International Journal of
Environmental Research and Public Health

MDPI

Brief Report

A Flexible Approach for Assessing Heterogeneity of Causal Treatment Effects on Patient Survival Using Large Datasets with Clustered Observations

Liangyuan Hu [1,*], Jiayi Ji [1], Hao Liu [1,2] and Ronald Ennis [2,3]

[1] Department of Biostatistics and Epidemiology, Rutgers University, New Brunswick, NJ 07102, USA
[2] Cancer Institute of New Jersey, Rutgers University, New Brunswick, NJ 07102, USA
[3] Robert Wood Johnson Medical School, Rutgers University, New Brunswick, NJ 07102, USA
* Correspondence: lh707@sph.rutgers.edu

Abstract: Personalized medicine requires an understanding of treatment effect heterogeneity. Evolving toward causal evidence for scenarios not studied in randomized trials necessitates a methodology using real-world evidence. Herein, we demonstrate a methodology that generates causal effects, assesses the heterogeneity of the effects and adjusts for the clustered nature of the data. This study uses a state-of-the-art machine learning survival model, riAFT-BART, to draw causal inferences about individual survival treatment effects, while accounting for the variability in institutional effects; further, it proposes a data-driven approach to agnostically (as opposed to a priori hypotheses) ascertain which subgroups exhibit an enhanced treatment effect from which intervention, relative to global evidence—average treatment effects measured at the population level. Comprehensive simulations show the advantages of the proposed method in terms of bias, efficiency and precision in estimating heterogeneous causal effects. The empirically validated method was then used to analyze the National Cancer Database.

Keywords: causal inference; survival data analysis; machine learning; treatment effect heterogeneity; clustering

Citation: Hu, L.; Ji, J.; Liu, H.; Ennis, R. A Flexible Approach for Assessing Heterogeneity of Causal Treatment Effects on Patient Survival Using Large Datasets with Clustered Observations. *Int. J. Environ. Res. Public Health* **2022**, *19*, 14903. https://doi.org/10.3390/ijerph192214903

Academic Editor: Paul B. Tchounwou

Received: 9 September 2022
Accepted: 9 November 2022
Published: 12 November 2022

1. Introduction

Knowing which subgroup of individuals could best benefit from a treatment is crucial to evidence-based medicine. It is critical to understand the treatment effect heterogeneity (TEH), which reflects the variability in patient response to treatment [1]. The population average treatment effect is less useful for personalized medicine because its estimate could potentially average out treatment benefits and harms.

In health research, patient survival is of most clinical relevance. There is a pressing need for robust causal inference methods to evaluate heterogeneous treatment effects on clustered patient survival. Existing methods for assessing heterogeneous treatment effects on patient survival are largely focused on subgroup analysis with a priori hypotheses about either subpopulations who might depart from the population average or interactions between treatment and pre-selected covariates. These methods are prone to multiple testing concerns and estimation bias [2]. More importantly, when patients are clustered, and when there are multiple treatments, special statistical considerations are needed to address the implications of the multilevel data structure for drawing causal inferences about multiple treatment comparisons.

2. Methods

We developed a random-intercept accelerated failure time model leveraging a probabilistic machine learning technique, Bayesian additive regression trees (BART) [3–5], for causal inferences about multiple treatments and clustered survival outcomes. This

method, termed riAFT-BART [6], flexibly and accurately captures the relationships among the patient survival times, treatment and covariates via a sum of the tree models; it also accounts for the cluster-specific main effects using the random intercepts. Regularizing priors are placed on the parameters of riAFT-BART to ensure that the model is flexible in capturing nonlinearity and interactions but not overfitted [7,8]. An efficient and stable Markov chain Monte Carlo algorithm is developed for posterior inferences about the parameters. The formal statistical methodology was described in our earlier work [6] and in the Supplementary Material.

