A simulation study for comparison of piece-wise exponential additive mixed-effects models with Cox proportional hazards shared frailty models

May 21, 2021

1 Introduction

In this simulation study we aim to show that the piece-wise exponential additive mixed-effects models (PAMMs) when assuming proportional hazards, give the same estimates as the Cox proportional hazards (CPH) frailty models. We do this for both gap and calendar timescales and stratified and unstratified models. Stratified models refer to event-specific baseline hazards being modelled for the different event numbers and unstratified models refer to a common baseline hazard assumed for all events. Note: The additive value of PAMMs in case of, for instance, non-proportionality is given in the main text. The simulation set-up to generate recurrent event data is motivated by the approaches outlined by [1] and [2]. `R` and `RStudio` are used for the constructing the simulations and the analysis [3, 4]. The `coxme` package [5] is used to analyze the simulated data using the CPH frailty model framework, and the `mgcv` [6–11] and `pammtools` [12–14] packages are used to analyze the simulated data using the PAMM framework. In this supplement, we will first describe the simulation setup and results in the gap timescale in section 3 followed by the calendar timescale in section 4.

For the R code for this simulation study see https://github.com/jordache-ramjith/PAMM/ and for access to the simulation data files, figures and results files see [15].

2 Model comparison statistics

In this section we will define the measures used to summarize the simulation results and thus the measures used for comparing the estimations between the CPH frailty model and the PAMM. The simulations over the different sections are each summarized into a table and two figures. The first figure shows smoothed histograms of the estimated β coefficients from both the CPH frailty model and the PAMM, with a dashed black line indicating the true β. In the second figure this is done for the estimated frailty variance. In the table, the mean and standard error (SE) of the estimated β's across the different simulations from the CPH frailty model and the PAMM are shown for each scenario. Moreover, the coverage (cov.) of the true β in the 95% confidence intervals for the estimated β are given for both models. Since the distribution of the estimated frailty variances is skewed when the true frailty variances are small, the sample medians are presented. In the last two columns of the table the mean of the pair-wise differences between the β estimates in the CPH frailty model and the PAMM are presented as well as the SE for these difference.

3 Gap timescale

In the gap timescale we will make comparisons between the unstratified CPH frailty model and the unstratified PAMM followed by a comparison between the stratified CPH frailty model and the stratified PAMM.

3.1 Simulating data

3.1.1 Generating gap times

We simulate survival times such that the baseline hazard is increasing non-linearly over time, i.e. $\lambda_0(t) = \alpha_j\ t^2$ where

$$\alpha_j = \begin{cases} 1 & \text{if unstratified baseline hazards} \\ j & \text{if } j \leq 4 \text{ \& stratified baseline hazards} \\ 4 & \text{if } j > 4 \text{ \& stratified baseline hazards} \end{cases}$$

where j is the event number. The hazard over time for the j^{th} event conditional on covariate x and frailty z is thus

$$\lambda_j(t|x,z) = \alpha_j\ t^2\ \exp(\beta x + z).$$

We place the restriction on $\alpha_j = \min(j,4)$ to avoid issues of extremely small (and many) simulated gap times that may rise in the data analysis. For the hazard for the j^{th} event given in the previous display, the cumulative distribution for the event time conditional the covariate x and the frailty z, equals

$$F_j(t|x,z) = 1 - \exp\left(-\int_0^t \lambda_j(u|x,z)du\right) = 1 - \exp\left(-\tfrac{1}{3}\alpha_j t^3 \exp(\beta x + z)\right).$$

Let C be the maximum individual follow-up time (see subsections 3.1.2 and 3.1.3) and $T_{ij} \sim F_j$ be the j^{th} gap time for an individual i. Using the fact that $F_j(T_{ij})$ follows a uniform distribution at the interval $[0,1]$, the gap times for individual i are simulated using the following algorithm:

1. Simulate the frailty variable $z_i \sim \text{N}(0,\sigma^2)$.
2. Simulate a binary covariate $x_i \sim \text{Bernoulli}(p)$.
3. Set $t_{i0} = 0$ and $j = 1$.
4. Simulate $u_{ij} \sim \text{Uniform}(0,1)$.
5. Set the gap time as $t_{ij} = F_j^{-1}(u_{ij}) = \left(\frac{-3\ln(1-u_{ij})}{\alpha_j \exp(\beta x_i + z_i)}\right)^{\frac{1}{3}}$ where α_j is as defined before.
6. if $\sum_{k=1}^{j} t_{ik} < C$, set $j := j+1$ and go to step 4, to simulate the next event time. Otherwise, the j^{th} event time is censored by the maximum follow-up time C and $t_{ij} = C - \sum_{k=1}^{j-1} t_{ik}$.

3.1.2 Scenarios: Unstratified baseline hazards

We consider in total eight different scenarios. For each of these scenarios we simulate 500 datasets consisting of recurrent event data of 100 individuals. Across the scenarios, we take $\beta = -1$ and $p = 0.5$ for simulating the binary covariate, we vary the level of the frailty variance, σ^2, as well as the maximum follow-up time (maxtime) and we set the baseline hazard equal to $\alpha_j \equiv 1, \forall j$. The eight scenarios are given in Table 1.

scenario	1	2	3	4	5	6	7	8
σ^2	0.0	0.2	0.5	1.0	0.0	0.2	0.5	1.0
maxtime	5	5	5	5	20	20	20	20

Table 1: Overview of the eight scenarios in the unstratified baseline setting.

3.1.3 Scenarios: Stratified baseline hazards

The eight scenarios are as described in 3.1.2, except that we simulate stratified baseline hazards where α_j for stratified baseline hazards is defined in section 3.1.1, and we used 10 as our maximum follow-up time criteria for scenarios 5-8 instead of 20 to limit the number of events when using stratified baselines (remember that the hazards are higher in the stratified setting). The eight scenarios are described in Table 2.

scenario	1	2	3	4	5	6	7	8
σ^2	0.0	0.2	0.5	1.0	0.0	0.2	0.5	1.0
maxtime	5	5	5	5	10	10	10	10

Table 2: Overview of the eight scenarios in the stratified baseline setting.

3.2 Results

For the simulated gap timescale data from the unstratified baseline hazards model, we estimate the unstratified CPH frailty model and the unstratified PAMM, and compare the results. Thereafter, we do the same for the stratified models. We present the results separately in the sections 3.2.1 and 3.2.2 respectively. Remind that we aim to show that the unstratified models yield similar results, and the same for the stratified models.

