Utilization of High-Dimensional Propensity Score and Targeted Learning With Machine Learning in Complex, Real-World Data to Improve Causal Effect Estimation

Poster Code: MSR48

Scan QR

Di Zhang,[1] Ying Zhang,[1] Sung-Woo Ahn,[2] Susan Gruber,[3] Mark van der Laan,[4] Ravi Iyer,[1] Shoshana Reshef,[5] Marc Tian[1]

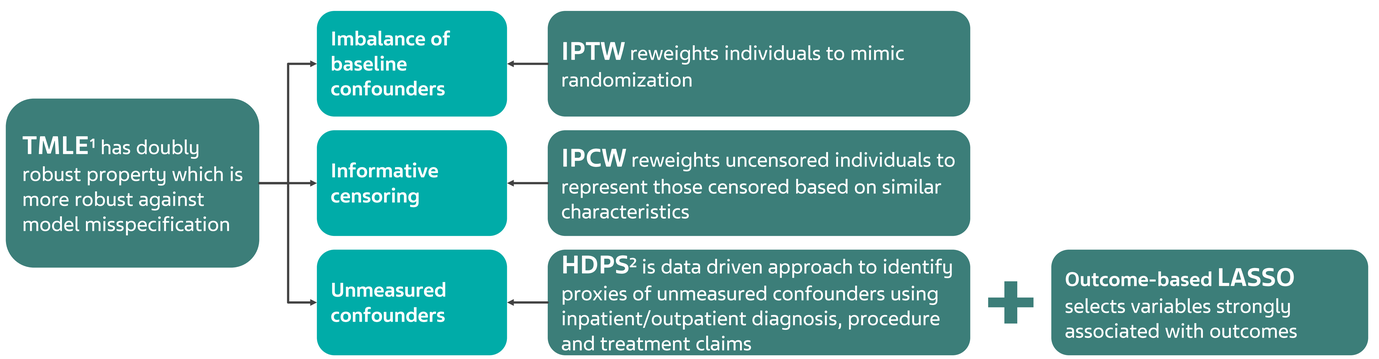
[1] Teva Branded Pharmaceutical Products R&D, Inc., West Chester, PA, United States; [2] KMK Consulting, Morristown, NJ, United States; [3] TL Revolution, Cambridge, MA, United States; [4] University of California, Berkeley, CA, United States; [5] Teva Branded Pharmaceutical Products R&D, Inc., Parsippany, NJ, United States

Objective: To examine whether integrating high-dimensional propensity score (HDPS) and/or targeted maximum likelihood estimation (TMLE) with machine learning (ML) produces robust causal estimates, specifically in patients with non-valvular atrial fibrillation (NVAF) and hypertension, on cardiovascular risks, using the MarketScan® database

Background

- Common issues in causal effect estimation using complex real-word data (RWD) include:
  - Lack of randomization → Imbalance of baseline confounders
  - Differential follow-ups between comparison groups → Informative censoring
  - Secondary use of data for other purposes → Unmeasured confounders
  - Non-linear relationship and interactions among variables → Model misspecification
- IPTW, IPCW, HDPS, and TMLE offer robust approaches to address various biases, enhancing the validity of causal inference estimate in observational studies

Statistical Methods to Account for Multiple Sources of Bias



IPTW = inverse probability treatment weight, IPCW = inverse probability censoring weight, LASSO = least absolute shrinkage and selection operator.

Conclusion

HDPS and TMLE with ML may reduce bias and produce robust causal estimates when using healthcare claims data

Study Design

Retrospective Cohort Study Using MarketScan® Claims

Population	Patients with NVAF and hypertension on stable dose of Calcium Channel Blocker (CCB)
Endpoint	Time-to-hospitalization due to composite of ischemic stroke or systemic embolism
Treatments	Group A: Anticoagulant A/CCB (at specified dosage levels), N=3471 Group B: Anticoagulant B/CCB (at specified dosage levels), N=2357
Index date	Rx claims date of anticoagulant A or B
Follow-up	1 year post index
Intercurrent events	Treatment discontinuation, treatment switching, dosage change (While on treatment strategy)
Population-level summary	Average treatment effect (ATE) measured by hazard ratio (Cox), or cumulative risk ratio/difference (TMLE)

Results

Data Characteristics

- Patients in comparator group B were older and had more comorbidities
- We observed shorter average follow-up time in comparator group B due to intercurrent events