Putting riAFT-BART in the causal framework (Rubin causal model) [9,10], we can obtain the individual treatment effect by contrasting the counterfactual survival time to a treatment with the counterfactual survival time to another treatment (or control in a binary treatment setting) for each individual. These individual effects, bearing a causal interpretation, can then be used to identify TEH via a hypothesis-free and data-driven procedure, leveraging the Random Forest (RF) model. The "fit-the-fit" procedure [11,12] follows the following steps: (1) using the individual treatment effects as the responses, fit a sequence of RF models, where covariates are sequentially added in a stepwise manner to improve the model fit, as measured by R^2; (2) at each step, select the variable producing the largest R^2 improvement; (3) stop the fitting process when the percent improvement in R^2 is less than 1%. The final RF model will be interpreted using the "inTrees" technique [13]. The tree branch decision rules sending individuals to different end nodes define the combination rules of covariates that form different subpopulations having differentiable treatment effects. Subgroup treatment effects are estimated by averaging the individual effects in each end note. A detailed description of the method is provided in Hu (2022) [14].

3. Results

3.1. Data Examples

We first used simulated data to evaluate the proposed method. An expansive and representative simulation was conducted following the state-of-the-art guidance [6,10,15] on generating data adhering to the structure of multiple treatments with heterogeneous treatment effects on clustered survival outcomes. We based our simulation procedures on real data from the National Cancer Database (NCDB) [16]. We compared our proposed approach to three current methods popularly used in clinical research: (1) inverse probability of treatment weighting with the random-intercept Cox regression model (IPW-riCox) [9]; (2) doubly robust random-intercept additive hazards model (DR-riAH) [17]; and (3) the random-intercept generalized additive proportional hazards model (riGAPH) [18]. Complete details for the simulation design are presented in the Supplement.

In both scenarios of proportional hazards (PH) and nonproportional hazards (nPH), our method, riAFT-BART, yielded the smallest biases (Figure S1) and root-mean-squared errors (Figure S2), and achieved the highest accuracy in estimating the the, indicated by the smallest PEHE values across all scenarios (Tables S1 and S3). The violation of the PH assumption had the least impact on the performance of our method but bore heavily on the random-intercept Cox regression model, which requires PH. Figure 1 shows the Kaplan–Meier survival curves for three simulated treatment groups, and the mean counterfactual survival curves estimated by the four methods under PH and nPH. The survival curves estimated by riAFT-BART are closest to the true survival curves for each treatment group under both PH and nPH, corroborating the accuracy of our proposed method in estimating the individual treatment effects (Table 1).

