70

Enhancing Treatment Effect Estimation in Clinical Trials using Machine Learning: A Within-Study Prognostic Score Approach

Antigoni Elefsinioti[1], Maike Ahrens[2], Sebastian Voss[2], Karl Koechert[1], Bohdana Ratitch[3]
[1]BAYER AG, Berlin, Germany. [2]Chrestos Concept GmbH& Co KG, Essen, Germany. [3]BAYER AG, Montreal, Canada

Antigoni Elefsinioti
Please provide a brief biography for the Presenting author(s)
Since 2018, Antigoni has focused on applying statistical methodologies and machine learning techniques to extract meaningful insights from clinical data, particularly in areas such as covariate adjustment, treatment effect heterogeneity, and predictive biomarkers. With over 20 years of experience in applied quantitative methods, she specializes in bioinformatics, statistics, and machine learning.

Her academic background includes a Ph.D. in Cellular Networks and Systems Biology and postdoctoral research in genomics, providing her with a solid foundation for analyzing complex biological systems. In the pharmaceutical industry, she has been involved in planning and analyzing biomarkers across various drug development programs, as well as working on knowledge management and the integration of genomic and clinical data for translational research.
Antigoni enjoys using Python and R to develop applications for data integration and analysis, and finds collaboration with diverse teams and engaging with stakeholders to be particularly rewarding.

Maike Ahrens
Please provide a brief biography for the Presenting author(s)
Maike is a trained biostatistician with over a decade of experience in applied biomarker research and data analysis, as well as in teaching on various topics. She has refined her methodological and programming skills both during her PhD and through her work in the pharmaceutical industry.
With the aim of making data science more accessible, her work in recent years has focused on delivering convenient and GxP-compliant solutions to complex data science problems within the pharmaceutical industry. This ranges from supporting interdisciplinary teams in generating new insights in exploratory settings to the implementation of machine learning analyses for clinical reports.
Maike has a passion for education and knowledge-sharing to help deepen the understanding of the methodological background of applied statistical and machine learning methods, but also to promote good programming practices and R package development. She has helped both novices and experienced professionals enhance their skills and apply data science techniques effectively in their respective fields.

Sebastian Voss
Please provide a brief biography for the Presenting author(s)
Sebastian is a seasoned statistician and data scientist with over a decade of experience in biomarker research within the pharmaceutical industry. Currently, he serves as the Head of the Biomarker Statistics & Data Science group at Chrestos GmbH in Essen. In this role, he leads a team dedicated to advancing the role of biomarkers in the drug development process and developing user-friendly and GxP compliant data science solutions, with a strong emphasis on R development.
With a robust background in both theoretical and applied statistics, Sebastian is passionate about leveraging data to drive innovation in biomarker research and improve outcomes in the pharmaceutical sector.
He has a deep commitment to advancing the field of data science through education,

having taught various courses on Shiny, R package development, machine learning, and other essential data science tools.

Karl Koechert
Please provide a brief biography for the Presenting author(s)
Since January 2025, Karl is a Statistical Project Leader at Sanofi's Statistical Innovation Hub. Prior to joining Sanofi, Karl was Head of Biomarker & Data Insights at Bayer AG (Berlin, Germany) leading a team of data scientists / statisticians. With a background in omics data science and precision medicine, his team's major objective is to advance drug development by enabling holistic understanding of complex biological systems - namely the patients in need. In this vein, their current endeavor is to understand how applied AI/ML can help to detect highly complex safety and efficacy signals in interventional clinical trials and how that can be utilized as a basis for creating virtual twins of specific disease indications.
Karl, being a Biochemist by training, discovered his passion for mathematical modelling of complex biological systems during his PhD at Humboldt University (Berlin, Germany) and subsequently devoted himself to applied machine learning during postdocs at TU-Dresden (Germany) and the Max-Delbrück-Center for Molecular Medicine (Berlin, Germany) before joining Bayer AG in 2014. At Bayer he has held positions of increasing responsibility as Study and Project Biomarker statistician for interventional trials of all phases in oncology.

Bohdana Ratitch
Please provide a brief biography for the Presenting author(s)
Bohdana Ratitch is an Expert Statistician and a Science Fellow at Bayer. She has over 19 years of experience in biopharmaceutical industry including statistical methodology research and consulting, with expertise in missing data, subgroup identification, machine learning, and digital health. Bohdana holds a BSs in Applied Mathematics from Ivan Franko National University of Lviv, Ukraine, and a MSc and a PhD in Machine Learning from McGill University in Montreal, Canada.

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

A single presentation/poster

Single presentation or poster submission

We present a simulation study investigating the performance of an approach for estimating treatment effects in randomized clinical trials adjusting for a prognostic score, where a prognostic model is obtained from within-study control treatment data. This method builds on and extends the original Prognostic Covariate Adjustment (PROCOVA) methodology qualified by the European Medicines Agency, which aims to enhance the efficiency of estimating treatment effects by employing machine learning techniques to develop prognostic scores that condense the prognostic information from multiple baseline covariates. While PROCOVA utilizes data from historical trials, our approach focuses on estimating prognostic models from within-study data. By using Random Forests and out-of-bag predictions, we ensure that prognostic scores for each participant—including those in the placebo arm—are derived from models trained on independent datasets, thus mitigating biases related to model selection. This is especially beneficial in situations where historical data is scarce, of insufficient quality, or when population characteristics have shifted over time, compromising the prognostic model's suitability for new studies.

Our extensive simulations indicate that the within-study approach can enhance the precision of treatment effect estimates without introducing bias or compromising the Type 1 error control across various scenarios, thereby supporting the usefulness of this method. Although the historical study approach may offer greater benefits when prognostic factors are consistent and historical databases are large, our within-study method demonstrates superior performance when these factors differ, making it a robust alternative in diverse clinical contexts.