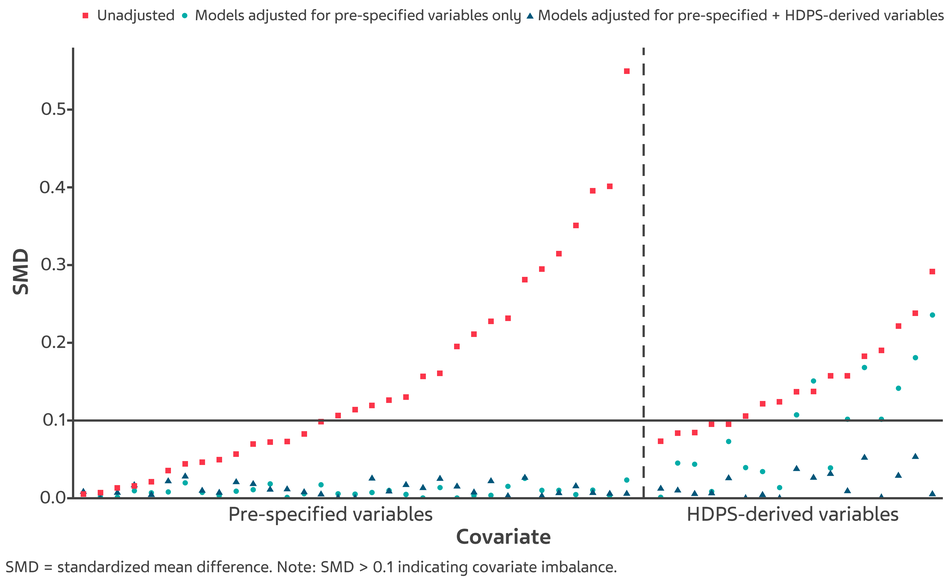
Table 1. Treatment Effect Estimates by Statistical Methods

Methods	Covariate Adjustment	Effect Estimate	95% CI
Unadjusted	NA	HR: 0.48	(0.33, 0.70)
IPTW	36 pre-specified	HR: 0.65	(0.41, 1.02)
IPTW + IPCW	36 pre-specified	HR: 0.73	(0.46, 1.16)
	36 pre-specified + 33 key HDPS derived variables	HR: 0.81	(0.51, 1.28)
TMLE	36 pre-specified + 33 key HDPS derived variables	CRR* at 1 year: 0.82	(0.46, 1.17)
		CRD at 1 year: −0.005	(−0.015, 0.006)

CRR = cumulative risk ratio, CRD= cumulative risk difference, CI = confidence interval, HR = hazard ratio, NA = not applicable.
*When rare events, CRR can approximate HR numerically; TMLE was implemented using R package *survtmle*.

- When accounting for each additional source of confounding, the HRs shift towards null
- Estimates produced by TMLE and IPTW+IPCW with HDPS-derived variables from RWD align with findings from a previous clinical trial with similar endpoints[3]
- HDPS-derived variables reveal potential (proxies of) unmeasured confounders that are imbalanced at baseline and are not accounted for in pre-specified confounders:
  - Iron deficiency anemia
  - Prior electrocardiogram procedure
  - Long-term (current) drug therapy for chronic or long-term conditions
  - Hospital procedure for arterial catheterization

Figure 1. Covariate Balance for Pre-Specified and HDPS-Derived Variables



SMD = standardized mean difference. Note: SMD > 0.1 indicating covariate imbalance.

Acknowledgments

The authors would like to thank Nitzan Shahar, Filipa Markotic and Irmela Gabriel from Teva Pharmaceuticals for clinical insights regarding the interpretation of the study results.

Disclosures

This study was supported by funding from Teva Branded Pharmaceutical Products R&D, Inc. Sung-Woo Ahn, Susan Gruber, and Mark van der Laan have received consulting fees and/or honoraria from Teva Pharmaceuticals. Di Zhang, Ying Zhang, Ravi Iyer, Shoshana Reshef, and Marc Tian are employees and stockholders of Teva Pharmaceuticals.

Abbreviations

ATE = average treatment effect, CCB = calcium channel blocker, CI = confidence ratio, CRD = cumulative risk difference, CRR = cumulative risk ratio, HDPS = high-dimensional propensity score, HR = hazard ratio, IPTW = inverse probability treatment weight, IPCW = inverse probability censoring weight, LASSO = least absolute shrinkage and selection operator, ML = machine learning, NA = not applicable, NVAF = non-valvular atrial fibrillation, RWD = real-world data, Rx = prescription, SMD = standard mean difference, TMLE = targeted maximum likelihood estimation.

References

1. Schuler MS, Rose S. *Am J Epidemiol*. 2017;185:65-73.
2. Schneeweiss S, et al. *Epidemiology*. 2009;20:512-522. Erratum in: *Epidemiology*. 2018; 29:e63-e64.
3. Patel, M. R., et al. *New England Journal of Medicine. 2011;*365(10), 883-891.

Presented at the Annual Meeting of the International Society for Pharmacoeconomics and Outcomes Research; May 5–8, 2024; Atlanta, GA.