



## Special feature: Recent statistical methods for survival analysis

Takeshi Emura<sup>1</sup> · Il Do Ha<sup>2</sup>

Published online: 18 October 2021

© Japanese Federation of Statistical Science Associations 2021

In this special feature of *JJSD*, we collected papers on the state-of-the-art techniques for analyzing *survival data* in all areas of application. We started the call for papers in early 2020, after having swift approval of our project from Editor-in-Chief, Prof. Makoto Aoshima. We then received a number of submissions from Brazil, Canada, Denmark, Germany, Greece, India, Ireland, Japan, South Korea, UK, USA, Taiwan, and the Netherlands. We closed the call in early 2021, and finalized all the peer review in the middle of 2021.

All the submitted manuscripts were reviewed anonymously by two or three leading experts, and underwent multiple rounds of revisions. We found that invited reviewers enthusiastically wrote their reports and scientifically contributed to the eventual outcomes.

Seventeen papers were finally accepted for publication by meeting the criterion of two positive reports set by *JJSD*. We are delighted to achieve this number that is far greater than our original target “at least ten papers”. We sincerely thank the tremendous efforts for all the authors for their painstaking work to meet the rigorous publication criterion. Last but not least, we thank Editor-in-Chief, Prof. Makoto Aoshima, who cooperated with us.

*Survival analysis* deals with data on time-to-event and duration of time until one or more events occur. Survival analysis has been widely applied to biomedical, industrial, econometric, ecological, and other studies. As the coordinating editors of this special feature dealing with a number of excellent papers, we witnessed the tremendous growth in the new methodologies and flexible models for dealing with complex survival data. Readers will see a variety of topics within our special feature:

- Clustered survival data (Ha & Lee, 2021; Kim et al., 2021; Zhang & Jeong, 2021)

---

✉ Takeshi Emura  
takeshiemura@gmail.com

<sup>1</sup> Biostatistics Center, Kurume University, 67 Asahi-machi, Kurume, Japan

<sup>2</sup> Department of Statistics, Pukyong National University, Busan, South Korea

- Clinical trial and sample size (Bakoyannis et al., 2021; Hu et al., 2021; Zhang & Jeong, 2021)
- Computational method (de Freitas Costa et al., 2021; Huang et al., 2020; Kawakami et al., 2021; Kim et al., 2021; Mitra et al., 2021; Ota & Kimura, 2021)
- Competing risk (Bakoyannis et al., 2021; Ha & Lee, 2021; Lipowski et al., 2021; Zhang & Jeong, 2021)
- Copula (Huang et al., 2020; Lipowski et al., 2021; Ota & Kimura, 2021; Wang & Emura, 2021)
- Counting or Poisson process (Dörre, 2021; Su & Lin, 2020)
- Dependence modeling (Kim & Lee, 2021; Ota & Kimura, 2021; Wang & Emura, 2021)
- Dynamic prediction (Kawakami et al., 2021; Mitra et al., 2021)
- Ecological application (de Freitas Costa et al., 2021)
- Econometric application (Dörre, 2021; Hsu et al., 2021; Lipowski et al., 2021)
- Flexible parametric model (Lipowski et al., 2021; MacKenzie et al., 2021)
- Frailty and random-effects models (Ha & Lee, 2021; Kawakami et al., 2021; MacKenzie et al., 2021; Wang & Emura, 2021; Zhang & Jeong, 2021)
- Interval censoring (Hsu et al., 2021; Kim & Lee, 2021)
- Joint frailty-copula model (Huang et al., 2020; Kawakami et al., 2021; Wang & Emura, 2021)
- Left-truncation or double-truncation (Dörre, 2021; Mitra et al., 2021)
- Meta-analysis or multi-center analysis (Ha & Lee, 2021; Kawakami et al., 2021; Zhang & Jeong, 2021)
- Reliability (Hu et al., 2021; Mitra et al., 2021; Ota & Kimura, 2021)
- Recurrent event times (Huang et al., 2020; Lipowski et al., 2021; Su & Lin, 2020)
- Regression (Bakoyannis et al., 2021; de Freitas Costa et al., 2021; Ha & Lee, 2021; Hsu et al., 2021; Kim et al., 2021; Lipowski et al., 2021; MacKenzie et al., 2021; Su & Lin, 2020)

The seventeen papers in the special feature are summarized as follows:

Su and Lin (2020) propose marginal rate models for multiple counting processes. The novelty of their models is to account for *cyclic* or *periodic* patterns of the evolution of the baseline rates over time. The model is a variant of the model proposed by Cai and Schaubel (2004) for multiple type recurrent events without cyclic patterns. The proposed estimator is shown to be more efficient than the estimator of Cai and Schaubel (2004) if the cyclic pattern exists.