3.2.1 Model comparisons: Unstratified models

A summary of the results is shown in Table 3 and figures 1 and 2. From the difference column in Table 3 we see that across all scenarios, the mean difference between the CPH frailty model estimated β and the PAMM estimated β are less than 0.01, with low standard errors. For scenarios 1-4, in Table 3 we see that the coverage of the true effect in the 95% CIs for the CPH frailty model and the PAMM are just below 95%. When the maximum follow-up time is increased to 20, i.e. scenarios 5-8, we notice an improvement in both models in terms of reduced bias in the estimated effect β, and for the PAMM in particular, we see that the coverage of the true effects in the 95% CIs are closer to 95% than the CPH frailty model.

Scenario	CPH frailty model				PAMM				Difference	
	$\hat{\beta}$	SE	cov.	$\hat{\sigma}^2$	$\hat{\beta}$	SE	cov.	$\hat{\sigma}^2$	mean	SE
1	-1.02	0.14	93.6	0.00	-1.02	0.14	93.4	0.00	-9.1×10^{-4}	1.1×10^{-2}
2	-0.99	0.17	90.4	0.18	-0.99	0.17	90.6	0.19	2.8×10^{-3}	1.0×10^{-2}
3	-0.97	0.20	91.8	0.45	-0.97	0.20	92.8	0.46	9.1×10^{-5}	1.0×10^{-2}
4	-0.95	0.24	93.2	0.91	-0.94	0.24	93.8	0.89	-5.9×10^{-3}	1.1×10^{-2}
5	-1.00	0.06	94.4	0.00	-1.00	0.06	94.6	0.00	-3.0×10^{-3}	4.3×10^{-3}
6	-0.99	0.11	93.2	0.19	-0.99	0.11	95.4	0.19	-1.2×10^{-4}	2.1×10^{-3}
7	-0.98	0.15	92.0	0.47	-0.98	0.15	95.8	0.48	-7.5×10^{-4}	2.3×10^{-3}
8	-0.97	0.20	92.4	0.90	-0.97	0.20	95.0	0.95	2.0×10^{-3}	6.8×10^{-3}

Table 3: Summary of the estimated effects, from the CPH frailty model and the PAMM, and their differences for the eight different scenarios. Here we used the unstratified baseline hazards approach in a gap timescale. The summary measures are described in section 2.

In the figures 1 and 2 smoothed histograms of the estimated regression coefficient and frailty variance in both unstratified models for the 8 scenarios are shown. For all scenarios presented, we see that the results of the CPH frailty model and the PAMM are very similar. Looking specifically at the scenarios 1-4 where the data are simulated with a maximum follow-up time of 5, we see that estimated effects and frailty variances from the unstratified CPH frailty model and unstratified PAMM are very similar but somewhat underestimate the true frailty variance (Figure 2).

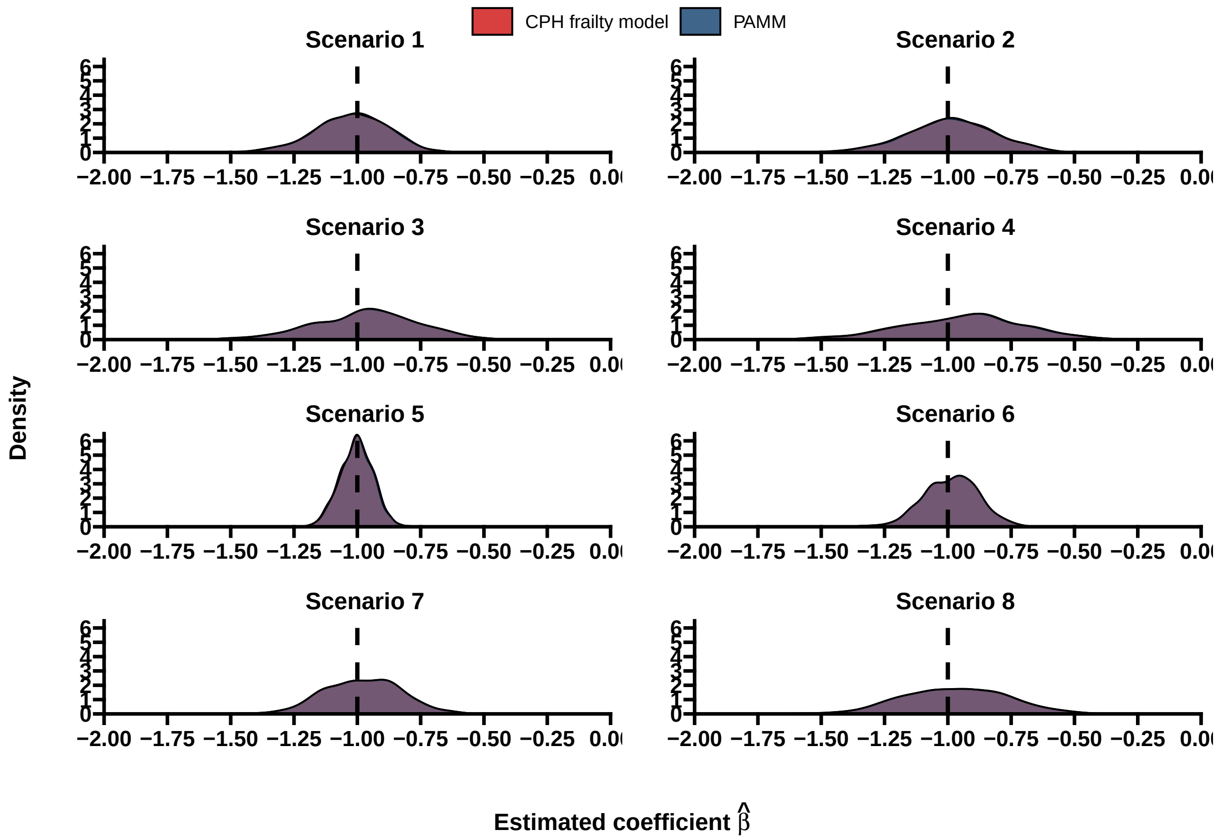

Figure 1: Smoothed histograms of the estimated beta coefficients from the unstratified CPH frailty model and the unstratified PAMM across the simulated data for the eight different scenarios. The overlap of the histograms from both models is shown in purple (there is a perfect overlap).

