An Empirical Bayesian Method for Subgroup Identification in Personalized Medicine

Jian Yin[1], Zihang Zhong[1], Senmiao Ni[1], Yang Zhao[1], **Jingwei Wu[2]**, Hao Yu[1], Jianling Bai[1]

[1]*Department of Biostatistics, School of Public Health, Nanjing Medical University*

[2]*Department of Epidemiology and Biostatistics, Barnett College of Public Health, Temple University*

ABSTRACT

Personalized medicine tailors therapies to patient-specific characteristics, creating a growing need for reliable subgroup identification. Data-driven approaches that search for subgroups with enhanced efficacy or safety using predictive biomarkers are especially valuable when mechanistic knowledge is limited. However, single-trial methods often suffer from selection inaccuracies and instability due to insufficient information. To address these limitations, we propose PEMBA (normalized Power prior based on the EMpirical BAyesian method), a subgroup identification framework that improves detection of treatment-effect heterogeneity by incorporating evidence from multiple historical studies together with current trial data. PEMBA applies a normalized power prior within an empirical Bayesian structure to integrate information and compute posterior treatment-effect distributions through a grid search for optimal splits. A permutation test is used to control the overall false positive rate. Simulation studies show that PEMBA improves subgroup classification accuracy compared with existing approaches while maintaining false positive rates at pre-specified levels. The method also remains robust in the presence of heterogeneity across historical trials. A real-data application in breast cancer further demonstrates PEMBA's ability to leverage multi-trial information to identify a clinically meaningful subgroup. By integrating evidence across studies, PEMBA provides a more reliable approach to detecting treatment-effect heterogeneity. This method can advance personalized medicine by improving clinical study efficiency, increasing trial success rates, and enabling more patients to receive appropriately targeted treatments.

Keywords: Subgroup Identification, Data borrowing, Normalized power prior

Dynamic and Concordance-Assisted Learning for Risk Stratification

Jing Ning

Department of Biostatistics/1MC 12.3557

The University of Texas M.D. Anderson Cancer Center

7007 Bertner Ave Houston, TX 77030 USA

ABSTRACT

Dynamic prediction models that adapt over time and maintain accuracy can play a crucial role in monitoring disease progression in clinical practice. In biomedical studies with long-term follow-up, participants are typically monitored through periodic clinical visits with repeated measurements until the occurrence of the event of interest (e.g., disease onset) or study completion. Recognizing the dynamic nature of disease risk and the information captured by longitudinal markers, we propose an innovative concordance-assisted learning algorithm to derive a real-time risk stratification score. Our approach avoids the need to fit regression models, such as joint models of longitudinal markers and time-to-event outcomes, thereby offering robustness to model misspecification. Simulation studies demonstrate that the proposed method performs well in dynamically monitoring disease risk and distinguishing high- from low-risk populations over time. We further apply the method to data from the Alzheimer's Disease Neuroimaging Initiative to develop a dynamic risk score for Alzheimer's disease in patients with mild cognitive impairment, incorporating multiple longitudinal markers and baseline prognostic factors.

Keywords: concordance-assisted learning; dynamic prediction; longitudinal markers; risk stratification.

Model Based Multiple Imputation in Censored Quantile Regression

Zhaozhi Fan, Ummay Nayeema Islam

Department of Mathematics and Statistics, Memorial University of Newfoundland

ABSTRACT

Censored quantile regression models provide a global description of the association between the censored response and potential risk factors, through proper selection of quantiles. But when the censoring rate goes high, the model would face identifiability issues, especially for the extreme quantiles. In this article we propose a multiple imputation method based on the AFT model to handle the censored values. Quantile regression parameters are estimated based on the imputed data. The imputation was done multiple times through randomly selecting residuals from AFT regression. The estimators of the regression parameters are consistent and having asymptotic multivariate normal distribution. Simulation results show that our method performs equally well with existing approaches when the models are identifiable and can also provide reliable estimation beyond the identifiable range of quantile levels of existing methods.

Keywords: survival analysis; quantile regression; multiple imputation