

A Comparison of Estimation Methods for Shared Gamma Frailty Models[†]

Tingxuan Wu^{1,2}, Cindy Feng³ and Longhai Li^{1*}

^{1*}Department of Mathematics and Statistics, University of Saskatchewan, 106 Wiggins Rd, Saskatoon, S7N 5E6, SK, Canada.

²School of Public Health, University of Saskatchewan, 104 Clinic Place, Saskatoon, S7N 5E5, SK, Canada.

³Department of Community Health and Epidemiology, Dalhousie University, 5790 University Ave., Halifax, B3H 1V7, NS, Canada.

*Corresponding author(s). E-mail(s): longhai.li@usask.ca;

Abstract

This paper compares six different estimation methods for shared frailty models via a series of simulation studies. A shared frailty model is a survival model that incorporates random effects, where the frailties are common or shared among individuals within specific groups. Several estimation methods are available for fitting shared frailty models, such as penalized partial likelihood (PPL), expectation-maximization (EM), pseudo full likelihood (PFL), hierarchical likelihood (HL), maximum marginal likelihood (MML), and maximization penalized likelihood (MPL) algorithms. These estimation methods are implemented in various R packages, providing researchers with various options for analyzing clustered survival data using shared frailty models. However, there is a limited amount of research comparing the performance of these estimation methods. Consequently, it can be challenging for users to determine the most appropriate method for analyzing clustered survival data. To address this gap, this paper aims to conduct a series of simulation studies to compare the performance of different estimation methods implemented in R packages. We will evaluate several key aspects, including the performance of parameter estimators, rate of convergence, and computational time. Through this systematic evaluation, our goal is to provide a comprehensive understanding of the advantages and limitations associated with each estimation method.

[†]Published version: <https://doi.org/10.1007/s12561-024-09444-7>

Keywords: shared frailty models, random effects models, survival analysis, unobserved heterogeneity

1 Introduction

In survival analysis, conventional Cox proportional hazards models [1] and accelerated failure time models [2] assume that subjects are independent of one another. However, many research problems involve data with a multilevel structure, such as biomedical data or data on genetically related individuals, which exhibit correlation [3, 4]. The hazard of events differs from one cluster to another induced by unobserved cluster-level factors [4, 5]. Random effects can be incorporated into conventional survival models to account for cluster-level heterogeneity [6]. Such heterogeneity is often called frailty [6, 7] in the context of survival analysis. Frailty models extend the classic survival models by incorporating random effects (frailties) acting multiplicatively on the baseline hazard function [5, 6]. In cases where the frailty is greater than one, subjects experience an increased failure hazard. A shared frailty model is a model where the frailties are common or shared among individuals within a cluster or group [5, 7–10].

Various estimation methods have been developed for fitting shared frailty models. These methods include penalized partial likelihood (PPL) [11–13], expectation-maximization (EM) [14, 15], pseudo full likelihood (PFL) [16, 17], hierarchical likelihood (HL) [18], maximum marginal likelihood (MML) [19, 20], and maximization penalized likelihood (MPL) [21, 22] algorithms. These estimation methods have been implemented in various R packages, providing researchers with various options for fitting shared frailty models. The most widely used package for fitting shared frailty models is the `survival` [23] package, which estimates the parameters by maximizing the penalized partial likelihood. The `frailtyEM` [24] package implements the general expectation-maximization (EM) algorithm, considering the frailty as a latent variable. The `frailtySurv` [25] package adopts a pseudo full likelihood approach, and the `frailtyHL` [26] package estimates the parameters using a hierarchical-likelihood approach. The `survival` [23], `frailtyEM` [24], `frailtySurv` [25], and `frailtyHL` [26] packages are commonly employed to implement semi-parametric survival models with frailties. For fitting parametric shared frailty models, the `parfm` [27] package can be used. It supports various distributions such as exponential, Weibull, inverse Weibull, Gompertz, lognormal, log-skewNormal, loglogistic, and others. The `parfm` package uses the maximum marginal likelihood (MML) approach. In addition, the `frailtypack` [28] package fits flexible parametric frailty models. It accommodates scenarios with shared frailty, nested frailty, joint frailty and additive frailty. The `frailtypack` package is based on the maximization of the penalized log-likelihood.