Huang et al. (2020) propose a copula-based Markov chain model for recurrent events and a terminal event. The model adopts two copulas: one for the serial dependence in recurrent events, and the other for dependence for the terminal event and initial recurrence. This model is regarded as a deviation from the joint frailty-copula model of recurrent events (Emura et al., 2017; Li et al., 2019). Readers interested in copula-based Markov models are referred to the book of Sun et al. (2020).

Lipowski et al. (2021) revisit the competing risks model of Lo et al. (2020). This is a nested model involving two copulas: one for the dependence of risks, and the other for the dependence among spells (time periods). The cause-specific hazard

models by the piecewise constant and other parametric hazards are fitted to a large dataset of maternity leave periods of mothers.

Bakoyannis et al. (2021) investigate the issues of the omission and categorization of covariates from the Fine–Gray model of competing risks. They show that the omission and categorization of an important covariate leads to attenuated estimates for treatment effect even if the treatments are randomized. The practical consequence is illustrated using data from HIV-infected individuals.

Hsu et al. (2021) develop a functional quantile regression analysis for interval censored data. This approach differs from the traditional quantile or median regression, where the quantile level is fixed at a specified percentage (e.g., 50% quantile). The proposed methodology is applied to a survey dataset on monthly salaries of Taiwanese workers, where the exact values of salaries are interval-censored.

Mitra et al. (2021) propose the Lehmann family of distributions to analyze left-truncated and right-censored lifetime data from the electric power transformers. They develop maximum likelihood estimators under the Weibull, Gompertz, and Lomax distributions that belong to the Lehmann family. The models and methods are illustrated by the transformer data presented in Emura and Shiu (2016).

Ota and Kimura (2021) consider the parameter estimation for the multivariate Farlie–Gumbel–Morgenstern copula (Johnson & Kotz, 1975). Owing to a large number of parameters and the complex parameter constraints, the full maximum likelihood (ML) estimation may not be feasible. To mitigate the issue, the authors develop a sequential ML estimation, and assess its asymptotic efficiency.

Hu et al. (2021) derive a sample size determination scheme to estimate the parameter  $P(X < Y)$ , where  $X$  follows a geometric distribution and  $Y$  follows an exponential distribution. Based on the idea of Mukhopadhyay and Zhuang (2016), the authors develop a sampling procedure to determine the sample size for a pre-specified accuracy. A survival dataset for prostate cancer is used for illustration.

Wang and Emura (2021) investigate a class of multivariate failure time models that include copula models and frailty models as special cases. Hence, their class is named as *frailty-copula models*, resembling to the *joint frailty-copula models* for clustered survival data (Emura et al., 2019). An engineering application of the proposed models is seen in their early publication (Wang et al., 2020).

de Freitas Costa et al. (2021) suggest zero-inflated Weibull and gamma regression models to fit right-censored dispersal data of the wild boars. The Weibull model is a special case of Louzada et al. (2018) and the gamma model is an extension of Lee et al. (2010). The methodological and computational details (including R codes) allow ecological researchers to apply the proposed methods.

Ha and Lee (2021) present an authentic review paper on the hierarchical likelihood (h-likelihood) methods for survival data. The paper provides not only an informative summary for survival data analysis via random-effects models, but also includes a tutorial for a newly established web-based software Albatross Analytics. Furthermore, the authors discuss recent and future extensions of the h-likelihood. Readers may also refer the book of Ha et al. (2017) for survival analysis with the h-likelihood.

Kim et al. (2021) review several different approaches for estimating the variance of rank-based estimators in the semiparametric marginal AFT model for

multivariate failure times. They conduct simulations comparing the variance estimators under correlated failure times, and discuss the results. These variance estimators are then illustrated with the data from a diabetic retinopathy study.

Dörre (2021) presents a statistical method for analyzing doubly truncated data from heterogeneous populations. This method extends the homogeneous population method previously considered by Dörre (2020). The proposed model consists of a nonparametric birth process and a parametric lifetime. A dataset on personal insolvencies is used as a case study, where the proposed model is fitted to estimate the mean age at insolvency and the birth distribution of the population. Readers interested in doubly truncated data analyses are referred to the book of Dörre and Emura (2019).

MacKenzie et al. (2021) revisit the parametric family of time-dependent logistic models that was first founded by the author in 1996. Their main contribution is to generalize the family by means of a Gamma frailty mixture as well as covariate effects, which is shown to accommodate crossing hazards. After a set of simulations, the new models are used to analyze two survival datasets.

