ASA Section on Lifetime Data Science

LiDS Newsletter

Volume 11, Number 2 – July 2026

ISSN 2473-5159

A section associated with the American Statistical Association

LiDS website: `https://community.amstat.org/lids/home/`
Published newsletters are archived under *Newsletters*.

In Brief

2026 Election Results
Chair-Elect 2027: Hongyuan Cao
Program Chair-Elect 2027: Wenbo Wu
Treasurer 2027–2029: Soyoung Kim
Charter revision approved

LiDS Sessions at 2026 JSM
Three invited sessions
Two topic-contributed sessions

LiDS Annual Business Meeting
At 2026 JSM, Boston, MA

2027 LiDS Conference
June 9–11, 2027: Milwaukee, WI
More information coming soon

LiDS Officers

Chair:	Michael Fay
Chair-Elect:	Ronghui (Lily) Xu
Past Chair:	Zhezhen Jin
Admin. Officer 2025–2027:	Fei Gao
Treasurer 2024–2026:	Yifei Sun
Program Chair:	Lu Mao
Program Chair-Elect:	Joel Dubin
Past Program Chair:	Mengling Liu
COS Representative:	Shanshan Zhao
Webinar Committee Chair:	Wenbo Wu
Webinar Committee Co-chair:	Zhe Fei
Communications Officer:	Steven Chiou

Chair's Message

One of my memories from graduate school stays with me to this day. I don't know why I still remember it. Maybe it is because so many times in grad school I was confused by things and was not sure I would be able to understand statistics at a deep level. I had been working on my dissertation about using permutation tests for clusters of individual data where each cluster is represented as a distribution of responses. I was trying to apply the same idea to right censored data, except now each individual was represented by their own personal time-to-event distribution. I tried many ideas, and did not keep my notes, so I do not remember what I tried and how many things were really stupid ideas. But I finally had a method for estimating individual time-to-event distributions that seemed to make sense to me. Then I tried to see how it related to what others had done. It turns out that if I averaged the individual distributions by "my" method, I got the Kaplan–Meier estimate. You might think that would make me despair that I would never think of something new. But for some reason, I only felt the joy that I had figured something out myself. Things were fitting together in ways that seemed to make sense to me. I was so excited by that discovery, that I had to get out of my basement cubicle and take a walk around the courtyard.

You may think that is such a non-story. Or a story that you could tell to only a small group of people that would get it. And folks, you readers are my best shot at an audience that would get it! The joy of figuring out something. Even if it is something that is "well known" in the literature. That is part of what I love about my job. And that is what is nice about forming a group that is interested in what you are interested in. You have people to bounce ideas off of. People that may actually care about what you have been thinking about for hours and hours over the course of the year. So welcome to the LiDS Newsletter with some ideas and news about Lifetime Data Science. Read to see what the section has planned in the coming months and year, and how you can be part of it.

Michael Fay
Chair 2026

Mathematical Statistician
National Institute of Allergy and Infectious Diseases

Scan to visit the LiDS website

Nominations Invited for Positions on the LiDS Executive Committee

In the upcoming 2026 LiDS Section election, the following positions will be open:

- Chair-Elect (2028)
- Program Chair-Elect (2028)
- Secretary (2028–2030)
- Council of Sections Representative (2028–2030)

Each position requires a three-year commitment. The Chair-Elect will serve as Chair-Elect in 2028, Section Chair in 2029, and Past Chair in 2030. The Program Chair-Elect will serve as Program Chair-Elect in 2028, Program Chair in 2029, and Past Program Chair in 2030. The Secretary and Council of Sections Representative will serve three-year terms from 2028 through 2030.

A brief description of the responsibilities for each position is available in the LiDS Section Charter. If you are interested in running for one of these positions, please complete the online interest survey (`https://docs.google.com/forms/d/e/1FAIpQLSdjFwQYer4HuWzx8kjNYeP4w81BhB07Sjlqz4uTqIKWr5wblQ/viewform?usp=header`) by **August 14, 2026**. We encourage all interested members to consider serving in a leadership role and helping shape the future of the LiDS Section. Thank you for your interest and support.

Zhezhen Jin
Nomination Committee Chair 2026

Professor
Columbia University

2026 Election Results

Congratulations to the newly elected members of the LiDS Executive Committee! The Charter revision of the Lifetime Data Science Section was also approved by the membership vote held alongside the election.