Despite the wide range of estimation methods available in R packages for fitting shared frailty models, it remains unclear if these methods have similar

or different performances in terms of precision and efficiency of the parameter estimators, computational speed and convergence rate. Previous research [29] compared three estimation methods for fitting shared frailty models through simulation studies. However, with the development of new estimation methods in recent years, an updated comparison is warranted. This study aims to fill this gap by providing a general overview of estimation methods for fitting shared frailty models and comparing their performances through simulation studies. Our simulation studies demonstrated that all the estimation methods implemented in the six considered packages for fitting shared frailty models yielded very similar and unbiased parameter estimates for the fixed-effect regression coefficients, regardless of sample size, cluster sizes and censoring rates. However, differences were observed in the performance of estimating the frailty variance, convergence rate, and computational time. Furthermore, statistical inference for the frailty variance is not straightforward. Some packages do not provide an estimate of the standard error of the estimated frailty variance. Most packages assume that the distribution of the estimated frailty variance is approximately normally distributed. However, we observed that this distribution is highly skewed, hence, a symmetric confidence interval for the frailty variance is not ideal. To address this issue, we developed an asymmetric confidence interval by accounting for this skewness and demonstrated its superior performance compared to the conventional confidence interval provided by most packages.

The remaining sections of the article are structured as follows. Section 2 briefly reviews shared frailty models. Section 3 introduces the various estimation methods and the corresponding packages. Section 4 presents the design and results of the simulation study for comparing the performance of the estimation methods in the packages. Finally, the paper concludes with a discussion of the advantages and limitations of each estimation method in the packages for fitting a shared frailty model in Section 5. Additionally, recommendations for selecting a estimation method for fitting shared frailty models are provided in Section 5.

2 Shared Frailty Models

A shared frailty model is one where the frailties are common or shared among individuals within groups [3, 5, 10]. The formulation of a frailty model for clustered failure survival data is defined as follows. Suppose there are g groups of individuals with n_i individuals in the i th group, $i = 1, 2, \dots, g$. If the number of subjects n_i is 1 for all groups, then the univariate frailty model is obtained [3, 9]. Otherwise, the model is called the shared frailty model [5, 7, 10, 30, 31] because all subjects in the same cluster share the same frailty value z_i . Suppose t_{ij} is the true failure time for the j th individual of the i th group, which we assume to be a continuous random variable in this article, where $j = 1, 2, \dots, n_i$. In many practical problems, we may not be able to observe t_{ij} exactly, but we can observe that it is greater than a value c_{ij} ,

where c_{ij} is the corresponding censoring time. The observed failure times are denoted by the pair (y_{ij}, δ_{ij}) , where $y_{ij} = \min(t_{ij}, c_{ij})$, $\delta_{ij} = I(t_{ij} < c_{ij})$. The observed data can be written as $y = (y_{11}, \dots, y_{gn_g})$ and $\delta = (\delta_{11}, \dots, \delta_{gn_g})$. This is called right-censoring. Since we only consider right-censoring in this article, we will use “censoring” as shorthand for “right-censoring”.

For the shared frailty models [5, 10], the hazard of an event at time t for the j th individual in the i th group, is given by

$$h_{ij}(t) = z_i \exp(\beta^T x_{ij}) h_0(t); \quad (1)$$

and the survival function [5, 10] for the j th individual of the i th group at time t follows:

$$S_{ij}(t) = \exp \left\{ - \int_0^t h_{ij}(\tau) d\tau \right\} = \exp \left\{ - z_i \exp(\beta^T x_{ij}) H_0(t) \right\}, \quad (2)$$

where x_{ij} is a vector of values of p explanatory variables for the j th individual in the i th group, β is the vector of regression coefficients; $h_0(t)$ is the baseline hazard function, $H_0(t)$ is the baseline cumulative hazard function, and z_i is the frailty term that is common to all n_i individuals within the i th group, let $z = (z_1, \dots, z_g)$. It is more convenient to work with an alternative representation of the frailty effect by setting $u_i = \log(z_i)$ as a random effect in the linear component of the proportional hazards model [5], because of that z_i cannot be negative, but u_i can be any real value. If all u_i 's are equal to zero, then correspondingly all z_i 's are equal to one. In such cases, the model does not have frailty. The form of the baseline hazard function is assumed to be unspecified as a semi-parametric model or fully specified to follow a parametric distribution.

