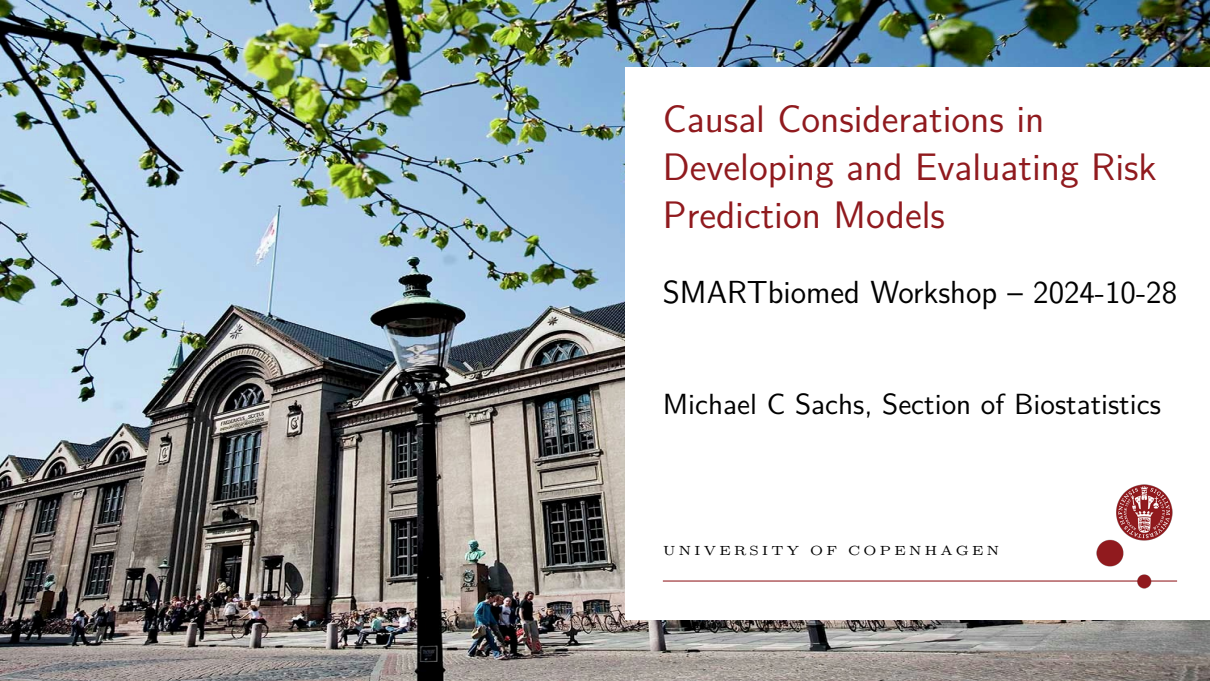


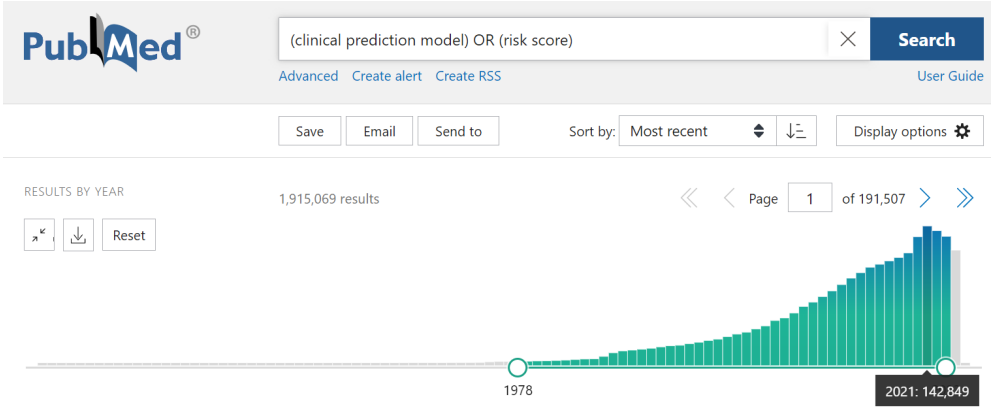
Causal Considerations in Developing and Evaluating Risk Prediction Models

SMARTbiomed Workshop – 2024-10-28

Michael C Sachs, Section of Biostatistics

UNIVERSITY OF COPENHAGEN

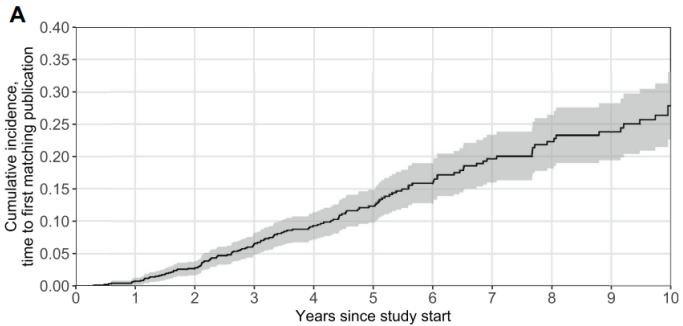
The prediction model business is booming ...



