A section associated with the American Statistical Association

LiDS website: `https://community.amstat.org/lids/home`

* * * * * * * * * * *

In Brief

Annual Business Meeting (virtual)
Thursday, August 19, 2021,
14:00–15:00 PDT (17:00–18:00 EDT)

Nomination for Election 2022
Chair-Elect and Program Chair-Elect

2021 Election Results
Chair-elect: Dr. Grace Yi
Program chair-elect: Dr. Ying Ding
Secretary: Dr. Sharon Xie
COS Representative: Dr. Ronghui Xu

Member Awards
3 ASA fellows and 3 IMS fellows

LiDS Activities at the 2021 JSM
One invited sessions and one topic-contributed session.

LiDS Short Course in September
Causal Inference with Survival Data

LiDS Webinar in October
Career Development: Publishing and Grant Writing

Software Review
Multistate Models in the **survival** Package

* * * * * * * * * * *

LiDS Officers

Chair:	Mimi Y. Kim
Chair-Elect:	Douglas Schaubel
Past Chair:	Nicholas P. Jewell
Secretary 2019–2021:	Joan Hu
Treasurer 2021–2023:	Yu Cheng
Program Chair:	Haoda Fu
Program Chair-Elect:	Jing Ning
Past Program Chair:	Zhezhen Jin
COS Representative:	Xiaonan Xue
Webmaster:	Ker-Ai Lee
Newsletter Editor:	Wenjie Wang

Chair's Message

We are in the middle of summer 2021 and life, at least in the US, seems to be returning to some semblance of normalcy. COVID-19 cases are beginning to rise again in some areas, though, and it remains unclear when we'll be completely out of the woods with this pandemic. But on the positive side, I am pleased to report that the ASA LiDS section is thriving because of the incredible dedication and enthusiasm of the section's members and officers. As of the end of June, the LiDS membership stood at over 530. We successfully launched the webinar series with presentations in January by Noah Simon on statistical learning with time-to-event data, and in April by Doug Schaubel, our Chair-elect, on the analysis of competing risks data. Yi Li also taught in June a LiDS-sponsored four hours online short course on high dimensional survival data. These events attracted close to 350 registrants in total. A huge thanks to Noah, Doug and Yi for offering these valuable and stimulating continuing education opportunities to our statistical community! I am also extremely grateful to the LiDS webinar committee, Shanshan Zhao and Ying Ding, and to Rick Peterson of the ASA for their assistance in organizing these activities.

We have some exciting LiDS events lined up for the fall, too, including a short-course on causal inference that will be led by Stijn Vansteelandt, and a career development workshop in which Malka Gorfine and Jeremy Taylor will offer helpful tips for our junior LiDS members on manuscript and grant writing. We are also eager to begin planning the scientific program of the 2023 LiDS conference and welcome your input on topics for sessions and short-courses.

The Joint Statistical Meetings are just a few weeks away and LiDS has sponsored one invited session and one topic contributed session that will be dedicated to presentations by the six winners of the 2021 LiDS Student Paper Competition. Special thanks to Haoda Fu and Zhezhen Jin, our current and past Program Chairs, for putting together the JSM LiDS program and to the members of the student award committee (Zhezhen Jin, Guoqing Diao, Haoda Fu, Zhuqing Liu, Tony Sit, Wenjie Wang) for their hard work in selecting the winners. Also during JSM, our very own LiDS Past-Chair, Nick Jewell, will be honored as the recipient of the 2021 Nathan Mantel Award. It is wonderful that he is being recognized by the ASA for his lifetime contributions to the development and application of statistical science to problems and issues in epidemiology. Congratulations, Nick, on this well-deserved honor!

Finally, please mark your calendars for the LiDS Annual Business meeting that is scheduled via Zoom for August 19th (14:00–15:00 PT), the week after JSM. You will have the opportunity to learn more about our section's activities and how to get involved, to meet the current and newly elected section officers, and to share your feedback and ideas for future events. The meeting should be both informative and fun so I hope you can make it!

Thank you for your interest in the LiDS section and I look forward to connecting with many of you during JSM!