Chair-Elect 2027 Dr. Hongyuan Cao is a professor in the Department of Biostatistics and Medical Informatics at the University of Wisconsin–Madison. She received her Ph.D. in Statistics from the University of North Carolina at Chapel Hill. Her research interests include dynamic risk stratification and prediction, genetic epidemiology and statistical genetics, and high-dimensional clinical AI. She has served the community in numerous roles within the ASA and other professional societies and is honored to serve as 2026 Chair-Elect of the ASA Lifetime Data Analysis Section. She is an elected Fellow of the American Statistical Association.

Hongyuan Cao
Professor
University of Wisconsin–Madison

Program Chair-Elect 2027 Dr. Wenbo Wu is a tenure-track Assistant Professor in the Department of Health Policy and Management, the Center for Population Health Information Technology, and the Department of Biostatistics at the Johns Hopkins Bloomberg School of Public Health, an affiliated faculty member in the Armstrong Institute for Patient Safety and Quality at the Johns Hopkins School of Medicine, and a faculty member in the Data Science and AI Institute and a core faculty member in the Hopkins Business of Health Initiative of the Johns Hopkins University. Previously, he was an Assistant Professor in the Division of Biostatistics of the Department of Population Health and the Division of Nephrology of the Department of Medicine at the New York University (NYU) Grossman School of Medicine, as well as an affiliated faculty member of the NYU Center for Data Science. His research synthesizes state-of-the-art methods from statistics, causal inference, machine learning, optimization, and computational science to address critical and far-reaching issues in health equity, outcomes and services, and clinical practice, leveraging disease registries, administrative claims, electronic health records, and randomized controlled trials. In 2022, Wenbo received his joint Ph.D. in Biostatistics and Scientific Computing from the University of Michigan.

Wenbo Wu
Assistant Professor
Johns Hopkins Bloomberg School of Public Health

Treasurer 2027–2029 Dr. Soyoung Kim is an Associate Professor and Graduate Director of the PhD Program in the Division of Biostatistics within the Data Science Institute at the Medical College of Wisconsin. She received her Ph.D. in Biostatistics from the University of North Carolina at Chapel Hill and completed postdoctoral training at the Fred Hutchinson Cancer Center. Dr. Kim's research focuses on survival analysis, competing risks, causal inference, missing data, case-cohort studies, and machine learning methods for biomedical research. Her methodological work centers on the development of statistical and computational approaches for analyzing complex time-to-event data, including data with missing information and large-scale biomedical datasets. She collaborates extensively with clinical and translational investigators, particularly through the Center for International Blood and Marrow Transplant Research (CIBMTR) at the Medical College of Wisconsin, to develop innovative statistical methods that address important challenges in hematopoietic cell transplantation, cellular therapies, and public health research.

Soyoung Kim
Associate Professor
Medical College of Wisconsin

Outstanding Service Award

The recipient of the 2026 ASA LiDS Outstanding Service Award is Dr. **Mei-Cheng Wang**, recognized "for her outstanding contributions to the Section over the past 12 years, including service as a founding Executive Committee member, Chair of the LiDS Interest Group, organizer of the inaugural LiDS Conference, sustained contributions to LiDS conferences, and numerous other roles benefiting LiDS members."

Dr. Wang is a Professor of Biostatistics at the Johns Hopkins Bloomberg School of Public Health and an internationally recognized leader in survival analysis, truncated and length-biased data, and recurrent event modeling. Her pioneering research has had a profound impact on statistical methodology and its applications to public health, epidemiology, and clinical research. Beyond her outstanding scholarly achievements, Dr. Wang has distinguished herself through extraordinary service to the statistical community, particularly to the ASA Lifetime Data Science (LiDS) Section. Over the past 12 years, she has played a pivotal role in the growth and success of the Section. As a founding member of the Executive Committee, Chair of the LiDS Interest Group, organizer of the inaugural LiDS Conference, and a dedicated contributor to subsequent LiDS conferences and numerous other Section activities, she has provided visionary leadership and unwavering commitment. Her efforts have helped build a vibrant LiDS community and advanced its mission to promote research, education, and collaboration in lifetime data science.

Please join us in congratulating Dr. Mei-Cheng Wang on this well-deserved honor as the recipient of the 2026 ASA Lifetime Data Science Outstanding Service Award.

Zhezhen Jin
Nomination Committee Chair 2026

Professor
Columbia University

Report from the Administrative Officer

As of July 10, 2026, our section had a total of 876 members.

Our official website, maintained by Communication Officer Steven Chiou (email: `schiou@smu.edu`), can be accessed at: `https://community.amstat.org/lids/home/`.

