SUPPLEMENTAL DIGITAL CONTENT

Title: Towards Precision in Critical Care Research: Methods for Observational and Interventional Studies

Authors: Emma J. Graham Linck, MSc[1], Ewan C. Goligher, MD, PhD, FRCPC[2-4], Matthew W. Semler, MD, MSc[5,6], Matthew M. Churpek, MD, MPH, PhD, ATSF[1,7]

Affiliations:

1. Department of Biostatistics and Medical Informatics, UW-Madison, Madison, USA
2. Interdepartmental Division of Critical Care Medicine, Temerty Faculty of Medicine, University of Toronto, Toronto, Canada
3. Department of Physiology, University of Toronto, Toronto, Canada
4. Toronto General Hospital Research Institute, University Health Network, Toronto, Canada
5. Division of Allergy, Pulmonary, and Critical Care Medicine, Vanderbilt University Medical Center, Nashville, USA
6. Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, USA
7. Division of Pulmonary and Critical Care, Department of Medicine, University of Wisconsin-Madison, Madison, USA

Corresponding author: Matthew Churpek, mchurpek@medicine.wisc.edu

Table of Contents

eTable 1: Brief description of each approach for heterogeneity of treatment effect (HTE) discovery, including examples of methods used for each approach, as well as a selection of published applications of each method in the medical domain. Note: this is not an exhaustive list of available methods or their applications.

Approaches for HTE discovery	**Description**	**Example method (method's publication)** **representative applications of method in medicine**
Expert-derived subgrouping	1. Consult expert to define subgroups. 2. Compare treatment effect estimates between subsets of individuals with and without a particular characteristic.	Likelihood ratio test (1) [e.g., (2), (3)]
Supervised data-driven subgrouping	1. Use statistical method to find subgroups of individuals that differentially respond to a treatment. 2. Compare treatment effect between subgroups.	Virtual Twins (4) [(5), (6), (7), (8)] Model based recursive partitioning (9) [(10), (11), (12), (13), (14)] PRIM (15) [(16), (17), (18), (19)] SIDES (20) [(21), (22), (23), (24)] Berger, 2014 (25) 7/1/24 11:33:00 AM
Unsupervised data-driven subgrouping	1. Use statistical method to find subgroups of individuals with different baseline characteristics. 2. Compare treatment effect between subgroups.	K-means (26) [(27), (28), (29), (30)] LCA (31) [(27), (32), (33), (34), (35)]
Risk-based modeling	1. Obtain off-the-shelf risk model or train risk model in training set of data. 2. Compare treatment effect in subgroups defined by quantiles of predicted risk.	**Off-the-shelf risk model:** APACHE (37) or SOFA (38) [(39)] **Internally derived risk model:** Logistic regression or Bayesian logistic regression [(40), (41), (42), (43)]
Treatment effect modeling	1. Train model to predict individualized treatment effect (ITE) based on baseline characteristics, treatment and outcome. 2. In held-out testing set, predict ITE based only on an individual's baseline characteristics.	**Meta-learners:** S-learner (44) [(45)] T-learner (44) [(46), (45)] X-learner (44) R-learner (47) [(48)]

		Models that directly estimate individualized treatment effect: Causal Forest (49) (50), (41), (51), (52), (53), (54) BART (55)
Individualized Treatment Rule modeling	1. Train model to predict the optimal treatment for an individual given their baseline characteristics, treatment and outcome. 2. In held-out testing set, predict optimal treatment based only on an individual's baseline characteristics.	**Direct methods:** Qian, 2011 (56) Zhang, 2012 (57) (58) **Indirect methods:** Q-learning (59) (60), (61) Dynamical System Models (62)

References

1. Wang R, Lagakos SW, Ware JH, et al.: Statistics in Medicine — Reporting of Subgroup Analyses in Clinical Trials. *N Engl J Med* 2007; 357:2189–2194

2. McMurray John J.V., Packer Milton, Desai Akshay S., et al.: Angiotensin–Neprilysin Inhibition versus Enalapril in Heart Failure. *N Engl J Med* 2014; 371:993–1004

3. Bhatt Deepak L., Fox Keith A.A., Hacke Werner, et al.: Clopidogrel and Aspirin versus Aspirin Alone for the Prevention of Atherothrombotic Events. *N Engl J Med* 2006; 354:1706–1717

4. Foster JC, Taylor JMG, Ruberg SJ: Subgroup identification from randomized clinical trial data. *Stat Med* 2011; 30:2867–2880