Mimi Kim, Chair 2021

AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

Nominations Invited for Positions on the Executive Committee

This fall, the preparation for the annual election of officers for Sections of the ASA will take place, and, as Chair of the Nominations Committee, I am pleased to announce that we are seeking nominations for two positions on the LiDS Executive Committee (EC). In Spring 2022 there will be an election for Chair-Elect, and Program Chair-Elect. The candidate elected in 2022 for the position of Chair-Elect will serve as Chair-Elect in 2023, Chair in 2024, and Past-Chair in 2025. The candidate elected in 2022 for Program Chair-Elect will be Program Chair-Elect in 2023, Program Chair in 2024, and Past-Program Chair in 2025.

In accordance with the charter of the Section on Lifetime Data Science, the Nominating Committee consists of the Past Chair of LiDS (Nicholas P Jewell), the Past Program Chair (Zhezhen Jin), and designated member (Paul Albert, NCI); the Past Chair of the Executive Committee serves as Chair of the Nominating Committee. The Nominating Committee shall be responsible for nominating a slate of officers for each annual election which is to be submitted to the ASA by mid-November.

There is the opportunity for members of the Section on Lifetime Data Science to nominate candidates for open positions. Such nominations must be signed by at least five members of the Section on LiDS. Nominators should ensure that the nominee is (a) a member of the ASA, (b) a member of the Section on Lifetime Data Science, and (c) is willing to stand for the nomination.

Nominations should be sent to the Chair of the Nominations Committee (Nicholas P Jewell, `nicholas.jewell@lshtm.ac.uk`) with a copy to the Secretary (Joan Hu, `joan hu@sfu.ca`). Please make the subject heading of any nominations "Nomination for 2022 Election for LiDS EC". The deadline for nominations is September 15, 2021. The election is coordinated by the Committee on Sections (COS) of the American Statistical Association.

Nicholas P. Jewell, Nomination Committee Chair 2021

2021 Election Results
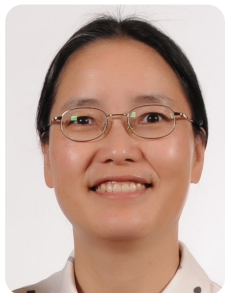
Chair-Elect 2022 Dr. Grace Yi is a professor at the University of Western Ontario where she currently holds a Tier I Canada Research Chair in Data Science. Dr. Yi received her Ph.D. in Statistics from the University of Toronto in 2000 and joined the University of Waterloo as a postdoctoral fellow (2000–2001), Assistant Professor (2001–2004), Associate Professor (2004–2010), Professor (2010–2019), and University Research Chair (2011–2018). Dr. Yi's research interests focus on developing methodology to address various challenges concerning Data Science, public health, cancer research, epidemiological studies, environmental studies, and social science. Professor Yi's recent research has been centered around investigating machine learning and statistical methods to tackle problems concerning imaging data, missing data, measurement error in variables, causal inference, high dimensional data, survival data, and longitudinal data. Dr. Yi is an elected Fellow of the American Statistical Association.
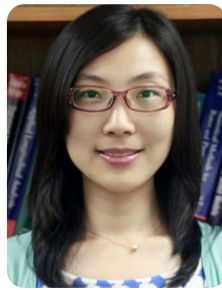
Program Chair-Elect 2022 Dr. Ying Ding is an Associate Professor of in the Department of Biostatistics at the University of Pittsburgh. Dr. Ding received her Ph.D. in Biostatistics from the University of Michigan in 2010 and joined Eli Lilly and Company after her graduation. With four years' industry experience, Dr. Ding joined the University of Pittsburgh as an Assistant professor in 2013. Dr. Ding's primary research interests include semiparametric methods and inferences, especially for time-to-event data; subgroup analysis such as simultaneous inference and biomarker/subgroup identification. Her recent collaborative research focuses on proteomic experiment design, network analysis for psychiatric disorders and progression analysis of AMD (Age-related Macular Degeneration).

