Survival Bump Hunting For Identification and Characterization of Informative Prognostic Subgroups

Jean-Eudes Dazard, PhD[1]*†, Michael Choe[1]*, J. Sunil Rao, PhD[3]

[1]Division of Bioinformatics, Center for Proteomics and Bioinformatics, School Of Medicine, Case Western Reserve University; [2]Department of Biostatistics, School of Public Health, University of Washington, Washington, Public Health Sciences, [2]Fred Hutchinson Cancer Research Center; [3]Division of Biostatistics, Department of Epidemiology and Public Health, The University of Miami, Florida.

*Equal contributions, †Corresponding authors.

Abstract

Background: The heterogeneous phenotypic nature of patients with complex disease has made the identification of patients subgroups evasive and the validation of classification or prognostic biomarkers difficult. This has created the need to develop and apply search algorithms for these purposes in low and high-dimensional settings.

Methods: Survival Bump Hunting (SBH) is a rule induction method for the induction of simple decision rules on the most relevant predictor variables of extreme survival groups. We propose the utilization of SBH in a proof of concept endeavor to identify extreme prognostic groups among 21 publically available clinical and genomic datasets where the number of covariates either remains small in comparison to the number of observations ($p < n$) or dominates it ($p >> n$). Datasets included studies of various pathologies (11 breast cancer, 2 lung cancer, 1 prostate cancer, 1 multiple myeloma, 1 Hodgkin's lymphoma, 1 bladder cancer, 1 follicular cell lymphoma, 1 primary biliary cirrhosis, 2 HIV) where the response variables included either overall survival, cause specific survival, disease free survival, progression free survival, or metastasis free survival.

Results: We report peeling trajectories against subgroup supports of selected covariates, hazard ratios, log-rank statistics and prediction-error statistics as well as event-free times and probabilities and median time-to-events. Trace curves of covariates variable importance and usage as well as Kaplan-Meier survival probability curves with log-rank *p*-values for these subgroups were also evaluated. SBH was able to identify distinct survival groups in 11 of the 21 datasets, which remained robust after replicated cross-validation ($p < 0.01$). The clinical implications of the corresponding survival predictor signatures are discussed and the applications of SBH is further explored.

Conclusions: Survival Bump Hunting (SBH) is effective at identifying extreme survival patients subgroups and characterizing them in the predictor space of any dimensionality. This makes it potentially useful in clinical practice as a clinical stratification scheme to distinguish between high and low-risk patients and tailor treatments. Moreover, the subsequent derivation of survival predictor signatures in these patients subgroups highlight potential therapeutic interventions. An R package `PRIMsrc` is available on CRAN and GitHub.

Keywords

Bump Hunting, Patient Rule Induction Method, Non-Parametric Methods, Survival/Risk Analysis, Prognostication, Predictive Prognosis, Precision Medicine