Causal prediction for medical decision making: Methods and Practice

1 Background

The development of risk prediction algorithms has exploded in recent years in the medical research literature, yet few of these make their way into routine clinical practice. A primary purpose of prediction models is often stated as being to inform clinical decision making, such as whether to give a treatment to a particular patient. However, it has been shown that prediction models are typically not designed in a way that makes them a valid tool for informing such decisions, in particular because they do not appropriately accommodate the treatment decision at their design stage. Similarly, the evaluation of prediction models does not focus on how well the model would perform for the task of informing clinical decision making. A useful model should be evaluated from the perspective of its intended use and the intended users. Evaluation should consider not only predictive performance, but should also be based on costs and benefits which are defined by the clinical context and the available resources.

2 Learning objectives

By the end of the course, participants will:

- Be able to easily identify and explain to others the pitfalls of using prognostic predictions for treatment decision support.
- Be able to identify the difference between prognostic and causal prediction, and be able to outline and explain the differences to others.
- Be aware of methods for causal prediction and how these methods differ from those used for prognostic prediction.
- Understand the concept and importance of clinical utility of using these predictions for medical decision making.
- Have ideas about how to evaluate the performance of and compare causal prediction methods.

3 Target Participants

PhD students, postdocs, and faculty in biostatistics, machine learning, epidemiology, bioinformatics, and medicine interested in using risk predictions for decision making. In particular, the target participants are those involved in (or planning to be involved in) research projects involving risk prediction for decision making and who would like to learn how causal thinking can improve the development and evaluation of their models.

Participants should have a basic knowledge of statistics or machine learning and have a working understanding of R. Participants should have some familiarity with regular prediction methods, including performance metrics, as reviewed in these papers [Efthimiou et al., 2024, Cowley et al., 2019, Gerds et al., 2008].

4 Content and Structure

The course will consist of lectures and hands-on participation in the form of group work on a risk prediction project based on simulated data.

Everyone will be given the same simulated data set to work with and a briefing on the goals of the research project including steps that they should be taking along the way. Lectures and resources will be provided to prepare the groups to work and make progress on their projects.

5 Case studies

We encourage students who are developing risk prediction models as part of their research to provide information about those projects during the registration confirmation. The details of the project will be possibly used as case studies to facilitate discussion of the methods in a concrete example that is relevant to the students. If you have such an example please answer the following questions:

1. Describe the population for which the prediction model is intended to be used. Who are they, and what information is available to base the predictions on? What outcome is being predicted?

2. What decisions will be made on the basis of the prediction? What is likely to happen or change after the prediction is made?

3. What are the main challenges in the development, evaluation, or use of the prediction model?

6 Teachers

- Erin Gabriel, Professor the Section of Biostatistics at the University of Copenhagen. Her research interest is in causal inference, specifically partial identification.

- Ruth Keogh, Professor of Biostatistics and Epidemiology in the Medical Statistics Department and Co-Director of the Centre for Data and Statistical Science for Health (DASH). Ruth's research interests are in statistical methods for analysis of observational data, particularly in causal inference for time-to-event outcomes, and in applications in a range of areas of health research.

- Thomas Alexander Gerds, Professor at the Section of Biostatistics, University of Copenhagen and Steno Diabetes Center Copenhagen. Thomas' research is about the theory and the applications of statistical methods for binary, longitudinal and time-to-event data.

- Michael Sachs, Associate Professor at the Section of Biostatistics at the University of Copenhagen. His research interests are the development and evaluation of risk prediction models and biomarkers, statistical computing, and causal inference.

- Nan van Geloven, Assistant professor of Biostatistics in the department of Biomedical Data Sciences at Leiden Univeristy Medical Center, Leiden, the Netherlands. Her research interests include causal prediction.

- Karla Diaz-Ordaz is a Professor of Biostatistics at University College London, Department of Statistical Science. She is interested in causal inference, machine learning and non-parametric methods, motivated by epidemiology and clinical trials applications.

References

Ankur Ankan and Johannes Textor. Expert-in-the-loop causal discovery: Iterative model refinement using expert knowledge. In *The 41st Conference on Uncertainty in Artificial Intelligence*, 2025.

Christopher B. Boyer, Issa J. Dahabreh, and Jon A. Steingrimsson. Estimating and evaluating counterfactual prediction models, 2024. URL `https://arxiv.org/abs/2308.13026`.

Amanda Coston, Alan Mishler, Edward H Kennedy, and Alexandra Chouldechova. Counterfactual risk assessments, evaluation, and fairness. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pages 582–593, 2020.

Laura E Cowley, Daniel M Farewell, Sabine Maguire, and Alison M Kemp. Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature. *Diagnostic and prognostic research*, 3(1):16, 2019.

Johanna AAG Damen, Lotty Hooft, Ewoud Schuit, Thomas PA Debray, Gary S Collins, Ioanna Tzoulaki, Camille M Lassale, George CM Siontis, Virginia Chiocchia, Corran Roberts, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *British Medical Journal*, 353, 2016.

Orestis Efthimiou, Michael Seo, Konstantina Chalkou, Thomas Debray, Matthias Egger, and Georgia Salanti. Developing clinical prediction models: a step-by-step guide. *Bmj*, 386, 2024.

