Standardization over disease risk score vs. propensity score for confounding control using random forests for model fitting

Yi Li, PhD Candidate

Department of Epidemiology, Biostatistics and Occupational Health, McGill University

ICPE at Halifax, Nova Scotia,
Aug. 27, 2023

Coauthors: Tibor Schuster (McGill), Kazuki Yoshida (Takeda), Robert Platt (McGill)

Disclosure

Funding institution: McGill University.

Potential conflict of interest

Robert Platt serves at

- Biogen: Consultant/Advisory Board
- Boehringer Ingelheim: Speaker/Honoraria (includes speakers bureau, symposia, and expert witness)
- Merck: Speaker/Honoraria (includes speakers bureau, symposia, and expert witness)
- Pfizer: Consultant/Advisory Board

Kazuki Yoshida is the Medical Director of Clinical Science – Rheumatology GI & Inflammation (GI^2) Therapeutic Area Unit at Takeda Pharmaceutical Company Limited.

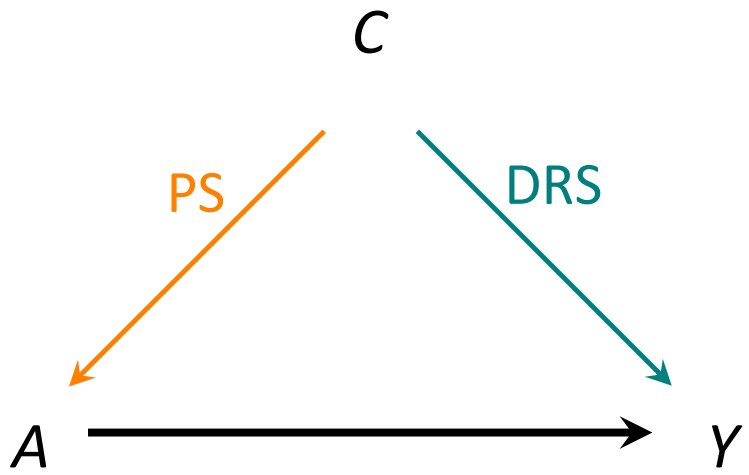
The other authors declare no conflict of interest.

Confounding summarizing score:

propensity score & prognostic score

Propensity score & Prognostic score

1. Propensity Score (PS)
   - Rosenbaum & Rubin (1983) in *Biometrika*
2. Prognostic Score / Disease Risk Score (DRS)
   - Hansen (2008) in *Biometrika*



A: treatment, *Y* : outcome, *C*: confounders + outcome predictors.

Propensity score

- PS estimation:

$$PS = P(A = 1 \mid C).$$

- Assumptions:

$$A \perp C \mid PS\ ,$$

$$(Y_{a=0}, Y_{a=1}) \perp A \mid C \Rightarrow (Y_{a=0}, Y_{a=1}) \perp A \mid PS\ .$$

- Identification of Average Treatment Effect (ATE):

$$E(Y_{a=1}) - E(Y_{a=0}) = E_{ps}\{E[Y_{a=1} \mid PS] - E[Y_{a=0} \mid PS]\}$$

$$= E_{ps}\{E[Y|do(A = 1), PS] - E_{ps}[Y|do(A = 0), PS]\}.$$

Disease risk score: more complicated

DRS is more complicated, regarding the balance it aims to achieve. Hansen (2008):

> *"If, in advance of studying a new experimental manipulation, an investigator conducts tests without the new manipulation in order better to understand the accompanying conditions and their influence on the outcome, then it is this second ideal that his or her procedure seeks to attain."*

DRS is not $P(Y = 1 \mid C)$, but $P(Y_{a=0} = 1 \mid C)$, i.e., "had the subject not been treated".

Disease risk score: Hansen (2008)

So, the balance in the DRS is will be checked in the control group.

- Definition of DRS:

$$Y_{a=0} \perp C \mid \text{DRS}.$$

Note that this holds for both control ($A = 0$) and treatment ($A = 1$) groups.