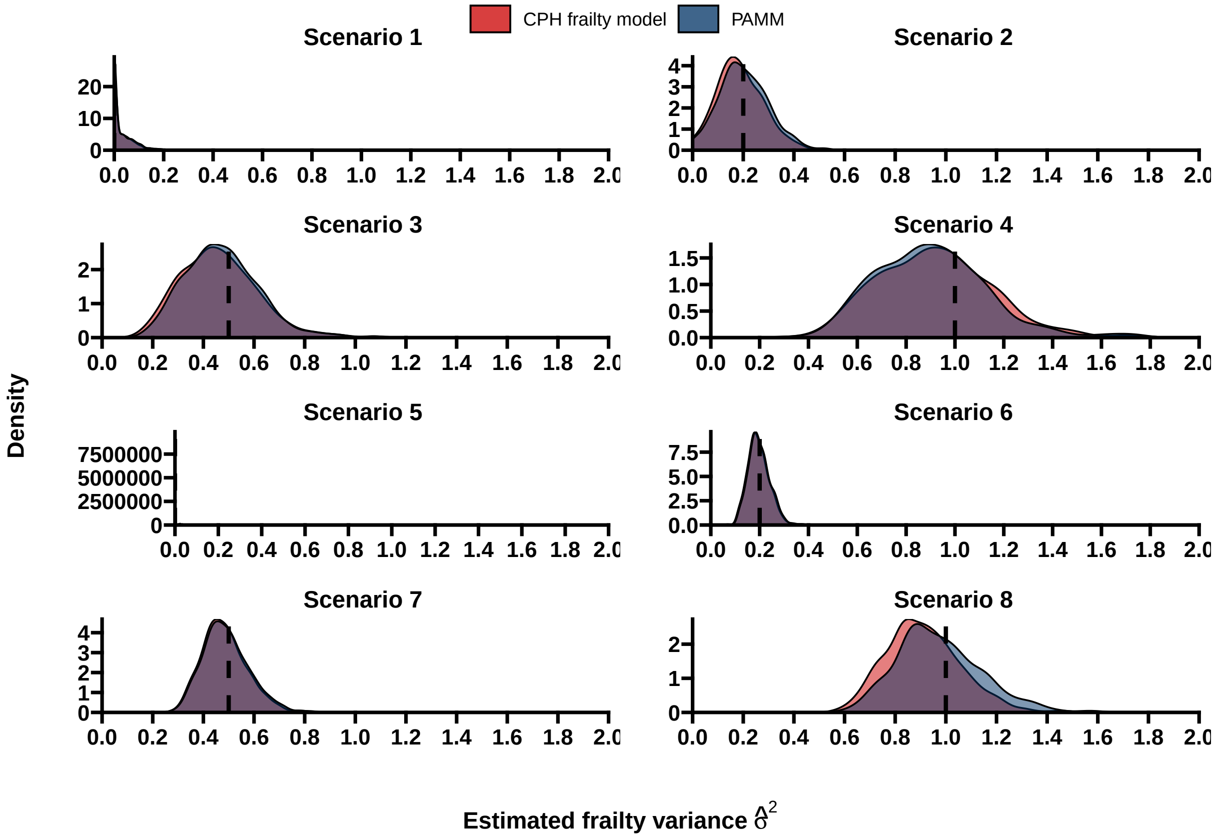

Figure 2: Smoothed histograms of the estimated frailty variances from the unstratified CPH frailty model and the unstratified PAMM for the eight different scenarios. The overlap of the histograms from both models is shown in purple. In scenario 5 there is a high spike at zero.

3.2.2 Model comparisons: Stratified models

A summary of the results for the stratified models is shown in Table 4, and smoothed histograms of the estimated regression coefficient and frailty variance for both stratified models for the 8 scenarios are given in Figures 3 and 4. From the difference column in Table 4 we see that across all scenarios, the mean difference between the CPH frailty model estimated β and the PAMM estimated β are slightly higher than for the differences in the unstratified models (Table 3), but are still relatively small (mean difference is < 0.05 with low standard errors). For the shorter follow-up time, i.e. scenarios 1-4, we see that the PAMM estimates the regression parameter β and the frailty variance more accurately than the CPH frailty model when the true frailty variances are larger (scenarios 3 & 4). However, when the true frailty variance is very small (scenario 1), the CPH frailty model estimates the regression parameter β and the frailty variance more accurately than the PAMM. For true frailty variances $\sigma^2 > 0$, we see that the coverage of the true effect ($\beta = -1$) in the PAMM 95% CIs are slightly higher than that of the CPH frailty model. When the maximum follow-up time is increased to 10, i.e. scenarios 5-8, we notice an improvement in both models in terms a higher accuracy of the estimated effect β, and for the PAMM in particular, we see that the coverage of the true effects in the 95% CIs are closer to 95% than the CPH frailty model except for scenario 5 where the true frailty variance is 0. The estimated frailty variances from the PAMM also become closer to the true frailty variances than the CPH frailty model - again, except for scenario 5. Overall we also see that the smoothed histograms of the estimated beta coefficients and the frailty variances become narrower when the follow-up time is longer.

Scenario	CPH frailty model $\hat{\beta}$	SE	cov.	$\hat{\sigma}^2$	PAMM $\hat{\beta}$	SE	cov.	$\hat{\sigma}^2$	Difference mean	SE
1	-1.03	0.13	94.6	0.00	-1.06	0.13	91.6	0.02	3.6×10^{-2}	1.9×10^{-2}
2	-0.98	0.16	92.2	0.18	-1.02	0.17	93.0	0.22	4.3×10^{-2}	1.7×10^{-2}
3	-0.95	0.19	92.2	0.45	-0.99	0.19	93.6	0.50	3.6×10^{-2}	1.8×10^{-2}
4	-0.94	0.23	92.6	0.89	-0.96	0.24	93.8	0.95	2.7×10^{-2}	1.9×10^{-2}
5	-1.01	0.09	93.6	0.00	-1.04	0.09	91.6	0.01	3.3×10^{-2}	1.2×10^{-2}
6	-0.98	0.12	92.6	0.18	-1.01	0.12	94.8	0.21	3.7×10^{-2}	1.1×10^{-2}
7	-0.97	0.16	92.8	0.46	-1.00	0.16	95.4	0.51	3.5×10^{-2}	1.2×10^{-2}
8	-0.95	0.20	93.4	0.88	-0.99	0.21	95.2	1.00	3.7×10^{-2}	1.6×10^{-2}

Table 4: Summary of the estimated effects, from the CPH frailty model and the PAMM, and their differences for the eight different scenarios. Here we used the stratified baseline hazards approach in a gap timescale. The summary measures are described in section 2.