Our study primarily focuses on the shared gamma frailty model [5], as the gamma distribution is commonly used for modelling the frailty effect. The gamma distribution [5, 32] makes it easy to obtain a closed-form representation of the observable survival, cumulative density, and hazard functions due to the simplicity of the Laplace transform. All the packages considered in this paper use a one-parameter gamma distribution with a shape parameter equal to $1/\theta$ and a scale parameter equal to θ . This ensures that the distribution for the frailty z_i has a mean of one [5, 32] and a variance of θ . We will call θ by frailty variance throughout this paper. The frailty z_i has a multiplicative effect on the baseline hazard function. Therefore, using a gamma frailty distribution with a mean equal to one does not change the overall hazard rate of events. When we parameterize the frailty z_i with the random effect $u_i = \log(z_i)$, setting the random effect to have a mean equal to zero is equivalent to setting the mean of the frailty z_i approximately equal to one.

3 Estimation and Inference for Shared Frailty Models

In this section, we provide a brief review of six distinct estimation methods utilized for fitting shared frailty models. These estimation methods vary significantly in terms of the employed likelihoods, the methods for estimating the baseline hazard functions, and the methods for handling the frailty term.

3.1 Penalized Partial Likelihood (PPL) Algorithm (**R** package: **survival**)

The most popular package for fitting semiparametric shared frailty models is the **survival** package [23]. The function `coxph` in the **survival** package offers a way of fitting shared frailty models via the penalized partial likelihood (PPL) method. The key arguments are the terms including fixed effects of the model, random effects and the data. The frailty distribution can be gamma, Gaussian, or t distribution. It accommodates the clustered failures and recurrent events data with right, left, and interval censoring types. When the `coxph` function fits shared frailty models with clustered failures data, cluster size should be above five. Otherwise, the random effects will be treated as fixed effects [23]. The PPL algorithm will be described as follows.

The PPL [7, 11–13, 33] approach can be used to estimate parameters in a shared frailty model. This estimation is based on maximizing the penalized partial log-likelihood, which consists of two parts. The first part is the conditional likelihood of the data given the frailties. The second part corresponds to the frailties distribution in which the likelihood is considered a penalty term. The PPL [7, 11] for the shared frailty model is then given by

$$l_{ppl}(\beta, u, \theta; y, \delta) = l_{part}(\beta, u; y, \delta) + l_{pen}(\theta; u), \quad (3)$$

where $u = (u_1, \dots, u_g)$ is the vector of random effects and β is the vector of regression coefficients. Here $l_{part}(\beta, u; y, \delta)$ [7, 11] is the partial log-likelihood for the Cox model that includes the random effects, as follows:

$$l_{part}(\beta, u; y, \delta) = \sum_{i=1}^g \sum_{j=1}^{n_i} \delta_{ij} \left\{ \eta_{ij} - \log \left[\sum_{(q,w) \in R(y_{ij})} \exp(\eta_{qw}) \right] \right\}, \quad (4)$$

where $\eta_{ij} = \beta^T x_{ij} + u_i$, $\eta = (\eta_{11}, \dots, \eta_{gn_g})$, and $y_{(1)} < \dots < y_{(r)}$ are the ordered distinct event time among the y_{ij} 's. $R(y_{(v)}) = \{(i, j) : y_{ij} \geq y_{(v)}\}$ is the risk set at $y_{(v)}$. In the penalty function $l_{pen}(\theta; u)$, the random effect u_i is equal to $\log(z_i)$, where z_i usually follows either a lognormal or a gamma distribution. The penalty function [7, 11] can be written as,

$$l_{pen}(\theta; u) = \sum_{i=1}^g \log f_U(u_i | \theta), \quad (5)$$

where $f_U(u_i | \theta)$ denotes the density function of the random effect u_i given θ , which is the variance of $z_i = \exp(u_i)$.