Kim and Lee (2021) consider the estimation of a cross-ratio function for bivariate lifetimes to understand their local dependence structure. They cleverly extend the pseudo-partial likelihood estimator for bivariate censored data to the one for bivariate interval-censored data. The methods are illustrated by two datasets: one for the dependence pattern of cow's udder infection times, and the other for the dependence pattern of children's tooth eruption times.

Zhang and Jeong (2021) propose a U-statistic to estimate the win ratio for terminal and non-terminal event times under cluster-randomized trials. They define the *clustered win ratio* as the target for estimation; this is actually the conditional win ratio given a trial-specific frailty value; the idea of conditioning is explained in Emura et al. (2021). The asymptotic distribution of the proposed estimator is derived and the hypothesis test is founded. The method is illustrated with breast cancer data.

Kawakami et al. (2021) consider the Gumbel copula and the FGM copula to simplify the dynamic prediction formula of the joint frailty-copula model. Unlike the numerical approximation approach suggested by Emura et al. (2018), the proposed prediction formulas are written explicitly without any approximation. This gives a novel tool for assessing approximation errors by the numerical integration.

We strongly believe that any reader of this special feature will find the merit of each paper and the trend of recent statistical methods for survival analysis.

## References

- Bakoyannis, G., Chu, F. I., Babiker, A. G. A., et al. (2021). Impact of covariate omission and categorization from the Fine-Gray model in randomized-controlled trials. *Japanese Journal of Statistics and Data Science*. <https://doi.org/10.1007/s42081-021-00111-5>
- Cai, J., & Schaubel, D. E. (2004). Marginal means/rates models for multiple type recurrent event data. *Lifetime Data Analysis*, 10, 121–138.

- de Freitas Costa, E., Schneider, S., Carlotto, G. B., et al. (2021). Zero-inflated-censored Weibull and gamma regression models to estimate wild boar population dispersal distance. *Japanese Journal of Statistics and Data Science*. <https://doi.org/10.1007/s42081-021-00124-0>
- Dörre, A. (2020). Bayesian estimation of a lifetime distribution under double truncation caused by time-restricted data collection. *Statistical Papers*, 61(3), 941–965. <https://doi.org/10.1007/s00362-017-0968-7>
- Dörre, A. (2021). Semiparametric likelihood inference for heterogeneous survival data under double truncation based on a Poisson birth process. *Japanese Journal of Statistics and Data Science*. <https://doi.org/10.1007/s42081-021-00128-w>
- Dörre, A., & Emura, T. (2019). *Analysis of doubly truncated data: an introduction*. JSS Research Series in Statistics, Springer, Singapore.
- Emura, T., Matsui, S., & Rondeau, V. (2019). *Survival analysis with correlated endpoints: joint frailty-copula models*, JSS Research Series in Statistics, Springer, Singapore.
- Emura, T., Nakatochi, M., Matsui, S., Michimae, H., & Rondeau, V. (2018). Personalized dynamic prediction of death according to tumour progression and high-dimensional genetic factors: Meta-analysis with a joint model. *Statistical Methods in Medical Research*, 27(9), 2842–2858.
- Emura, T., Nakatochi, M., Murotani, K., & Rondeau, V. (2017). A joint frailty-copula model between tumour progression and death for meta-analysis. *Statistical Methods in Medical Research*, 26(6), 2649–2666.
- Emura, T., & Shiu, S. K. (2016). Estimation and model selection for left-truncated and right-censored lifetime data with application to electric power transformers analysis. *Communications in Statistics-Simulation and Computation*, 45(9), 3171–3189.
- Emura, T., Sofeu, C., & Rondeau, V. (2021). Conditional copula models for correlated survival endpoints: individual patient data meta-analysis of randomized controlled trials. *Statistical Methods in Medical Research*. <https://doi.org/10.1177/09622802211046390>
- Ha, I. D., Jeong, J. H., & Lee, Y. (2017). *Statistical modelling of survival data with random effects: H-likelihood approach*. Berlin: Springer.
- Ha, I. D., & Lee, Y. (2021). A review of h-likelihood for survival analysis. *Japanese Journal of Statistics and Data Science*. <https://doi.org/10.1007/s42081-021-00125-z>
- Hsu, C. Y., Wen, C. C., & Chen, Y. H. (2021). Quantile function regression analysis for interval censored data, with application to salary survey data. *Japanese Journal of Statistics and Data Science*. <https://doi.org/10.1007/s42081-021-00113-3>
- Hu, J., Zhuang, Y., & Goldiner, C. (2021). Fixed-accuracy confidence interval estimation of  $P(X < Y)$  under a geometric–exponential model. *Japanese Journal of Statistics and Data Science*. <https://doi.org/10.1007/s42081-021-00122-2>
- Huang, X. W., Wang, W., & Emura, T. (2020). A copula-based Markov chain model for serially dependent event times with a dependent terminal event. *Japanese Journal of Statistics and Data Science*. <https://doi.org/10.1007/s42081-020-00087-8>
- Johnson, N. L., & Kotz, S. (1975). On some generalized Farlie–Gumbel–Morgenstern distributions. *Communications in Statistics*, 4(5), 415–427.
- Kawakami, R., Michimae, H., & Lin, Y. H. (2021). Assessing the numerical integration of dynamic prediction formulas using the exact expressions under the joint frailty-copula model. *Japanese Journal of Statistics and Data Science*. <https://doi.org/10.1007/s42081-021-00133-z>
- Kim, K., Ko, J., & Kang, S. (2021). Comparison of variance estimation methods in semiparametric accelerated failure time models for multivariate failure time data. *Japanese Journal of Statistics and Data Science*. <https://doi.org/10.1007/s42081-021-00126-y>
- Kim, Y. J., & Lee, N. (2021). Time-dependent association for bivariate interval censored data. *Japanese Journal of Statistics and Data Science*. <https://doi.org/10.1007/s42081-021-00132-0>
- Lee, A. H., Zhao, Y., Yau, K. K., & Xiang, L. (2010). How to analyze longitudinal multilevel physical activity data with many zeros? *Preventive Medicine*, 51(6), 476–481.
- Li, Z., Chinchilli, V. M., & Wang, M. (2019). A Bayesian joint model of recurrent events and a terminal event. *Biometrical Journal*, 61(1), 187–202.
- Lipowski, C., Lo, S. M. S., Shi, S., et al. (2021). Competing risks regression with dependent multiple spells: Monte Carlo evidence and an application to maternity leave. *Japanese Journal of Statistics and Data Science*. <https://doi.org/10.1007/s42081-021-00110-6>
- Lo, S. M. S., Mammen, E., & Wilke, R. A. (2020). A nested copula duration model for competing risks with multiple spells. *Computational Statistics and Data Analysis*, 150, 106986.