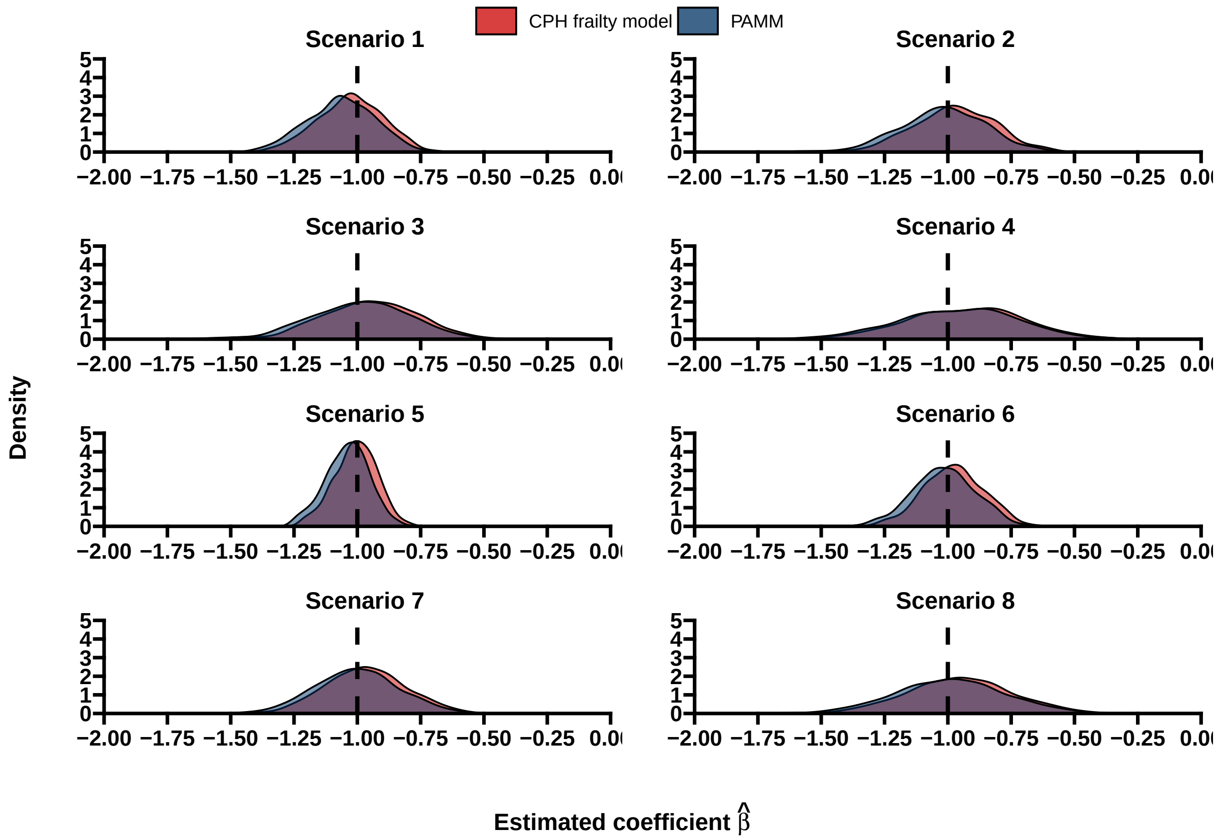

Figure 3: Smoothed histograms of the estimated beta coefficients from the stratified CPH frailty model and the stratified PAMM for the eight different scenarios. The overlap of the histograms from both models is shown in purple.

