ASA Section on Lifetime Data Science

LiDS Newsletter

Volume 7, Number 1 – January 2022

A section associated with the American Statistical Association

ISSN 2473-5159

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

* * * * * * * * * * * *

In Brief

2023 LiDS Conference
May 31–June 2, 2023 in Raleigh, North Carolina

2022 Election Candidates
Chair-Elect;
Program Chair-Elect

Secretary Report
A new record high: 561 LiDS members as of December, 2021

LiDS Sessions at 2022 JSM
Three invited sessions;
Two topic-contributed sessions

Software Review
The `flexsurv` package: recent developments motivated by applications to COVID-19

* * * * * * * * * * * *

LiDS Officers

Chair:	Douglas Schaubel
Chair-Elect:	Grace Y. Yi
Past Chair:	Mimi Y. Kim
Secretary 2022–2024:	Sharon Xie
Treasurer 2021–2023:	Yu Cheng
Program Chair (2022):	Jing Ning
Program Chair-Elect:	Ying Ding
Past Program Chair:	Haoda Fu
COS Representative:	Ronghui Xu
Webinar Committee Chair:	Esra Kurum
Communications Officer:	Wenjie Wang

AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

Chair's Message

I am excited and honored to serve as Chair for the Lifetime Data Science (LiDS) section Executive Committee for 2022! This marks the fourth year of LiDS as a full-fledged ASA section, having initially started as an ASA Interest Group in 2014. The LiDS section has grown tremendously since its inception and currently consists of more than 560 members. The LiDS community is indebted to those whose tireless work made rising to, and thriving in Section status possible. In particular, this list is anchored by the former Chairs of LiDS: Mimi Kim, Nicholas Jewell, Jianwen Cai, Richard Cook, Mei-Cheng Wang, Jack Kalbfleisch, Mei-Ling Lee and Ross Prentice.

I offer heartfelt thanks, on behalf of myself, the EC and the Section as a whole, to Mimi Kim, who is transitioning from 2021 LiDS Chair to Past-Chair. Mimi is a pleasure to work with and carried out her Chair duties with vigor and an abundance of enthusiasm. Among other things, Mimi will be remembered for spearheading the organization of the highly successful LiDS webinar series, which proved a very effective means of keeping LiDS members connected in response to the pandemic. Mimi also organized the inaugural LiDS Trivia Game, held at the conclusion of the August 2021 Annual Business Meeting. In addition, Mimi selected the Section's unofficial theme song, Gloria Gaynor's "I Will Survive", inspired in part by the Section's focus on survival analysis and the ongoing pandemic related challenges. The EC looks forward to Mimi's continued contributions in her 2022 role as Past-Chair.

The 2023 LiDS conference will be held May 31–June 2, 2023 in Raleigh, North Carolina at the Marriott City Center. The theme of the conference is "Making an impact in the data science era", retained from the planned 2021 event which was unfortunately postponed due to the pandemic. The format for 2023 will be similar to that of the previous conferences held at the University of Pittsburgh (2019) and the University of Connecticut (2017). Day 1 (May 31) will consist of up to three short courses, with an opening mixer in the evening. Days 2 and 3 will each begin with keynote talks, followed by parallel sessions, with a dinner held during Day 2. We are thrilled to have Mei-Cheng Wang and Per Andersen as our keynote speakers! Planning is well underway for the conference, with Shanshan Zhao and Wenbin Lu kindly agreeing to stay on as Chair and Co-Chair of the Conference Organizing Committee, respectively.We look forward to continuing the tradition of academically enriching and fun events that have characterized the previous LiDS conferences. Mark your calendars!

Between now and the conference, the LiDS Webinar Committee is working to organize several events for 2022. We thank Esra Kurum for taking on the role of Webinar Committee Chair. The year 2021 will be a difficult act to follow, with last year's Committee (consisting of Shanshan Zhao and Ying Ding) having organized 5 events: two short courses (by Yi Li, and jointly by Stijn Vansteelandt, Torbin Martinussen and Oliver Dukes); two webinars (by Noah Simon and myself); and a professional development webinar (by Malka Gorfine and Jeremy Taylor). A big thank you is owed to last year's Webinar Committee, Ying Ding (who is transitioning to LiDS Program Chair-Elect) and Shanshan Zhao (who continues her pivotal role, as mentioned above) for the considerable time and effort they put to recruit the speakers and organize the events, which were a great success! We are also grateful to Rick Peterson of the ASA for his invaluable assistance with these activities.