... with many more beneath the surface ...

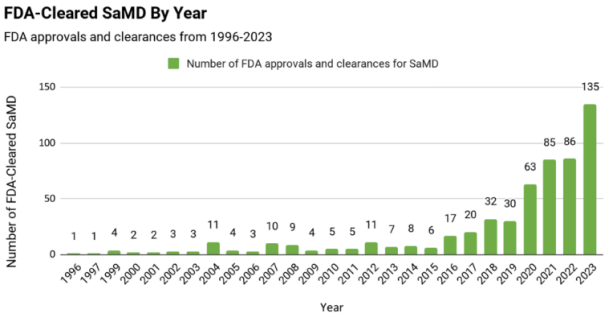
- Out of about 1000 prediction model studies registered on ClinicalTrials.gov
- Less than 1/3 of them were published after 10 years [White et al., 2024]

N. White et al. / Journal of Clinical Epidemiology 173 (2024) 111433



... and almost none of these models are used in practice.

- Graph from a company called Orthogonal, which pulled the data from FDA public documents. SaMD = Software as a medical device
- About 2/3 of these are image processing tools in Radiology



Use cases for clinical prediction models

Intended Use	**Examples**
Diagnose	eGFR, cardiac monitors
Determine treatment	HER2, Mammaprint, OncotypeDX
Inform decisions	Framingham, SCORE2
Research only	CCI, Inflammatory burden score [Axelrad et al., 2020]

- High stakes settings require a high level of rigor and high quality evidence that using the model benefits patients, on average, compared to the standard of care (clinical utility, effectiveness).
- In low stakes settings, or when the use case is unclear, a publication is well-within reach, no matter the level of rigor.

Key features of clinical utility

- **Outcomes** - health related outcomes that matter to patients. Balances both benefits and risks.
- **Actions** - a clear space of possible actions taken in response to the model
- **Comparative** - defined relative to another clearly defined strategy or the standard of care
- **Implementation** - based on things that are measurable in the clinic, able to be used in the appropriate timeframe, transparent, clear plan for updating, ...

Clearly we are talking about a *causal effect*: the effect of using the model compared to doing something else.
The statistical task of developing a model that is a good predictor of the outcome (high AUC) is not obviously linked to this goal.

Causal thinking in model development

Let X be the vector of covariates under consideration for the model, A be the treatments/actions, and Y be the outcome.
I will use $Y(A = a)$ or just $Y(a)$ to denote the potential outcome if A were intervened on to take value a.
Assuming people want to minimize their Y, the optimal action is for covariate vector x is

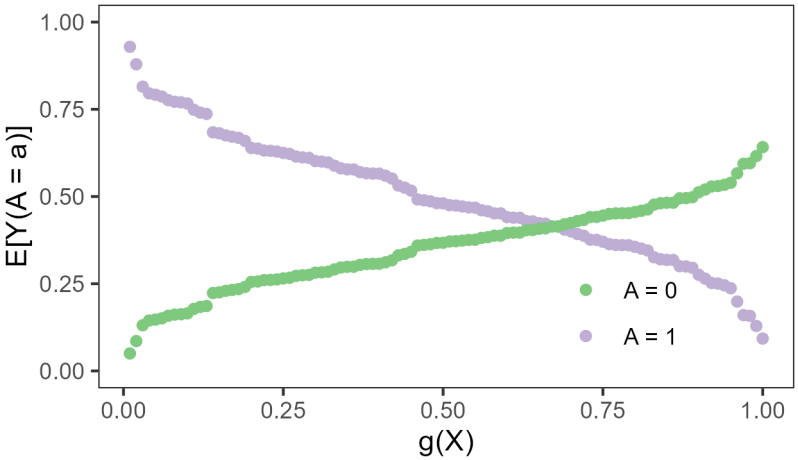
$$\text{argmin}_a\{E[Y(A = a)|X = x]\}.$$