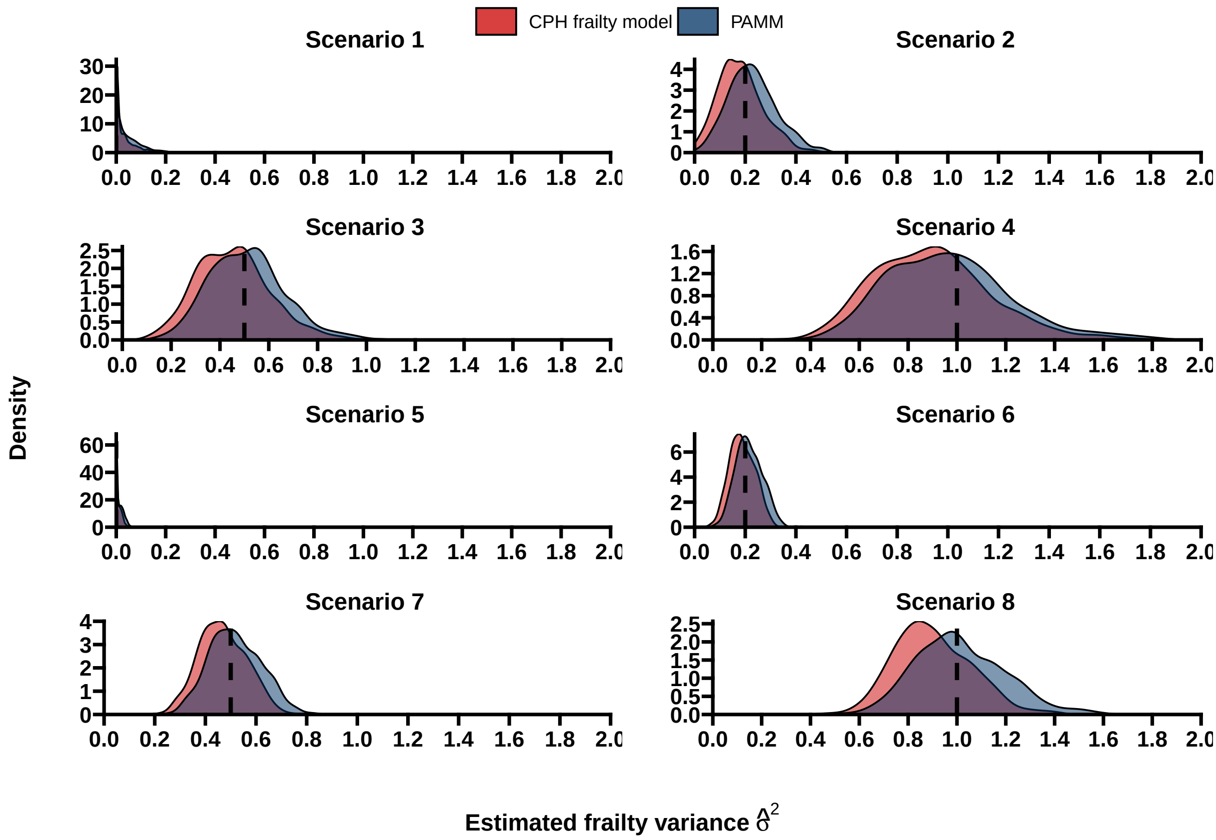

Figure 4: Smoothed histograms of the estimated frailty variances from the stratified CPH frailty model and the stratified PAMM for the eight different scenarios. The overlap of the histograms from both models is shown in purple.

4 Calendar timescale

In the calendar timescale, we will make comparisons between the unstratified CPH frailty model and PAMMs.

4.1 Simulating data

4.1.1 Generating calendar times

For the simulation of calendar times, we used Weibull distributed baseline hazards with parameters $\lambda = 1$ and $\gamma = 0.5$, a baseline hazards that is high early on in calendar time and then decreases over calendar time. This is an unstratified baseline hazard pattern that is seen in recurrent childhood infectious diseases, i.e. as children grow older, their event rate for childhood infectious diseases decrease. The hazard over time conditional on a covariate x and frailty z is thus

$$\lambda(t|x,z) = \frac{\exp(\beta x + z)}{\sqrt{t}}.$$

[2] shows that the cumulative incidence function for the time between the $(j-1)^{th}$ and the j^{th} event, $w = t_j - t_{j-1}$ is

$$F(w|t_{j-1}, x, z) = 1 - \exp\left(-\int_{t_{j-1}}^{t_{j-1}+w} \lambda(u|x,z)\, du\right).$$

For the baseline hazard we chose, the calendar time for the j^{th} event is thus

$$t_j = \left(\frac{-\ln(1 - F(w|t_{j-1}, x, z))}{2\exp(\beta x + z)} + \sqrt{t_{j-1}} \right)^2.$$

For C the maximum individual follow-up time (see subsection 4.1.2), the recurrent event times in calendar timescale for individual i are simulated according to the following algorithm:

1. Simulate the frailty variable $z_i \sim \mathrm{N}(0, \sigma^2)$.
2. Simulate a binary covariate $x_i \sim$ Bernoulli(0.5).
3. Set $t_{i0} = 0$ and set $j = 1$.
4. For the j^{th} event, simulate $u_{ij} \sim$ Uniform(0,1).
5. Set $t_{ij} = \left(\frac{-\ln(1-u_{ij})}{2\exp(\beta x_i + z_i)} + \sqrt{t_{i(j-1)}} \right)^2$.
6. if $t_{ij} < C$, set $j := j+1$ and go to step 4, to simulate the next event. Otherwise, the j^{th} event time is censored by the maximum follow-up time C and $t_j = C$.

For more information on the simulation details for simulating recurrent events in calendar time, see [2, 16].

4.1.2 Scenarios

Again, we consider eight different settings. For each setting, we simulate 500 datasets each consisting of recurrent event data of 100 individuals. Across all scenarios, we take $\beta = -1$ and $p = 0.5$ for the simulation of the binary covariate, we vary the level of frailty variance, σ^2, and the maximum follow-up time. The eight scenarios are described in Table 5.

scenario	1	2	3	4	5	6	7	8
σ^2	0.0	0.2	0.5	1.0	0.0	0.2	0.5	1.0
maxtime	25	25	25	25	100	100	100	100

Table 5: Overview of the eight scenarios in the calendar timescale setting.

4.2 Results

For the calendar timescale simulated data, we analyze each simulated dataset with the unstratified CPH frailty model and the unstratified PAMM, both using a calendar time approach. A summary of the analysis is shown in Table 6, and the smoothed histograms of the estimated regression coefficient and frailty variance for both the CPH frailty model and PAMMs for the eight scenarios are shown in Figures 5 and 6. For all scenarios presented, we see that the estimates in the CPH frailty model and the PAMM are very similar (difference column, Table 6). Both models tend to estimate the regression parameter and frailty variance better when there is a longer follow-up time - due to a larger number of events. When the true frailty variance is more than zero, both models slightly underestimate the frailty variance; and this slight underestimation is larger when the true frailty variance is large and the maximum follow-up time is low. Overall, the bias in estimates are relatively low for all scenarios with the highest (7%) for scenario 4, from both models, where there is more censoring because individuals with very low frailties will more likely exceed the maximum time.

	CPH frailty model				PAMM				Difference	
Scenario	$\hat{\beta}$	SE	cov.	$\hat{\sigma}^2$	$\hat{\beta}$	SE	cov.	$\hat{\sigma}^2$	mean	SE
1	-1.00	0.08	96.2	0.00	-1.00	0.08	96.2	0.00	1.7×10^{-3}	1.8×10^{-3}
2	-0.97	0.12	94.4	0.18	-0.97	0.12	94.8	0.18	9.1×10^{-4}	1.1×10^{-3}
3	-0.95	0.15	93.8	0.46	-0.95	0.15	93.6	0.44	-1.0×10^{-3}	1.5×10^{-3}
4	-0.93	0.20	94.8	0.92	-0.93	0.20	93.6	0.85	-3.1×10^{-3}	4.6×10^{-3}
5	-1.00	0.06	96.2	0.00	-1.00	0.06	96.4	0.00	8.6×10^{-4}	9.1×10^{-4}
6	-0.97	0.11	93.8	0.19	-0.97	0.11	93.8	0.19	3.3×10^{-4}	5.9×10^{-4}
7	-0.96	0.15	94.4	0.47	-0.96	0.15	93.8	0.46	-7.9×10^{-4}	7.1×10^{-4}
8	-0.95	0.20	94.6	0.94	-0.94	0.20	93.6	0.90	-1.4×10^{-3}	4.4×10^{-3}

Table 6: Summary of the estimated effects, from the CPH frailty model and the PAMM, and their differences for the eight different scenarios. Here we used the unstratified baseline hazards approach in a calendar timescale. The summary measures are described in section 2.

