LAY ABSTRACT

TITLE: Outcome adaptive propensity score methods for handling censoring and high-dimensionality: application to insurance claims

JOURNAL: Stat Methods Med Res. 2025 May

AUTHORS: Jiacong Du[1], Youfei Yu[1], Min Zhang[2], Zhenke Wu[1], Andrew M Ryan[3], Bhramar Mukherjee[1*]

* Corresponding author

INSTITUTIONS:
1 Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA
2 Vanke School of Public Health, Tsinghua University, Beijing, China
3 Department of Health Management and Policy, University of Michigan, Ann Arbor, MI, USA

This is attributed to the CEECR grant: UG3CA267907

LAY ABSTRACT

When scientists compare two medicines, they need the groups of people they are studying to be as similar as possible. To do this, scientists use a math tool called a propensity score. This tool matches similar people together to see which drug works better. For example, the tool would let scientists see how two different medicines would work for two people with similar health history details. However, choosing which details to use in the tool to match people is hard. If a scientist picks the wrong details to match people, the results may not actually tell them which medicine would work better for someone. The tool also does not always work well if the details are missing for some people in a large dataset.

In this paper, the researchers share a better way to choose which details to use in the propensity score tool in large datasets. This new method uses machine learning to choose which details to use in the propensity score tool. The researchers used "decision trees" - computer programs that act like filters - to sift through thousands of details in a large dataset and find only the details that could predict which cancer drug worked best.

The researchers tested this new method using insurance records from men with advanced prostate cancer. They compared different drugs to see which ones caused more side effects, even when some patient information was missing. This new method was better at comparing drug side effects than older methods. This work helps scientists use large datasets to give doctors and patients more reliable tools to compare medicines.