The Section's 2026 Annual Business Meeting will take place during JSM 2026 on Monday, August 3, from 5:30 PM to 6:30 PM in Room CC-102B at the Thomas M. Menino Convention & Exhibition Center. All members are welcome to attend.

Fei Gao
Administrative Officer 2025–2027

Associate Professor
Fred Hutchinson Cancer Center

Treasurer's Report

The beginning balance of the LiDS Section account on January 1, 2026, was $74,365, and the ending balance on June 30, 2026, was $70,553. The total income was $1,188, composed of membership dues and webinar registration fees. The total expenses were $5,000 for the 2026 JSM Student Paper Award, resulting in a net loss of $3,812 for the period.

Beginning Balance	**01/01/2026**	**$74,365**
Income		
Membership Dues		$1,038
Registration Fees		$150
Total Income		**$1,188**
Expense		
Awards/Plaques		$5,000
Total Expense		**$5,000**
Net Gain/Loss		**−$3,812**
Ending Balance	**06/30/2026**	**$70,553**

Yifei Sun
Treasurer 2024–2026

Associate Professor
Columbia University

LiDS Activities at the 2026 JSM

The 2026 Joint Statistical Meetings (JSM) will be held in Boston, Massachusetts, August 1–6, 2026, with the theme "Communities in Action: Advancing Society." Many thanks to the session organizers and speakers who chose LiDS as a primary sponsor or co-sponsor.

LiDS is sponsoring the following three invited sessions:

- *Recent Advances in the Use of Area Under the Curve Methods for Time-to-Event Analyses in Clinical Trials* Organizer: Bo Huang, Pfizer | Tuesday, August 4, 10:30am – 12:20pm
- *Causal Inference Meets Data Integration and Survival Analysis* Organizer: Guanbo Wang, Dartmouth College | Wednesday, August 5, 8:30am – 10:20am
- *Recent Theoretical and Methodological Advances in Causal Inference for Survival Data* Organizer: Limin Peng, Emory University | Wednesday, August 5, 2:00pm – 3:50pm

LiDS is also sponsoring the following topic-contributed sessions:

- *Lifetime Data Science Section Student Paper Awards* Organizer: Mengling Liu, New York University. Speakers: James Peng, Han Lu, Shike Xu, Yuchen Qi, Yue Zhan
- *Recent Advances in Survival Analysis for Correlated Time-to-event Data* Organizer: Mei-Ling Ting Lee, University of Maryland at College Park

Please mark your calendar and attend these sessions to support LiDS.

Lu Mao
Program Chair 2026

Associate Professor
University of Wisconsin–Madison

2027 JSM Call for Invited and Topic-Contributed Session Proposals

The 2027 JSM will be held in Chicago, Illinois, from August 7–12, 2027. The LiDS Program Committee is soliciting proposals for two invited sessions and a number of topic-contributed sessions to be sponsored by the Lifetime Data Science Section.

Invited Sessions

- **Duration:** 110 minutes
- **Formats:** Flexible; e.g., 2–3 speakers with a discussant, 4 speakers, or panel discussions with 3–5 panelists
- **Submission Opens:** Now open
- **Submission Deadline:** September 9, 2026 (11:59 p.m. ET)

Topic-Contributed Sessions Each topic-contributed session must include five presentations (which may include discussants). Details on the submission timeline will be announced later this year.

A link to the submission portal will be provided once it becomes available. You will need the following information to submit your session proposal:

- Session type (Invited Paper or Invited Panel)
- Sponsor: Select "Lifetime Data Science Section" as the primary sponsor (you may choose up to two additional co-sponsors)
- Session title
- Session description (including focus, content, timeliness, appeal, and format)
- Organizer's name, affiliation, and email
- Chair's name, affiliation, and email
- Speakers and any discussants (names, affiliations, and emails)

Please contact Joel Dubin at `jdubin@uwaterloo.ca` should you have any question regarding the LiDS-sponsored invited sessions.

Joel Dubin
Program Chair-Elect 2026

Professor
University of Waterloo

2026 LiDS Webinars

Recent LiDS Webinar:
The LiDS Section hosted a successful webinar on March 4, 2026, featuring Professor **James Zou** (Stanford University), an Associate Professor of Biomedical Data Science with courtesy appointments in Computer Science and Electrical Engineering. His presentation, *AI Agents to Accelerate Biomedical Discoveries*, explored how large language model–based agents can support scientific research. Dr. Zou discussed the Virtual Lab, in which AI scientist agents collaborate on open-ended research projects; the development and experimental validation of nanobody binders targeting recent COVID variants; and Paper2Agent, a framework that transforms research papers into interactive AI agents.
Upcoming LiDS Webinar:
An upcoming LiDS webinar is planned for the summer, featuring Dr. **Yanxun Xu**, Professor of Applied Mathematics and Statistics and Joseph & Suzanne Jenniches Faculty Scholar at Johns Hopkins University. The webinar is scheduled for **September 8, from 11:00 a.m. to 1:00 p.m.** The presentation topic and registration information will follow shortly.