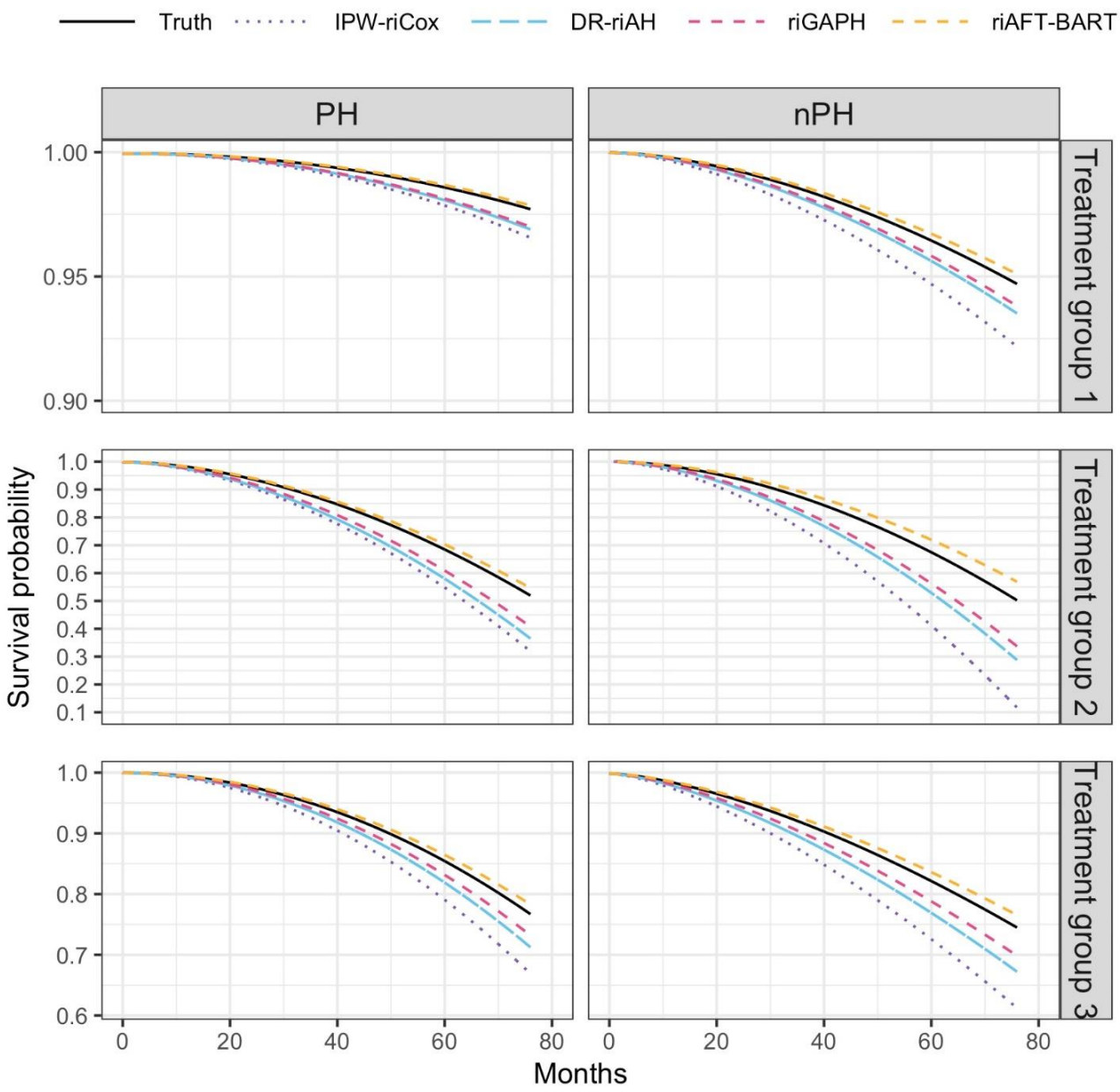

Figure 1. The Kaplan–Meier survival curves for three treatment groups in our simulation, and the mean counterfactual survival curves estimated by the four methods IPW-riCox, DR-riAH, riGAPH and riAFT-BART, under proportional hazards and nonproportional hazards.

Table 1. Mean (and standard deviation) precision in the estimation of heterogeneous effects (PEHE) for each of the four methods based on 5-year survival probability. The proposed method riAFT-BART delivered the highest accuracy in estimating the treatment effect heterogeneity, indicated by the smallest PEHE values across all simulation settings.

	Proportional Hazards			**Nonproportional Hazards**		
Method	**Trt 1 vs. 2**	**Trt 1 vs. 3**	**Trt 2 vs. 3**	**Trt 1 vs. 2**	**Trt 1 vs. 3**	**Trt 2 vs. 3**
IPW-riCox	0.093 (0.032)	0.070 (0.030)	0.075 (0.031)	0.112 (0.033)	0.090 (0.030)	0.094 (0.031)
DR-riAH	0.043 (0.024)	0.022 (0.023)	0.027 (0.022)	0.051 (0.022)	0.034 (0.021)	0.039 (0.022)
riGAPH	0.041 (0.022)	0.020 (0.021)	0.025 (0.022)	0.049 (0.023)	0.031 (0.023)	0.038 (0.023)
riAFT-BART	0.013 (0.011)	0.006 (0.009)	0.008 (0.009)	0.018 (0.014)	0.011 (0.011)	0.013 (0.012)