Secretary 2022–2024 Dr. Sharon Xie is a Professor of Biostatistics in the Department of Biostatistics, Epidemiology and Informatics at the University of Pennsylvania. Dr. Xie received her Ph.D. in Biostatistics from the University of Washington in 1997. Dr. Xie is a principal investigator of an NIH-funded statistical method R01 grant that aims to develop novel statistical methods for assessing dementia risk in Parkinson's disease and related disorders. Her current research goals are 1) to develop new statistical methods for survival analysis, missing data, measurement error problems, high dimensional data, biomarker evaluations, and longitudinal analysis in response to methodological problems that arise in her collaborative work in the research of neurodegenerative diseases, and 2) to study epidemiology of neurodegenerative diseases. Dr. Xie is an elected Fellow of the American Statistical Association.

Sections Representative 2022–2024 Dr. Ronghui Xu is a Professor in the Department of Family Medicine and Public Health, and Department of Mathematics at the University of California, San Diego (UCSD). Dr. Xu received her Ph.D. in Mathematics from UCSD in 1996. With the emergence of biomedical big data, Dr. Xu's recent research focus is to 1) apply machine learning methods to predict, as well as develop statistical inference, for complex data type such as competing risks of cancer versus non-cancer mortality, in the presence of high dimensional covariates. 2) causal inference methodology using propensity scores, instrumental variables, principal stratification, mediation or path specific analysis, for complex data type such as in the above, or for rare events in pregnancy safety data. Dr. Xu is an elected Fellow of the American Statistical Association.

Member Awards: ASA/IMS Fellows

Three out of the 48 newly elected 2021 ASA fellows and three out of the 41 newly elected 2021 IMS fellows are or were members of our section.

New ASA Fellows

Kellie Archer, The Ohio State University

For exemplary research associating high-throughput 'omic' data to clinical outcomes; for methodological contributions to high-dimensional data analysis; and for outstanding educational contributions to the statistical analysis of genomic data.

Li-Shan Huang, National Tsing Hua University

For influential contributions to smoothing methods; for blending creative statistical methodology with cross-disciplinary application across a variety of fields; for mentoring of graduate students; and for service to the profession.

Yichuan Zhao, Georgia State University

For contributions in the areas of survival analysis, nonparametric statistics, and empirical likelihood-based methods; for applications in biomedical research; and for service to the profession, including organizing conferences, editorial service, and mentoring graduate students.

New IMS Fellows

Kwun Chuen Gary Chan, University of Washington

For outstanding contributions to the methodology of preferential sampling design, observational data, and complex lifetime data; for substantive leadership in the application of statistics in public health and biomedical research.

Huazhen Lin, Southwestern University of Finance & Economics

For outstanding contributions to theory and methods for survival analysis and semiparametric/nonparametric modeling; for dedicated service to the profession; and for strong leadership in statistics education and development in China.

Zhengjun Zhang, University of Wisconsin

For significant contributions to extreme value statistics and risk management, including tail dependence measure and nonlinear dependence measure construction and inference, max-linear competing factor models, nonlinear time series models for high-frequency financial data; and for conscientious editorial and other services to the profession.

LiDS Activities at the 2021 JSM

The Joint Statistical Meetings 2021 will be held virtually from August 8–12, 2021. This year, LiDS has sponsored one invited session with topic on Causal, Robust, and Machine Learning for Survival Outcomes. The topic was selected by LiDS JSM program committee based on novelty and quality of the abstracts. We are very glad to have 4 speakers. Prof. Michael Kosorok will present a topic about finite-horizon reinforcement learning on time to event outcomes. Prof. Eric Tchetgen will present his work on instrumental variable estimation of marginal structure cox model for time-varying treatments. Stijn Vansteelandt will introduce a topic on assumption-learn inference for Cox regression, and Ronghui Xu will present on

doubly robust methods for Cox marginal structure model. Our sponsored session will be on August 11 from 3:30pm to 5:20pm. In addition, we will also have a special topic contributed session for LiDS student paper award. This year, a total number of 13 papers were submitted for the LiDS Student Paper Award Competition. Each paper was reviewed by all reviewers and six papers were selected. The six papers will be presented in a topic-contributed session at JSM 2021 at 10am–11:50pm on Thursday, August 12, 2021.