The beginning of the new year brings other rotations in and out of the LiDS Executive Committee. We owe great thanks to Nick Jewell for his dedicated and spirited efforts as Chair-elect (2019), Chair (2020) and Past-Chair (2021). We thank Nick for his insights and stellar leadership of LiDS during an especially challenging time, and we welcome Grace Yi to the EC as the LiDS 2022 Chair-Elect. Joan Hu completed a three-year term as LiDS Secretary at the end of 2021. We thank Joan for her tremendous organizational skills, and welcome Sharon Xie as our incoming LiDS secretary. LiDS thanks Nan Xue for her efforts as Council of Sections representative and welcomes Ronghui (Lily) Xu, who will be serving as our ASA COC representative. We thank Zhezhen Jin for his efforts to maintain solid representation by LiDS at the Joint Statistical Meetings during his three years in the Program Chair track. Zhezhen was succeeded by Haoda Fu as Program Chair for 2021, and we thank Haoda for his hard work to organize the 2021 JSM LiDS events. Ning Jing has taken over as 2022 Program Chair and is hard at work on the 2022 JSM program. Yu Cheng will continue her stellar work as LiDS Treasurer in 2022. In addition, LiDS is very fortunate to again retain the services of Wenjie Wang for his hard work as LiDS Communications Officer. Finally, LiDS was so fortunate to benefit from the expertise of Ker-Ai Lee, who has moved on from her role as LiDS Webmaster.

Big thanks to all incoming, outgoing and continuing LiDS Executive Committee members. Time is at a premium, especially during a pandemic. Your commitment to making LiDS such a success is greatly valued. In closing, the year ahead holds much uncertainty. The continued presence of COVID-19 in our lives kept 2021 from ever really getting on track for many of us. In uncertain times like these, much inspiration is offered by the pivotal role played by the statistical community towards efforts to understand the pandemic, with the undercurrent being an unyielding commitment to the truth. The dedicated work of statisticians towards this endeavor has been among the pandemic's most prominent silver linings.

In closing, I wish you all a happy and healthy 2022! I hope you renew your LiDS membership and encourage your colleagues to join. I welcome any comments or suggestions you care to send to `douglas.schaubel@pennmedicine.upenn.edu`. Thank you for your past and continued support of the Lifetime Data Science Section!

Douglas Schaubel, Chair 2022

Message from the Past Chair

Dear LiDS Members,

I hope you and your families are staying safe and healthy in the New Year as we try to get through this latest COVID-19 surge. Despite the ongoing challenges of the pandemic, the Lifetime Data Science (LiDS) section had another successful and productive year thanks to the outstanding efforts of the LiDS executive committee. It was such a privilege, as well as an incredibly enjoyable and rewarding experience, for me to work with this "dream team" of enthusiastic and dedicated LiDS officers who accomplished so much over the past year. The section continues to grow and reached an all-time high of more than 560 members as of December. The section also held a number of activities last year that generated significant interest and increased the visibility of LiDS in the statistical community. Shanshan Zhao and Ying Ding, our webinar co-chairs, with the help of Doug Schaubel, 2022 Chair, did an amazing job launching our webinar series at the beginning of 2021. They organized two webinars, two short courses and a career development workshop that attracted over 500 participants in total. Our section also had a strong presence at last year's JSM because of the contributions of Haoda Fu, 2021 Program Chair, who put together several LiDS sponsored scientific sessions, and Zhezhen Jin, 2021 Past-Program Chair, who led the student paper competition. Our 2022 Program Chair, Jing Ning has also been working hard on developing an exciting LiDS program for the upcoming JSM meeting.

I am extremely grateful to the individuals who finished three years of service to the LiDS executive committee: Nick Jewell, Joan Hu, Zhezhen Jin and Nan Xue. My job as LiDS chair was so much easier because of their collective experience, wisdom and knowledge that they so generously shared with me. I am very appreciative of Nick's outstanding leadership and guidance as Chair-Elect, Chair and Past-Chair, especially in helping us to navigate all the pandemic related issues that arose in our section. I also could not have managed without Joan's invaluable help as Secretary in planning and recording the minutes of our many meetings, Zhezhen's excellent work in developing the section's scientific programs, and Nan's dedication to representing LiDS on the ASA Council of Sections. The extensive efforts of our treasurer, Yu Cheng, who ensures that the section remains financially strong, and Wenjie Wang, who is responsible for both the LiDS newsletter and website as our communications officer, are also greatly appreciated. Wenjie is taking over the webmaster duties from Ker-Ai Lee who so kindly volunteered her services for the past three years.