If $a \in \{0, 1\}$, then it suffices to look at the difference

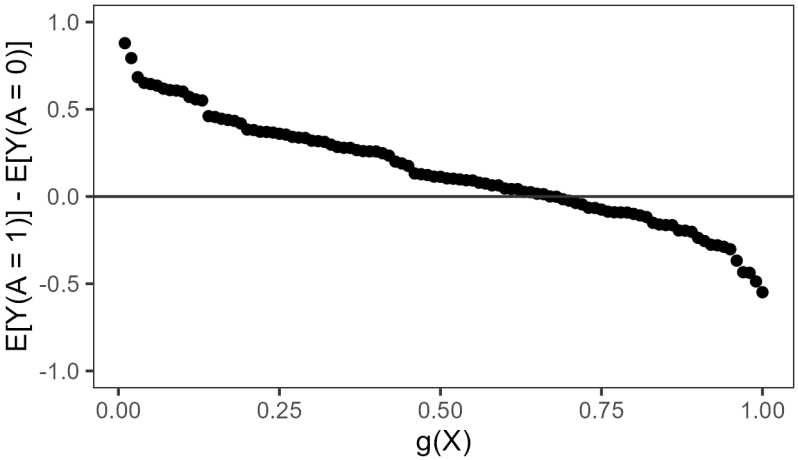
$$\text{CATE}(x) = E[Y(1)|X = x] - E[Y(0)|X = x]$$

and take action 1 if $\text{CATE}(x) < 0$ and action 0 otherwise.

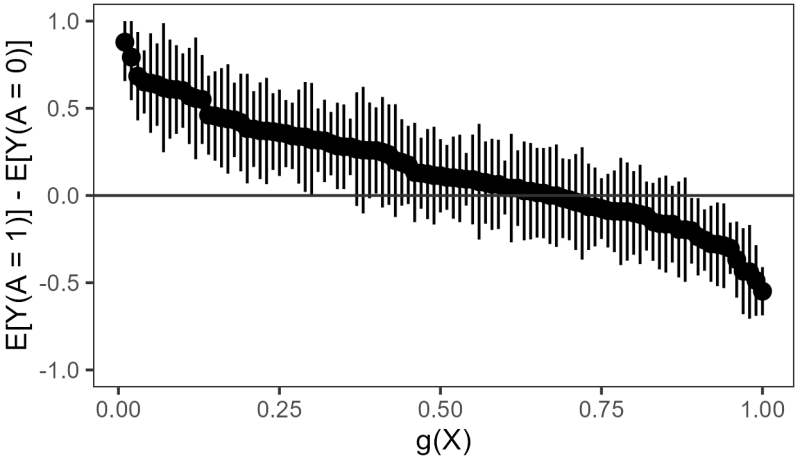
Graphically



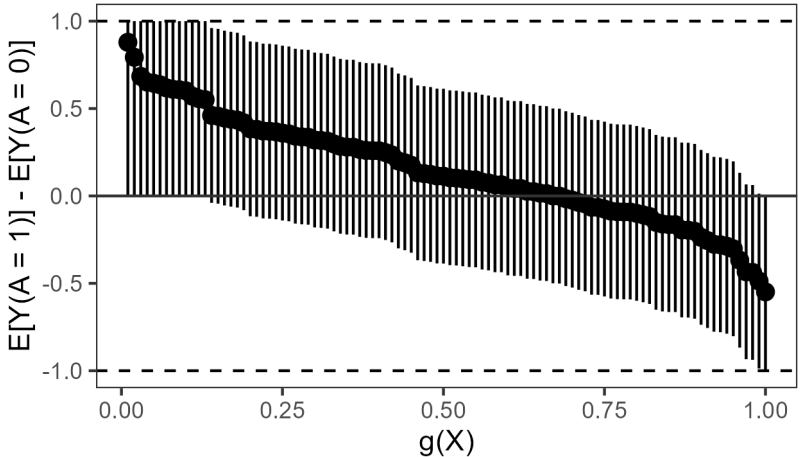
Graphically



Bars show uncertainty due to sampling variability. If the CATE is identified, this is all we have to worry about

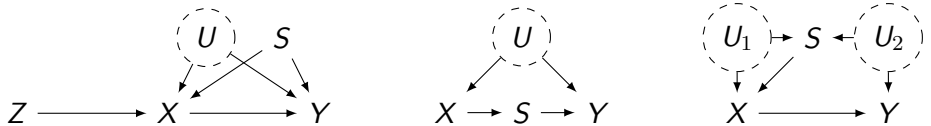


Bars show uncertainty due to *unmeasured confounders*. Without an instrument, bounds have width 1.



Idea 1: Causal variable selection

- It is now well-known that having an IV can narrow the bounds
- IV (the more the better) plus certain measured covariates can narrow further [Cai et al., 2007]
- Can we exploit the front-door criteria, surrogate experiments, others, to narrow the bounds on the CATE?



Main challenge is discovering the causal structure from observed data. One approach would be using observable constraints to falsify models, if there are any [Evans, 2012].

