An overview and empirical appraisal of recent methodological developments for the dynamic prediction of survival outcomes

Mirko Signorelli
mirkosignorelli.github.io
in/signorelli

Mathematical Institute
Leiden University

Joint work with Sophie Retif

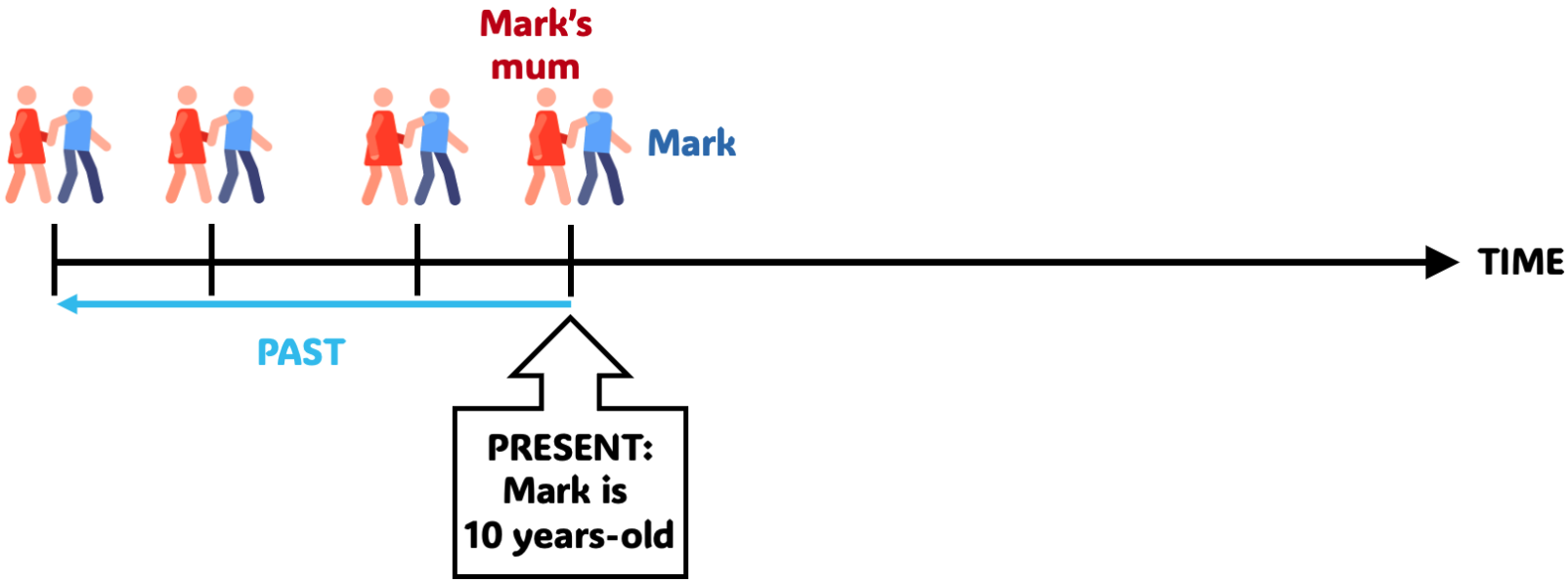
October 8, 2025
65th ISI World Statistics Congress

Universiteit Leiden

Preprint: arXiv:2403.14336v2

Preprint: arXiv:2403.14336v2

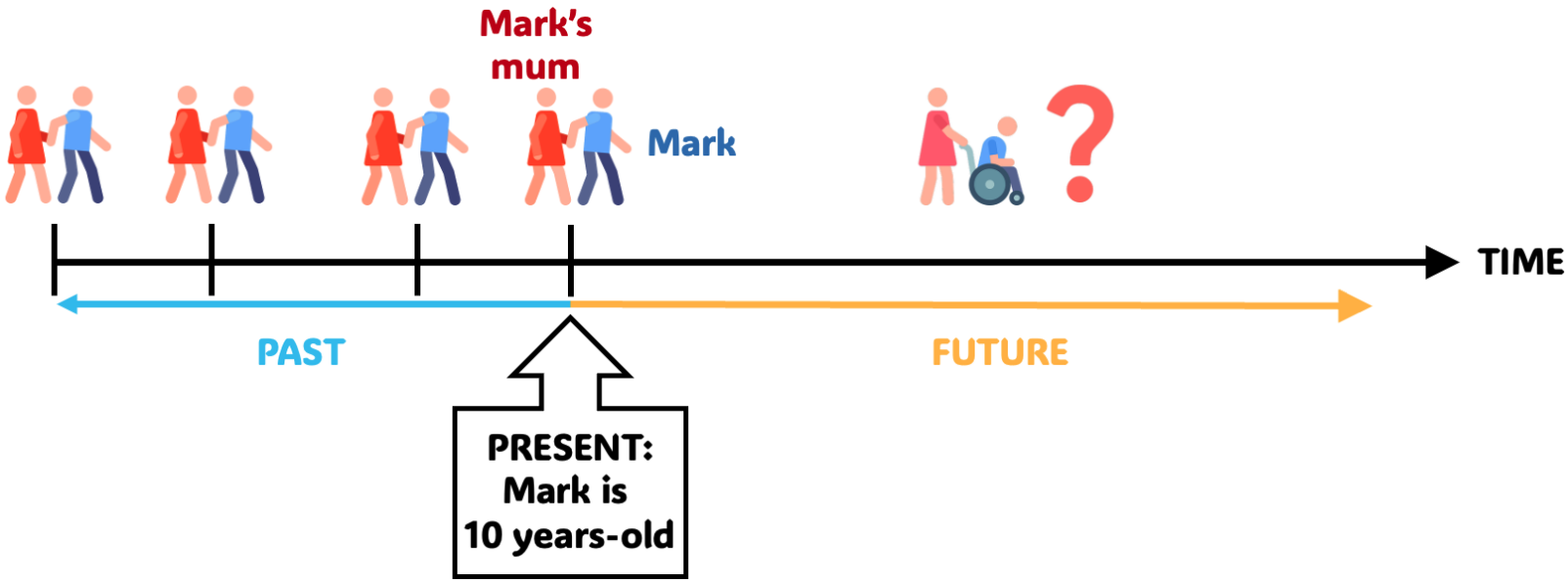
Dynamic prediction 101



- Mark suffers from Duchenne muscular dystrophy (DMD)
- Usually: loss of ambulation during adolescence

