UMC Utrecht

The use of prognostic scores for causal inference with general treatment regimes

Long Nguyen & Thomas Debray*

University Medical Center Utrecht

Disclaimer

The methods in this talk were conceived by Long

Causal inference in non-randomized studies

- Randomized trials
  - Generally preferred for assessing treatment effects
  - May not always be necessary, appropriate, possible or adequate (Black. BMJ 1996)

- The appeal of non-randomized studies of interventions
  - Good external validity
  - Access to long-term outcomes
  - Can provide evidence of “real world” effectiveness
  - Can provide insight into delivery of care
  - Potential to study rare diseases

Causal inference in non-randomized studies

Addressing confounding bias with binary treatment exposures

- **Propensity Score Analysis (PPS)**
  Restore the balance in the subjects' baseline covariate distributions across the different treatment exposures
- **Prognostic Score Analysis (PGS)**
  Restore the balance in the subjects' baseline prognosis, rather than their covariates *per se*

Here, we extend PGS to compare multiple treatment exposures

Case study

International stroke trial (RCT)

- 19 435 stroke patients
- 2x2 factorial design
  - Aspirin vs placebo
  - Heparin vs non-Heparin
- Death or dependence at 6 months

Subgroup analysis in 9720 patients of the Aspirin arm

- **48 treatment exposures**: 1 to 48h delay to Aspirin administration
- Severe covariate imbalance for delay < 10h and for delay >35h (w.r.t. reference delay of 24h)

Generalized Prognostic Score Analysis

The prognostic model $\psi_r(\mathbf{X})$ is estimated in a reference treatment group $\mathrm{Z} = z_r$ (e.g. usual care) with baseline covariates $\mathbf{X}$ and outcomes Y_r

- $\psi_r(\mathbf{X})$ indicates the expected outcome (risk) if treated according to z_r
- We assume absence of hidden bias: $Y_r \perp Z|\mathbf{X}$
  i.e. all important confounders are measured
- We assume (a relaxed form of) positivity: $0 < \Pr(\mathrm{Z} \mid \psi_r(\mathbf{X}))$

Generalized Prognostic Score Analysis

The prognostic model is applied to *all* other subjects $\mathrm{Z} \neq z_r$

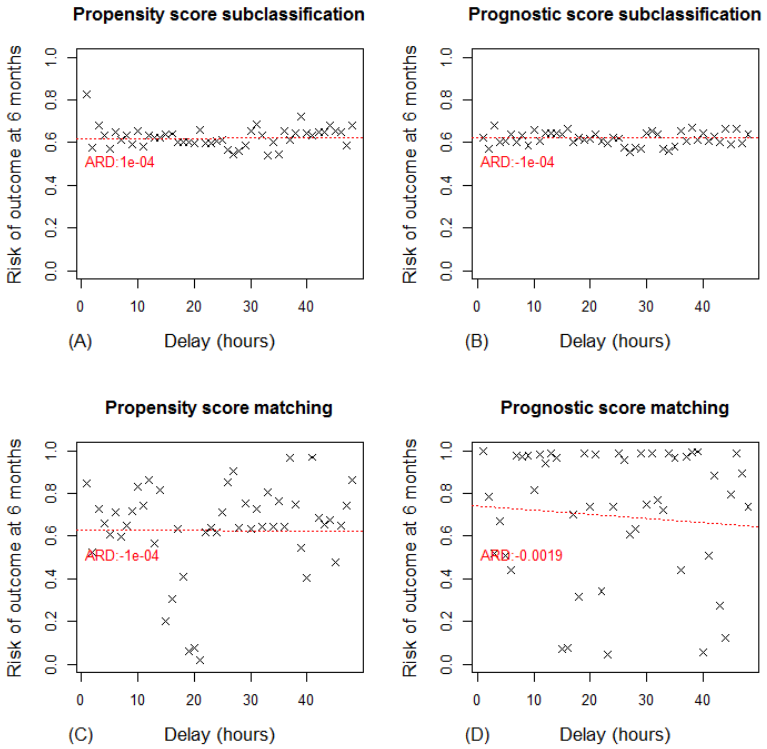
- Use $\psi_r(\mathbf{X})$ to estimate the potential outcome for $\mathrm{Z} = z_r$
- Treatment groups are matched/subclassified according to $\psi_r(\mathbf{X})$
- The ATT is simply given as $E(Y_s - Y_r | Z = z_s)$, and is calculated in the sample where every patient from z_s is matched to one or more patients from z_r
- For ATE and ATR, matching also needs to consider distribution of effect modifier(s)

