Generating Actionable Insights: Machine Learning for Causal Inference with Individual-Level Patient Generated Data

Shannon Wongvibulsin[1-4], Eric J. Daza[5]
[1]Department of Biomedical Engineering, Johns Hopkins University School of Medicine
[2]Johns Hopkins University Medical Scientist Training Program
[3]Johns Hopkins University School of Medicine, Johns Hopkins University
[4]Welch Center for Prevention, Epidemiology, Clinical Research, Johns Hopkins Medical Institutions
[5]Stanford Prevention Research Center, Stanford University School of Medicine
* Both authors contributed equally to this abstract

Background.

Individual-level patient generated data (PGD) from implantable or wearable devices ("implantables" or "wearables"), sensors, or mobile health applications (mHealth apps) are rapidly growing in scale, availability, and type. However, the clinical utility of the massive amounts of data generated by these devices has not yet been realized, and questions remain as to how we can use PGD to help guide personalized clinical decision-making. We propose that the advancement of rigorous statistical causal inference techniques using machine learning (ML) applied to PGD can help generate analytically robust and clinically actionable insights in an individualized, data-driven manner.

Many precision medicine approaches provide insights into an individual's health by applying powerful ML methods to dense population-level datasets. This top-down, population-based analytical approach is slowly being complemented by a bottom-up, individual-based approach using PGD. However, researchers and clinical practitioners have only begun applying ML techniques that use PGD to generate, test, and apply new idiographic (i.e., individual-focused) theories of diagnosis and treatment. Moreover, these idiographic studies often utilize data collected on much finer time scales than those of population-level datasets. These emerging streams of data, coupled with growing interest in generating personalized insights from PGD, will also require rigorous causal analysis (in both observational and experimental designs) for constructing sensible interventions. Unfortunately, such PGD-based causal inference techniques are largely nonexistent.

Methods.

Causal inference approaches strengthened with ML can help advance PGD-based health studies and applications. For example, the n-of-1 randomized trial (N1RT) counterfactual framework (Daza, 2018) and corresponding N1CPM (n-of-1 causal predictive modeling) R package enable personalized clinical investigation. This framework and package allow an individual to design personalized N1RTs for testing idiographic treatment effects discovered using his or her own observational time series data (i.e., a treatment can be a medication or a behavior change, etc.). However, for this approach to produce broad clinical impact, both the ML and the healthcare communities must be involved in its implementation and refinement. In general, partnerships between clinicians (i.e., the experts on individual patients), clinical researchers, public/population health researchers, epidemiologists, biostatisticians, and ML experts are required in order to build and improve analytic frameworks that will yield actionable clinical insights. Discussion among these partners regarding the rigor necessary for causal inference and ML applied to PGD can help accelerate the personalization of healthcare through clinically impactful interventions.

Conclusion.

Massive amounts of PGD are currently generated but lack clinical utility. The advancement of rigorous statistical causal inference techniques applied to PGD (e.g. N-of-1 studies) through partnerships between the ML and the healthcare communities can help improve and maintain an individual's health in a data-driven manner.

References.

Daza EJ. Causal Analysis of Self-tracked Time Series Data Using a Counterfactual Framework for N-of-1 Trials. Methods of information in medicine. 2018 Feb;57(01):e10-21.