Preprint: arXiv:2403.14336v2

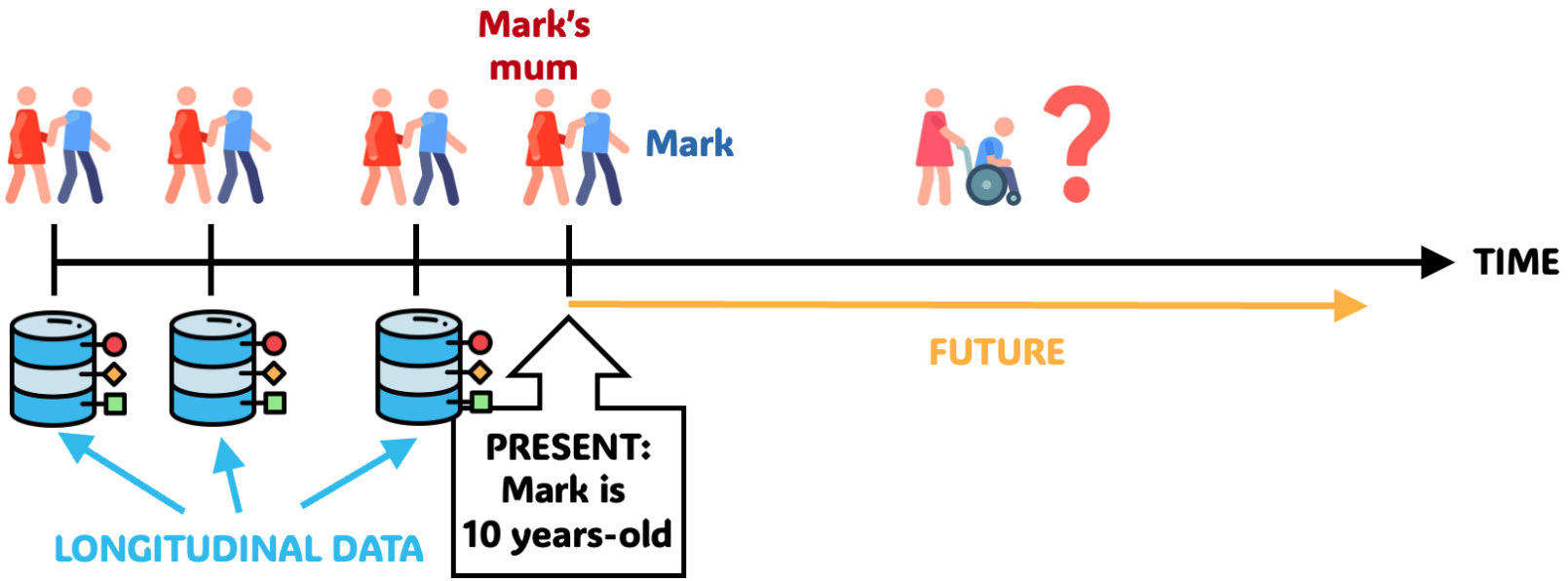
Dynamic prediction 101



- Mum: P(wheelchair within 2 years)?

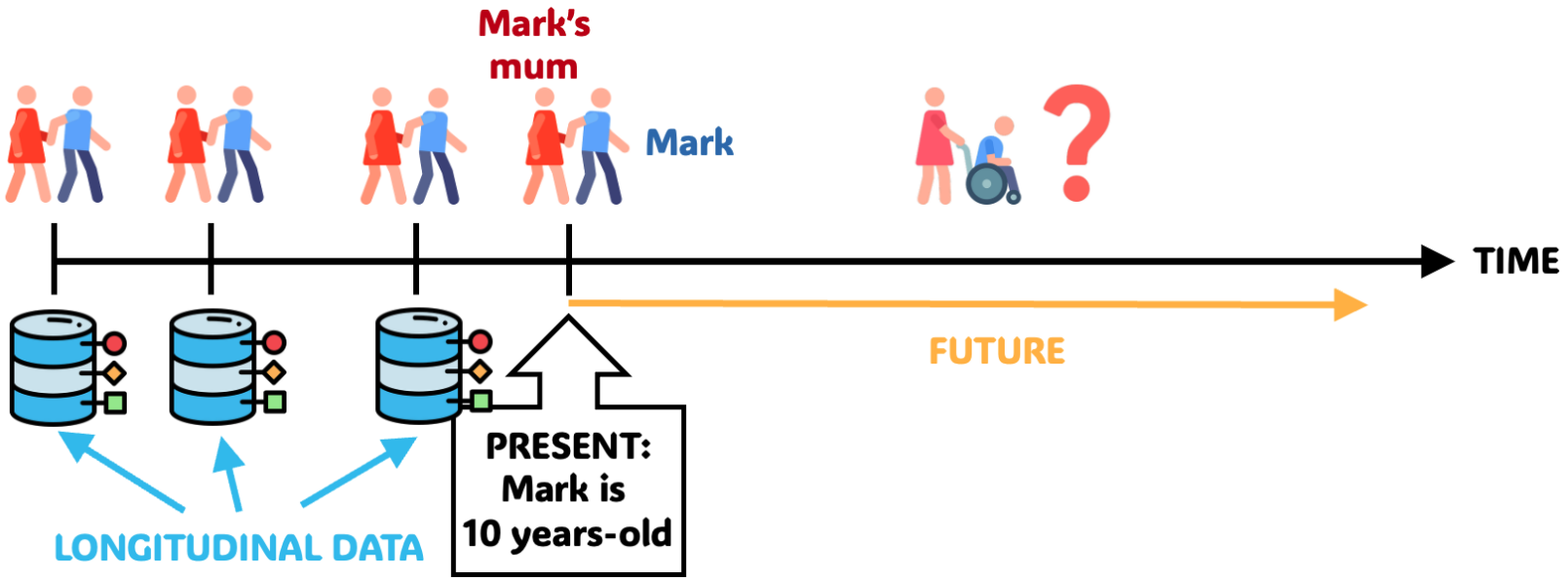
Preprint: arXiv:2403.14336v2

Dynamic prediction 101



Preprint: arXiv:2403.14336v2

Dynamic prediction 101



Dynamic prediction goal:

- use available longitudinal data to predict conditional survival probability

$$S(t|10, Y = \{y_1, y_2, y_3\}) = P(T > t|T > 10, Y = \{y_1, y_2, y_3\}),\ t > 10$$

Preprint: arXiv:2403.14336v2

The problem

- Limitations of traditional methods:
  1. landmarking with Last Observation Carried Forward (LOCF) discards longitudinal information
  2. joint modelling: very computationally-intensive (hours / days) + many estimation errors. Can't usually be estimated with more than 5-10 longitudinal covariates!

Preprint: arXiv:2403.14336v2

The problem

- Limitations of traditional methods:
  1. landmarking with Last Observation Carried Forward (LOCF) discards longitudinal information
  2. joint modelling: very computationally-intensive (hours / days) + many estimation errors. Can't usually be estimated with more than 5-10 longitudinal covariates!
- Nowadays, longitudinal studies can comprise tens, hundreds, or even thousands of longitudinal variables ("biomarkers")
- How to do dynamic prediction with "many" longitudinal covariates?

Preprint: arXiv:2403.14336v2

Preprint: arXiv:2403.14336v2

New dynamic prediction methods (2019-2023)

Several solutions proposed over the last 6 years:

1. Multivariate Functional Principal Component Cox model (MFPCCox, Li & Luo (2019))
2. Penalized Regression Calibration (pencal, Signorelli et al. (2021))
3. Functional Random Survival Forest (FunRSF, Lin et al. (2021))
4. Dynamic Random Survival Forest (DynForest, Devaux et al. (2023))

Preprint: arXiv:2403.14336v2

Modelling approach

Modelling steps:

1. model trajectories of longitudinal covariates over $[0, \ell]$[1]
2. obtain subject-specific summaries of the longitudinal covariates
3. use baseline covariates and summaries of longitudinal covariates to predict $S(t|\ell)$, $t \geq \ell$

[1] Technically, often data gathered after ℓ are also included. This is problematic for multi-step methods: selection bias after $\ell \Rightarrow$ lower predictive performance (Gomon et al., 2024) + use future to predict the future! ☹

Preprint: arXiv:2403.14336v2

Methods overview

- In a nutshell:

	Approach used to model...	
Method	**longitudinal covariates**	**survival outcome**
MFPCCox	multiv. functional PCA	Cox model
FunRSF	multiv. functional PCA	random survival forest
pencal	linear mixed models	penalized Cox model
DynForest	linear mixed models	random survival forest

Preprint: arXiv:2403.14336v2

Software

- Software implementations:

Method	Software	Details
MFPCCox	☹	
FunRSF	☹	
pencal	☺	`R` package on CRAN
DynForest	☺	`R` package on CRAN