Case study

- Derivation of the prognostic score $\psi_r(\mathbf{X})$
  - **Population**: patients with aspirin administration at 24 hours (z_r)
  - **Sample size**: 809 stroke patients (483 events)
  - **Outcome**: death or dependence at 6 months
  - **Covariates**: age (RCS), systolic blood pressure (RCS), sex, consciousness, previous computerised tomography, visible infarct at CT-scan, stroke subtype, atrial fibrillation, Aspirin intake within the previous 3 days, and all function deficits
  - **Analysis**: logistic regression (MLE)

- Optimal full matching and subclassification of the prognostic score

Case study



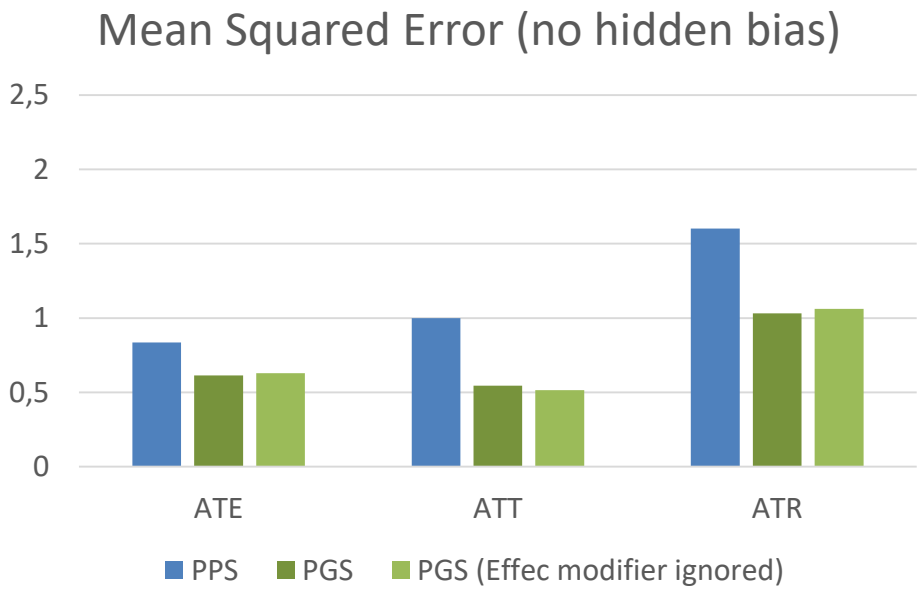
Estimated risk of death or dependence at 6 months, given delay to Aspirin administration in stroke patients, by propensity (LEFT) and prognostic (RIGHT) score analyses. ARD, absolute risk difference per hour of delay

No evident effect of the delay to Aspirin administration on the risk of death or dependence at 6 months.