5. Baird A, Cheng Y, Xia Y: Determinants of outpatient substance use disorder treatment length-of-stay and completion: the case of a treatment program in the southeast U.S. *Sci Rep* 2023; 13:13961

6. Colloca L, Dworkin RH, Farrar JT, et al.: Predicting Treatment Responses in Patients With Osteoarthritis: Results From Two Phase III Tanezumab Randomized Clinical Trials. *Clin Pharmacol Ther* 2023; 113:878–886

7. Baird A, Cheng Y, Xia Y: Use of machine learning to examine disparities in completion of substance use disorder treatment. *PloS One* 2022; 17:e0275054

8. Raja S, Rice TW, Lu M, et al.: Adjuvant Therapy After Neoadjuvant Therapy for Esophageal Cancer: Who Needs It? *Ann Surg* 2023; 278:e240

9. Seibold H, Zeileis A, Hothorn T: Model-Based Recursive Partitioning for Subgroup Analyses. *Int J Biostat* 2016; 12:45–63

10. Steinert JI, Sternberg H, Prince H, et al.: COVID-19 vaccine hesitancy in eight European countries: Prevalence, determinants, and heterogeneity. *Sci Adv* 2022; 8:eabm9825

11. Foster S, Mohler-Kuo M, Tay L, et al.: Estimating patient-specific treatment advantages in the ‘Treatment for Adolescents with Depression Study.’ *J Psychiatr Res* 2019; 112:61–70

12. Pirkle CM, Wu YY, Zunzunegui M-V, et al.: Model-based recursive partitioning to identify risk clusters for metabolic syndrome and its components: findings from the International Mobility in Aging Study. *BMJ Open* 2018; 8:e018680

13. Chen T, Xia E, Chen T, et al.: Identification and external validation of IgA nephropathy patients benefiting from immunosuppression therapy [Internet]. *eBioMedicine* 2020; 52[cited 2024 May 20] Available from: https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(20)30032-3/fulltext

14. Hatten KM, Amin J, Isaiah A: Machine Learning Prediction of Extracapsular Extension in Human Papillomavirus–Associated Oropharyngeal Squamous Cell Carcinoma. *Otolaryngol Neck Surg* 2020; 163:992–999

15. Chen G, Zhong H, Belousov A, et al.: A PRIM approach to predictive-signature development for patient stratification. *Stat Med* 2015; 34:317–342

16. Zouboulis CC, Okun MM, Prens EP, et al.: Long-term adalimumab efficacy in patients with moderate-to-severe hidradenitis suppurativa/acne inversa: 3-year results of a phase 3 open-label extension study. *J Am Acad Dermatol* 2019; 80:60-69.e2

17. Llano DA, Devanarayan P, Devanarayan V: VGF in Cerebrospinal Fluid Combined With Conventional Biomarkers Enhances Prediction of Conversion From MCI to AD. *Alzheimer Dis Assoc Disord* 2019; 33:307

18. Qi W, Abu-Hanna A, van Esch TEM, et al.: Explaining heterogeneity of individual treatment causal effects by subgroup discovery: An observational case study in antibiotics treatment of acute rhino-sinusitis. *Artif Intell Med* 2021; 116:102080

19. Devanarayan P, Devanarayan V, Llano DA, et al.: Identification of a Simple and Novel Cut-Point Based Cerebrospinal Fluid and MRI Signature for Predicting Alzheimer's Disease Progression that Reinforces the 2018 NIA-AA Research Framework. *J Alzheimers Dis* 2019; 68:537–550

20. Lipkovich I, Dmitrienko A, Denne J, et al.: Subgroup identification based on differential effect search--a recursive partitioning method for establishing response to treatment in patient subpopulations. *Stat Med* 2011; 30:2601–2621

21. Ohnmacht AJ, Stahler A, Stintzing S, et al.: The Oncology Biomarker Discovery framework reveals cetuximab and bevacizumab response patterns in metastatic colorectal cancer. *Nat Commun* 2023; 14:5391

22. Nagpal C, Wei D, Vinzamuri B, et al.: Interpretable subgroup discovery in treatment effect estimation with application to opioid prescribing guidelines [Internet]. In: Proceedings of the ACM Conference on Health, Inference, and Learning. Toronto Ontario Canada: ACM; 2020. p. 19–29.[cited 2024 May 20] Available from: https://dl.acm.org/doi/10.1145/3368555.3384456