3.2. Heterogeneous Treatment Effects for High-Risk Localized Prostate Cancer Patients

Next, we applied our method on 64,569 high-risk localized prostate cancer patients diagnosed between 2004 and 2015, drawn from the NCDB. We evaluated the TEH among three treatments: (i) radical prostatectomy (RP); (ii) external beam radiotherapy (EBRT) combined with androgen deprivation (AD) (EBRT + AD); and (iii) EBRT plus brachytherapy with or without AD (EBRT + brachy $\pm$ AD) [13]. Patients were naturally clustered within the institution. The pre-treatment risk factors included age, prostate-specific antigen (PSA), clinical T stage, Charlson–Deyo score, biopsy Gleason score, year of diagnosis, insurance status, median income level, education, race, and ethnicity. Table S5 summarizes the baseline characteristics of the patients.

Figure S3 demonstrates that, on average, the expected survival time for patients who underwent RP was 1.25 (1.15, 1.37) times as long as that of patients who underwent EBRT + brachy $\pm$ AD. However, among high-grade cancer patients with a Gleason score ≥ 9, there was no statistically significant treatment benefit associated with RP. When compared to EBRT + AD, RP led to a significantly longer survival time, and there was no directional TEH (Figure S4). Between EBRT + AD and EBRT + brachy $\pm$ AD, the population average treatment effect suggests a significant treatment benefit from EBRT + brachy $\pm$ AD; however, TEH analysis suggests that younger patients with lower PSA had a favorable treatment effect from EBRT + AD (Figure S5). Our method was able to identify the location (cluster-level) effects, which are displayed in Figure S6. Hospitals in New England had substantially better patient outcomes than hospitals in the South-Central area.

4. Discussion

We developed a machine learning-based method to evaluate the heterogenous causal effects on patient survival using real-world evidence data with clustered patient observations. This method provides a much-needed causal analysis tool for researchers to conduct in-depth analyses of multilevel data with a survival endpoint (overall and event-free survival). Our method can be used to gain insights into personalized treatment and institutional variation in treatment effects. Expansive and representative simulations provide strong empirical evidence that our method has better performance than existing methods in a wide range of data settings. Application to the NCDB data elucidates the importance of estimating heterogenous institutional and treatment effects. The developed methods provide an analysis apparatus for researchers working with large health datasets with clustered observations, and can aid in treatment effect discovery in subpopulations and inform the planning of future confirmatory trials.

5. Conclusions

For evidence-based medicine, it is difficult to apply evidence of population average effects to individual patients who might deviate from the population average. To reveal a potentially complex mixture of causal treatment effects (e.g., treatment benefit and treatment harm), robust analyses of TEH are needed. Large-scale clinical datasets collected from multiple institutions become common for generating real-world TEH evidence but introduce clustered observations. We developed a robust method to overcome these challenges and to provide accurate and efficient estimation of heterogenous causal treatment effects on patient survival. The TEH analysis of the NCDB prostate cancer data suggests that there may exist heterogeneous treatment effects between the EBRT-based treatments, which can inform individualized treatment strategies. The examination of variation due to clusters may stimulate further investigation into reasons why patient outcomes are different across clusters, which may lead to new insight into the quality of treatment delivery. Results from applying our methods to the NCDB prostate cancer data suggest a substantial variability in institutional effects, with hospitals in the New England area having much better patient outcomes than hospitals in the East Central region. Importantly, the results based on our methods are relevant to all stakeholders, including researchers, patients, clinicians and policymakers. For future research, an immediate extension of the proposed methods is to