For the latest updates, please check the LiDS Webinar Series page at `https://community.amstat.org/lids/events/webinar-series`.
We are also pleased to announce that LiDS webinars are now free for LiDS section members.

Zhe Fei
Webinar Committee Co-Chair

Assistant Professor
University of California, Riverside

2027 LiDS Conference Updates

The fifth Conference on Lifetime Data Science will be held June 9–11, 2027, in Milwaukee, Wisconsin. The hotel contract has been signed, and a Scientific Program Committee has been formed to recruit speakers and develop the program. The conference will consist of invited sessions. Further details, including the theme, keynote speakers, short courses, and registration, will follow in future newsletters.

Invited Session Proposals Requested

The LiDS Section is accepting invited session proposals for the 2027 LiDS Conference in Milwaukee, WI on June 9–11, 2027. An invited session consists of either three or four presenters. A session proposal needs to provide a title, a short description of the session, and information about the speakers, including their talk titles and abstracts, names, emails, and affiliations. A one-talk rule will be applied, so each speaker can only give one invited talk. It is required to confirm all speakers' availability before the proposal is submitted.

Proposals should be submitted by September 30 to the Co-chairs of the Scientific Program Committee, Michael Fay (`mfay@niaid.nih.gov`) and Ronghui (Lily) Xu (`rxu@health.ucsd.edu`).

News from Lifetime Data Analysis

Lifetime Data Analysis is the only journal dedicated to statistical methods and applications for lifetime data. The journal advances and promotes statistical science in various applied fields that deal with lifetime data, including actuarial science, economics, engineering, environmental sciences, management, medicine, operations research, public health, and social and behavioral sciences. The journal can be accessed at `https://link.springer.com/journal/10985`.

New Articles in Lifetime Data Analysis

Volume 32, Issue 2 of Lifetime Data Analysis has been published:

- Variable selection with broken adaptive ridge regression for interval-censored competing risks data *by* F. Mahmoudi, C. Li, K. Cai, & X. Lu.
- Reduction techniques for survival analysis *by* J. Piller, L. Orsini, S. Wiegrebe, S.H. Langbein, L. Burk, J. Zobolas, P. Studener, M. Goeswein, & A. Bender.
- Training-set conditionally valid prediction sets with right-censored data *by* W. Si, & H. Qiu.
- Kolmogorov–Smirnov and Cramér–von Mises tests for the k-sample problem with left-truncated and right-censored data *by* A. Lago, J.C. Pardo-Fernández, & J. de Uña-Álvarez.
- Gradient boosting-based discrete failure time model for selecting time-varying effects and interactions *by* L. Luo, K. He, & J.M.G. Taylor.
- Instrumental variable estimation of a hazard ratio for treatment with a waiting time without specifying its dependence on unmeasured confounders: application to a procedural registry *by* T.A. MacKenzie, P. Martínez-Camblor, H. Chen, J.A. Columbo, & A.J. O'Malley.
- A dependent and censored first hitting-time model with compound Poisson processes *by* M. Escobar-Bach, A. Popier, & M. Sahin.
- The semi-competing risk problem revisited *by* R.L. Prentice, & A.K. Aragaki.
- Q-learning via deep learning-based Buckley–James method for non-linear censored data *by* J. Lee, & J.-M. Kim.
- Comparing adaptive treatment strategies in two-stage randomized trials with grouped survival data *by* M. Sparkman, S. Halabi, & Z. Li.
- Nonparametric estimation of average effects of a continuous treatment for survival data with a cured fraction *by* H. Liu, & Y. Peng.
- Estimating attributable risk functions for censored time-to-event in disease prevention research *by* Y.Q. Chen, Y. Wang, X. Zhang, & R.L. Prentice.
- Nonparametric estimation of a state entry time distribution conditional on a "past" state occupation in a progressive multistate model with current status data *by* S. Anyaso-Samuel, & S. Datta.
- Nonparametric estimation of conditional survival function with time-varying covariates using DeepONet *by* B. Hu, & B. Nan.
- Partially linear Cox model with neural networks for left-truncated data *by* S. Li, L. Shao, & S. Li.
- Inference for cause-specific cox model absolute risk in cohort subsampling designs *by* L. Etiévant, & M.H. Gail.
- Semiparametric regression analysis of interval-censored competing risks data under additive hazards model with missing event types *by* R. Jia, Y. Lou, J. Sun, & P. Wang.
- Doubly robust g-estimation of structural nested cumulative survival time models with non-ignorable, non-monotone missing data in time-varying confounders *by* Y. Takeuchi, S. Komukai, A. Goto, & T. Shinozaki.
- Optimal designs for discrete-time survival models with competing risks *by* X. Zhou, Y. Wang, R. Yue, & W.K. Wong.