Haoda Fu, Program Chair 2021

Treasurer's Report
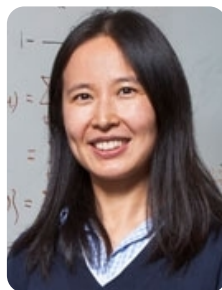
We received the updated numbers on membership dues, allocations, interests, bank charges and food expense for this year through 5/31/2021. Pending revenue from the two webinars by Dr. Doug Schaubel in April and by Dr. Yi Li in June will be added once detailed dollar amount is available.

Beginning Balance	**12/31/2020**	**$39,190.54**
Income		
Membership due and interest		$944.38
Registration		$1,147.50
Total Income		**$2091.88**
Expense		
Hotel Deposit		$4,000.00
Honorarium and Fee		$462.50
Bank Charges, Food		$204.14
Total Expense		**$4,666.64**
Net Total Income		**($2,574.76)**
Ending Balance	**12/31/2020**	**$36,615.78**

Yu Cheng, Treasurer 2021–2023

Call for Invited and Topic-Contributed Session Proposals
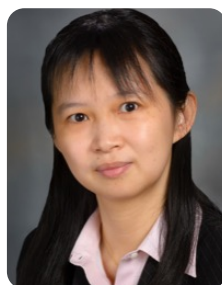
The 2022 Joint Statistical Meetings (JSM) will be held in Washington, D.C. during August 6–11, 2022. The LiDS Program Committee is soliciting proposals for an invited session and two topic-contributed sessions for the 2022 JSM. In addition, LiDS will submit up to two proposals for additional invited sessions that are open for competition among the ASA Sections, Interest Groups, the Leadership Support Council and the Council of Chapters. An invited session can include two to six participants with a variety of formats. The online submission of invited session proposals opens on July 21, 2021 and will close on September 8, 2021.

A topic-contributed session must have five presentations (including discussants). The online submission of topic contributed session proposals will open on November 11, 2021 and close on December 9, 2021. The regular contributed abstract submission will open on December 1, 2021 and close on February 1, 2022, and a draft manuscript must be received by May 31, 2022. All proposals should be submitted via the online submission process.

You will need to have the following information to submit your session proposal via the online submission process at `http://ww2.amstat.org/meetings/jsm/2022/submissions.cfm`:

- Session Type (invited)
- Session Subtype (e.g., papers or panel)
- Sponsor: Please select "Lifetime Data Science Section" as the sponsor, and a maximum of 2 additional potential sponsors can be selected.
- Title of Session
- Brief session description and information for each presenter (title of presentation, name, affiliation, address, phone, email, names of co-authors). Abstracts are NOT required for submission at this time.
- Theme (yes or no): Designate if the session has topics relevant to the JSM theme.
- Applied (yes or no): Designate if the session has topics relevant or will have special appeal to applied statisticians.
- Estimated audience size: Please select the estimated audience size that the session will attract.
- Organizer: name, affiliation, address, phone and email.
- Chair: name, affiliation, address, phone and email.
- Discussants (if any): name, affiliation, address, phone and email.

Please contact Dr. Jing Ning at `jning@mdanderson.org` should you have any question regarding the LiDS-sponsored invited sessions.

Jing Ning, Program Chair 2022

Webinars & Short Courses in 2021

The LiDS webinar and short course series has been a success since its launch in January 2021. Dr. Noah Simon (University of Washington) kicked off the series on January 26th with a 2-hour webinar on "Statistical Learning with Time-to-Event Outcome". A second webinar on "Competing Frameworks and Methods for Competing Risks Data" was presented by Dr. Douglas Schaubel (University of Pennsylvania) on April 30th. Dr. Yi Li (University of Michigan) gave a 4-hour short course on "Survival Outcome Data with High Dimensional Predictors" on June 24. All the events were well-received by participants.

Several additional events have been planned for the second half of 2021. In September, Drs. Stijn Vansteelandt (Ghent University), Oliver Dukes (Ghent University) and Torben Martinussen (University of Copenhagen) will give a short course on causal inference with survival data. On October 20, there will be a 1.5-hour career development webinar, during which Dr. Malka Gorfine (Tel Aviv University) will talk about publishing and Dr. Jeremy Taylor (University of Michigan) will discuss grant writing. Further details about these events will be available in future LiDS communications.

