Individual dynamic prediction of clinical endpoint from large dimensional longitudinal biomarker history: a landmark approach

Anthony Devaux[1], Robin Genuer[1,2], Karine Pérès[1] and Cécile Proust-Lima[1]

[1] Univ. Bordeaux, INSERM, BPH, U1219, F-33000 Bordeaux, France

[2] INRIA Bordeaux Sud-Ouest, Talence, France

E-mail for correspondence: anthony.devaux@u-bordeaux.fr

Abstract: The individual data collected throughout patient follow-up constitutes crucial information for assessing the risk of a health event. Some statistical methods were proposed to compute dynamic individual predictions from longitudinal information. However, they hardly handle a large number of markers, partly due to numerical issues. By relying on a landmark approach, we propose a methodology for computing individual dynamic predictions that may take into account a very large number of repeated markers over time while remaining computationally feasible. We model each marker trajectory up to the landmark time, and derive summary variables that best capture the individual trajectories. The summaries and additional covariates are then included in different prediction methods to predict the event from the landmark time. As we need to handle a possibly large dimensional history, we rely on machine learning methods adapted to survival data, namely regularized regressions and survival random forests, and show how they can be combined into a superlearner. We demonstrate in a simulation study the benefits of machine learning survival methods, especially in the case of multiple and/or nonlinear relationships between the predictors and the event. We also illustrate the methodology to predict death in the general elderly population at different ages.

Key words: Individual prediction; Landmark; Longitudinal data; Survival data; Machine learning methods