How about $Y_{a=1}$?

Disease risk score: Hansen (2008)

How about $Y_{a=1}$?

We need to consider effect measure modification.

- Assumption 1:

$$Y_{a=1} \perp C \mid DRS, M.$$

M: modifier(s).

Disease risk score: Hansen (2008)

- Assumption 2:

$$Y_{a=0} \perp A \mid \text{DRS}.$$

- Assumption 3: If there is no modifier,

$$Y_{a=1} \perp A \mid \text{DRS}.$$

What if there is modifier?

Disease risk score: when there is modifier

- Assumption 4: If there is modifier,

$$Y_{a=1} \perp A \mid DRS, M.$$

M: modifier(s).

Disease risk score

- DRS estimation:

$$DRS = P(Y_{a=0} = 1 \mid C).$$

The goal of this step is not to predict Y well, but to achieve DRS balance in Y = 0 and Y = 1 groups.

- Identification of ATE:

$$E(Y_1) - E(Y_0) = E_{drs,m} \{E[Y_1 \mid DRS, M] - E[Y_0 \mid DRS, M]\}.$$

If there is no modifier, the above will be simplified to the one without M .

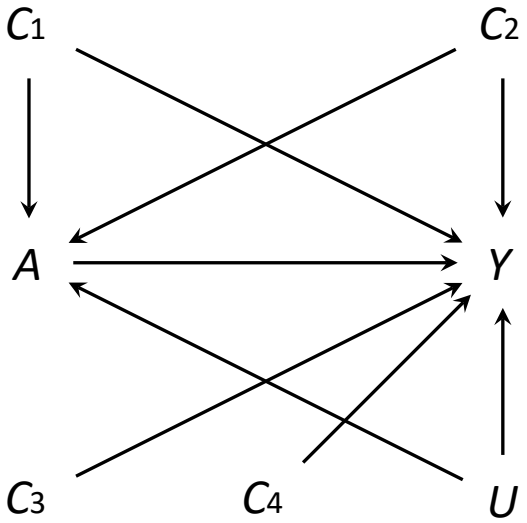
Disease risk score: Hansen (2008)

Which population will be used to fit DRS model?

- In the untreated sub-cohort: estimate E(Y | C) only in the untreated, then extrapolate to the treated cohort
  – Overfitting
- In the full cohort: estimate E(Y | A, C, M), then set everyone's A to 0 to get DRS
  – Prone to treatment effect misspecification (eg., omit A × M), the estimated DRS will carry information about the treatment effect
- Fit DRS in a separate/historical untreated cohort, then explicitly control for modification in the outcome model

Simulation Example

Data Generating Process (DGP)



C_1, C_4: modifiers

- The study population: $N = 3000$.

$$U \sim Bin(N, 0.3),$$
$$C_1 \sim Bin(N, 0.5),$$
$$C_2 \sim N(0.6, 0.2),$$
$$C_3 \sim Bin(N, 0.1),$$
$$C_4 \sim N(1, 0.2),$$
$$A \sim Bin(N, \text{expit}[-1 + 0.8\,C_1 + 0.5C_2 + 0.4U - 0.5C_1 \cdot C_2]),$$
$$Y \sim Bin(N, \text{expit}[-0.7 - 1.5A + 0.1C_1 - 0.5C_2^3 + 0.5C_3 - 0.5C_4 + 0.3U + 0.4\,A \cdot C_1 + 0.2\,A \cdot C_4 - 0.7C_1 \cdot C_2 \cdot C_4^2])$$