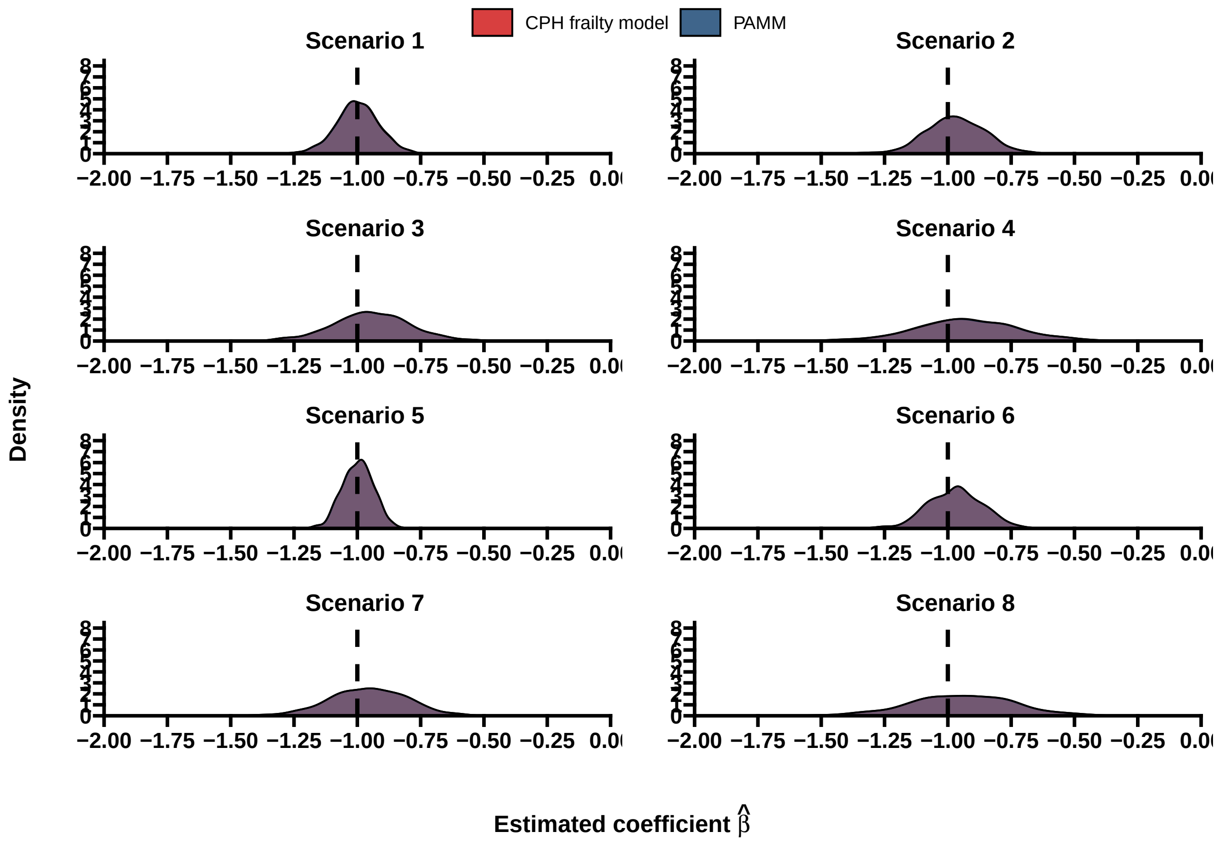

Figure 5: Smoothed histograms of the estimated beta coefficients from the unstratified CPH frailty model and the unstratified PAMM, both in the calendar timescale, for the eight different scenarios. The overlap of the histograms from both models is shown in purple.