We are so lucky to have an outstanding group of new officers who will be leading LiDS in 2022 and ensuring its continued growth and success. Doug Schaubel brings to the Chair position significant experience and accomplishments in the lifetime data science field. I really enjoyed working with and getting to know Doug last year through the LiDS executive committee, and I have no doubt the section will flourish under his effective leadership. I also welcome Grace Yi as Chair-Elect, Ying Ding as Program Chair-Elect, Sharon Xie as Secretary, Ronghui Xu as Council of Sections Representative, and Esra Kurum as the new chair of the webinar committee. Because of the efforts of our Nominations Committee, which included Nick, Zhezhen and Paul Albert, we also have an impressive slate of candidates for LiDS officers for the 2022 elections.

A sincere and heartfelt thanks to all of you for your support of the LiDS section! Please remember to renew your memberships, and also encourage your friends and colleagues to join LiDS and participate in our upcoming activities. I wish you all the best in 2022!

Mimi Kim, Chair 2021

2022 Election Candidates

This year we have two positions open on the Executive Committee of the Section on Lifetime Data Science, namely Chair-Elect and Program Chair-Elect. We are again in the very fortunate position of having a strong slate of candidates who have agreed to stand for election. They are as follows:

Chair-Elect
Rebecca Betensky, New York University
Jianguo (Tony) Sun, University of Missouri

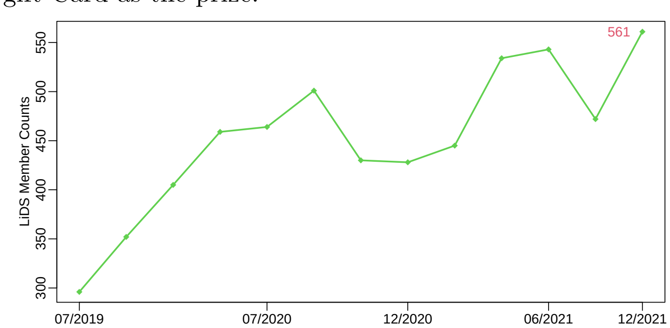
Program Chair-Elect
Sebastien Haneuse, Harvard University
Pamela Shaw, Kaiser-Permanente

The American Statistical Association will handle the election process, and we will hear more about the candidates including their biographies and personal statements in the coming weeks. On behalf of the LiDS Committee on Nominations, comprised of Zhezhen Jin, Paul Albert, and myself, we express thanks to section members for participating in the nomination process, and to the candidates for their commitment to the section and willingness to consider these leadership positions.

Nicholas P. Jewell, Chair of Committee on Nominations

Report from the Section Secretary

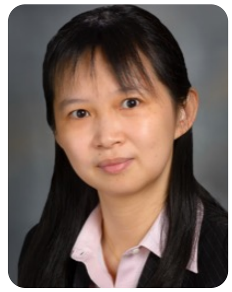
I am honored and excited to officially start my role of Section Secretary in January this year. I look forward to making important contributions to promoting and supporting the missions of LiDS Section. As of December 10, 2021, our section has a total of 561 members. Below is a graph displaying the member counts since July 2019. Our official website is maintained by our communications officer, Wenjie Wang (email: `wang@wwenjie.org`): its link is `https://community.amstat.org/lids/home/`. The section's 2021 annual meeting took place virtually on August 19, 2021. It was chaired by the section chair Mimi Kim. The section congratulated its three new IMS fellows and three new ASA fellows. Five section members won the Trivia Contest and each of them received a $50 Amazon E-gift Card as the prize.



Sharon Xie, Secretary 2022–2024

JSM 2022 Program Update

The 2022 Joint Statistical Meetings (JSM) will be held in Washington, D.C. during August 6–11, 2022. Despite that the details of the JSM 2022 Program is not yet fully decided, the following three invited sessions will be sponsored by LiDS:



1. Monday, August 8, 8:30–10:20 AM "New Statistical Methods for Survival Analysis in Complex Biomedical Studies";
2. Tuesday, August 9, 2:00–3:50 PM "Data integration and information synthesis in survival analysis";
3. Wednesday, August 10, 10:30 AM–12:20 PM "Real-world Survival Data with Multiple Events: Challenges, Opportunities, and Recent Advancements".

