On the aggregation of historical prognostic scores for causal inference

TL Nguyen[1,2,3,4], GS Collins [5], Pellegrini F [6], KGM Moons[1,7], TPA Debray[1,7]

[1]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands
[2]Laboratory of Biostatistics, Epidemiology, Clinical Research and Health Economics, Faculty of Medicine, University of Montpellier, Montpellier, France
[3]Department of Biostatistics, Faculty of Pharmacy, University of Montpellier, Montpellier, France
[4]Department of Pharmacy, Nîmes University Hospital Centre, Nîmes, France
[5]Centre for Statistics in Medicine, Botnar Research Centre, University of Oxford, Oxford, UK
[6]Biogen International GmbH, Zug, Switzerland
[7]Cochrane Netherlands, University Medical Center Utrecht, The Netherlands

UMC Utrecht

Background

- Causal inference in non-randomized studies requires statistical methods to adjust for confounding biases
- Propensity score (PPS) analysis aims to assess treatment effect among patients who have the same probability of receiving the treatment [1].
- Prognostic score (PGS) analysis aims to assess treatment effect among patients who have the same predicted prognosis for a given reference treatment [2].
- PPS and PGS assume absence of hidden bias, and therefore need to adjust for many confounders. This may be problematic when data are sparse.
- It has been recommended to develop PGS from a separate large sample of control individuals [2], rather than the (limited) data at hand. Because individual participant data (IPD) for historical controls are not always available, researchers may consider using published PGS rather then developing them from scratch.

Aim

We introduce the idea of aggregating multiple published PGS, and tailoring them to the control arm of a non-randomized treatment study. This approach allows to adjust for a large number of confounders in small non-randomized data sets, and may therefore reduce bias in treatment effect estimates.

Principles of PGS Analysis

Consider a non-randomized study with treatment Z, observed covariates $\mathbf{X}$ and effect modifier $m(\mathbf{X})$. The potential outcome for treatment $Z = z$ is defined as Y_z.

- The PGS is defined as $\hat{\psi}(\mathbf{X}) = \hat{E}(Y_0|\mathbf{X})$, and is estimated in patients where $Z = 0$.
- It is assumed that $Y_0 \perp Z|\mathbf{X}$ (i.e. absence of hidden bias) and that $Pr\left[0 < Pr(Z = 1|\hat{\psi}(\mathbf{X}), m(\mathbf{X})) < 1\right] = 1$.
- The average treatment effect in the entire population (**ATE**) is given as

$$E\left[E\left(Y_1|\hat{\psi}(\mathbf{X}), m(\mathbf{X})\right) - E\left(Y_0|\hat{\psi}(\mathbf{X}), m(\mathbf{X})\right)\right]$$

- The average treatment effect in the treated population (**ATT**) is given as

$$E\left[E\left(Y_1|\hat{\psi}(\mathbf{X})\right) - E\left(Y_0|\hat{\psi}(\mathbf{X})\right)|Z = 1\right]$$

- The ATE and ATT can be calculated by adopting (full) matching [3].

Aggregation of Published PGS

Consider a set of M published prognostic scores for predicting a binary outcome: $\hat{\Psi} = \left[\hat{\psi}_1, \hat{\psi}_2, \ldots, \hat{\psi}_M\right]$. Aggregation is achieved by optimizing the following function in the control subjects from the non-randomized study [4]:

$$argmin\left[Y \log(P) + (1 - Y)\log(1 - P)\right]$$

where

$$P = \mathrm{g}^{-1}\left(\theta_0 + \sum_m \theta_m g(\hat{\psi}_m(\mathbf{X}))\right)$$

with unknown parameters $\theta_0, \theta_1, \ldots, \theta_M$ and the constraint $\theta_m \geq 0$. Here, we use the logistic link function for $g(\cdot)$.

Case study

Subgroup analysis in a clinical trial to compare the effectiveness of controlled asthma (Ca) versus partly controlled asthma (PCa) on the 1-year risk of exacerbation in asthmatic patients.

- Total sample size: 147 (Ca) + 134 (PCa)
- Total number of events: 23 (Ca) + 21 (PCa)
- Total number of measured covariates: $p = 20$
- 3 published PGS: Schatz, Eisner, TENOR

	ATT	95% CI	p	dfr	EPV
Same-sample PGS	0.021	-0.054 to 0.097	7	8	3.3
Published PGS (Schatz)	0.018	-0.069 to 0.105	3	0	-
Published PGS (Eisner)	0.027	-0.054 to 0.108	2	0	-
Published PGS (TENOR)	-0.002	-0.088 to 0.085	10	0	-
Aggregated PGS	0.019	-0.060 to 0.098	15	4	7.7

p = number of covariates in the PGS; dfr = number of estimated parameters for calculating the ATT; EPV = Events per variable

Simulation Study

- We used a Gibbs sampler to simulate data from the case study dataset (binary outcome, 2 treatments, 9 covariates)
  - 1 non-randomized study for estimating the causal effect (sample size 100)
  - 3 historical cohorts for the 'published' PGS (sample size: 2000)
- PGS: All PGS were derived using LASSO. For published PGS, we allowed for bias in the published PGS by modeling a different outcome (rather than the primary outcome of interest), and/or by limiting the available covariates
- 5000 iterations for each simulation.

Estimation method	ATT (SD)	Bias	MSE (x100)
Same-sample PGS	0.074 (0.081)	-0.013	0.6692
Published PGS (scen 1)	0.080 (0.077)	-0.007	0.6047
Published PGS (scen 2)	0.053 (0.081)	-0.034	0.7686
Published PGS (scen 3)	0.058 (0.083)	-0.029	0.7664
Published PGS (scen 4)	0.034 (0.084)	-0.053	0.9829
Aggregated PGS (scen 1)	0.083 (0.075)	-0.004	0.5713
Aggregated PGS (scen 2)	0.084 (0.076)	-0.003	0.5835
Aggregated PGS (scen 3)	0.082 (0.077)	-0.005	0.6020
Aggregated PGS (scen 4)	0.080 (0.078)	-0.007	0.6202

scen 1: identical outcome & covariates, scen 2: identical outcome & different covariates, scen 3: different outcome & identical covariates, scen 4: different outcome & different covariates

Conclusions

- Aggregating published PGS allows to adjust for many covariates at a limited expense (in terms of required degrees of freedom)
- Aggregating published PGS allows an unbiased estimation of treatment effects from non-randomized study data, even if published PGS have omitted important covariates and/or estimate different (but related) outcomes.

References

1. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41–55.
2. Hansen BB. The prognostic analogue of the propensity score. Biometrika. 2008 Feb 4;95(2):481–8.
3. Hansen BB. Full Matching in an Observational Study of Coaching for the SAT. Journal of the American Statistical Association. 2004 Sep;99(467):609–18.
4. Debray TPA, Koffijberg H, Nieboer D, Vergouwe Y, Steyerberg EW, Moons KGM. Meta-analysis and aggregation of multiple published prediction models. Stat Med. 2014 Jun 30;33(14):2341–62.

Contact

Please do not hesitate to get in touch with us during the ISCB!

Thomas Debray, PhD
Assistant Professor
T.Debray@umcutrecht.nl

This work was supported by the Netherlands Organization for Scientific Research (91617050 and 91215058)