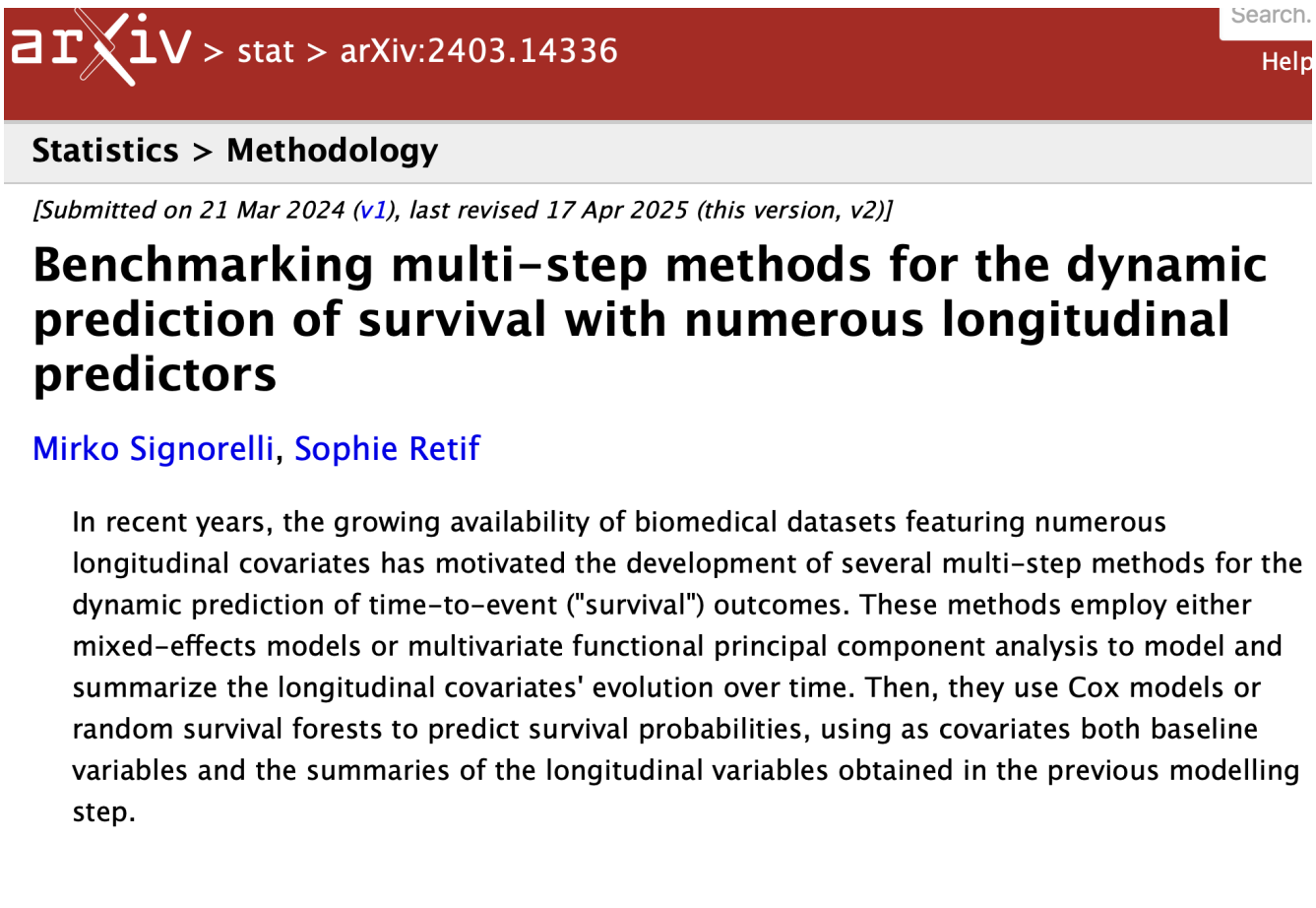
- Software articles:
  - `pencal`: Signorelli (2024)
  - `DynForest`: Devaux et al. (2024+)

Preprint: arXiv:2403.14336v2

Preprint: arXiv:2403.14336v2

Motivation

- Methods proposed quite recently: between 2019 and 2023 $\Rightarrow$ very little knowledge about their (relative) predictive performance, advantages and limitations
- Let's compare them using real-world data!



Preprint: arXiv:2403.14336v2

Longitudinal studies

- We considered data from three longitudinal studies:

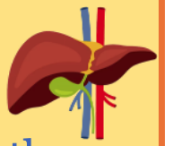
ROSMAP

- Event: Alzheimer's Disease diagnosis
- n = 3293
- 5 baseline covariates
- 30 longitudinal covariates
- Follow-up: [1, 29] years

ADNI

- Event: diagnosis of dementia
- n = 1643
- 5 baseline covariates
- 21 longitudinal covariates
- Follow-up: [0, 15.5] years

PBC2



- Event: death (primary biliary cirrhosis trial)
- n = 312
- 3 baseline covariates
- 8 longitudinal covariates
- Follow-up: [0, 14] years

- Methods included: MFPCCox, pencal, FunRSF, DynForest + static Cox + LOCF landmarking

Preprint: arXiv:2403.14336v2

Longitudinal studies

- We considered data from three longitudinal studies:

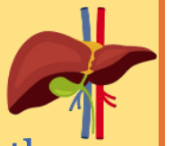
ROSMAP

- Event: Alzheimer's Disease diagnosis
- n = 3293
- 5 baseline covariates
- 30 longitudinal covariates
- Follow-up: [1, 29] years

ADNI

- Event: diagnosis of dementia
- n = 1643
- 5 baseline covariates
- 21 longitudinal covariates
- Follow-up: [0, 15.5] years

PBC2



- Event: death (primary biliary cirrhosis trial)
- n = 312
- 3 baseline covariates
- 8 longitudinal covariates
- Follow-up: [0, 14] years

- Methods included: MFPCCox, pencal, FunRSF, DynForest + static Cox + LOCF landmarking
- Performance evaluated at multiple landmark times
- Performance measures: C index, tdAUC, Brier score
- 10-fold cross-validation, repeated 10 times

Preprint: arXiv:2403.14336v2

ADNI dataset: C index (higher = better)

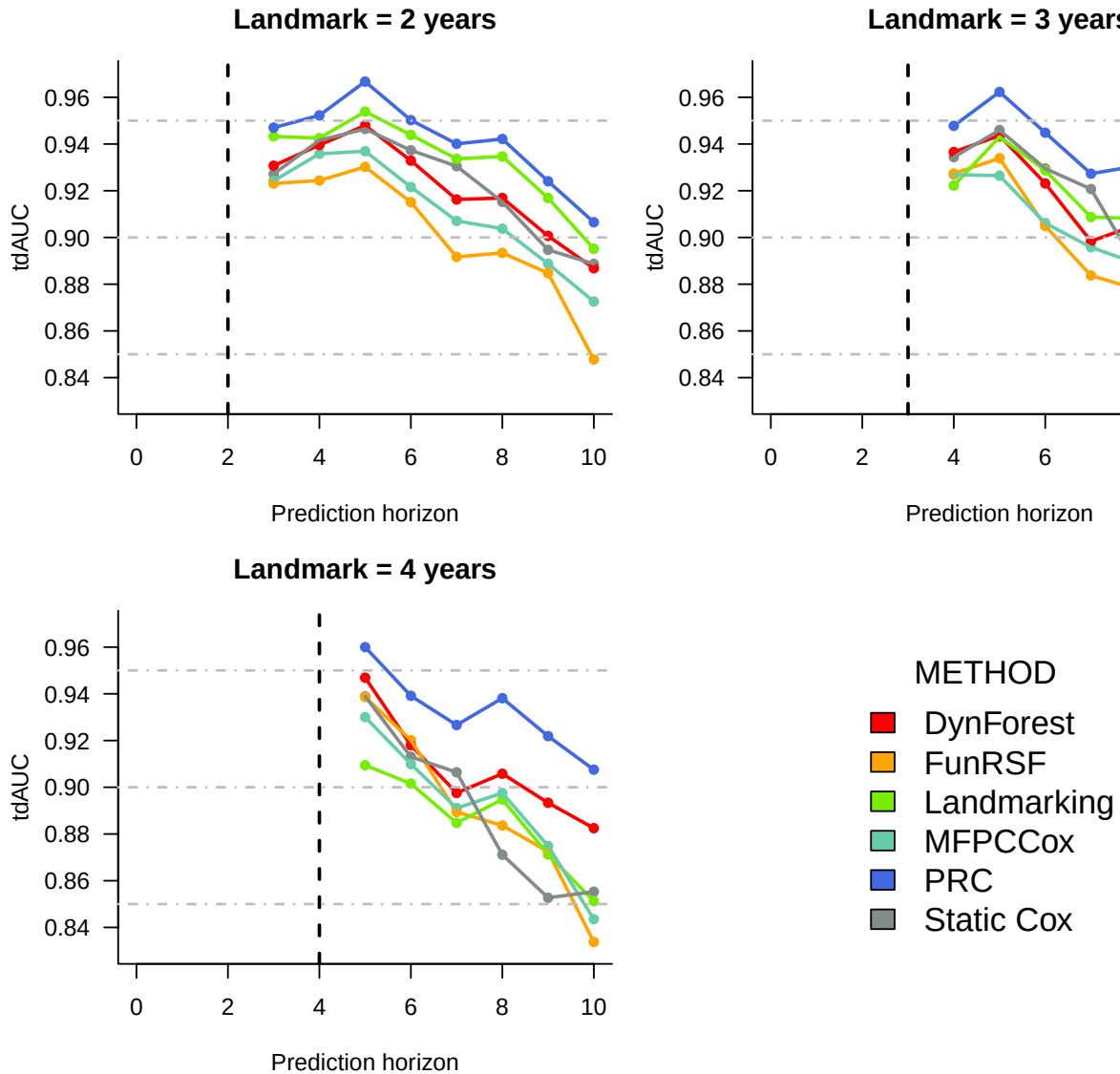
Method	Landmark 2	3	4
Static Cox	0.901 (0.002)	0.885 (0.006)	0.856 (0.008)
Landmarking	**0.906** (0.001)	**0.89** (0.004)	**0.855** (0.009)
MFPCCox	0.889 (0.003)	0.872 (0.009)	0.859 (0.008)
pencal	0.913 (0.001)	0.908 (0.003)	0.904 (0.003)
FunRSF	0.873 (0.004)	0.858 (0.011)	0.845 (0.012)
DynForest	0.891 (0.003)	0.883 (0.005)	0.871 (0.011)