Simulation study

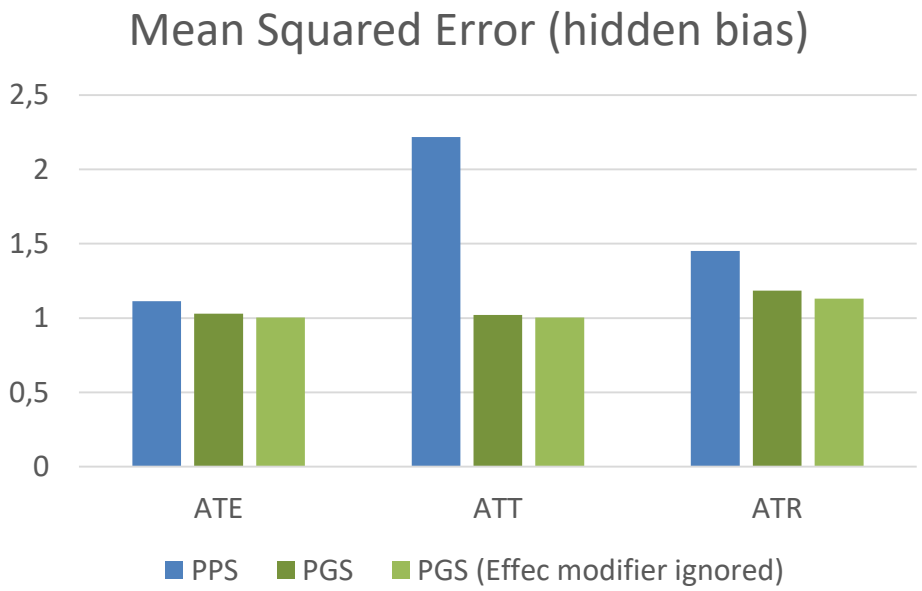
- 1000 non-randomized studies of N=500
- 10 independent variables (5 continuous & 5 binary)
  - 1 effect modifier
  - 9 prognostic variables
- 3 treatment exposures A, B, C
  - Exposure is defined by a selection of the covariates (different confounder set for each treatment)
- Reference treatment effects were estimated in a “super-population” of N = 5,000,000
- Comparison of generalised propensity score analysis versus generalised prognostic score analysis

Simulation study



All comparisons are based on optimal matching.
Treatment effects are B vs A (similar results for other comparisons)

Simulation study



All comparisons are based on optimal matching.
Treatment effects are B vs A (similar results for other comparisons)

Concluding remarks

Generalised PPS and PGS analysis tend to yield similar results. However:

- Generalised PPS analysis
  - Generally leads to higher variance in treatment effect estimates
  - Does not need to account for effect modifiers

- Generalised PGS analysis
  - Less dependent on the positivity assumption
  - Possibly more prone to over-fitting
  - Estimation of standard errors time-consuming (as it requires bootstrapping)

Questions?

We also have a poster on the development of more robust prognostic scores:

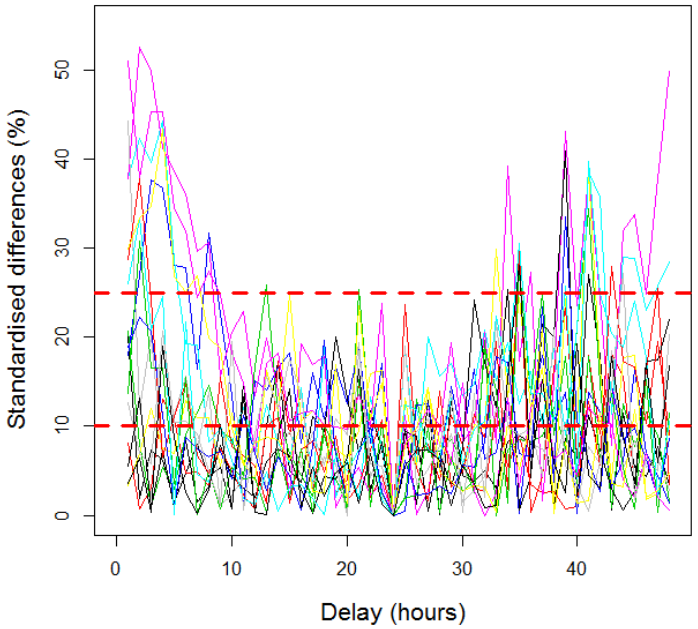
“On the aggregation of historical prognostic scores for causal inference”

@TPA_Debray

Extra slide - Case study

Confounding

- Covariate imbalance as expressed in standardised mean difference (reference = delay of 24-hours)
- Strong covariate imbalance for delay < 10 and for delay >35
- Treatment effects may particularly be prone to bias for these exposures



Generalized Prognostic Score Analysis to adjust for confounding