

**TITLE:           LEVERAGING HISTORICAL DATA FOR HIGH-DIMENSIONAL COVARIATE-  
ADAPTIVE RANDOMIZATION, A MACHINE LEARNING APPROACH.**

---

**Scientific Presentation made on June 2019 at the Promoting Statistical Insight  
Conference, London (UK).**

**Abstract**

There is a continuous growth in data collected in clinical trials. Many of those patient's characteristics are potential confounding factors. Ideally, these factors should be accounted for in the randomization process to balance study arms and reduce the variability of the estimated treatment effect. However, the efficiency of the randomization decreases very fast with the number of factors, in particular when the sample size is not very large. Furthermore, balancing for multiple covariates may not account for their multivariate effects. The purpose of this talk is to present a machine learning solution to improve high-dimensional randomization.

Our solution comes from the observation that a good randomization does not require balancing all covariates individually. The problem is to balance patients while considering all covariate effects together. However, those multivariate interactions are difficult to model/estimate with a limited sample size. To solve this issue, we propose to estimate them a priori within a composite covariate. Such composite covariate could be learned with machine learning on historical data from similar studies. Then, even in small studies, a covariate adaptive randomization could be applied on this single composite covariate to account for them all.

The composite covariate approach was first presented at the 2018 PSI conference in the context of adjusted analyses. Extending our results to covariate-adaptive randomization, we showed its particular interest with complex data (high-dimensional, non-linear, etc). Indeed, limiting the number of covariates to one has a direct positive impact on the efficiency of the randomization. We also put this efficiency gain into perspective with the quality of the learning process.

**Authors**

---

*Samuel Branders, PhD; Guillaume Bernard, PhD; Alvaro Pereira, PhD*