LEGEND:

- green = better than **LOCF landmarking**
- red = worse than **LOCF landmarking**

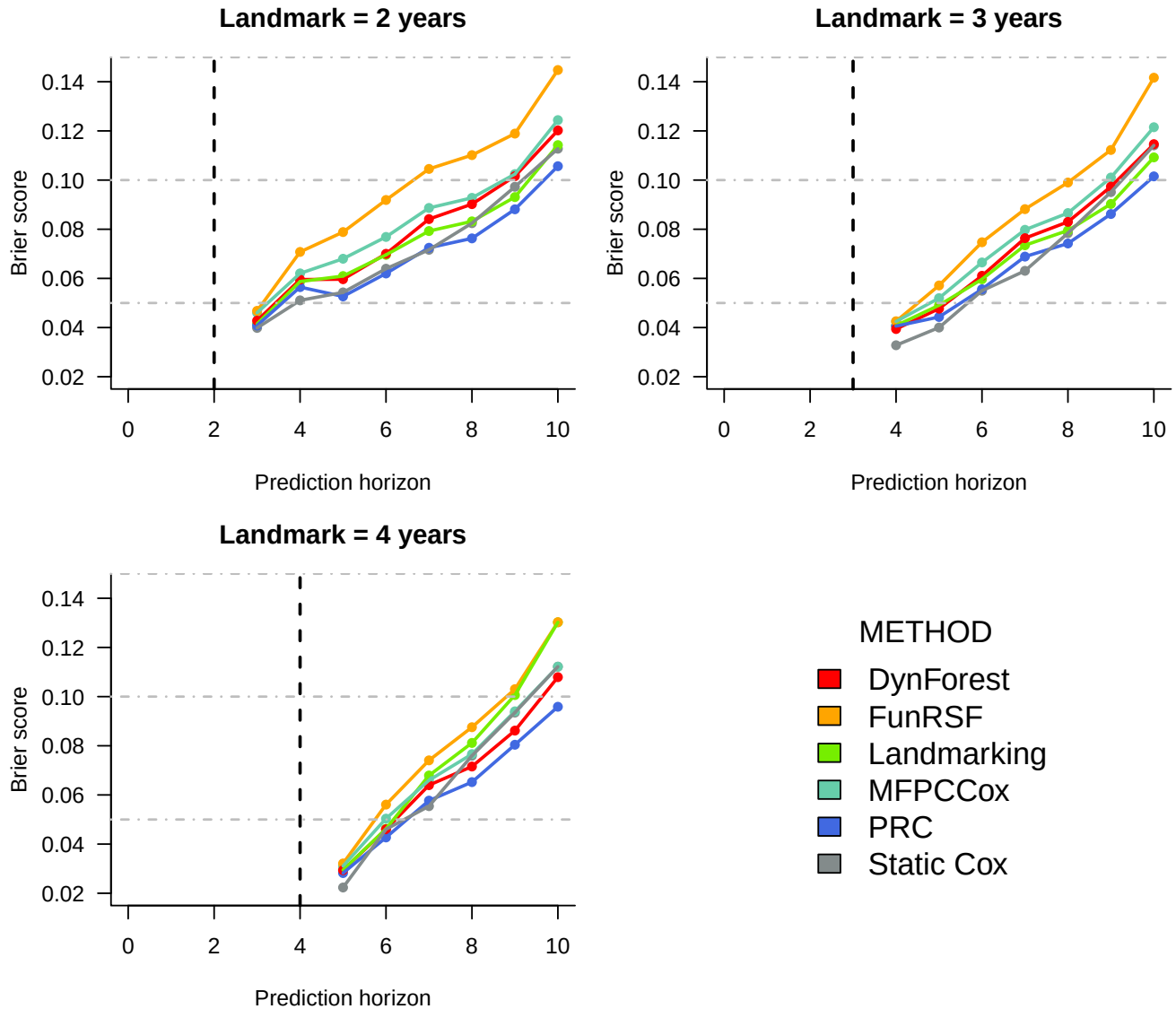
Preprint: arXiv:2403.14336v2

ADNI dataset: time-dependent AUC (higher = better)



Preprint: arXiv:2403.14336v2

ADNI dataset: Brier score (higher = worse)



Preprint: arXiv:2403.14336v2

ADNI dataset: computing time (higher = worse)

- Average computing time per CV fold (in **minutes**):

	Landmark			
Method	2	3	4	Average
Static Cox	0.009	0.007	0.006	0.007
Landmarking	0.010	0.007	0.006	0.008
MFPCCox	0.080	0.046	0.046	0.057
pencal	0.776	0.482	0.453	0.571
FunRSF	0.240	0.122	0.125	0.163
DynForest	12.501	9.077	8.099	9.892
n at risk	1226	954	721	

Preprint: arXiv:2403.14336v2

Preprint: arXiv:2403.14336v2

Results overview

- Similar results across the 3 datasets
- pencal, landmarking, DynForest $>$ MFPCCox, static Cox, FunRSF
  1. Methods that use LMMs $>$ methods using MFPCA
  2. Conditionally on method used to model longitudinal covariates (MFPCA / LMMs), methods that use Cox model $>$ methods that use RSF

Preprint: arXiv:2403.14336v2

Results overview

- Similar results across the 3 datasets
- pencal, landmarking, DynForest > MFPCCox, static Cox, FunRSF
  1. Methods that use LMMs > methods using MFPCA
  2. Conditionally on method used to model longitudinal covariates (MFPCA / LMMs), methods that use Cox model > methods that use RSF
- Relative performance of landmarking and static Cox worsens with higher landmark & horizon times $\Rightarrow$ important to make use of all available longitudinal data!
- DynForest estimation particularly slow (due to re-estimation of LMMs in each node)
- More details: arXiv:2403.14336v2 (Signorelli & Retif, 2025+)

Limitations of multistep methods

- MFPCA-based methods:
  1. regular measurement grid "needed" for MFPCA estimation $\rightarrow$ unrealistic & unflexible
  2. lack software implementation
- LMM-based methods: only LMMs. Using GLMMs would allow for more modelling flexibility
- Methods using RSF: need to choose value of multiple tuning parameters ☹