Articles in Volume 32, Issue 1 are:

- Robust functional Cox regression model *by* G.B. Sezer, & U. Beyaztas.
- Estimating treatment effects on duration with disease: a principal stratification framework *by* E.T. Parner
- Deep learning for the change-point Cox model with current status data *by* Q. Huang, A. Feng, Q. Wu, & X. Tong.
- Generalized win-odds regression models for composite endpoints *by* B. Wang, Z. Wang, & Y. Cheng.
- Wasserstein GAN-based estimation for conditional distribution function with current status data *by* W. Su, C. Liu, G. Yin, & J. Huang.
- Estimation of the interpretable heterogeneous treatment effect with causal subgroup discovery in survival outcomes *by* N. Bo, & Y. Ding.
- Deep tobit model: an integrated framework for high-dimensional censored regression with variable selection *by* T. Wu, J. Hu, Z.-S. Ye, & N. Chen.
- On Multiple Time Scales and Collapsibility *by* D. Oakes
- Beyond Bonferroni: new multiple contrast tests for time-to-event data under non-proportional hazards *by* I. Dormuth, C. Herrmann, F. Konietschke, M. Pauly, M. Wirth, & M. Ditzhaus.
- Confidence intervals for high-dimensional accelerated failure time models under measurement errors *by* Q. Yu, X. Zhou, J. Zhou, & Z. Zheng.
- Two-stage recurrent events random effects models *by* T.H. Scheike
- Continuously updated estimation of conditional hazard functions *by* D. Aurouet, & V. Patilea.
- Bayesian semiparametric partially linear cure models with partly interval-censored data *by* Y. Guo, C. Wang, & X. Liu.
- Integrating high-dimensional censored data under privacy constraints via localized computations *by* B. Huang, Y. Liu, & X. Ye.
- A flexible copula model for bivariate survival data with dependent censoring *by* R. Adatorwovor, & Y. Pan.
- Reliability and estimation of the zero-inflated transmuted geometric distribution with applications and actuarial insights *by* K. Sharma, P.J. Hazarika, M.S. Eliwa, & M. El-Morshedy.

Articles in Volume 31, Issue 4 are:

- Pseudo-observations and super learner for the estimation of the restricted mean survival time *by* A. Cwiling, V. Perduca, & O. Bouaziz. Pages 713–746
- A comparison of Kaplan–Meier-based inverse probability of censoring weighted regression methods *by* M. Overgaard Pages

747–783

- Assessing delayed treatment benefits of immunotherapy using long-term average hazard: a novel test/estimation approach *by* M. Horiguchi, L. Tian, K.L. Kehl, & H. Uno. Pages 784–809
- Statistical methods for composite analysis of recurrent and terminal events in clinical trials *by* Y. Huang, D. Schaubel, & M. Zhang. Pages 810–829
- Simultaneous clustering and joint modeling of multivariate binary longitudinal and time-to-event data *by* S. Chattopadhyay, S. Basu, S. Bhattacharyya, M.P. Gogoi, & K. Das. Pages 830–851
- Analysis of interval censored survival data in sequential multiple assignment randomized trials *by* Z. Li Pages 852–868
- Multi-source analyses of average treatment effects with failure time outcomes *by* L. Wen, J.A. Steingrimsson, S.E. Robertson, & I.J. Dahabreh. Pages 869–897
- Causal effect estimation on restricted mean survival time under case-cohort design via propensity score stratification *by* W.-E. Lu, & A. Ni. Pages 898–931
- Bayesian joint models for longitudinal, recurrent, and terminal event data *by* E.M. Damone, M.A. Psioda, & J.G. Ibrahim. Pages 932–949
- Bayesian joint analysis of longitudinal data and interval-censored failure time data *by* Y. Mao, L. Wang, & X. Sui. Pages 950–969
- Bayesian generalized method of moments applied to pseudo-observations in survival analysis *by* L. Orsini, C. Brard, E. Lesaffre, G. Yin, D. Dejardin, & G. Le Teuff. Pages 970–993
- Modelling dependent censoring in time-to-event data using boosting copula regression *by* A. Strömer, N. Klein, I. Van Keilegom, & A. Mayr. Pages 994–1016