The maximization of the PPL consists of an inner and an outer loop[7, 11]. For the gamma frailties (Gaussian frailties as well), the penalized likelihood in the inner loop can be maximized with the Newton-Raphson method. The maximization process proceeds iteratively by starting with a provisional θ and finding the estimates of the β 's and the u 's that maximize $l_{ppl}(\beta, u, \theta)$. In the outer loop, it is based on maximizing a profiled version of the marginal likelihood for θ . Given a specific value of θ , the estimates for β and u are determined as the values that maximize the likelihood function $l_{ppl}(\beta, u, \theta)$ with respect to θ . Using these estimated parameters, we can calculate the estimates for the baseline hazard function.

The partial likelihood is not a true likelihood in general. Therefore, the maximizer of the penalized partial likelihood is independent of the baseline hazard function. The Breslow approximation is the first option to estimate the baseline hazard function in nearly all the packages for fitting Cox regression models with or without frailties. The Breslow estimator [34] is a nonparametric maximum likelihood estimator for the cumulative baseline hazard function. It has been implemented in all major statistical software packages. The baseline cumulative hazard function is $H_0(t) = \int_0^t h_0(s) ds$. Breslow (1972) suggested estimating the cumulative baseline hazard via maximizing likelihood function. After getting the estimates $\hat{\beta}$ and \hat{u}_i , it can provide the nonparametric maximum likelihood estimate of $\hat{H}_0(t)$ [34]:

$$\hat{H}_0(t) = \sum_{\{v: y_{(v)} \leq t\}} \left\{ \frac{d_{(v)}}{\sum_{(i,j) \in R(y_{(v)})} \exp(\hat{\beta}^T x_{ij} + \hat{u}_i)} \right\}, \quad (6)$$

where $d_{(v)}$ is the number of events at $y_{(v)}$.

3.2 Expectation-maximization (EM) Algorithm (R package: frailtyEM)

The **frailtyEM** package was written by Balan, T.A. and Putter, H. [24]. It provides maximum likelihood estimation of semiparametric shared frailty models using the expectation-maximization (EM) algorithm. The main model fitting function in **frailtyEM** is **emfrail**, and the user has to define the main arguments formula, data set, distribution and control. This formulation is common to most survival analysis packages, allowing for several scenarios, including possibly left truncated clustered failures and recurrent events in both calendar time and gap time formulation. The distribution argument determines the frailty distribution; the gamma, stable and power variance function family distributions are supported. The package can access predicted survival and cumulative hazard curves, both for an individual and on a population level. As can be seen in our simulation study of section 4, results from the **frailtyEM**

package are very close to the `survival` package. The description of the EM algorithm is outlined below.

The EM algorithm [7, 14] is an iterative method for performing maximum likelihood estimation when the model involves latent variables (missing values). The expectation (E) step attempts to evaluate conditional expectations of functions of the latent variables in the log-likelihood function given the observed data and the estimates of the parameters in the previous step. In the maximization (M) step, we treat these expected values as true information and new estimates of the parameters of interest are obtained by maximizing the likelihood given the expected values. We will discuss the procedure in details below.

We first consider the complete data log-likelihood [7, 14] in which the frailties z_i are regarded as another set of parameters:

$$l_{full}(\theta, \beta, z) = \log f(y, \delta, z \mid \hat{h}_0, \beta, \theta) = l_{full,1}(\beta; y, \delta, \hat{h}_0) + l_{full,2}(\theta; z), \quad (7)$$

where,

$$l_{full,1}(\beta; y, \delta, \hat{h}_0) = \sum_{i=1}^g \sum_{j=1}^{n_i} \left\{ \delta_{ij} \log \left[\hat{h}_0(y_{ij}) z_i \exp(\beta^T x_{ij}) \right] - \hat{H}_0(y_{ij}) z_i \exp(\beta^T x_{ij}) \right\} \quad (8)$$

