21. Improving Propensity Scoring through Machine Learning

John A Rigg,[1] Huma Lodhi,[1] John Gregson,[2] Joseph Kim.[1,2] [1]*Real-world Evidence Solutions, IMS Health, London, England, United Kingdom;* [2]*Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, England, United Kingdom.*

Background: Propensity scores (PS) are a popular method to adjust for confounding bias in estimation of treatment effects. PS are almost always based on logistic regression (LR) despite its well-known potential limitations concerning model specification. Moreover, consideration is seldom given to the impact of exposure misclassification to results using PS based on LR.

Objectives: To assess whether the accuracy of treatment effects may be improved through a range of advanced machine learning algorithms, Support Vector Machines (SVMs), used to compute PS; to assess whether the optimal SVM specification differs according to the nature of confounding; to assess whether treatment effects are more accurate with SVMs than LR with differential and non-differential exposure misclassification.

Methods: A hypothetical study cohort was simulated (N = 2,000) using Monte Carlo sampling for a binary outcome, a binary exposure and ten binary or continuous covariates. Nine scenarios were created; seven scenarios differing by non-linear and/or non-additive confounding and two scenarios differing by differential or non-differential exposure misclassification. PS were based on LR and SVMs – single, boosted and bagged specifications. Optimal SVM parameter settings were identified using cross-validation. Performance metrics included the percent bias in the treatment effect and the average standardized difference in absolute mean of covariates between treated and non-treated (a measure of covariate balance).

Results: SVMs provided more accurate treatment effect estimates and better covariate balance than LR, especially in the presence of non-linear and non-additive confounding. There was generally no substantive difference between both ensemble SVM methods, except for a notably superior performance by bagged compared to boosted SVMs with differential exposure misclassification.

Conclusions: SVMs provide a highly promising solution to reducing treatment effect bias and ensuring covariate balance in the presence of confounding bias and differential exposure misclassification. These advanced algorithms out-performed LR and pharmacoepidemiology would benefit from their more widespread adoption.

22. Propensity Score Matching and Unmeasured Covariate Imbalance: A Simulation Study

M Sanni Ali,[1,2] Rolf HH Groenwold,[1,2] Svetlana V Belitser,[1] Arno W Hoes,[2] A de Boer,[1] Olaf H Klungel.[1,2] [1]*Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, Netherlands;* [2]*Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands.*

Background: Selecting covariates for adjustment or inclusion in propensity score (PS) analysis is a trade-off between reducing confounding bias and a risk of amplifying residual bias by unmeasured confounders.

Objectives: To assess the covariate balancing properties of PS matching with respect to unmeasured covariates and its impact on bias.

Methods: Simulation studies were conducted in binary covariates, treatment and outcome data. In different scenarios, instrumental variables (IV, i.e., variables related to treatment but not to the outcome or other covariates), risk factors (variables related only to the outcome), unmeasured covariates, and confounders with various associations among each other were considered.Treatment effects estimates (risk ratio) were derived after PS matching using Poisson models; balance for each covariate was checked before and after matching using the absolute standardized difference.The choice of covariates for the PS model was compared with respect to bias in the treatment-outcome relation and balance of (unobserved) covariates.

Results: PS matching improved balance of measured covariates included in the PS model but exacerbated the imbalance of the unmeasured covariate that was unrelated to measured covariates compared to the full unmatched sample. Inclusion of instrumental variables, independent of unmeasured covariates, exacerbated the imbalance in unmeasured covariates and amplified the residual bias. However, including instrumental variables that were associated with unmeasured covariates improved the balance of unmeasured covariates and reduced bias. When the PS model included variables related to the outcome, exclusion of instrumental variables that were related to unmeasured covariates exacerbated the balance of unmeasured covariates and increased the bias.

Conclusions: In choosing covariates for a PS model, the pattern of association among covariates has substantial

impact on other covariates' balance and the bias of the treatment effect.Investigators should not rely only on covariate association with treatment or outcome but should take into account possible associations among covariates and explore the balance of other covariates after PS matching.