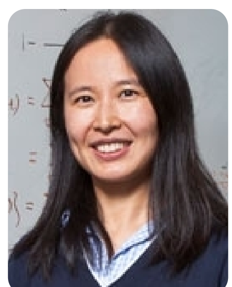
We also recommended the following two topic-contributed sessions sponsored by LiDS:

1. "Lifetime Data Science Student Awards";
2. "Recent Advances in Multivariate Correlated Time-to-event Data".

Thanks all the session organizers for their support and participation. All speakers in a Topic-Contributed paper session must submit abstracts and register for JSM by February 1, 2022 through the official JSM abstract submission system.

Jing Ning, Program Chair 2022

Treasurer's Report



The balance by 12/31/2020 was 39,190.54. Between January 1, 2021 to December 31, 2021, the LiDS offered four successful fee-based Webinars and short courses generating an income of $5,185. The additional revenue of $1,218.38 (through 11/30/2021) came from membership dues, interest, and dividends. During this period, the LiDS spent $3,250 on awards/plaques for student awardees presented at JSM and and $750 to sponsor a free career development workshop for our members. The current balance by December 31, 2021 is $41,350.94. This number may change slightly once pending income and expenses in this period are tabulated.

Beginning Balance	**12/31/2021**	**$39,190.54**
Income		
Membership due and interest		$1218.38
Registration	**11/30/2021**	$15,195.00
Net share—CE/Conf/Processdings		$(5185.00)
Total Income		**$11,228.38**
Expense		
Honorarium		$3,375.00
Meeting overhead		$2,200.00
Bank/Paypal charges & credit card fees		$242.98
Awards/plaques		$3,250.00
Total Expense		**$9,067.98**
Ending Balance	**12/31/2021**	**$41,350.94**

Yu Cheng, Treasurer 2021–2023

New Articles 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`.

The January 2022 issue (Volume 28, number 1) of Lifetime Data Analysis has been published:

- Prognostic accuracy for predicting ordinal competing risk outcomes using ROC surfaces *by* S. Zhang, Y. Qu, Y. Cheng, O. L. Lopez, & A. S. Wahed. Pages 1–22
- Nonparametric inference in the accelerated failure time model using restricted means *by* M. C. Giurcanu & T. G. Karrison. Pages 23–39
- Sample size calculation for clustered survival data under subunit randomization *by* J. Li & S.-H. Jung. Pages 40–67
- Maximum likelihood estimation for length-biased and interval-censored data with a nonsusceptible fraction *by* P. Shen, Y. Peng, H.-J. Chen & C.-M. Chen. Pages 68–88
- Semiparametric analysis of multivariate panel count data with nonlinear interactions *by* W. Wang, Y. Wang & X. Zhao. Pages 89–115
- An additive hazards frailty model with semi-varying coefficients *by* Z. Zhang, X. Wang & Y. Peng. Pages 116–138
- Bayesian analysis under accelerated failure time models with error-prone time-to-event outcomes *by* Y. Tang, X. Song. G. Y. Yi. Pages 139–168

Articles in the October 2021 issue (Volume 27, number 4) are:

- Instrumental variable estimation of early treatment effect in randomized screening trials *by* S. Saha, Z. Liu, & O. Saarela. Pages 537–560
- Weighted Lindley frailty model: estimation and application to lung cancer data *by* A. Mota, E. A. Milani, V. F. Calsavara, V. L. D. Tomazella, J. Leão, P. L. Ramos, P. H. Ferreira, & F. Louzada. Pages 561–587
- A generalized theory of separable effects in competing event settings *by* M. J. Stensrud, M. A. Hernán, E. J. T. Tchetgen, J. M. Robins, V. Didelez, & J. G. Young. Pages 588–631
- Conditional screening for ultrahigh-dimensional survival data in case-cohort studies *by* J. Zhang, H. Zhou, Y. Liu, & J. Cai. Pages 632–661
- The MLE of the uniform distribution with right-censored data *by* Q. Yu. Pages 662–678
- An efficient Gehan-type estimation for the accelerated failure time model with clustered and censored data *by* L. Fu, Z. Yang, Y. Zhou, & Y. G. Wang. Pages 679–709
- Continuous and discrete-time survival prediction with neural networks *by* H. Kvamme, & Ørnulf Borgan. Pages 710–736
- A hybrid landmark Aalen-Johansen estimator for transition probabilities in partially non-Markov multi-state models *by* N. Maltzahn, R. Hoff, O. O. Aalen, I. S. Mehlum, H. Putter, & J. M. Gran. Pages 737–760

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

The flexsurv package: recent developments

`flexsurv` is an R package for parametric survival modelling. The principle of the package is to enable any parametric distribution to be used as part of a time-to-event model. This article gives a quick summary of what it can do, highlighting some recently-developed features that were motivated by applications to COVID-19.