Shanshan Zhao & Ying Ding
LiDS Webinar Committee Co-Chairs

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`.

The July 2021 issue (Volume 27, number 3) of Lifetime Data Analysis has been published:

- Information measures and design issues in the study of mortality deceleration: findings for the gamma-Gompertz model *by* M. Böhnstedt, J. Gampe, & H. Putter. Pages 333–356
- The semiparametric accelerated trend-renewal process for recurrent event data *by* C. L. Su, R. J. Steele, & I. Shrier. Pages 357–387
- Estimating duration distribution aided by auxiliary longitudinal measures in presence of missing time origin *by* Y. Xiong, W. J. Braun, & X. J. Hu. Pages 388–412
- Regression analysis of current status data with latent variables *by* C. Wang, B. Zhao, L. Luo, & X. Song. Pages 413–436
- A varying-coefficient model for gap times between recurrent events *by* J. E. Soh, & Y. Huang. Pages 437–459
- Augmented likelihood for incorporating auxiliary information into left-truncated data *by* Y. Shi, L. Zeng, M. E. Thompson, & S. L. Tyas. Pages 460–480
- OptBand: optimization-based confidence bands for functions to characterize time-to-event distributions *by* T. Chen, S. Tracy, & H. Uno. Pages 481–498
- Factor copula models for right-censored clustered survival data *by* E. Campos, R. Braekers, D. J. de Souza, & L. M. Chaves. Pages 499–535

Articles in the April 2021 issue (Volume 27, number 2) are:
- The ROC of Cox proportional hazards cure models with application in cancer studies *by* Y. Zhang, X. Han, & Y. Shao. Pages 195–215
- Testing equivalence of survival before but not after end of follow-up *by* J. K. Furberg, C. B. Pipper, & T. Scheike. Pages 216–243
- An additive hazards cure model with informative interval censoring *by* S. Wang, C. Wang, & J. Sun. Pages 244–268
- Semiparametric regression based on quadratic inference function for multivariate failure time data with auxiliary information *by* F. Yan, L. Zhu, Y. Liu, J. Cai, & H. Zhou. Pages 269–299
- Optimal designs for discrete-time survival models with random effects *by* X. D. Zhou, Y. J. Wang, & R. X. Yue. Pages 300–332

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

Multistate Models in the survival Package

Over the last few years, multistate survival models have become an increasing part of our work, which in turn led to a lot of thought about how to make these models easier to use. This article highlights some of the features available in the R **survival** package (3.2 and later). More details about the topic are available in the survival vignette (Therneau 2021a).

The overall goals for the additions were to:

- Make probability-in-state curves as easy to create as Kaplan-Meier (KM) curves

- Make fitting multistate models as easy as fitting a Cox model

- Provide access to other estimates of absolute risk, e.g., the restricted mean time in state (RMTS).

Creating the correct dataset, comprised of multiple rows per subject, is key to running these models, and can be the most time-consuming part of the project. Requirements for the data include:

- Subject id

- Time interval (time1, time2]

- Covariates that apply over the time interval

- Transition state, if any, at the end of the interval. This is a factor variable whose first level is "no transition," but is otherwise unrestricted.

- Current state at the start of the interval. If not present, each subject is assumed to start in a common state "(s0)."

Counting process datasets are often used for modeling time-dependent covariates. The primary differences here are the presence of an `id` variable and the use of a factor for the status variable.