Preprint: arXiv:2403.14336v2

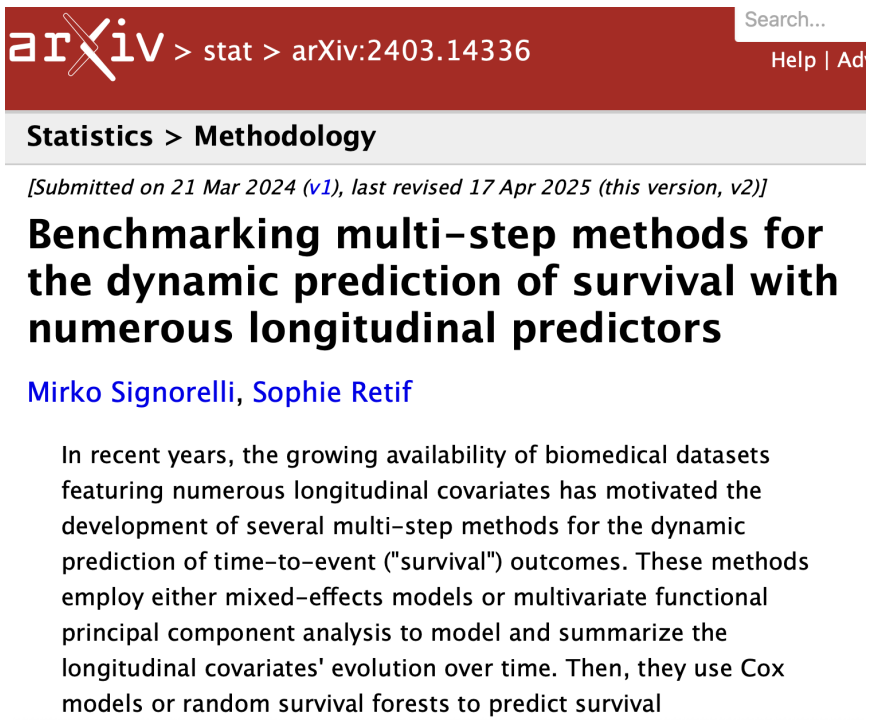
Limitations of multistep methods

- MFPCA-based methods:
  1. regular measurement grid "needed" for MFPCA estimation $\rightarrow$ unrealistic & unflexible
  2. lack software implementation
- LMM-based methods: only LMMs. Using GLMMs would allow for more modelling flexibility
- Methods using RSF: need to choose value of multiple tuning parameters ☹
- Dealing with more complex survival outcomes:
  1. competing risks: only in DynForest
  2. interval censoring: none of the methods

Preprint: arXiv:2403.14336v2

The End

- Preprint: arXiv:2403.14336v2


arXiv > stat > arXiv:2403.14336

Search...

Help | Ad

Statistics > Methodology

[Submitted on 21 Mar 2024 (v1), last revised 17 Apr 2025 (this version, v2)]

Benchmarking multi-step methods for the dynamic prediction of survival with numerous longitudinal predictors

Mirko Signorelli, Sophie Retif

In recent years, the growing availability of biomedical datasets featuring numerous longitudinal covariates has motivated the development of several multi-step methods for the dynamic prediction of time-to-event ("survival") outcomes. These methods employ either mixed-effects models or multivariate functional principal component analysis to model and summarize the longitudinal covariates' evolution over time. Then, they use Cox models or random survival forests to predict survival

: mirkosignorelli.github.io

: m.signorelli@math.leidenuniv.nl

: in/signorelli

Preprint: arXiv:2403.14336v2

References I

Devaux, A., Helmer, C., Genuer, R., & Proust-Lima, C. (2023). Random survival forests with multivariate longitudinal endogenous covariates. *Statistical Methods in Medical Research*, *32*(12), 2331–2346.

Devaux, A., Proust-Lima, C., & Genuer, R. (2024+). *Random Forests for time-fixed and time-dependent predictors: The DynForest R package*. arXiv:2302.02670.

Gomon, D., Putter, H., Fiocco, M., & Signorelli, M. (2024). Dynamic prediction of survival using multivariate functional principal component analysis: A strict landmarking approach. *Statistical Methods in Medical Research*, *33*(2), 256–272.

Li, K., & Luo, S. (2019). Dynamic prediction of Alzheimer's disease progression using features of multiple longitudinal outcomes and time-to-event data. *Statistics in Medicine*, *38*(24), 4804–4818.

References II

Lin, J., Li, K., & Luo, S. (2021). Functional survival forests for multivariate longitudinal outcomes: Dynamic prediction of Alzheimer's disease progression. *Statistical Methods in Medical Research*, *30*(1), 99–111.

Signorelli, M. (2024). Pencal: An R Package for the Dynamic Prediction of Survival with Many Longitudinal Predictors. *The R Journal*, *16*(2), 134–153.

Signorelli, M., & Retif, S. (2025+). Benchmarking multi-step methods for the dynamic prediction of survival with numerous longitudinal predictors. *arXiv Preprint arXiv:2403.14336*.

Signorelli, M., Spitali, P., Szigyarto, C. A., 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. *Statistics in Medicine*, *40*(27), 6178–6196.

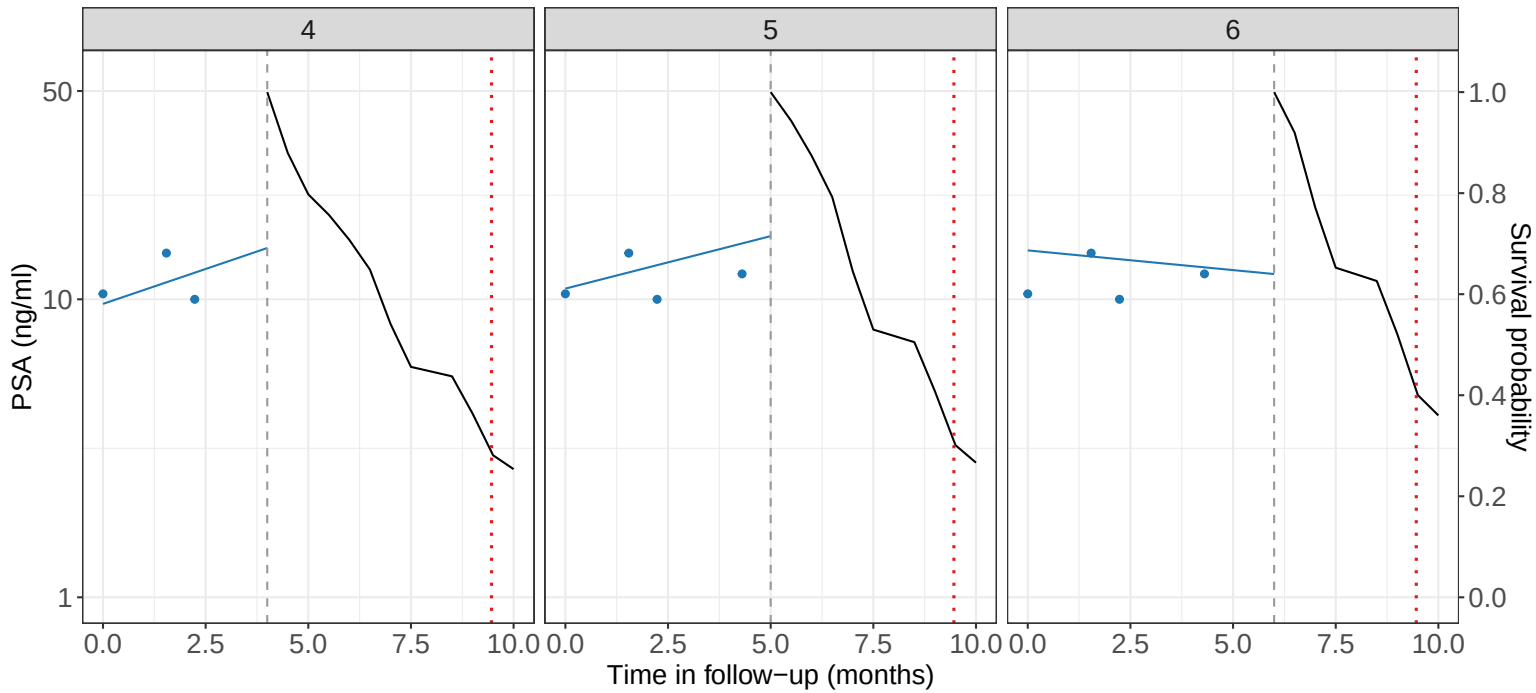
Preprint: arXiv:2403.14336v2

Benchmarking vs simulations

- Why benchmarking instead of MC simulations?
  1. Interest in how these methods perform on real datasets with complex and messy data
  2. Difficult to simulate realistic complex longitudinal + survival data
  3. Choices in simulation strategy could unfairly favour some methods over others

