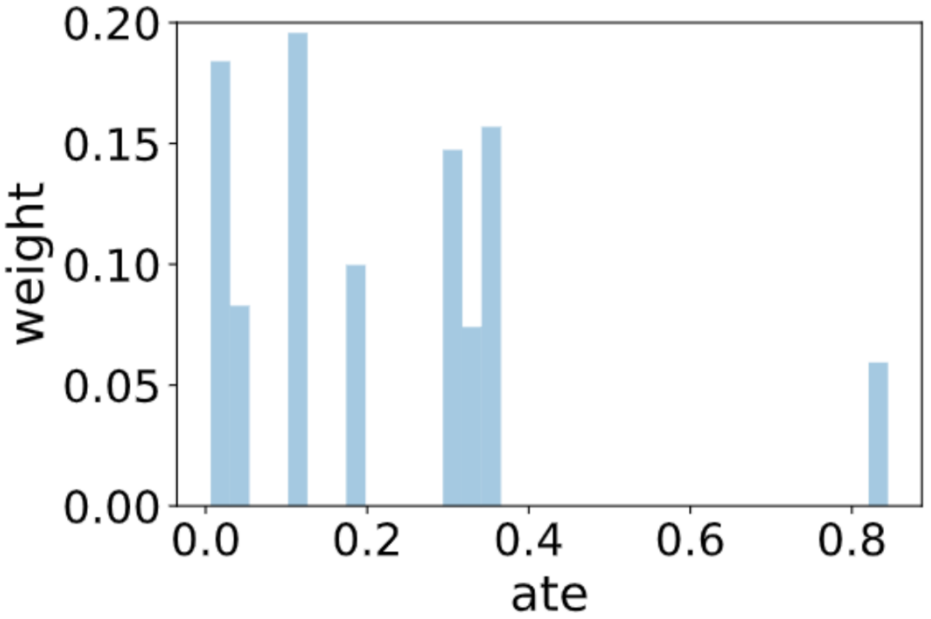
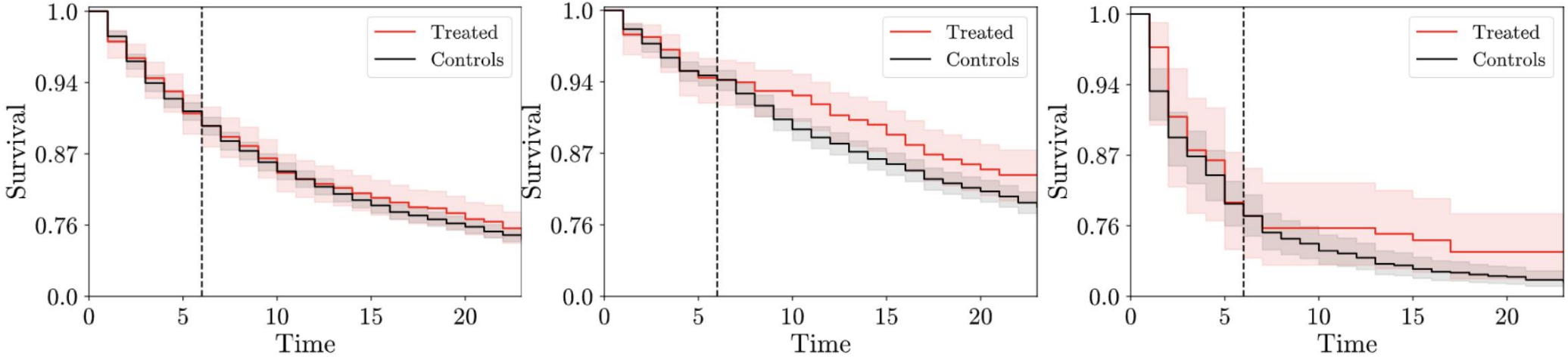


Fig: Clusters of ATE (18 Month RMST) retrieved by PCM

Results and Discussion (Effect-Level Subgroup Analysis)



(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 ($p = 0.08$).

Results and Discussion (Effect-Level Subgroup Analysis)

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)

Conclusion

- Our work extends the causal analysis to non-targeted health interventions and clinical trials -
  - treated population can consist of subpopulations exhibiting different effects to the treatment
- Novel PCM strategy finds three subpopulations with significantly different effects
- Strength of PCM was showcased on an appropriate case study
- Essential if one is to best understand the benefits and side-effects of a treatment