Standard parametric survival models

Given the familiar syntax of the `survival` package (Therneau, 2021), we can fit parametric models to right-censored data of the following form.

```
  days agegroup event
1  0.6    [0,45)     0
2  8.8   [45,65)     0
3  3.3    [0,45)     1
```

These are three rows from a simulated dataset based on a real example of length of stay in hospital for COVID-19, where `days` is the time from admission to discharge, and `event` is an indicator for whether the discharge time was observed or right-censored.

Several standard distributions for the time to the event are built in, e.g. the Weibull and Gamma. A family of distributions based on splines is also supported, which can have any number of parameters depending on how much flexibility is required. Users can also program their own distributions by supplying R functions to evaluate the density, distribution and/or hazard functions. Any parameter of any distribution can be modelled as a function of covariates.

In the following example, the three-parameter generalized gamma model is fitted, where the first two parameters, but not the third parameter, vary by age group.

```
flexsurvreg(Surv(days, event) ~ agegroup + sigma(agegroup),
            data=hosp, dist="gengamma")
```

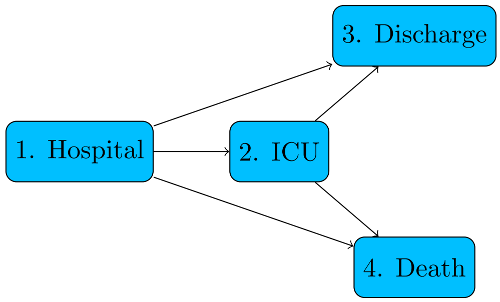
Left-censored and interval-censored times to events, and left and right-truncated distributions, are also supported. Models with right truncation were recently added to `flexsurv` to model survival given a particular form of data that arose in the early phase of the COVID-19 pandemic, where deaths were only observed if they had occurred before a certain calendar time (Seaman et al., 2021).

Parameter estimates, survival and hazard functions and other summaries can easily be produced from fitted models, as described in the package documentation. Confidence intervals for any function of the model parameters (including user-defined functions) can be produced by random sampling from the asymptotic distribution of the maximum likelihood estimates.

Multistate modelling

A lesser-known feature of `flexsurv` is multistate modelling of data consisting of times between different kinds of event. In the last couple of years, these capabilities of `flexsurv` have been extended, motivated by modelling the burden of severe COVID-19 disease. For example, policy-makers wanted to know the probability that someone who is admitted to hospital with

COVID-19 would die, the probability that they would be admitted to an intensive care unit (ICU), and the probability that they would recover and be discharged from hospital. They also wanted to know the average length of stay for people who died, and for people who recovered. The dependence of these quantities on patient characteristics was also of interest. This suggests a multistate model, with states and permitted direct transitions as illustrated in the following diagram:



Parametric multistate models in `flexsurv` can be constructed in two ways: either using *transition-specific hazards*, or as a *mixture model*.

Multistate modelling using transition-specific hazards.

A multistate model can be defined by the hazard λ_{rs} of transition from each state r to each state s. λ_{rs} can be interpreted as the hazard governing the latent time T_{rs} to the event s for someone who has just entered state r. For example, for $r = (Hospital)$, the competing events are $s = (ICU, Death, Discharge)$, and for $r = (ICU)$ the competing events are $s = (Death, Discharge)$. The event that actually happens on leaving state r is $\min_s T_{rs}$. Thus for each r, it is only possible to observe the T_{rs} for one of the competing s.