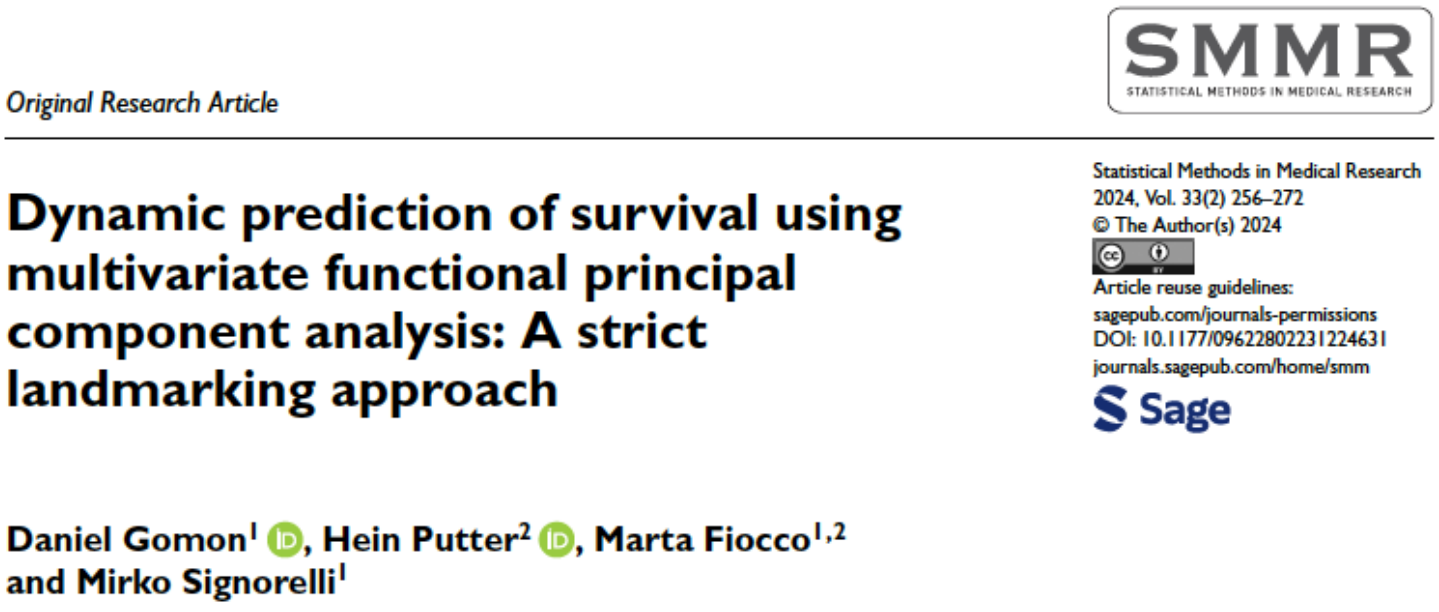
Preprint: arXiv:2403.14336v2

Dynamic prediction example



Preprint: arXiv:2403.14336v2

Strict vs relaxed landmarking with two-step methods



Original Research Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Dynamic prediction of survival using multivariate functional principal component analysis: A strict landmarking approach

Statistical Methods in Medical Research
2024, Vol. 33(2) 256–272

DOI: 10.1177/09622802231224631
journals.sagepub.com/home/smm
Sage

Daniel Gomon[1], Hein Putter[2], Marta Fiocco[1,2] and Mirko Signorelli[1]

Preprint: arXiv:2403.14336v2

Strict vs relaxed landmarking with two-step methods

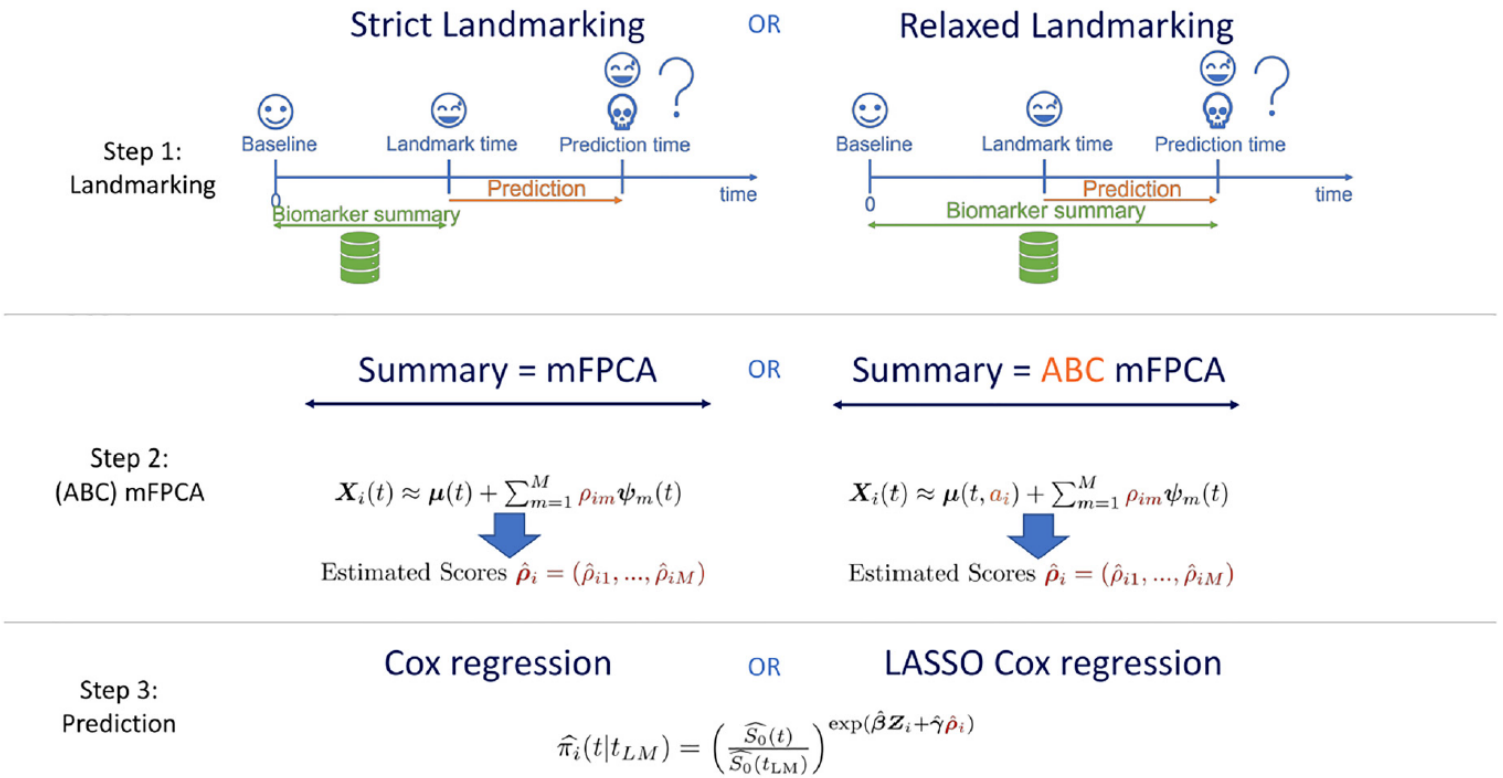


Figure 1. Graphical summary of the methods proposed in Section 2. See also Section 2.6.