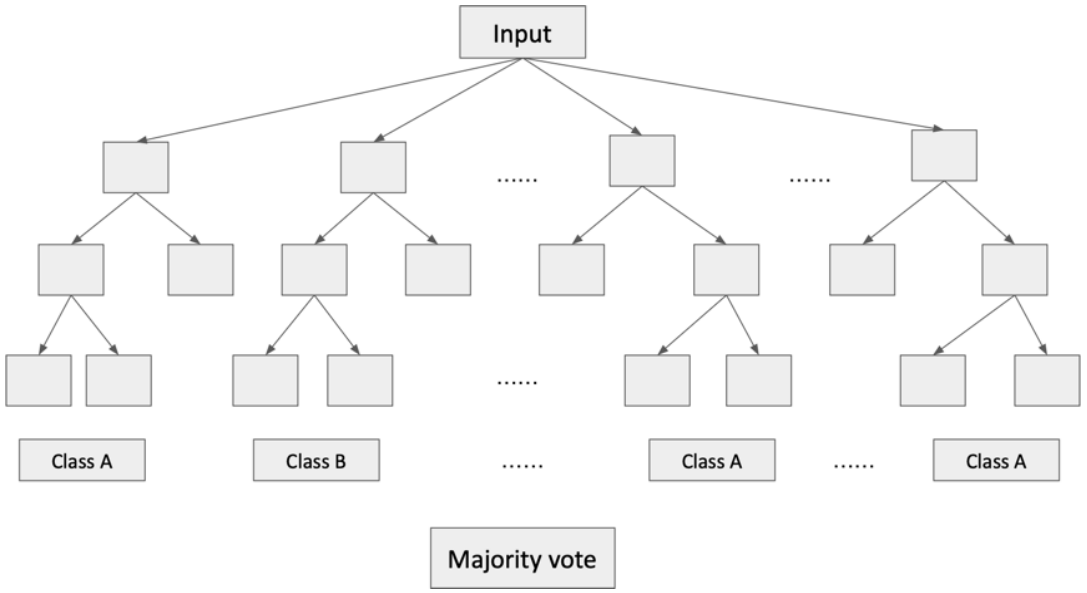
- $P(A = 1) = 0.42$, $P(Y = 1) = 0.16$.
- The separate cohort (for DRS model fit): $N = 2000$. Different parameters and parametric forms are used to generate this cohort.

Models for comparison

- PS approach:
  (1) PS model: $P(A = 1 \mid C_1, C_2)$,
  Outcome model: $P(Y = 1 \mid A, PS, C_3, C_4, A \times C_1, A \times C_4)$
  (2) Omit modification terms in outcome model
  (3) Omit modification terms, but replace with $A \times PS$ in outcome model

- DRS approach:
  (1) In the separate cohort, among the untreated, fit DRS model $P(Y = 1 \mid C_1, C_2, C_3, C_4)$. Then in the study population, fit $P(Y = 1 \mid A, DRS, A \times C_1, A \times C_4)$
  (2) Omit modification terms in outcome model
  (3) Omit modification terms , but replace with $A \times DRS$ in outcome model

