Víctor Velasco-Pardo[1], Nathan Constantine-Cooke[1], Evangelin Shaloom Vitus[1,2], Begoña Bolós[1,3], Charlie W Lees[1,2], Catalina A Vallejos[1,4,5]

[1] Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, United Kingdom. [2] Aalborg University, Copenhagen, Denmark. [3] Alan Turing Institute, London, United Kingdom. [4] Usher Institute, University of Edinburgh, Edinburgh, United Kingdom. [5] Health Data Research UK, London, UK

landmaRk: A modular R package for landmark time-to-event analysis

Introduction

When performing **dynamic prediction**, where risk estimates are updated as more data becomes available, there are two popular approaches: joint models and landmarking.

We focus here on landmarking models, which are computationally efficient, making them ideal for working with large datasets (e.g. electronic health record data[1]).

Traditional Landmarking

Landmarking uses a two-step approach.

Survival models (typically Cox proportional hazards) are fitted for time-to-event outcomes – at multiple pre-specified *landmark* times – summarising any dynamic covariates (e.g. biomarkers) observed up to each landmark time.

Typically, the last observation carried forward (LOCF) is used to summarise dynamic covariates. It is also popular to model dynamic covariates using linear mixed effects models[1,2] (LME).

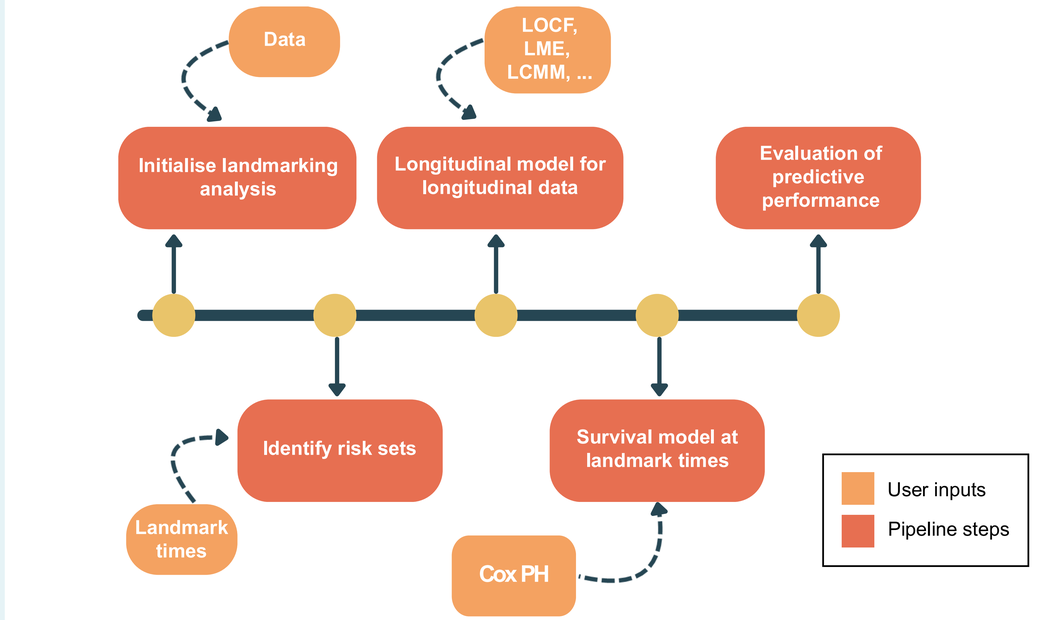
Latent class extension

Linear mixed effects models fail to account for heterogeneity in longitudinal trajectories of dynamic covariates. For example, this can correspond to latent groups of patients that share similar biomarker trajectories[3].