23. Use of Propensity Score Methodology to Assess Comparability of Treatment Groups in a Registry Program

Kenneth J Rothman,[1] Menno V Huisman,[2] Gregory YH Lip,[3] Hans-Christoph Diener,[4] Sergio J Dubner,[5] Chang-Sheng Ma,[6] Kristina Zint,[7] Christine Teutsch,[7] Nils Schoof,[7] Miney Paquette,[8] Eva Kleine,[7] Dorothee B Bartels,[7] Jonathan Halperin.[9] *[1]Epidemiology, Boston University School of Public Health, Boston, MA, United States; [2]Department of Thrombosis and Hemostasis, Leiden University Medical Center, Leiden, Netherlands; [3]Haemostasis Thrombosis & Vascular Biology Unit, Centre for Cardiovascular Sciences, University of Birmingham, Birmingham, United Kingdom; [4]Department of Neurology, Universit Hospital Essen, Essen, Germany; [5]Arrhythmia and Electrophysiology Service, Clinica y Maternidad Suizo Argentina, Buenos Aires, Argentina; [6]Department of Cardiology, Beijing AnZhen Hospital, Capital Medical University, Beijing, China; [7]Boehringer Ingelheim Pharma GmbH & Co. KG, Ingelheim am Rhein, Germany; [8]TCM GLORIA-AF Registry, Boehringer Ingelheim Canada Limited, Burlington, ON, Canada; [9]Clinical Cardiology Services, Mount Sinai School of Medicine, New York, United States.*

Background: Propensity scores (PS) are commonly used to account for measured confounding factors in comparative studies of treatment options. In the Global Registry on Long-Term Oral Antithrombotic Treatment in Patients with Atrial Fibrillation (GLORIA-AF), we used propensity scores in Phase II of the registry program to assess comparability of patients receiving different treatments regarding their baseline characteristics, before initiating longitudinal data collection to perform comparative analyses between the treatments.

Objectives: During Phase II we evaluated comparability of baseline patient characteristics for newly diagnosed AF patients receiving either dabigatran or vitamin-K antagonists (VKA) by examining PS distributions.

Methods: Specifically, we measured the proportion of patients within the region of overlap of the PS distributions. When data collection was started dabigatran was a new treatment and it was expected that patient characteristics might initially differ, resulting in limited overlap. The proportion of patients in the region of overlap was a key determinant in the decision when to start the large scale safety and effectiveness comparison in Phase III of GLORIA-AF. If 95% of patients were in the overlapping region, Phase III would be initiated for a specific region. The first analysis was done for North America.

Results: In the North American analysis, 536 patients initiating dabigatran were compared with 488 patients beginning VKA. The proportion of patients in the overlapping region of PS was 99.3% for the PS model containing a pre-specified subset of risk factors for stroke and bleeding. When we included all baseline characteristics in the model, the proportion of patients in the region of overlap was only slightly lower (96.3%). Either result was sufficient to proceed to Phase III.

Conclusions: We employed PS to assess comparability of anticoagulant treatment groups in this large registry program to help determine when to begin a comparative study assessing safety and effectiveness. This approach may have applications in other types of population-based registry studies as well.

24. Applying High Dimensional Propensity Score (HDPS) in a Exploratory Data Analysis with a US Claims Database for Recent Medicinal Products

Bing Cai, Sundaresan Murugesan, Jamie Geier, Andrew Bate. *Epidemiology, Pfizer Inc, New York, United States.*

Background: HDPS is a semi-automatic method used in active surveillance to evaluate potential safety signals. The Observational Medical Outcomes Partnership (OMOP) experiment applied and evaluated the method using well-established drug-outcome pairs with extensive literature support. Currently, it is unknown how well such evaluations would generalize to recently approved and marketed products.

Objectives: To conduct a pilot study for implementing the HDPS method for exploratory data analysis, investigating the performance of the method for adalimumab (Humira®), a biologic disease-modifying antirheumatic drug (bDMARD) which is approved for the treatment of rheumatoid arthritis.

Methods: We applied the HDPS method to the original Optum database and the database in OMOP common database model (CDM), employing a new user design with a 12-month washout period. Other bDMARDs