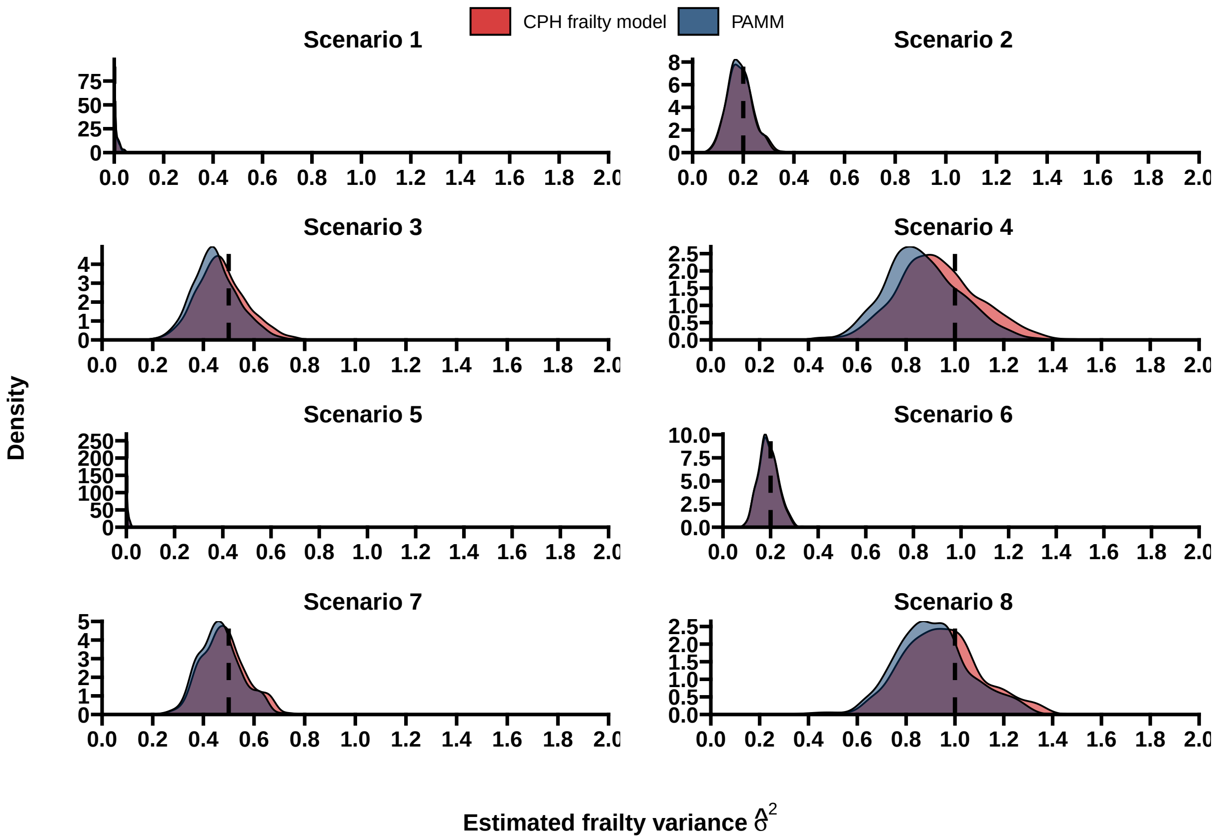

Figure 6: Smoothed histograms of the estimated frailty variances from the unstratified CPH frailty model and the unstratified PAMM, both in the calendar timescale, for the eight different scenarios. The overlap of the histograms from both models is shown in purple.

4.2.1 Additional simulations: N=200

We run 200 additional simulations for each of the scenarios described in section 4.1.2, but we increase the sample size to 200. We see similar results but with smaller standard errors, as expected (Table 7, Figures 7 & 8).

Scenario	CPH frailty model				PAMM				Difference	
	$\hat{\beta}$	SE	cov.	$\hat{\sigma}^2$	$\hat{\beta}$	SE	cov.	$\hat{\sigma}^2$	mean	SE
1	-0.99	0.06	94.0	0.00	-0.99	0.06	94.5	0.00	8.9×10^{-4}	8.0×10^{-4}
2	-0.97	0.09	91.0	0.19	-0.97	0.09	91.0	0.18	4.2×10^{-4}	5.9×10^{-4}
3	-0.95	0.12	89.5	0.47	-0.95	0.12	89.5	0.44	-1.6×10^{-3}	1.1×10^{-3}
4	-0.94	0.16	92.5	0.93	-0.94	0.16	91.5	0.86	-3.9×10^{-3}	2.5×10^{-3}
5	-0.99	0.05	95.0	0.00	-0.99	0.05	96.0	0.00	3.7×10^{-4}	3.5×10^{-4}
6	-0.98	0.08	91.0	0.19	-0.98	0.08	91.0	0.19	-7.2×10^{-5}	2.8×10^{-4}
7	-0.97	0.11	92.0	0.48	-0.97	0.11	92.0	0.46	-1.3×10^{-3}	6.0×10^{-4}
8	-0.96	0.15	94.0	0.95	-0.96	0.15	93.5	0.90	-2.0×10^{-3}	2.4×10^{-3}

Table 7: Summary of the estimated effects, from the CPH frailty model and the PAMM, and their differences for the eight different scenarios. Here we used the unstratified baseline hazards approach in a calendar timescale and a larger sample size ($N = 200$). The summary measures are described in section 2.

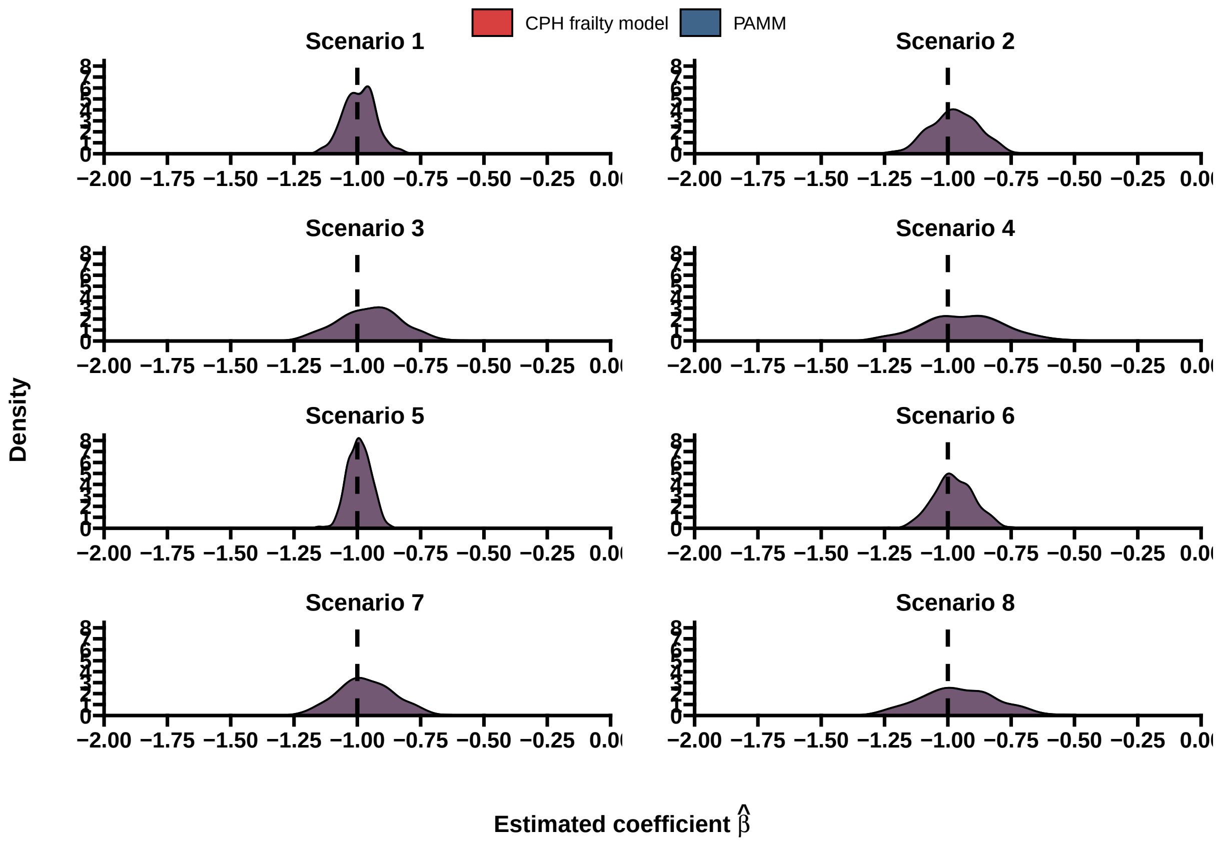

Figure 7: Smoothed histograms of the estimated beta coefficients from the unstratified CPH frailty model and the unstratified PAMM, both in the calendar timescale, for the eight different scenarios with sample size N=200. The overlap of the histograms from both models is shown in purple.