23. Kim DDH, Novitzky-Basso I, Kim TS, et al.: Optimal duration of imatinib treatment/deep molecular response for treatment-free remission after imatinib discontinuation from a Canadian tyrosine kinase inhibitor discontinuation trial. *Br J Haematol* 2021; 193:779–791

24. Puskarich MA, Jennaro TS, Gillies CE, et al.: Pharmacometabolomics identifies candidate predictor metabolites of an L-carnitine treatment mortality benefit in septic shock. *Clin Transl Sci* 2021; 14:2288–2299

25. Berger JO, Wang X, Shen L: A Bayesian approach to subgroup identification. *J Biopharm Stat* 2014; 24:110–129

26. Jin X, Han J: K-Means Clustering [Internet]. In: Sammut C, Webb GI, editor(s). Encyclopedia of Machine Learning. Boston, MA: Springer US; 2010. p. 563–564.[cited 2024 May 16] Available from: https://doi.org/10.1007/978-0-387-30164-8_425

27. Soussi S, Sharma D, Jüni P, et al.: Identifying clinical subtypes in sepsis-survivors with different one-year outcomes: a secondary latent class analysis of the FROG-ICU cohort. *Crit Care* 2022; 26:114

28. Duggal A, Kast R, Van Ark E, et al.: Identification of acute respiratory distress syndrome subphenotypes de novo using routine clinical data: a retrospective analysis of ARDS clinical trials. *BMJ Open* 2022; 12:e053297

29. Shehabi Y, Serpa Neto A, Howe BD, et al.: Early sedation with dexmedetomidine in ventilated critically ill patients and heterogeneity of treatment effect in the SPICE III randomised controlled trial. *Intensive Care Med* 2021; 47:455–466

30. Zampieri FG, Costa EL, Iwashyna TJ, et al.: Heterogeneous effects of alveolar recruitment in acute respiratory distress syndrome: a machine learning reanalysis of the Alveolar Recruitment for Acute Respiratory Distress Syndrome Trial. *Br J Anaesth* 2019; 123:88–95

31. Sinha P, Calfee CS, Delucchi KL: Practitioner's Guide to Latent Class Analysis: Methodological Considerations and Common Pitfalls. *Crit Care Med* 2021; 49:e63

32. Sinha P, Delucchi KL, Thompson BT, et al.: Latent class analysis of ARDS subphenotypes: a secondary analysis of the statins for acutely injured lungs from sepsis (SAILS) study. *Intensive Care Med* 2018; 44:1859–1869

33. Wendel Garcia PD, Caccioppola A, Coppola S, et al.: Latent class analysis to predict intensive care outcomes in Acute Respiratory Distress Syndrome: a proposal of two pulmonary phenotypes. *Crit Care* 2021; 25:154

34. Calfee CS, Delucchi K, Parsons PE, et al.: Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med* 2014; 2:611–620

35. Gårdlund B, Dmitrieva NO, Pieper CF, et al.: Six subphenotypes in septic shock: Latent class analysis of the PROWESS Shock study. *J Crit Care* 2018; 47:70–79

36. Tsoi KKF, Chan NB, Yiu KKL, et al.: Machine Learning Clustering for Blood Pressure Variability Applied to Systolic Blood Pressure Intervention Trial (SPRINT) and the Hong Kong Community Cohort. *Hypertension* 2020; 76:569–576

37. Knaus WA, Draper EA, Wagner DP, et al.: APACHE II-A Severity of Disease Classification System: Reply. *Crit Care Med* 1986; 14:755

38. Ferreira FL, Bota DP, Bross A, et al.: Serial Evaluation of the SOFA Score to Predict Outcome in Critically Ill Patients. *JAMA* 2001; 286:1754–1758

39. Santhakumaran S, Gordon A, Prevost AT, et al.: Heterogeneity of treatment effect by baseline risk of mortality in critically ill patients: re-analysis of three recent sepsis and ARDS randomised controlled trials. *Crit Care* 2019; 23:156

40. McKown AC, Huerta LE, Rice TW, et al.: Heterogeneity of Treatment Effect by Baseline Risk in a Trial of Balanced Crystalloids versus Saline. *Am J Respir Crit Care Med* 2018; 198:810–813

41. Goligher EC, Lawler PR, Jensen TP, et al.: Heterogeneous Treatment Effects of Therapeutic-Dose Heparin in Patients Hospitalized for COVID-19. *JAMA* 2023; 329:1066–1077