Diagram

The first step is to diagram the multistate process we intend to model. We will use the `myeloid` dataset as an example, which was a trial with two treatments (A and B). The outcomes were complete response (CR), relapse, and death. At this step, its important to consider the research questions and what arrows and boxes help address those questions.

```
library(survival)
states <- c("Entry", "CR", "Relapse", "Death")
cmat <- matrix(0L,4,4, dimnames=list(states, states))
cmat[1,2:4] <- 1; cmat[2,3:4] <- 1; cmat[3,4] <- 1
statefig(c(1,2,1), cmat, cex=.6)
```
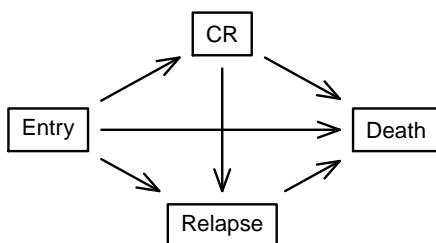

Build data

Once the diagram is created, the next step is to create a dataset that corresponds to the diagram. The code below prepares the `myeloid` dataset for analysis using the `tmerge()` function by breaking the follow-up into time intervals based on the different events. Subject 1 has CR at day 44, relapse at day 113, and died on day 235; subject 2 is censored on day 286 without CR or relapse.

```
m1 <- tmerge(myeloid[,1:3], myeloid, id=id,
            death=event(futime, death),
            cr=event(crtime),
            relapse=event(rltime))
m1$event <- with(m1,factor(3*death + 2*relapse +1*cr,
                0:3, c("censor", "CR",
                        "relapse", "death")))
m1[1:4,]
```

	id	trt	sex	tstart	tstop	death	cr	relapse
1	1	B	f	0	44	0	1	0
2	1	B	f	44	113	0	0	1
3	1	B	f	113	235	1	0	0
4	2	A	m	0	286	1	0	0

	event
1	CR
2	relapse
3	death
4	death

A key feature of the data is that it needs to describe a valid path for each subject: they cannot be two places at once (no overlapping intervals) and they cannot spend time nowhere (no time gaps). The `survcheck` routine is use to verify data validity and will return error messages if the data has problems. In this case it shows that 454/646 subjects achieved CR as their first transition, 20 initially relapsed, 102 died, and 70 were still alive at last follow-up without any observed transitions.

```
ck <- survcheck(Surv(tstart, tstop, event) ~1,
        data=m1, id=id)
ck$transitions
```

	to			
from	CR	relapse	death	(censored)
(s0)	454	20	102	70
CR	0	206	50	198
relapse	0	0	168	58
death	0	0	0	0

```
ck$flag
```

overlap	gap	jump	teleport	duplicate
0	0	0	0	0

Perhaps most importantly, no subject transitioned from death to another state or had further follow-up after death, nor were there any warning messages with respect to overlapping intervals and/or gaps as seen in `ck$flag`. It is also a good idea to check this transition matrix against the study diagram to see if there are any unexpected transitions.

Probability-in-state

A plot of probability-in-state is shown below; it is based on the Aalen-Johansen (AJ) estimate. The Kaplan-Meier is a special case of the AJ. More treatment B subjects (dashed lines) than treatment A subjects achieved CR, while fewer treatment B subjects died by 3 years compare to treatment A subjects. Relapse is a transient state, and has less difference between the two treatment arms.

```
fit1 <- survfit(Surv(tstart, tstop, event) ~
                trt, data=m1, id=id)
plot(fit1, lty=1:2, col= c(1,1,2,2,3,3),
     xmax= 3*365, xscale=365, lwd=2,
     xlab="Years", ylab="Probability in state")
text(c(260, 500, 800), c(.67, .06, .58),
     c("CR", "Relapse", "Death"),
     col=1:3)
```
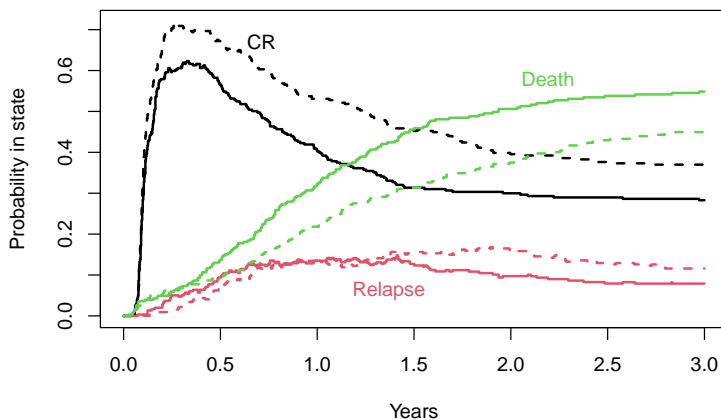