Using `flexsurv`, we can define a parametric model for each r, s transition-specific hazard by fitting a standard time-to-event model to a specially-constructed r, s-specific dataset. This is illustrated here for the transition from the hospital state to the ICU state. In the following extract from the dataset `dat_hosp_icu`,

- each row represents the time (`days`) from entering state r (hospital admission) to entering a different state (or censoring in state r).
- transitions ending in state $s = ICU$ correspond to observed times to events (where `event_icu` is 1)
- transitions ending in states other than $s = ICU$, i.e. deaths or discharges, correspond to censoring (where `event_icu` is 0).

```
  days event_icu agegroup
1  3.3         0   [0,45)
2  1.1         0  [45,65)
3  6.6         0  [65,75)
```

```
mod_hosp_icu <- flexsurvreg(
    Surv(days,event_icu) ~ agegroup,
    data=dat_hosp_icu, dist="gengamma"
)
```

We can fit the whole multistate model by independently fitting five standard parametric models of this form to appropriately-constructed datasets representing each of the five transitions. Different parametric families, and different covariates can be used for different transitions.

Suppose these five models have been respectively named `mod_hosp_icu`, `mod_hosp_death`, `mod_hosp_disc`, `mod_icu_death` and `mod_icu_disc`. An object containing the whole fitted multistate model can then be constructed using the function `fmsm` as follows,

```
tmat <- rbind(c(NA,1,2,3), c(NA,NA,4,5),
              rep(NA,4), rep(NA,4))
rownames(tmat) <- colnames(tmat) <- c("Hospital","ICU",
                                      "Discharge","Death")
covid_msm <- fmsm(mod_hosp_icu, mod_hosp_disc,
                  mod_hosp_death, mod_icu_disc,
                  mod_icu_death, trans=tmat)
```

The matrix `tmat` defines the *transition structure* of the model, where the r, s entry contains the integer i if the ith argument of `fmsm` is the model for the transition from state r to state s. This matrix format is also used by the `mstate` package for semiparametric multistate modelling (de Wreede et al., 2011).

The `fmsm` object, here named `covid_msm`, can then be supplied to various functions in `flexsurv` that make predictions from the multistate model. For example, for a person just admitted to hospital, we can estimate

- the probability that they will be in ICU 10 days later (the "transition probability"),
- the probability that they will go to ICU (at any time),
- the time to their next event,
- the probability that they will ultimately die in hospital, or be discharged (averaged over the chance of ICU admission) and the time to the ultimate event of each type.

Many of these quantities do not have known analytic forms, and are computed by simulating and summarising a large number of individual survival times. This is done in the following example, where a large population of people admitted to hospital are simulated from the model until their final outcome of death or discharge. The probability of each final outcome, and the distribution (mean and quantiles over individuals) of times to each outcome, given that particular outcome occurs, are presented. This is done separately for two age groups, which define covariate values in the fitted models.

```
library(tidyverse)
nd <- data.frame(agegroup=c("[65,75)", "[75,85)"))
(sf <- simfinal_fmsm(covid_msm, newdata=nd) %>%
  pivot_wider(names_from="quantity", values_from="val"))
```

```
# A tibble: 4 x 7
  agegroup state     `2.5%` `50%` `97.5%`  mean  prob
  <chr>    <chr>      <dbl> <dbl>   <dbl> <dbl> <dbl>
1 [65,75)  Death       1.37  9.72    83.7  18.7 0.396
2 [65,75)  Discharge   1.15  9.25    71.3  16.2 0.604
3 [75,85)  Death       1.14  8.59    77.0  17.0 0.447
4 [75,85)  Discharge   1.40 10.2     75.1  17.4 0.553
```

Confidence intervals could also be obtained, at a greater computational expense.

Note that the model defined above is a "semi-Markov" model, since the risk of transition out of ICU depends on the time since ICU entry, but not the time since hospital admission. Markov models are also supported, where the specified parametric distribution describes the hazard as a function of time since the start of the process, rather than the time since entry to the current state. While Markov models can be simulated from more efficiently, I have found these to be harder to interpret.

Multistate modelling using mixtures.

A lesser-known way of setting up a multistate model is as a mixture model (Cox, 1959; Larson and Dinse, 1985). This feature was added in version 2.0 of `flexsurv`. Instead of transition-specific hazards, the model is structured in terms of

(a) the probability p_{rs} that the next state is s, for someone in state r

(b) a parametric distribution for the time until entry to state s, for someone in state r who transitions to state s.

An advantage of this model in our application to COVID-19 hospitalisation was that these are the quantities that the policy-makers wanted to know. To obtain them, we just need to fit the model, and do not need to undertake an expensive second stage of simulation as in the cause-specific hazards model.