- Louzada, F., Moreira, F. F., & de Oliveira, M. R. (2018). A zero-inflated non default rate regression model for credit scoring data. *Communications in Statistics Theory and Methods*, 47(12), 3002–3021.
- MacKenzie, G., Blagojevic-Bucknall, M., Al-tawarah, Y., et al. (2021). The XGTDL family of survival distributions. *Japanese Journal of Statistics and Data Science*. <https://doi.org/10.1007/s42081-021-00129-9>
- Mitra, D., Kundu, D., & Balakrishnan, N. (2021). Likelihood analysis and stochastic EM algorithm for left truncated right censored data and associated model selection from the Lehmann family of life distributions. *Japanese Journal of Statistics and Data Science*. <https://doi.org/10.1007/s42081-021-00115-1>
- Mukhopadhyay, N., & Zhuang, Y. (2016). On fixed-accuracy and bounded accuracy confidence interval estimation problems in Fisher's "Nile" example. *Sequential Analysis*, 35, 516–535.
- Ota, S., & Kimura, M. (2021). Effective estimation algorithm for parameters of multivariate Farlie–Gumbel–Morgenstern copula. *Japanese Journal of Statistics and Data Science*. <https://doi.org/10.1007/s42081-021-00118-y>
- Su, C. L., & Lin, F. C. (2020). Analysis of cyclic recurrent event data with multiple event types. *Japanese Journal of Statistics and Data Science*. <https://doi.org/10.1007/s42081-020-00088-7b>
- Sun, L. H., Huang, X. W., Alqawba, M. S., Kim, J. M., & Emura, T. (2020). *Copula-based Markov models for time series: Parametric inference and process control*. JSS Research Series in Statistics, Springer, Singapore.
- Wang, Y. C., & Emura, T. (2021). Multivariate failure time distributions derived from shared frailty and copulas. *Japanese Journal of Statistics and Data Science*. <https://doi.org/10.1007/s42081-021-00123-1>
- Wang, Y. C., Emura, T., Fan, T. H., Lo, S. M., & Wilke, R. A. (2020). Likelihood-based inference for a frailty-copula model based on competing risks failure time data. *Quality and Reliability Engineering International*, 36(5), 1622–1638.
- Zhang, D., & Jeong, J. H. (2021). Inference on win ratio for cluster-randomized semi-competing risk data. *Japanese Journal of Statistics and Data Science*. <https://doi.org/10.1007/s42081-021-00131-1>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.