17

Powering RCTs for marginal effects with GLMs using prognostic score adjustment

Emilie Hojbjerre-Frandsen[1,2,3], Alejandro Schuler[3]
[1]Novo Nordisk A/S, Aalborg, Denmark. [2]Aalborg university, Aalborg, Denmark. [3]UC Berkeley, Berkeley, USA

Emilie Hojbjerre-Frandsen
Please provide a brief biography for the Presenting author(s)
Emilie holds a Master of Science in Statistics from Aalborg University, completed in 2022. Her thesis, undertaken in collaboration with Novo Nordisk A/S, focused on utilizing historical data through prognostic score adjustment for linear models. Following the completion of her studies, she worked as a trial statistician before she in February 2023 initiated an industrial Ph.D. program in partnership with Aalborg University and Novo Nordisk A/S, supported by the Innovation Fund Denmark. Her research focuses on prognostic score adjustment to increase power in randomized clinical trials. Additionally, she embarked on a one-year study visit to UC Berkeley's Department of Biostatistics as part of the CTML (Center for Targeted Machine Learning) research group from February 2024 to February 2025.

Alejandro Schuler
Please provide a brief biography for the Presenting author(s)
Alejandro Schuler is an Assistant Professor in Residence at UC Berkeley Biostatistics. His research focus is on developing methods for clinical decision-making in the real world that are economically or clinically necessary, statistically rigorous, and frictionless from the user perspective. He completed his Ph.D. at Stanford in 2018 and worked as a postdoc with CTML before starting on the faculty. Dr. Schuler is known for developing the Selectively Adaptive Lasso, NGBoost, and prognostic covariate adjustment methods, among others. In addition, he often collaborates with domain experts to translate their questions to mathematical formalisms and bring the right methods to bear on them. His experiences working as a data scientist at Kaiser Permanente's Division of Research and as an early employee of a health tech startup helped shape his research agenda into something with relevance beyond academia. Dr. Schuler is also passionate about pedagogy and making good statistics accessible to everyone regardless of background or experience.

Single topic, multi-speaker session, Workshop or Single presentation submission

A single presentation/poster

Single presentation or poster submission

Estimating causal effects from randomized experiments is crucial in clinical research, and enhancing the precision is a key goal for statisticians. Historical data from registries, prior clinical trials, and health records offer a rich resource for understanding patient outcomes under standard-of-care and should be leveraged to increase study power. However, many existing methods using historical data trade off reduced variance for less stringent type-I-error rate control. We presents a novel approach to leveraging historical data, specifically focusing on covariate adjustment for generalized linear models (GLMs) to enhance the efficiency of analyses without introducing bias. Our method involves training a prognostic model using historical data and then estimating the marginal effect using the plug-in GLM procedure proposed by Rosenblum & van der Laan 2009, while adjusting for the trial subjects' predicted outcomes, known as their prognostic scores, within the linear predictor. This extends the approach of Schuler 2021 beyond the linear model. Under certain conditions, this prognostic score adjustment procedure achieves the minimum possible variance among a broad class of estimators. Even when these conditions are not fulfilled, prognostic score adjustment remains more efficient than raw covariate adjustment, with the efficiency gains depending on the prognostic model's predictive accuracy beyond the linear relationship with the raw covariates. We validate our approach through simulations and a reanalysis of a trial conducted by Novo Nordisk A/S, demonstrating notable variance reductions of the marginal effect estimate. Finally, we present a simplified formula for asymptotic variance, facilitating power calculations that account for these efficiency gains.