Prognostic score methods for the estimation of the average causal effect

Chamika Porage and Ingeborg Waernbaum

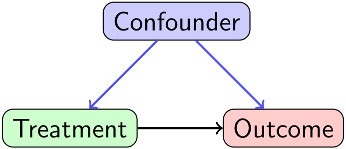
Department of Statistics, Uppsala University

25 March 2026

Accepted International Journal of Biostatistics

Background

- Goal: estimating causal effect
- Ideal: randomized experiment
- Reality: observational study
- Challenge: confounding variables
- Solution: adjust for confounders



Limitation of a Single Prognostic Score

- Previous approach: predict outcome without treatment (Hansen, 2008)
- Treatment effects may vary across individuals
- Important information may be lost
- Not sufficient for estimating a standard parameter: average causal effect (ACE)

Contribution

- Propose full prognostic score (FPGS)
- Uses two predictions: control and treatment
- Enables better adjustment for confounding when estimating the ACE
- Applicable with both standard and machine learning (ML) -based estimators

Prognostic score methods for the estimation of the average causal effect

Chamika Porage and Ingeborg Waernbaum

Abstract
The prognostic score (PGS) is a function of observed covariates that summarizes covariates' association with potential responses. In the current study, we propose a full prognostic score (FPGS), an extension of the PGS that integrates individual prognostic scores to account for confounding adjustments in causal inference. Under effect modification, we show that FPGS and a version of FPGS using conditional expectations of the outcomes, meet the sufficiency condition for confounding adjustment to estimate the average causal effect. We present a general algorithm to implement the FPGS approach for estimation by applying linear regression, random forest regression, and XGBoost regression. When determining the average causal effect, we incorporate FPGS into semiparametric estimators including regression imputation, simple stratification, and targeted maximum likelihood estimation (TMLE). The finite-sample properties

Outline

1. Theory and proposed method
2. Estimation methods
3. Simulation study
4. Empirical Analysis
5. Key Insights

Potential Outcomes and ACE

For each individual:

Treatment	**Outcome**
Treatment ($T = 1$)	$Y(1)$
No Treatment ($T = 0$)	$Y(0)$

Key challenge: We observe only one outcome

Goal: Estimate the ACE

$$\text{ACE} = \text{E}\left(Y(1) - Y(0)\right)$$

Key Assumptions

- All individual characteristics are measured (No unmeasured confounding)
- Each individual could receive either treatment (Overlap)

Prognostic Scores, $\psi_t(X)$

$$Y(t) \perp X \mid \psi_t(X), \quad t = 0, 1$$

- Prognostic score summarizes outcome information
- Individuals with same score are comparable
- Can be used to estimate treatment effects

Full Prognostic Score (FPGS)

Definition

$$\psi(X) = \{\psi_0(X), \psi_1(X)\}$$

- Combine control and treatment prognostic scores
- Captures both potential outcomes
- Provides richer information for each individual

Full Prognostic Score (FPGS)

Using conditional expectations

$$\psi_\mu(X) = \big(\mu_0(X), \mu_1(X)\big)$$

- $\mu_0(X)$: expected outcome without treatment
- $\mu_1(X)$: expected outcome with treatment

Proposed Method

Observed data: outcome, treatment, covariates

↓

First-step regression

Fit separate outcome models in treated and control groups

↓

Predicted outcome under treatment and under control

↓

Construct the FPGS

(some estimators can use the FPGS in this stage, e.g. matching on the FPGS)

↓

Second-step regression using FPGS

Fit separate models in treated and control groups

↓

Fitted treated and control outcomes for each individual

↓

Estimated ACE

Estimation Methods

Why Multiple Methods?

- Different methods have different strengths
- Some may be robust to model misspecification
- FPGS can be used across all these methods

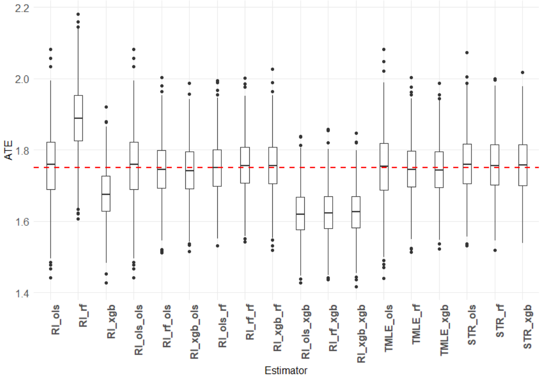
- Regression Imputation (RI)
- Targeted Maximum Likelihood Estimation (TMLE)
- Stratification
- Combined with both standard and ML methods; OLS, random forest, XGBoost

Simulation Design

- All simulations include: confounders, effect modifiers
- **Setup:** $n \in \{500, 1000, 3000\}$, 1000 replications
- **Simulation 1:**
  - Linear relationships
  - Continuous covariates
- **Simulation 2:**
  - Discrete covariates
- **Simulation 3:**
  - Complex version of first simulation
  - Nonlinear complex relationships

Simulation Results

- OLS-based estimators show consistently low bias and MSE
- ML-based estimators have higher bias in small samples
- Performance improves with larger sample sizes
- FPGS combined with ML improves performance in complex settings
- Nonlinear setting shows clear differences between methods



Empirical Analysis

- Data: NHANES (2007–2008)
- Effect of smoking on blood lead levels
- Study population: Males ($N = 1392$)
- Treatment:
  - Smokers ($N = 386$): daily smokers
  - Non-smokers ($N = 1006$): < 100 cigarettes lifetime, no smoking in past 30 days
- Outcome:
  - Blood lead level ($\mu g/dL$)
- Covariates:
  - Age, education, income, family size
  - Marital status, army service, birth country

Empirical Results and Sensitivity Analysis

- Estimated effects are consistently positive across methods
- Effect size ranges from 0.80 to 1.00 μg/dL
- FPGS estimators show higher variability
- Sensitivity analysis using negative control outcome
  - Estimated effects close to zero
  - Suggests limited unobserved confounding

Key Insights

- FPGS introduced
- Sufficient for confounding adjustment under effect modification to estimate ACE
- Works with both classical and ML methods
- No single estimator is uniformly optimal
- Linear settings: OLS-based estimators perform well
- Complex settings: flexible methods can better capture structure

References

Anoke, S. C., Normand, S.-L., and Zigler, C. M. (2019). Approaches to treatment effect heterogeneity in the presence of confounding. *Statistics in Medicine*, 38(15):2797–2815.

Bahamyirou, A., Schnitzer, M. E., Kennedy, E. H., Blais, L., and Yang, Y. (2022). Doubly robust adaptive lasso for effect modifier discovery. *The International Journal of Biostatistics*, 18(2):307–327.

Hansen, B. B. (2008). The prognostic analogue of the propensity score. *Biometrika*, 95(2):481–488.

Lee, M.-J. and Lee, S. (2022). Review and comparison of treatment effect estimators using propensity and prognostic scores. *The International Journal of Biostatistics*, 18(2):357–380.

Naimi, A. I., Mishler, A. E., and Kennedy, E. H. (2023). Challenges in obtaining valid causal effect estimates with machine learning algorithms. *American Journal of Epidemiology*, 192(9):1536–1544.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688.

Waernbaum, I. and Pazzagli, L. (2023). Model misspecification and bias for inverse probability weighting estimators of average causal effects. *Biometrical Journal*, 65(2):2100118.