

Penalized regression calibration: a method for the prediction of survival outcomes using complex longitudinal and high-dimensional data

Mirko Signorelli¹, Pietro Spitali², Roula Tsonaka¹

¹ Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, NL

² Department of Human Genetics, Leiden University Medical Center, Leiden, NL

E-mail for correspondence: m.signorelli@lumc.nl

Abstract:

Longitudinal and high-dimensional measurements have become increasingly common in biomedical research. However, methods to predict survival outcomes using covariates that are both longitudinal and high-dimensional are currently missing. We propose penalized regression calibration (PRC, Signorelli et al., 2021), a method that can be employed to predict survival in such situations. The method is implemented in the R package `pencal`, available from CRAN.

PRC comprises three modelling steps: first, the trajectories described by the longitudinal predictors are flexibly modelled through the specification of multivariate latent process mixed models. Second, subject-specific summaries of the longitudinal trajectories are derived from the fitted mixed effects models. Third, the time to event outcome is predicted using the subject-specific summaries as covariates in a penalized Cox model.

To ensure a proper internal validation of the fitted PRC models, we furthermore develop a cluster bootstrap optimism correction procedure that allows to correct for the optimistic bias of apparent measures of predictiveness.

After studying the behaviour of PRC via simulations, we conclude by illustrating an application of PRC to data from an observational study that involved patients affected by Duchenne muscular dystrophy, where the goal is predict time to loss of ambulation using longitudinal blood biomarkers.

Key words: survival analysis; risk prediction modelling; longitudinal data analysis; high-dimensionality; optimism correction.

Signorelli M, Spitali P, Al-Khalili Szigyarto C, The MARK-MD Consortium, Tsonaka R (2021). Penalized regression calibration: a method for the prediction of survival outcomes using complex longitudinal and high-dimensional data. arXiv preprint number: arXiv:2101.04426.