Preprint: arXiv:2403.14336v2

Strict vs relaxed landmarking with two-step methods

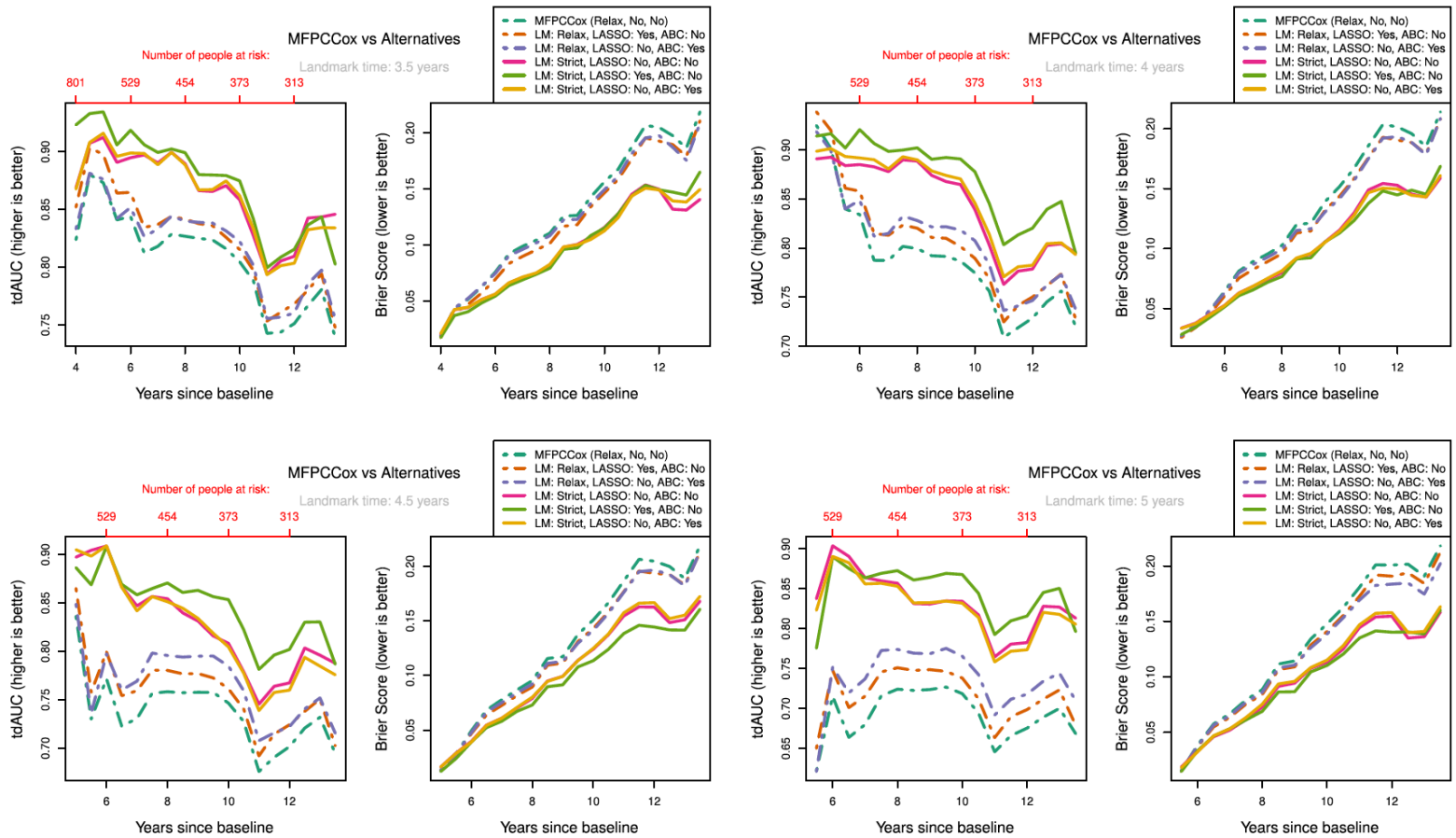
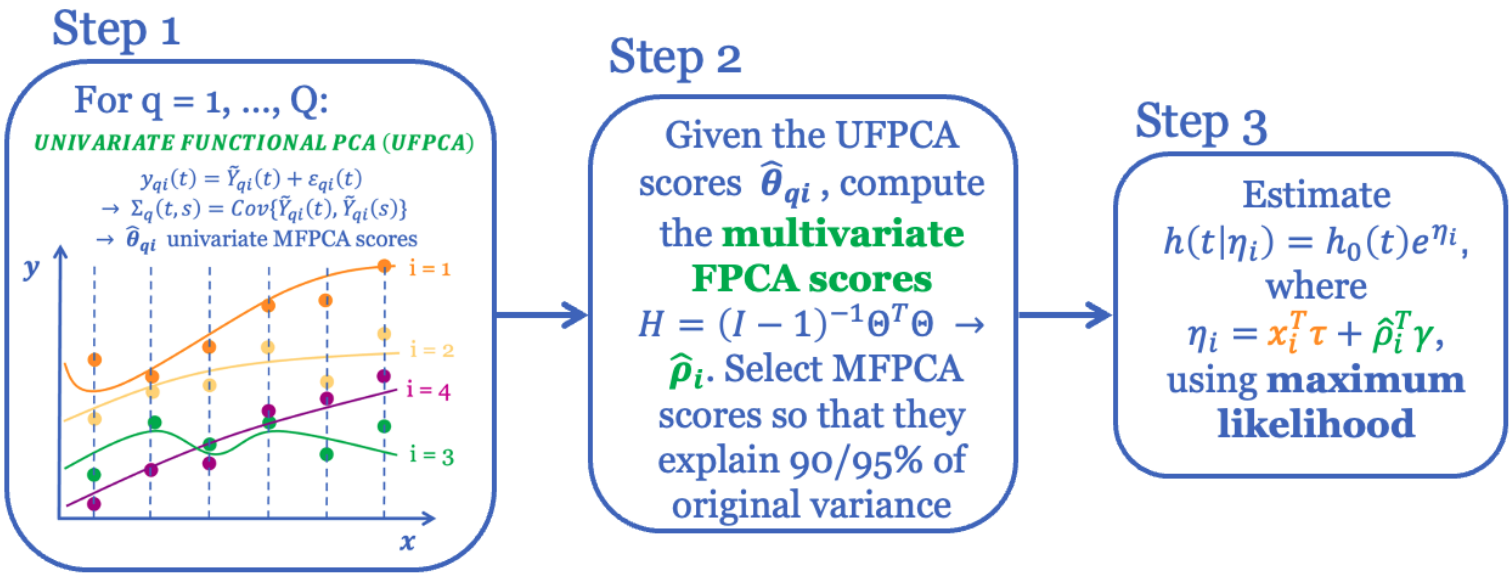


Figure 5. Measure of performance for LM, LASSO regularization (LASSO) and ABC methods at different landmark times on ADNI data. Validation scores were determined by using 20 times repeated 5-fold cross validation. Dashed lines: Relaxed landmarked methods. Solid lines: Strict landmarked methods. MFPCCox (LM: Relax, LASSO: No, ABC:No)[8] used as reference method. (a) Landmark time: 3.5 years; (b) Landmark time: 4 years; (c) Landmark time: 4.5 years; (d) Landmark time: 5 years. LM: landmark; ABC: age-based centered; ADNI: Alzheimer Disease Neuroimaging Initiative.

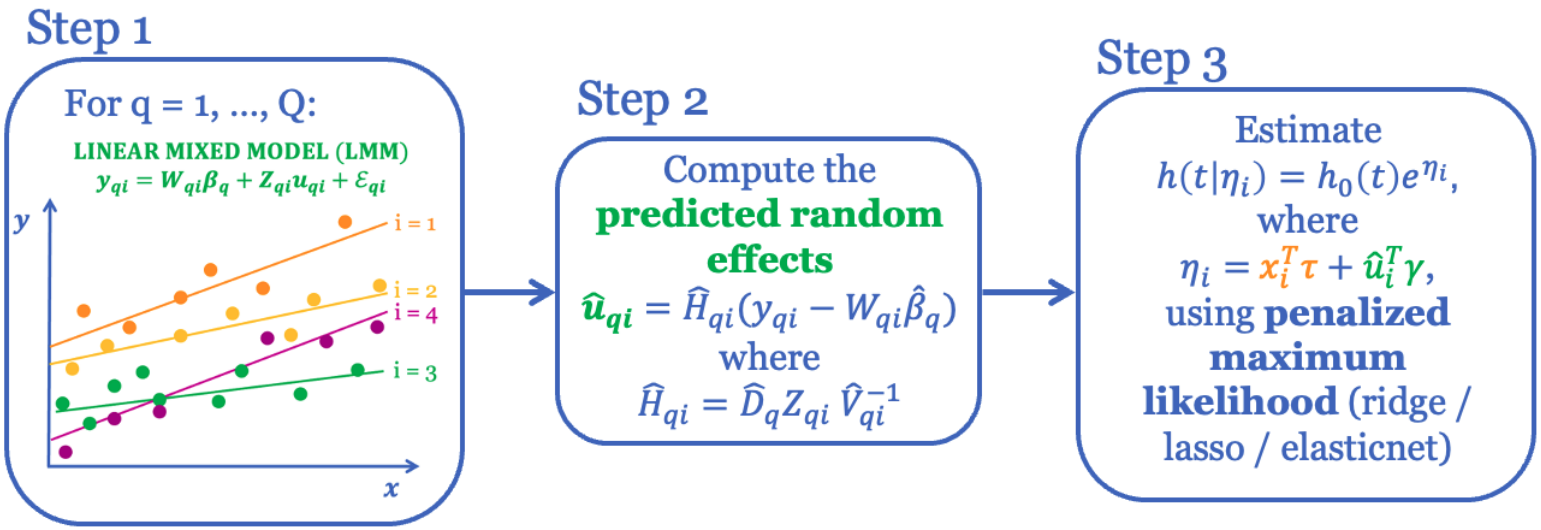
Preprint: arXiv:2403.14336v2

Multivariate Functional Principal Component Cox model (MFPCCox, Li & Luo (2019))