The RMTS is available by specifying the desired end time using the `rmean` option. Here, subjects in arm B spend 1.41 of their first 3 years in the CR state versus 1.09 for subjects in arm A.

```
print(fit1, rmean=3*365, scale=365, digits=2)
```

```
                 n nevent rmean*
trt=A, (s0)    632      0   0.50
trt=B, (s0)    694      0   0.41
trt=A, CR      632    206   1.09
trt=B, CR      694    248   1.41
trt=A, relapse 632    109   0.29
trt=B, relapse 694    117   0.35
trt=A, death   632    171   1.12
trt=B, death   694    149   0.84
   *mean time in state, restricted (max time = 3 )
```

Pseudovalues

Pseudovalues can be used to further explore the relationship of RMTS to covariates. (Formally one should use a working independence variance, available in the packages `geese` and `svyglm`, but we have found little impact when only a single time point is assessed.)
More details are available in the pseudo vignette (Therneau 2021b).

```
p1 <- pseudo(fit1, time=365*3, type="RMTS")

# data p1 has 1 row/subj, 1 column/outcome
colnames(p1)

[1] "(s0)"    "CR"       "relapse" "death"

summary(glm(p1[,2] ~ trt + sex, data=myeloid))

            Estimate Std. Error t value Pr(>|t|)
(Intercept)  398.086       15.9 25.0829 2.40e-97
trtB         115.329       19.4  5.9560 4.26e-09
sexm          -0.815       19.5 -0.0418 9.67e-01
```

Fit multistate models

The multistate hazard model is fit using the `coxph()` function. The default printout in this case is fairly long, showing the coefficients, exp(coefficient), standard error, z, and p separately for each of the 6 possible transitions between the 4 states. For instance, the column 1:2 refers to the transition from the entry state "(s0)" to the "CR" state. The ordering of the states (stored in the attributes) is printed after the coefficients. Below we have used a shorter summary giving only the coefficients.

```
cfit <- coxph(Surv(tstart, tstop, event) ~ trt + sex,
              data= m1, id=id)
round(coef(cfit, matrix=TRUE), 3)

        1:2    1:3    2:3    1:4    2:4    3:4
trtB  0.215 -0.521 -0.162 -0.097 -0.653 -0.300
sexm  0.077  1.156 -0.234  0.352  0.092  0.201
attr(,"states")
[1] "(s0)"    "CR"       "relapse" "death"
```

One downside to a multistate model is that there are so many coefficients. We do a second fit adding a constraint that the coefficients are identical for the 3 transitions to death. Shared coefficients for the transitions to death is a fairly common constraint; it is not particularly sensible in this particular case but provides an illustration. We see that treatment B has a higher transition rate to CR, and a lower one to relapse and death.

```
cfit2 <- coxph(list(Surv(tstart, tstop, event) ~
                    trt + sex,
                    1:4 + 2:4 + 3:4 ~
                    trt + sex / common),
               data = m1, id = id)
round(coef(cfit2, matrix=TRUE), 3)

        1:2    1:3    2:3    1:4    2:4    3:4
trtB  0.215 -0.521 -0.162 -0.294 -0.294 -0.294
sexm  0.077  1.156 -0.234  0.228  0.228  0.228
attr(,"states")
[1] "(s0)"    "CR"       "relapse" "death"
```

In the formula list, the first formula is the default for all the transitions, while the second and later are pseudo formulas containing a list of transitions on the left and covariates on the right along with optional modifiers. The covariates act like

the `update.formula()` function and can add or subtract terms from the base formula.

A score test for proportional hazards (PH) is obtained using the `cox.zph` function, just as in a standard `coxph` model.

```
cox.zph(cfit2)
```

```
          chisq df     p
trt_1:2  0.8465  1 0.358
sex_1:2  2.3974  1 0.122
trt_1:3  0.0666  1 0.796
sex_1:3  0.8907  1 0.345
trt_2:3  3.4882  1 0.062
sex_2:3  1.4994  1 0.221
trt_1:4  0.3161  1 0.574
sex_1:4  1.6168  1 0.204
GLOBAL  11.1414  8 0.194
```

Predicted curves

Predicted survival curves from a multistate PH model are obtained using the `survfit()` function. A dummy data set giving the covariate patterns for which a prediction is desired is created first, and then the curves.