include the random slopes in the model so that the variability of the covariate effects across different clusters can be incorporated. Developing a sensitivity analysis to potentially have no unmeasured confounding factors could also be a worthwhile contribution.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijerph192214903/s1. S1: The proposed method riAFT-BART; S2: Simulation design; Table S1: Subclasses of the true generalized propensity scores; Table S2: True survival probabilities, restricted mean survival time (RMST), and treatment effects under both proportional hazards and nonproportional hazards; S3: Simulation results; Figure S1: Biases results among subclasses defined by distributions of true generalized propensity scores under proportional hazards (PH) and nonproportional hazards (nPH) for each of four methods, IPW-riCox, DR-riAH, riGAPH and riAFT-BART. Treatment effects were estimated based on 5-year survival probability. Each boxplot visualizes the distribution of biases for 45 subclasses across 250 simulation runs; Figure S2: RMSE results among subclasses defined by distributions of true generalized propensity scores under proportional hazards (PH) and nonproportional hazards (nPH) for each of four methods, IPW-riCox, DR-riAH, riGAPH and riAFT-BART. Subgroup treatment effects were estimated based on 5-year survival probability. Each boxplot visualizes the distribution of biases for 45 subclasses across 250 simulation runs; Table S3: Mean (and standard deviation) of precision in the estimation of heterogeneous effects (PEHE) for each of the 4 methods based on 5-year survival RMST in months; S4: Additional results for case study; Table S4: Descriptions of pre-treatment variables and hospital locations (clusters) for each of three treatment groups in NCDB data.; Figure S3: Final Random Forests model fit to the posterior mean of the individual survival treatment effect comparing radical prostatectomy with external beam radiotherapy plus brachytherapy with or without androgen deprivation. Values in each node correspond to the posterior mean and 95% credible intervals of the average treatment effect, in terms of the ratio of expected survival time, for the subgroup of individuals represented in that node; Figure S4: Final Random Forests model fit to the posterior mean of the individual survival treatment effect comparing radical prostatectomy (RP) with external beam radiotherapy combined with androgen deprivation (EBRT + AD). Values in each node correspond to the posterior mean and 95% credible intervals of the average treatment effect, in terms of the ratio of expected survival time, for the subgroup of individuals represented in that node; Figure S5: Final Random Forests model fit to the posterior mean of the individual survival treatment effect comparing external beam radiotherapy combined with androgen deprivation (EBRT + AD) to external beam radiotherapy plus brachytherapy with or without androgen deprivation (EBRT + brachy $\pm$ AD). Values in each node correspond to the posterior mean and 95% credible intervals of the average treatment effect, in terms of the ratio of expected survival time, for the subgroup of individuals represented in that node; Figure S6: The institutional (location) effects in terms of the expected survival months represented by the posterior mean and credible intervals of random intercept b, k, k = 1, 2, . . . , 9; References [6,11,17–20] are cited in the supplementary materials.