Random forests



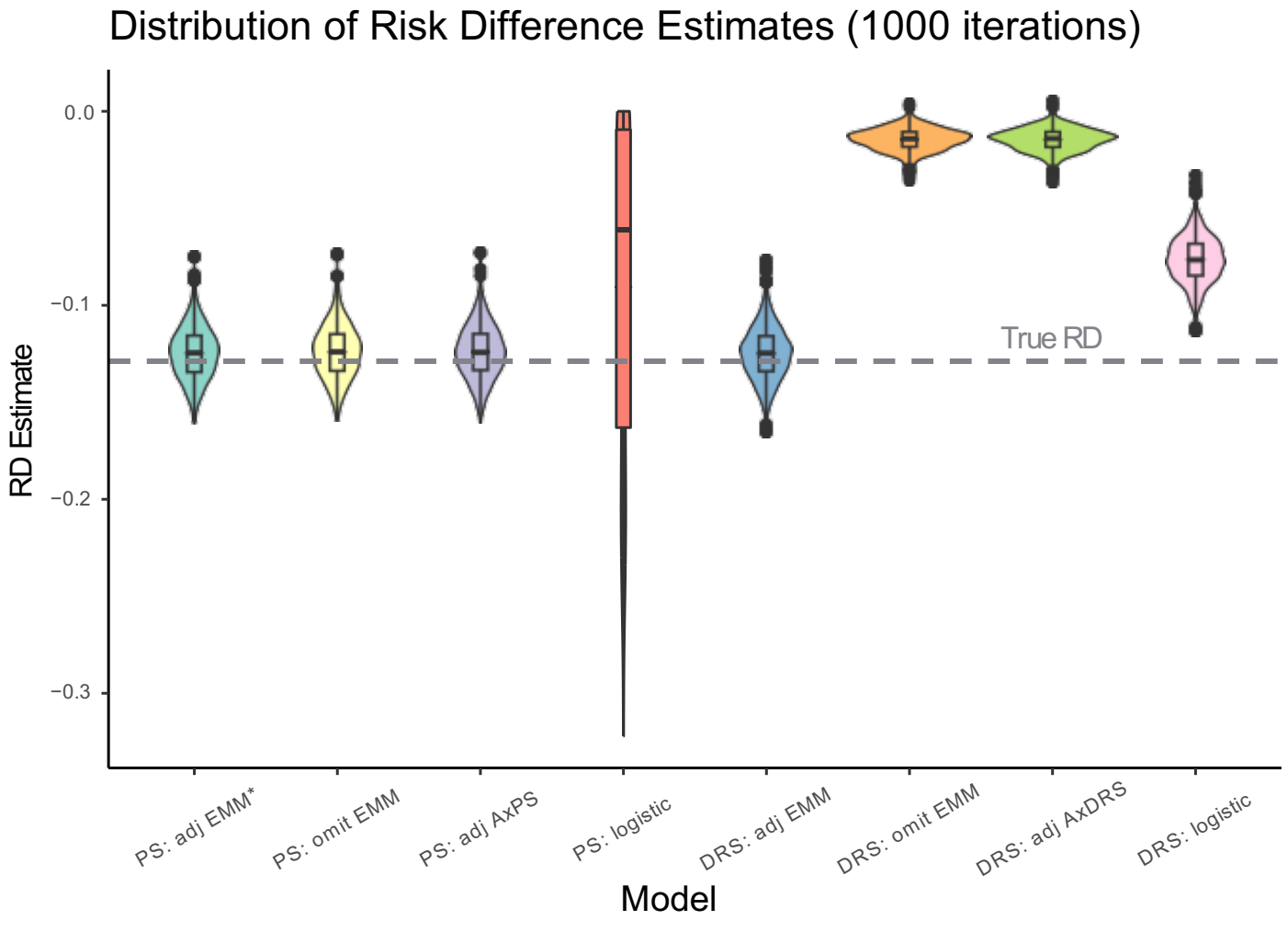
The predictions of the terminal ensemble of trees are combined by taking the most common predictions among all trees.

- Nonparametric
- Fewer parameters, easier to tune
- Often has the best classification performance among large collections of much fancier methods.

Will compare to logistic model:

– Adjust for $A \times C_1$ *and* $A \times C_4$, but no other complicated forms

Results †



†True RD = -0.129; increase sample size to sufficiently large (N = 1,000,000), fit the correct outcome model. Logistic model: use for both PS/DRS and outcome models, adjusting for $A \times C_1$ and $A \times C_4$, but no other complicated cubic or 3-way interaction terms. * EMM: effect measure modification.

Summary

- DRS is an alternative of PS.

  – It has been shown to have comparable performance as PS, and sometimes outperforms PS (e.g., when treatment is rare, a new therapy, ...).[§] [¶]
- Misspecifying modification terms has a larger impact on DRS than on PS.
- Future research: develop more robust approach to fit DRS in the same study sample.

[§] Glynn, Gagne, & Schneeweiss. PDS (2012)

[¶] Richardson, Keil, Kinlaw, & Cole. AJE (2019)