Mei-Ling Ting Lee
Editor-in-Chief, Lifetime Data Analysis

Professor
University of Maryland

Dynamic Prediction with the pencal Package

Dynamic prediction of time-to-event (survival) outcomes makes it possible to make use of repeated measurements data collected during follow-up to better predict the probability that an individual will experience an event of interest in the future. While joint models (Tsiatis and Davidian, 2004) provide an elegant framework for combining longitudinal and survival data, numerical errors during model estimation and excessive computing time have so far limited the applicability of joint models to prediction models with a handful (most often between 1 and 3) of longitudinal predictors. Recently, this limitation has motivated the development of several multi-step dynamic prediction methods that make it possible to use dozens, hundreds or even thousands of longitudinal covariates for dynamic prediction (see Signorelli and Retif (2025) for a review).

The `R` package `pencal` (Signorelli, 2024) provides a user-friendly `R` implementation of the Penalized Regression Calibration (PRC) method, which uses mixed-effects models to model the evolution over time of the longitudinal covariates, and a penalized Cox model to predict survival probabilities. `pencal` provides an integrated workflow for dynamic prediction and performance evaluation, automating the evaluation of several measures of predictive performance by means of a cluster bootstrap optimism correction procedure (Signorelli et al., 2021). With built-in parallelization and a modular design, the package offers a more scalable and robust alternative to joint modelling for dynamic prediction problems.

Example, goal, and data preparation

We illustrate the main functionalities of `pencal` using the PBC2 trial data, which can be loaded using `data(pbc2data)`:

```
library(pencal)
packageVersion("pencal")

# [1] '2.3.1'

data(pbc2data)
```

`pbc2data` is a list that contains two data frames: `baselineInfo`, a dataset in wide format containing the survival information and baseline covariates; and `longitudinalInfo`, a dataset in long format containing the longitudinal covariates. We extract these objects below:

```
sdata <- pbc2data$baselineInfo
ldata <- pbc2data$longitudinalInfo
```

Our goal is to predict the probability that an individual will be event-free at a time point $t > \ell$ given that they were event-free up until ℓ, using as predictors a vector of baseline covariates x_i and a matrix $\mathcal{Y}_i(\ell)$ that contains the repeated measurements of the longitudinal covariates up until ℓ:

$$S_i(t|\ell, x_i, \mathcal{Y}_i(\ell)) = P(T_i > t|T_i > \ell, x_i, \mathcal{Y}_i(\ell)). \quad (1)$$

In the example below, we choose $\ell = 2$ years since study entry as landmark time, focusing our interest on those subjects who are still at risk of death after two years. Furthermore, we retain only the longitudinal measurements collected up until ℓ, thus implementing the strict data landmarking approach recommended by Gomon et al. (2024):

```
# set the landmark time
lmark <- 2
# remove subjects who had event or
# were censored before landmark
sdata <- subset(sdata, time > lmark)
ldata <- subset(ldata, id %in% sdata$id)
# remove measurements taken after landmark:
ldata <- subset(ldata, futime <= lmark)
```

Several biomarkers in the PBC2 dataset exhibit substantial right-skewness. While their transformation is not mandatory, modelling highly skewed outcomes with linear mixed-effects models is more likely to lead to numerical problems during

the estimation of linear mixed models. Hereafter, we apply a log-transformation to 5 of the 7 longitudinal covariates that we will use to predict $S_i(t|\ell)$:

```
ldata <- dplyr::mutate(ldata,
                       lSB = log(serBilir),
                       lSC = log(serChol),
                       lAlk = log(alkaline),
                       lSGOT = log(SGOT),
                       lProt = log(prothrombin))
```

Model estimation

The estimation of PRC proceeds through three sequential steps. The first step models the longitudinal covariates using mixed-effects models. In the example below, the `fit_lmms` function is used to estimate a linear mixed model of the form $y_{ij} = \beta_1 + u_{i1} + (\beta_2 + u_{2i})a_{ij} + \varepsilon_{ij}$, where y_{ij} is the response variable and a_{ij} the age at which y_{ij} was recorded, to each longitudinal covariate:

```
lmms <- fit_lmms(
  y.names = c('lSB', 'lSC', 'albumin',
    'lAlk', 'lSGOT', 'platelets', 'lProt'),
  fixefs = ~ age, ranefs = ~ age | id,
  t.from.base = fuptime,
  long.data = ldata, surv.data = sdata,
  verbose = FALSE)
```