Author Contributions: Concept and design: H.L.; acquisition and statistical analysis of data and simulations: H.L. and J.J.; interpretation of data: all authors; drafting of the manuscript: H.L.; critical revision of the manuscript for important intellectual content: all authors.; obtained funding: H.L.; supervision: H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Institute of Health under R21CA245855 and 1R01HL159077-01A1, and by award ME-2017C3-9041 and ME-2021C2-23685 from the Patient-Centered Outcomes Research Institute.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The proposed method can be implemented through the R package riAFT-BART, freely available on CRAN. The NCDB data used in the case study is publicly available upon approval of the NCDB Participant User File application.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kravitz, R.L.; Duan, N.; Braslow, J. Evidence-Based Medicine, Heterogeneity of Treatment Effects, and the Trouble with Averages. *Milbank Q.* **2004**, *82*, 661–687. [CrossRef] [PubMed]
2. VanderWeele, T.J.; Knol, M.J. Interpretation of Subgroup Analyses in Randomized Trials: Heterogeneity Versus Secondary Interventions. *Ann. Intern. Med.* **2011**, *154*, 680–683. [CrossRef] [PubMed]
3. Chipman, H.A.; George, E.I.; McCulloch, R.E. BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **2010**, *4*, 266–298. [CrossRef]
4. Hu, L.; Gu, C. Estimation of causal effects of multiple treatments in healthcare database studies with rare outcomes. *Health Serv. Outcomes Res. Methodol.* **2021**, *21*, 287–308. [CrossRef]
5. Hu, L.; Joyce Lin, J.Y.; Ji, J. Variable selection with missing data in both covariates and outcomes: Imputation and machine learning. *Stat. Methods Med. Res.* **2021**, *30*, 2651–2671. [CrossRef] [PubMed]
6. Hu, L.; Ji, J.; Ennis, R.D.; Hogan, J.W. A flexible approach for causal inference with multiple treatments and clustered survival outcomes. *Stat. Med.* **2022**, *41*, 4982–4999. [CrossRef] [PubMed]
7. Hu, L.; Zou, J.; Gu, C.; Ji, J.; Lopez, M.; Kale, M. A flexible sensitivity analysis approach for unmeasured confounding with multiple treatments and a binary outcome with application to SEER-Medicare lung cancer data. *Ann. Appl. Stat.* **2022**, *16*, 1014–1037. [CrossRef]
8. Hill, J.L. Bayesian Nonparametric Modeling for Causal Inference. *J. Comput. Graph. Stat.* **2011**, *20*, 217–240. [CrossRef]
9. Hu, L.; Hogan, J.W. Causal comparative effectiveness analysis of dynamic continuous-time treatment initiation rules with sparsely measured outcomes and death. *Biometrics* **2019**, *75*, 695–707. [CrossRef] [PubMed]
10. Holland, P.W. Statistics and Causal Inference. *J. Am. Stat. Assoc.* **1986**, *81*, 945–960. [CrossRef]
11. Hu, L.; Ji, J.; Li, F. Estimating heterogeneous survival treatment effect in observational data using machine learning. *Stat. Med.* **2021**, *40*, 4691–4713. [CrossRef] [PubMed]
12. Logan, B.R.; Sparapani, R.; McCulloch, R.E.; Laud, P.W. Decision making and uncertainty quantification for individualized treatments using Bayesian Additive Regression Trees. *Stat. Methods Med. Res.* **2019**, *28*, 1079–1093. [CrossRef] [PubMed]
13. Ngufor, C.; Van Houten, H.; Caffo, B.S.; Shah, N.D.; McCoy, R.G. Mixed effect machine learning: A framework for predicting longitudinal change in hemoglobin A1c. *J. BioMed. Inform.* **2019**, *89*, 56–67. [CrossRef] [PubMed]
14. Hu, L. A new tool for clustered survival data and multiple treatments: Estimation of treatment effect heterogeneity and variable selection. *arXiv* **2022**, arXiv:2206.08271.
15. Hu, L.; Gu, C.; Lopez, M.; Ji, J.; Wisnivesky, J. Estimation of causal effects of multiple treatments in observational studies with a binary outcome. *Stat. Methods Med. Res.* **2020**, *29*, 3218–3234. [CrossRef] [PubMed]
16. Ennis, R.D.; Hu, L.; Ryemon, S.N.; Lin, J.; Mazumdar, M. Brachytherapy-Based Radiotherapy and Radical Prostatectomy Are Associated With Similar Survival in High-Risk Localized Prostate Cancer. *J. Clin. Oncol.* **2018**, *36*, 1192–1198. [CrossRef] [PubMed]
17. Li, F.; Zaslavsky, A.M.; Landrum, M.B. Propensity score weighting with multilevel data. *Stat. Med.* **2013**, *32*, 3373–3387. [CrossRef] [PubMed]
18. Hastie, T.J.; Tibshirani, R.J. *Generalized Additive Models*; Chapman &Hall: London, UK, 1990.
19. Royston, P.; Parmar, M.K. Restricted mean survival time: An alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol.* **2013**, *13*, 152. [CrossRef] [PubMed]
20. Lu, M.; Sadiq, S.; Feaster, D.J.; Ishwaran, H. Estimating Individual Treatment Effect in Observational Data Using Random Forest Methods. *J. Comput. Graph. Stat.* **2018**, *27*, 209–219. [CrossRef] [PubMed]