42. Granholm A, Marker S, Krag M, et al.: Heterogeneity of treatment effect of prophylactic pantoprazole in adult ICU patients: a post hoc analysis of the SUP-ICU trial. *Intensive Care Med* 2020; 46:717–726

43. Grolleau F, Porcher R, Barbar S, et al.: Personalization of renal replacement therapy initiation: a secondary analysis of the AKIKI and IDEAL-ICU trials. *Crit Care* 2022; 26:64

44. Künzel SR, Sekhon JS, Bickel PJ, et al.: Metalearners for estimating heterogeneous treatment effects using machine learning. *Proc Natl Acad Sci* 2019; 116:4156–4165

45. Xu Y, Bechler K, Callahan A, et al.: Principled estimation and evaluation of treatment effect heterogeneity: A case study application to dabigatran for patients with atrial fibrillation. *J Biomed Inform* 2023; 143:104420

46. Patel S, Green A, Wolfe Y, et al.: The Impact of Positive Fluid Balance on Sepsis Subtypes: A Causal Inference Study. *Crit Care Res Pract* 2023; 2023:e2081588

47. Nie X, Wager S: Quasi-Oracle Estimation of Heterogeneous Treatment Effects [Internet]. 2020; [cited 2023 Jan 25] Available from: http://arxiv.org/abs/1712.04912

48. Buell KG, Spicer AB, Casey JD, et al.: Individualized Treatment Effects of Oxygen Targets in Mechanically Ventilated Critically Ill Adults. *JAMA* 2024; 331:1195–1204

49. Wager S, Athey S: Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *J Am Stat Assoc* 2018; 113:1228–1242

50. Sadique Z, Grieve R, Diaz-Ordaz K, et al.: A Machine-Learning Approach for Estimating Subgroup- and Individual-Level Treatment Effects: An Illustration Using the 65 Trial. *Med Decis Making* 2022; 42:923–936

51. Seitz KP, Spicer AB, Casey JD, et al.: Individualized Treatment Effects of Bougie versus Stylet for Tracheal Intubation in Critical Illness. *Am J Respir Crit Care Med* 2023; 207:1602–1611

52. Wiemken TL, Furmanek SP, Carrico RM, et al.: Effectiveness of oseltamivir treatment on clinical failure in hospitalized patients with lower respiratory tract infection. *BMC Infect Dis* 2021; 21:1106

53. Desai RJ, Glynn RJ, Solomon SD, et al.: Individualized Treatment Effect Prediction with Machine Learning — Salient Considerations. *NEJM Evid* 2024; 3:EVIDoa2300041

54. Raghavan S, Josey K, Bahn G, et al.: Generalizability of heterogeneous treatment effects based on causal forests applied to two randomized clinical trials of intensive glycemic control. *Ann Epidemiol* 2022; 65:101–108

55. Hill JL: Bayesian Nonparametric Modeling for Causal Inference. *J Comput Graph Stat* 2011; 20:217–240

56. Qian M, Murphy SA: PERFORMANCE GUARANTEES FOR INDIVIDUALIZED TREATMENT RULES. *Ann Stat* 2011; 39:1180–1210

57. Zhang B, Tsiatis AA, Laber EB, et al.: A Robust Method for Estimating Optimal Treatment Regimes. *Biometrics* 2012; 68:1010–1018

58. Kahkoska AR, Shah KS, Kosorok MR, et al.: Estimation of a Machine Learning-Based Decision Rule to Reduce Hypoglycemia Among Older Adults With Type 1 Diabetes: A Post Hoc Analysis of Continuous Glucose Monitoring in the WISDM Study. *J Diabetes Sci Technol* 2023; 19322968221149040

59. Schulte PJ, Tsiatis AA, Laber EB, et al.: Q- and A-learning Methods for Estimating Optimal Dynamic Treatment Regimes. *Stat Sci Rev J Inst Math Stat* 2014; 29:640–661

60. Rashid NU, Luckett DJ, Chen J, et al.: High-Dimensional Precision Medicine From Patient-Derived Xenografts. *J Am Stat Assoc* 2020; 116:1140–1154

61. Hu X, Hsueh P-YS, Chen C-H, et al.: A First Step Towards Behavioral Coaching for Managing Stress: A Case Study on Optimal Policy Estimation with Multi-stage Threshold Q-learning. *AMIA Annu Symp Proc* 2018; 2017:930–939

62. Chakraborty B, Murphy SA: Dynamic Treatment Regimes. *Annu Rev Stat Its Appl* 2014; 1:447–464