Thomas A Gerds, Tianxi Cai, and Martin Schumacher. The performance of risk prediction models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 50(4):457–479, 2008.

Julia Hippisley-Cox, Carol Coupland, Yana Vinogradova, John Robson, and P Brindle. Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: a validation study. *Heart*, 94(1):34–39, 2008.

Ruth H Keogh and Nan Van Geloven. Prediction under interventions: evaluation of counterfactual performance using longitudinal observational data. *Epidemiology*, 35(3):329–339, 2024.

Iftikhar J Kullo, Hayan Jouni, Erin E Austin, Sherry-Ann Brown, Teresa M Kruisselbrink, Iyad N Isseh, Raad A Haddad, Tariq S Marroush, Khader Shameer, Janet E Olson, et al. Incorporating a genetic risk score into coronary heart disease risk estimates: effect on low-density lipoprotein cholesterol levels (the MI-GENES clinical trial). *Circulation*, 133(12):1181–1188, 2016.

Lijing Lin, Matthew Sperrin, David A Jenkins, Glen P Martin, and Niels Peek. A scoping review of causal methods enabling predictions under hypothetical interventions. *Diagnostic and prognostic research*, 5:1–16, 2021.

J. Pearl. *Causality: Models, Reasoning, and Inference.* New York: Cambridge University Press, 2nd edition, 2009.

Audrey Poinsot, Panayiotis Panayiotou, Alessandro Leite, Nicolas CHESNEAU, Özgür Şimşek, and Marc Schoenauer. Position: Causal machine learning requires rigorous synthetic experiments for broader adoption. In *Forty-second International Conference on Machine Learning Position Paper Track*, 2025. URL `https://openreview.net/forum?id=Nr2ulBN50q`.

Michael C. Sachs, Arvid Sjölander, and Erin E. Gabriel. Aim for clinical utility, not just predictive accuracy. *Epidemiology*, 31(3):359–364, May 2020.

Barbora Silarova, Stephen Sharp, Juliet A Usher-Smith, Joanne Lucas, Rupert A Payne, Guy Shefer, Carmel Moore, Christine Girling, Kathryn Lawrence, Zoe Tolkien, et al. Effect of communicating phenotypic and genetic risk of coronary heart disease alongside web-based lifestyle advice: the INFORM randomised controlled trial. *Heart*, 105(13):982–989, 2019.

Wouter AC van Amsterdam, Nan van Geloven, Jesse H Krijthe, Rajesh Ranganath, and Giovanni Ciná. When accurate prediction models yield harmful self-fulfilling prophecies. *Patterns*, 6(4), 2025.

Nan van Geloven, Sonja A Swanson, Chava L Ramspek, Kim Luijken, Merel van Diepen, Tim P Morris, Rolf HH Groenwold, Hans C van Houwelingen, Hein Putter, and Saskia le Cessie. Prediction meets causal inference: the role of treatment in clinical prediction models. *European journal of epidemiology*, 35:619–630, 2020.

Nan van Geloven, Ruth H Keogh, Wouter van Amsterdam, Giovanni Cinà, Jesse H Krijthe, Niels Peek, Kim Luijken, Sara Magliacane, Paweł Morzywołek, Thijs van Ommen, et al. The risks of risk assessment: causal blind spots when using prediction models for treatment decisions. *arXiv preprint arXiv:2402.17366*, 2024.

Jack Wilkinson, Kellyn F Arnold, Eleanor J Murray, Maarten van Smeden, Kareem Carr, Rachel Sippy, Marc de Kamps, Andrew Beam, Stefan Konigorski, Christoph Lippert, et al. Time to reality check the promises of machine learning-powered precision medicine. *The Lancet Digital Health*, 2(12):e677–e680, 2020.

Schedule

Day 1

Morning: **Overview and introduction to risk prediction modeling.** Lectures and practical.

Afternoon: **Prediction modeling issues and methods.** Lectures and group exercises.
Validation methods, missing data, time-to-event outcomes, motivating examples.

Introduction to causal prediction. Lecture.

Day 2

Morning: **Introduction to causal inference**. Lectures, practical, group exercises.
Causal graphs, estimands, estimation methods.

Afternoon: **Formalization of causal prediction.** Lectures and group exercises.
When non-causal prediction methods fail, causal prediction estimands.

Day 3

Morning: **Estimation of causal prediction models.** Lectures and practical.
Assumptions, adjustment sets and confounding, weighting and standardisation methods, doubly-robust methods.

Afternoon: **Evaluation of causal prediction models.** Lectures and practical.
Performance metrics and their estimation.

Day 4

Morning: **Advanced topics.** Lectures.
Time to event outcomes and competing events, time varying treatments and confounding.

Case studies. Group discussion.

Afternoon: **Case studies continued.** Group discussion.

Advanced topics continued: clinical utility. Lecture.

Summary and review. With Q&A.

Exercises and practicals

Students will have time to work individually and/or in groups to apply the methods to a synthetic dataset. For several exercises, small groups will produce prediction models and send them to the teachers for evaluation.