Preprint: arXiv:2403.14336v2

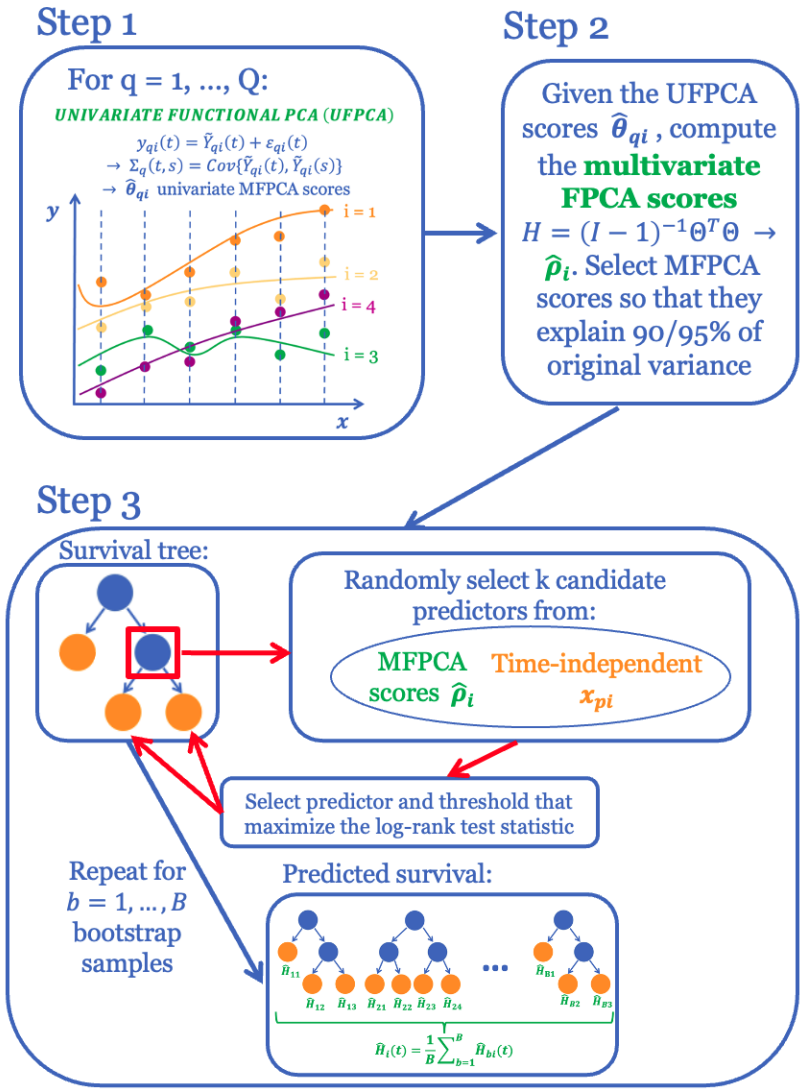
Penalized Regression Calibration (pencal, Signorelli et al. (2021))



- Steps 1-2: multivariate version with MLPMM possible

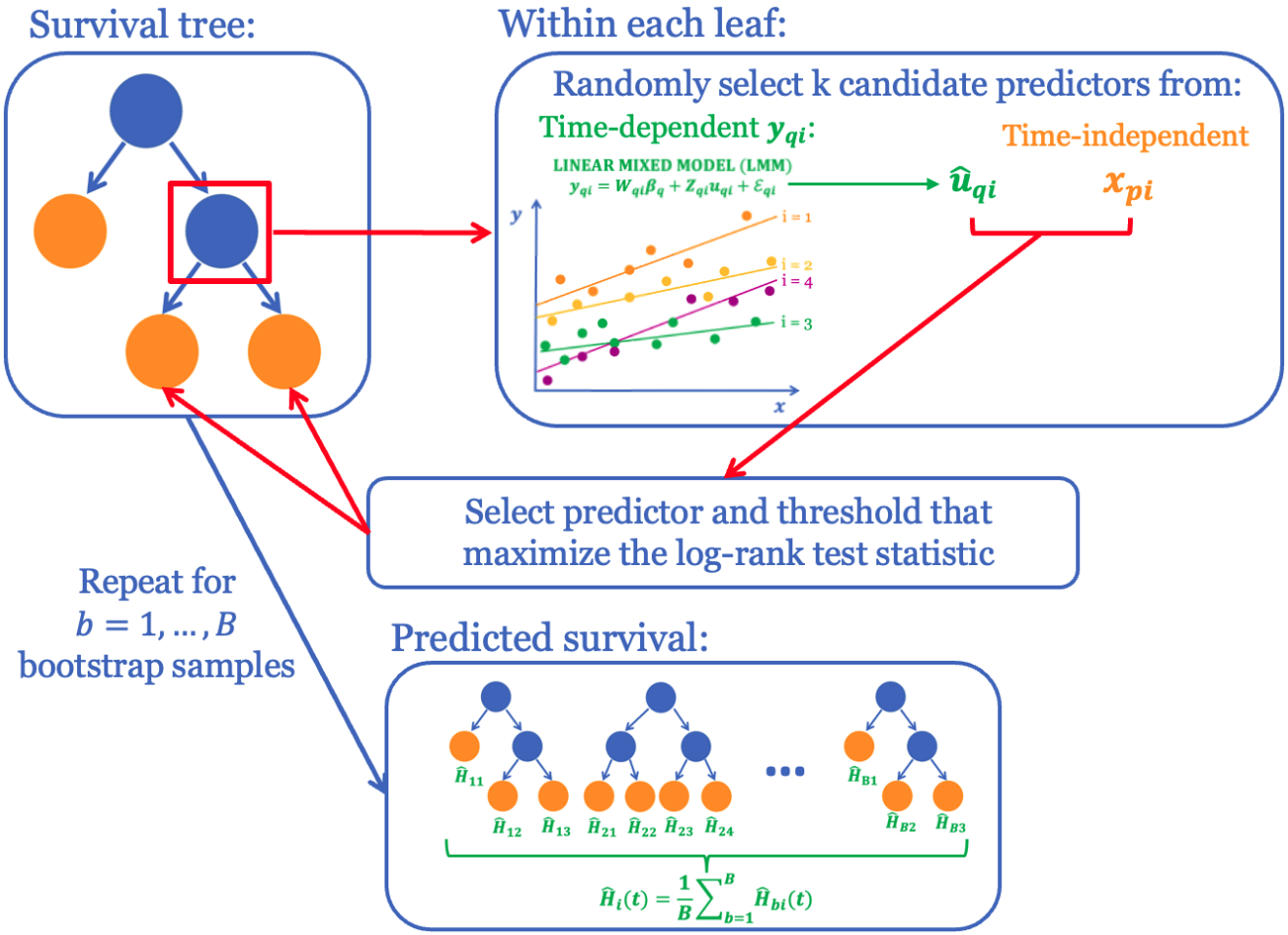
Preprint: arXiv:2403.14336v2

Functional Random Survival Forest (FunRSF, Lin et al. (2021))



Preprint: arXiv:2403.14336v2

Dynamic Random Survival Forest (DynForest, Devaux et al. (2023))



Preprint: arXiv:2403.14336v2