The function `flexsurvmix` is used to fit this model. Instead of one model per transition, one model per *transition starting state* is fitted, each representing all competing risks from that state. In this example, there are two starting states (hospital and ICU) and we make two calls to `flexsurvmix`, the first for the competing risks of ICU, death and discharge following hospital admission, and the second for the competing risks of death and discharge following ICU admission. The data for each call are formatted with one row per individual, as in the following extract from the dataset `simh_mixdat` for the events following hospital admission,

```
  timem    eventm statusm agegroup
1   3.3 Discharge       1   [0,45)
2   1.1 Discharge       1  [45,65)
```

In these data,

- `timem` represents the time to an event or censoring,
- `eventm` represents the competing event which we know to have occurred for this individual. This should be `NA` in cases where we do not know which event has happened or will happen.
- `statusm` is an indicator for whether the time to that event is observed or censored. This allows models where we know which of the competing events will happen, but not the time to that event. Interval censoring is also supported.

The `dists` argument to `flexsurvmix` specifies the parametric distribution that is used for the times to each of the competing events, with all choices available in `flexsurvreg` also available here. Covariates may modify parameters of the time-to-event distributions, using a similar syntax to `flexsurvreg`. Covariates can also modify the event probabilities through an additional argument `pformula`.

```
fm <- flexsurvmix(Surv(timem, statusm) ~ 1, event=eventm,
                  data= simh_mixdat,
                  dists= c("gengamma", "gamma", "gamma"))
```

Maximum likelihood estimates of the event probabilities and the parameters of the time-to-event distributions are computed using an expectation-maximisation algorithm. This is generally more expensive than fitting the analogous cause-specific hazards models, since the parameter space is larger. Printing the fitted model object shows the estimates of the event probabilities p_{rs} and parameters of the distributions governing the times T_{rs} for each s, here shown for $r = (Hospital)$.

```
fm
```

```
Call:
flexsurvmix(formula = Surv(timem, statusm) ~ 1,
    data = simh_mixdat, event = eventm,
    dists = c("gengamma", "gamma", "gamma"))

Estimates:
   component      dist  terms    est    est.t     se
1        ICU            prob1 0.1911       NA 0.0443
2      Death            prob2 0.2756   0.3662 0.0403
3  Discharge            prob3 0.5333   1.0265 0.0361
4        ICU  gengamma     mu 0.1996   0.1996 0.0492
5        ICU  gengamma  sigma 0.9829  -0.0173 0.0221
6        ICU  gengamma      Q 0.1130   0.1130 0.0786
7      Death     gamma  shape 0.8664  -0.1434 0.0334
8      Death     gamma   rate 0.0497  -3.0019 0.0471
9  Discharge     gamma  shape 1.2871   0.2524 0.0242
10 Discharge     gamma   rate 0.1381  -1.9797 0.0303

Log-likelihood = -21760, df = 9
AIC = 43538
```

Several functions are provided to summarise quantities of interest from the fitted models.

More theoretical details about these two different frameworks for multistate modelling are given in Jackson et al. (2021). Further information about their implementation in `flexsurv` is given in the package help and the package vignettes, including one devoted to multistate models.

References

D. R. Cox. The analysis of exponentially distributed life-times with two types of failure. *Journal of the Royal Statistical Society: Series B (Methodological)*, 21(2):411–421, 1959.

L. C. de Wreede, M. Fiocco, and H. Putter. mstate: an R package for the analysis of competing risks and multi-state models. *Journal of Statistical Software*, 38(1):1–30, 2011.

C. Jackson, B. Tom, P. Kirwan, S. Mandal, S. Seaman, K. Kunzmann, A. Presanis, and D. De Angelis. A comparison of two frameworks for multi-state modelling, applied to outcomes after hospital admissions with COVID-19, 2021. URL `https://doi.org/10.17863/CAM.73551`. Accepted for publication in *Statistical Methods in Medical Research*.

M. G. Larson and G. E. Dinse. A mixture model for the regression analysis of competing risks data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 34(3):201–211, 1985.

S. Seaman, C. Jackson, and A. Presanis. Estimating a time-to-event distribution from right-truncated data in an epidemic: a review of methods, 2021. URL `https://doi.org/10.17863/CAM.72880`. Accepted for publication in *Statistical Methods in Medical Research*.

T. M. Therneau. *survival: a package for survival analysis in R*, 2021. URL `https://CRAN.R-project.org/package=survival`. R package version 3.2-13.

Christopher Jackson
Senior Investigator Statistician
MRC Biostatistics Unit
University of Cambridge
Email:
`chris.jackson@mrc-bsu.cam.ac.uk`