Next, we use the function `summarize_lmms` to compute the predicted random effects $\hat{u}_i$, which we will then use as subject-specific summaries of the longitudinal trajectories to predict survival:

```
predRanefs <- summarize_lmms(object = lmms, verbose = FALSE)
```

In the last estimation step, we estimate a Cox model using as covariates the predicted random effects $\hat{u}_i$ obtained in the previous step, together with a vector of baseline covariates x_i:

$$h(t \mid x_i, \hat{u}_i) = h_0(t) \exp\{\gamma^\top x_i + \delta^\top \hat{u}_i\}.$$

We employ penalized likelihood methods for the estimation via `fit_prclmm`; in the example below, we rely on the ridge penalty:

```
pencox <- fit_prclmm(object = predRanefs,
           surv.data = sdata,
           baseline.covs = ~ baselineAge + sex + treatment,
           penalty = 'ridge',
           standardize = TRUE, verbose = FALSE)
```

`summary` methods are implemented for each modelling step, making it easier to have a look at each of the different components of the PRC model. For example:

```
summary(lmms, yname = 'albumin', what = 'tTable')
```

```
#               Value Std.Error  DF t-value  p-value
# (Intercept)  3.8227    0.1062 566    36.0 1.6e-148
# age         -0.0057    0.0021 566    -2.7  6.4e-03
```

```
summary(predRanefs)
```

```
# Number of predicted random effect variables: 14
# Sample size: 278
```

```
summary(pencox)
```

```
# Fitted model: PRC-LMM
# Penalty function used: ridge
# Tuning parameters:
#   lambda alpha
# 1   0.21     0
# Sample size: 278
# Number of events: 107
# Bootstrap optimism correction: not computed
# Penalized likelihood estimates (rounded to 4 digits):
#   baselineAge sexfemale treatmentD-penicil lSB_b_int
# 1       0.046     -0.32             -0.016      0.43
#   lSB_b_age lSC_b_int lSC_b_age albumin_b_int albumin_b_age
# 1       111     0.097       -10          -1.1         23105
#   lAlk_b_int lAlk_b_age lSGOT_b_int lSGOT_b_age
# 1      0.087        -12        0.23         271
#   platelets_b_int platelets_b_age lProt_b_int lProt_b_age
# 1         -0.0011           -0.21         2.8        -552
```

Dynamic prediction

Next, we can proceed to compute the predicted survival probabilities in equation (1) using `survpred_prclmm`. By setting `times = 3:5`, we can compute $\hat{S}_i(t|\ell = 2)$, $t = 3, 4, 5$:

```
preds <- survpred_prclmm(step1 = lmms, step2 = predRanefs,
                         step3 = pencox, times = 3:5)
ls(preds)
```

```
# [1] "call"               "predicted_survival"
```

The predicted probabilities for the first 3 subjects in the dataset are:

```
head(preds$predicted_survival, 3)
```

```
#   id     S(3)     S(4)     S(5)
# 2  2   0.9399   0.8868   0.8330
# 3  3   0.8556   0.7392   0.6316
# 4  4   0.8138   0.6710   0.5451
```

For example, subject 3 has a predicted probability equal to 85.6% to survive at least until 3 years, and 63.2% to survive at least up until 5 years since study entry. Furthermore, we can use the function `survplot_prc` to visualize the predicted survival over time for selected subjects:

```
survplot_prc(step1 = lmms, step2 = predRanefs,
             step3 = pencox,
             ids = c(3, 54, 111, 173, 271), tmax = 10)
```

The outputted plot is presented in Figure 1. Predictions for new subjects, which were not used to estimate the model, can be obtained by specifying their corresponding data with the arguments `new.basecovs` and `new.longdata` inside `survpred_prclmm`.

Evaluation of predictive performance

Besides model estimation and prediction, `pencal` automates also the evaluation of the model's predictive performance via bootstrap. This procedure makes it possible to obtain unbiased estimates of the model's predictive performance. To compute these, it is first necessary to rerun the estimation steps with the argument `n.boots` set to a value equal to the desired number of bootstrap samples inside `fit_lmms`; we set `n.boots = 100` in

the example below. To speed computations up, we additionally set the argument `n.cores` equal to 8, so that the bootstrap computations are automatically parallelized over 8 cores:

```
step1 <- fit_lmms(
  y.names = c('lSB', 'lSC', 'albumin',
    'lAlk', 'lSGOT', 'platelets', 'lProt'),
  fixefs = ~ age,
  ranefs = ~ age | id,
  t.from.base = fuptime,
  long.data = ldata,
  surv.data = sdata,
  n.boots = 100, n.cores = 8, verbose = F)
step2 <- summarize_lmms(object = step1, n.cores = 8,
  verbose = F)
step3 <- fit_prclmm(object = step2, surv.data = sdata,
  baseline.covs = ~ baselineAge + sex + treatment,
  penalty = 'ridge', n.cores = 8, verbose = F)
```