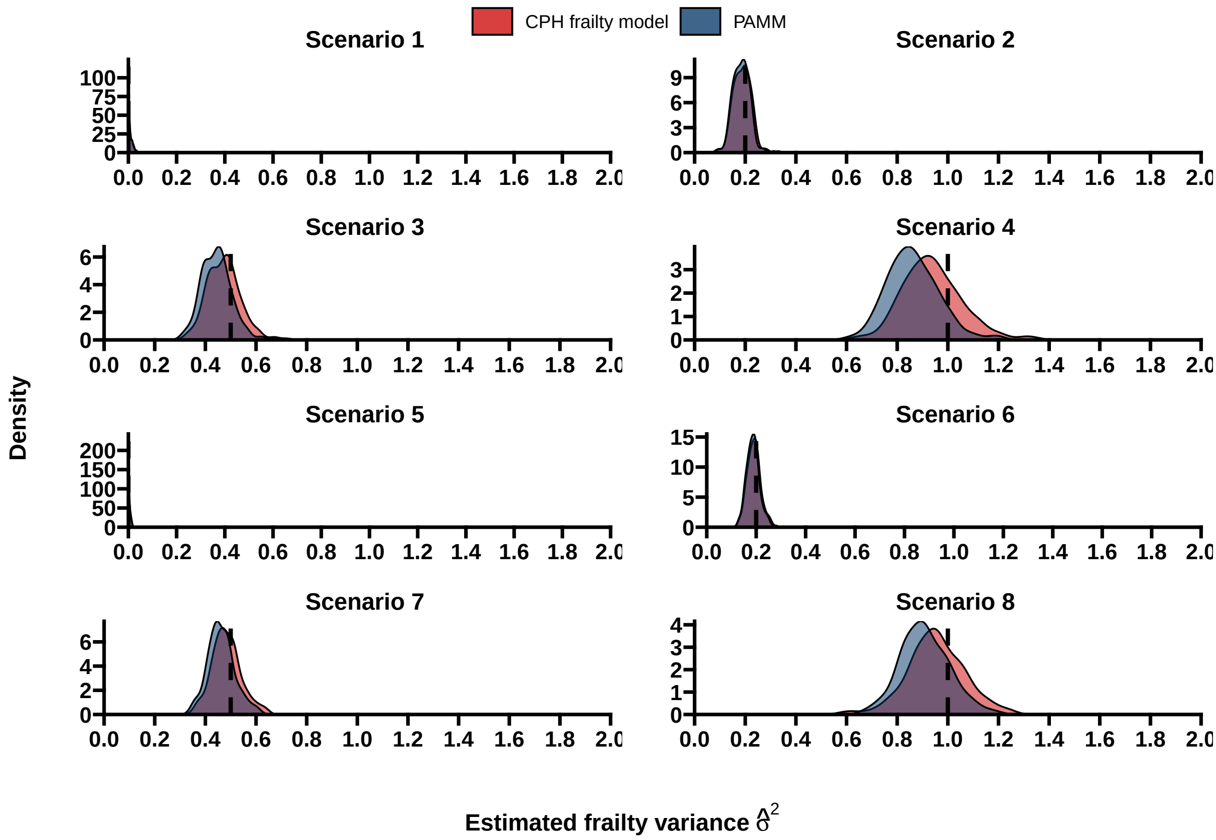

Figure 8: Smoothed histograms of the estimated frailty variances from the unstratified CPH frailty model and the unstratified PAMM, both in the calendar timescale, for the eight different scenarios with sample size N=200

References

[1] Box-Steffensmeier JM, De Boef S. Repeated events survival models: the conditional frailty model. Statistics in medicine. 2006;25(20):3518–3533.

[2] Pénichoux J, Moreau T, Latouche A. Simulating recurrent events that mimic actual data: a review of the literature with emphasis on event-dependence. arXiv preprint arXiv:150305798. 2015;.

[3] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2020. Available from: `https://www.R-project.org/`.

[4] RStudio Team. RStudio: Integrated Development Environment for R. Boston, MA; 2020. Available from: `http://www.rstudio.com/`.

[5] Therneau TM. coxme: Mixed Effects Cox Models; 2020. R package version 2.2-16. Available from: `https://CRAN.R-project.org/package=coxme`.

[6] Wood SN. Thin-plate regression splines. Journal of the Royal Statistical Society (B). 2003;65(1):95–114.

[7] Wood SN. Stable and efficient multiple smoothing parameter estimation for generalized additive models. Journal of the American Statistical Association. 2004;99(467):673–686.

[8] Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B). 2011;73(1):3–36.

[9] Wood SN, Scheipl F, Faraway JJ. Straightforward intermediate rank tensor product smoothing in mixed models. Statistics and Computing. 2013;23(3):341–360.

[10] Wood SN, N , Pya, Säfken B. Smoothing parameter and model selection for general smooth models (with discussion). Journal of the American Statistical Association. 2016;111:1548–1575.

[11] Wood SN. Generalized additive models: an introduction with R. CRC press; 2017.

[12] Bender A, Groll A, Scheipl F. A generalized additive model approach to time-to-event analysis. Statistical Modelling. 2018;18(3-4):299–321.

[13] Bender A, Scheipl F. pammtools: Piece-wise exponential Additive Mixed Modeling tools. arXiv:1806.01042 [stat]; 2018.

[14] Bender A, Rügamer D, Scheipl F, Bischl B. A General Machine Learning Framework for Survival Analysis. In: Machine Learning and Knowledge Discovery in Databases. Springer International Publishing; 2021. p. 158–173.

[15] Ramjith J, Bender A, Roes KCB, Jonker MA. A simulation study for comparison of piece-wise exponential additive mixed-effects models with Cox proportional hazards shared frailty models; 2021. Available from: doi.org/10.6084/m9.figshare.14638353.v1.

[16] Wang W, Fu H, Yan J. reda: Recurrent Event Data Analysis; 2019. R package version 0.5.2. Available from: https://github.com/wenjie2wang/reda.