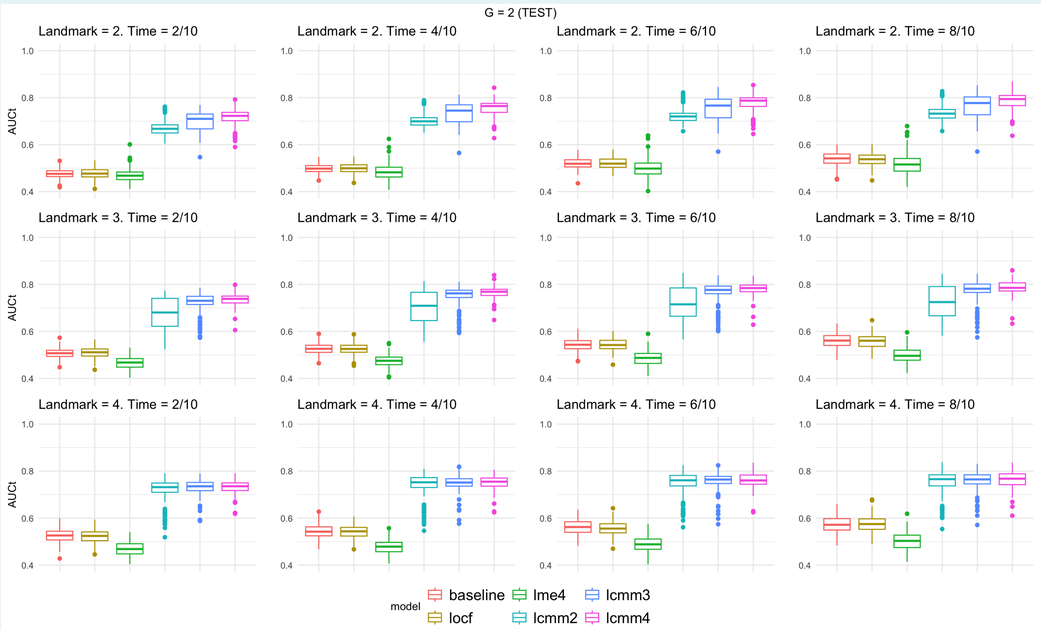
We propose using latent class mixed models[4] (LCMM) to model underlying the underlying latent substructure of model covariates, acting as a computationally feasible approximation to joint latent class models[2].

Longitudinal component: LCMM with both global and class-specific fixed effects, and random intercept and slope.

Survival component: Cox PH models stratified according to clusters estimated by LCMM, and interaction terms for covariates with respect to cluster membership.



Schematic representation of the landmaRk package pipeline



Discriminatory performance in out-of-sample simulated data of the following models: baseline (survival model without dynamic covariates), locf (last observation carried forward), lme4 (linear mixed effect model), and lcmm assuming 2, 3 and 4 clusters.

The landmaRk R package

The landmark R package implements a **modular framework** for dynamic risk prediction. Beyond, our latent class extension to landmarking, it is possible for users to extend the landmaRk package with their own components for longitudinal and time-to-event sub-models.

Simulation study

We used a modified version of synthetic datasets simulated by Andrinopoulou et al.[5], simulating scenarios with 1, 2, and 3 'ground-truth' clusters.

We evaluate out-of-sample predictive performance of (a) survival model with no covariates, (b) LOCF, (c) LME and (d-f) LCMMs with 2-4 clusters.

We found that the LCMM outperforms baseline, LOCF, and LME specifications, and performance does not substantially decrease if the number of clusters is miss-specified.

Future work

Having uncovered latent subgroups which describe previously uncharacterised heterogeneity in inflammatory bowel disease (IBD), our group is now using landmaRk to predict poor disease outcomes in Inflammatory Bowel Disease (IBD) using routinely collected biomarker data.

References

1. Paige E, et al. Landmark models for optimizing the use of repeated measurements of risk factors in electronic health records to predict future disease risk. American journal of epidemiology. 2018 Jul 1;187(7):1530-8.

2. Rizopoulos D, et al.. Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. Biometrical Journal. 2017 Nov;59(6):1261-76.

3. Constantine-Cooke N, et al. Longitudinal fecal calprotectin profiles characterize disease course heterogeneity in Crohn's disease. Clinical Gastroenterology and Hepatology. 2023 Oct 1;21(11):2918-27.

4. Proust-Lima C, et al. Joint latent class models for longitudinal and time-to-event data: a review. Statistical methods in medical research. 2014 Feb;23(1):74-90.

5. Andrinopoulou ER, Rizopoulos D. Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assuming different association structures. Statistics in medicine. 2016 Nov 20;35(26):4813-23.