No universal best decision with partial identification

- Balancing benefit with risks and/or costs
- Individuals may be different decision strategies – e.g., minimize worst-case versus maximize possible benefit

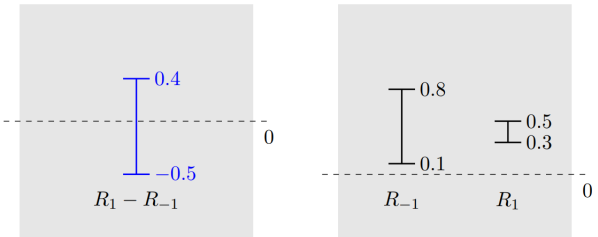


Figure 1. A toy example on slot machines. The left panel: the possible range of $R_1 - R_{-1}$; the right panel: the possible ranges of R_{-1} and R_1, respectively.

[Cui, 2021]

Idea 2: Evaluating clinical decision support in observational data

- Maybe it is too ambitious to aim for a model that directs medical action from data
  - It may be considered a high-risk medical device subject to more scrutiny
  - We do not want to take the human out of the loop
- Instead aim for a clinical decision support, provide information and let the patient/doctor decide course of action
- How to evaluate the clinical utility of such a model?

Requires some data collection.

Conduct a survey

- Compile a list of scenarios (covariate vectors) and their model output.
- Ask some (a random sample) doctors what action they would take when presented with the information.
- Estimate probability of each action for a series of covariate vectors.

Treat the use of the decision support as a stochastic intervention [Haneuse and Rotnitzky, 2013].

Example

Two scenarios and two treatments.

$g(X)$	$\hat{P}(A=1)$	$\hat{P}(A=0)$
$< .5$	.3	.7
$\geq .5$	.9	.1

Probabilities are estimated as the proportion of clinicians who recommend that action when presented with that information (or subjective probabilities).
The clinical utility of the decision support system is that of the conditional treatment policy described in the table.
This can feasibly be estimated or bounded if $E[Y(A=a)|g(X)]$ can be and then compared to the standard of care.

Summary and discussion

⋆ Causal thinking in prediction modeling can help us ask the right questions.
⋆ Close collaboration with practitioners is needed so that we can get the information we need.
⋆ Partial identification is highly relevant in both developing and evaluating prediction models.
⋆ Quantify the limits of the information in our data so that we have better go/no go criteria than AUC > 0.8

Lot of relevant work in this area covering both tools and principles: [Pu and Zhang, 2021], [Laurendeau et al., 2024], [van Geloven et al., 2020], [Rytgaard et al., 2023], [Vansteelandt and Lancker, 2024]

References I

Jordan E Axelrad, Michael C Sachs, Jonas F Ludvigsson, Ola Olén, and SWIBREG Study Group. A novel method for quantifying intestinal inflammatory burden in inflammatory bowel disease using register data. *Clinical Epidemiology*, pages 1059–1072, 2020.

Zhihong Cai, Manabu Kuroki, and Tosiya Sato. Non-parametric bounds on treatment effects with non-compliance by covariate adjustment. *Statistics in Medicine*, 26(16): 3188–3204, 2007. doi: https://doi.org/10.1002/sim.2766. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.2766`.

Yifan Cui. Individualized decision making under partial identification: ThreePerspectives, two optimality results, and one paradox. *Harvard Data Science Review*, 06 2021. doi: 10.1162/99608f92.d07b8d16. URL `https://doi.org/10.1162%2F99608f92.d07b8d16`.

Robin J Evans. Graphical methods for inequality constraints in marginalized dags. In *2012 IEEE International Workshop on Machine Learning for Signal Processing*, pages 1–6.

References II

Sebastian Haneuse and Andrea Rotnitzky. Estimation of the effect of interventions that modify the received treatment. *Statistics in medicine*, 32(30):5260–5277, 2013.

Julien D Laurendeau, Aaron L Sarvet, and Mats J Stensrud. Improved bounds and inference on optimal regimes. *arXiv preprint arXiv:2404.11510*, 2024.

Hongming Pu and Bo Zhang. Estimating Optimal Treatment Rules with an Instrumental Variable: A Partial Identification Learning Approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83(2):318–345, 03 2021. ISSN 1369-7412. doi: 10.1111/rssb.12413. URL `https://doi.org/10.1111/rssb.12413`.

Helene CW Rytgaard, Claus T Ekstrøm, Lars V Kessing, and Thomas A Gerds. Ranking of average treatment effects with generalized random forests for time-to-event outcomes. *Statistics in Medicine*, 42(10):1542–1564, 2023.

References III

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.

Stijn Vansteelandt and Kelly Van Lancker. Chasing shadows: How implausible assumptions skew our understanding of causal estimands, 2024. URL `https://arxiv.org/abs/2409.11162`.

Nicole White, Rex Parsons, David Borg, Gary Collins, and Adrian Barnett. Planned but ever published? a retrospective analysis of clinical prediction model studies registered on clinicaltrials. gov since 2000. *Journal of Clinical Epidemiology*, page 111433, 2024.