Then, we can proceed to estimate the values of the C index, time-dependent AUC and Brier score using `performance_prc`. Hereafter we show how to estimate the time-dependent AUC (tdAUC) at prediction times 3, 5 and 7:

```
perf <- performance_prc(
  step2 = step2, step3 = step3,
  metric = 'tdauc', times = c(3, 5, 7),
  n.cores = 8, verbose = F)
perf$tdAUC
```

```
#   pred.time tdAUC.naive optimism.correction tdAUC.adjusted
# 1         3      0.9439             -0.0096         0.9343
# 2         5      0.9366             -0.0138         0.9131
# 3         7      0.8831             -0.0109         0.8722
```

We can see from the output that the estimated model achieves a tdAUC of 0.938 at 3 years, and 0.872 at 7 years.

Conclusion

This short introduction to the `R` package `pencal` illustrates how `pencal` can be used to implement dynamic prediction of survival outcomes in settings with numerous longitudinal predictors. By pursuing an efficient model-based landmarking approach to dynamic prediction, `pencal` enables scalable and computationally robust prediction in settings with dozens or hundreds of longitudinal covariates, where the estimation of joint models is practically unfeasible. We refer readers to Signorelli (2024) for a more detailed description of PRC and of the `R` package `pencal`.

References

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

Signorelli, M., Spitali, P., Al-Khalili Sgyziarto, C., The Mark-MD Consortium, and 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, 6178–6196.

Signorelli, M. (2024), "pencal: an R package for the dynamic prediction of survival with many longitudinal predictors", *The R Journal*, 16, 134–153.

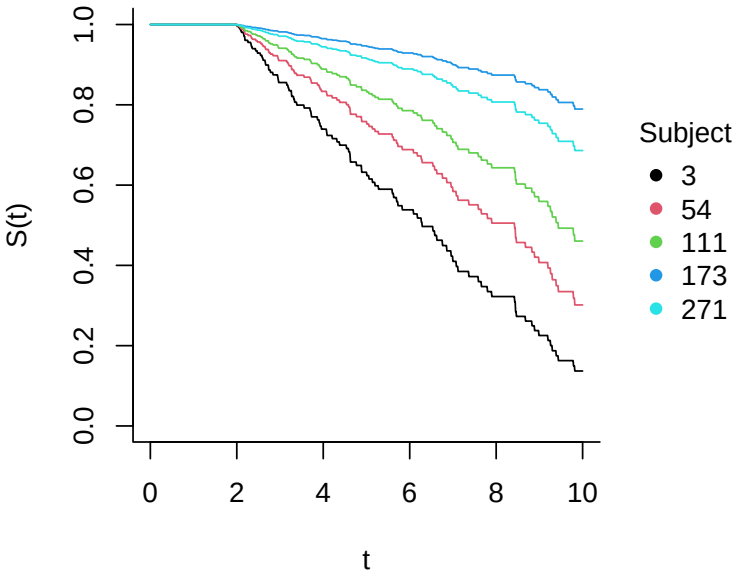

Figure 1: Predicted survival probabilities for subjects 3, 54, 111, 173, 271.

Signorelli, M., and Retif, S. (2025), "Benchmarking multi-step methods for the dynamic prediction of survival with numerous longitudinal predictors", *The International Journal of Biostatistics*, 22(1), 65–85.

Tsiatis, A. A., and Davidian, M. (2004), "Joint modeling of longitudinal and time-to-event data: An overview", *Statistica Sinica*, 14(3), 809–834.

Mirko Signorelli

Assistant Professor
Mathematical Institute
Leiden University

Call for Contributions to the Newsletter

We welcome contributions from section members for upcoming issues. Submissions may address a broad range of topics of interest to LiDS, including classroom examples, stimulating problems, perspectives on developments in the field, software or book reviews, practical experiences, and news about members or activities. Submissions will be published subject to approval by the Communications Officer and the LiDS Chair. To contribute, please contact Steven Chiou at `schiou@smu.edu`.

Sy Han (Steven) Chiou
Communications Officer

Associate Professor
Southern Methodist University