We see that the predicted survival from the model matches the Aalen-Johansen (AJ) curve fairly well. Note that although the predicted hazards for each transition in the fitted model satisfy the PH assumption, by design, the overall estimates of P(death) shown in the curve will not be perfectly proportional.

```
dummy <- data.frame(trt=c("A", "B"), sex = "f")
csurv2 <- survfit(cfit2, newdata=dummy)
dim(csurv2)
```

```
  data states
     2      4
```

```
plot(fit1[,4], lty=1:2, lwd=1, xmax= 3*365,
     xscale=365, xlab="Years", ylab="P(Death)")
lines(csurv2[,4], lty=1:2, lwd=2, col=2)
legend("bottomright", c("trt A, AJ", "trt A, Cox",
                        "trt B, AJ", "trt B, Cox"),
       lty=c(1,1,2,2), lwd=c(1,2,1,2),
       col=c(1,2,1,2), bty='n')
```
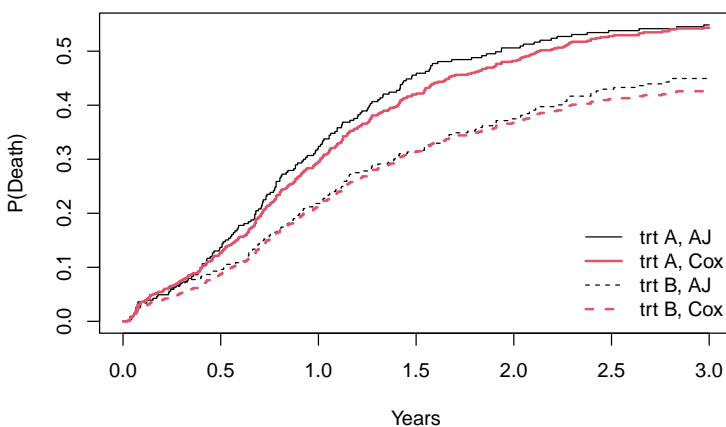

Conclusion

This has provided a very short introduction to the functions; further examples are found in the vignettes for the package. Many other R packages build on `survival` as a foundation, extending and enhancing its results, and we expect that the same will occur for multistate hazard models.

One question we commonly receive is how this code corresponds with the `mstate` package. Putter et al (2007) created an excellent tutorial highlighting the `mstate` package. Most of the `mstate` functionality is now available in the `survival` package. Details can be found in an online vignette (Therneau, Crowson, and Atkinson 2021).

References

Putter, H., M. Fiocco, and R. B. Geskus. 2007. "Tutorial in Biostatistics: Competing Risk and Multi-State Models." *Stat. In Medicine* 26: 2389–2430.

Therneau, T. M. 2021a. "A Package for Survival Analysis in r." `https://cran.r-project.org/web/packages/survival/vignettes/survival.pdf`.

———. 2021b. "Pseudovalues for Survival Data." `https://github.com/therneau/survival/blob/master/vignette2/pseudo.pdf`.

Therneau, T. M., C. Crowson, and E. J. Atkinson. 2021. "Companion to the Tutorial in Biostatistics." `https://github.com/therneau/survival/blob/master/vignette2/tutorial.pdf`.

Terry M. Therneau, Ph.D.
Professor
Division of Clinical Trials and
Biostatistics
Mayo Clinic
Email: `therneau@mayo.edu`

Elizabeth J. Atkinson, M.S.
Assistant Professor
Division of Clinical Trials and
Biostatistics
Mayo Clinic
Email: `atkinson@mayo.edu`

Cynthia S. Crowson, Ph.D.
Professor
Division of Clinical Trials and
Biostatistics
Division of Rheumatology
Mayo Clinic
Email: `crowson@mayo.edu`