and

$$l_{full,2}(\theta; z) = \sum_{i=1}^g \log f_Z(z_i \mid \theta). \quad (9)$$

We use $l_{full,1}(\beta; y, \delta, \hat{h}_0)$ to estimate β , and $l_{full,2}(\theta; z)$ to estimate θ . The Breslow estimator in equation (6) is used to estimate the baseline hazard function \hat{h}_0 , which is required in the expectation step. Within the framework of the EM algorithm, the expected value of the full log-likelihood needs to be maximized [7, 14]. In the E step, the “posterior” distribution of the frailties $p(z_i \mid y_i, \delta_i, \beta^{(k-1)}, \theta^{(k-1)})$ can be obtained. Then, the $E^{(k)}(z_i)$ and $E^{(k)}(\log z_i)$ can be calculated. In the M step, the log-likelihood in (7) is profiled to a partial log-likelihood by considering the frailties as fixed offset terms, then the $E^{(k)}(z_i)$ and $E^{(k)}(\log z_i)$ are considered to be the true value to replace the z_i ’s and $\log z_i$ ’s in the partial log-likelihood [7, 14] leading to

$$l_{part,1}^{(k)}(\beta) = \sum_{i=1}^g \sum_{j=1}^{n_i} \delta_{ij} \left\{ E^{(k)}(\log z_i) + \beta^T x_{ij} - \log \left(\sum_{(q,w) \in R(y_{ij})} E^{(k)}(z_q) \exp(x_{qw} \beta) \right) \right\}. \quad (10)$$

The new estimates $\beta^{(k)}$ can be obtained from the $l_{part,1}(\beta)$. A new estimate $\theta^{(k)}$ can be obtained immediately by maximization of $l_{full,2}(\theta; u)$, replacing z_i ’s and $\log z_i$ ’s in (9) by the current expected values at iteration step k . In the initialization E step, $\theta^{(0)}$ is set to one and an ordinary Cox model is fitted leading to estimates $\beta^{(0)}$. Next, we iterate between the expectation and

maximization steps until convergence. The marginal log-likelihood can be used for assessing the convergence of the algorithm

3.3 Maximum Marginal Likelihood (MML) Algorithm (R package: `parfm`)

The `parfm` package [27] estimates the parameters for parametric frailty models by maximizing the marginal log-likelihood (MML). The baseline hazard distributions can be exponential, Weibull, inverse Weibull (Frechet), Gompertz, lognormal, log-kewNormal, and loglogistic. The frailty distribution can be gamma, positive stable, inverse Gaussian, and lognormal distribution. The following will detail the MML algorithm.

The MML approach was proposed for estimating the parameters for shared frailty models [19, 20]. The frailties are integrated out by averaging the conditional likelihood with respect to the frailty distribution. This method can be applied to any frailty distribution with explicit Laplace transform.

For the right-censored clustered survival data, the observation for the j th individual in the i th group is the triplet $(y_{ij}, \delta_{ij}, x_{ij})$. Let ψ represent a vector of parameters for the baseline hazard function. The marginal log-likelihood [19, 20] can be written as

$$l_{\text{marg}}(\psi, \beta, \theta; y, \delta, x) = \sum_{i=1}^g \left\{ \left[\sum_{j=1}^{n_i} \delta_{ij} (\log(h_0(y_{ij} | \psi)) + \beta^T x_{ij}) \right] + \log \left[(-1)^{d_i} \mathcal{L}^{d_i} \left(\sum_{j=1}^{n_i} H_0(y_{ij} | \psi) \exp(\beta^T x_{ij}) \right) \right] \right\}, \quad (11)$$

where θ is used as the vector of parameters for the frailty distribution function, $d_i = \sum_{j=1}^{n_i} \delta_{ij}$ the number of events in the i -th cluster, and $\mathcal{L}^q(\cdot)$ is the q th derivative of the Laplace transform of the frailty distribution [19, 20], which is defined as,

$$\mathcal{L}(s) = \int_0^\infty \exp(-sz) f(z) dz, \quad (12)$$

where $f(z)$ is the density function of the frailty term z . If the higher-order derivatives $\mathcal{L}^q(\cdot)$ of the Laplace transform up to $q = \max\{d_1, \dots, d_g\}$ are computable, the estimates of ψ , β , θ , can be obtained by maximising the marginal log-likelihood (11). The parametric estimation approach is applicable for modelling the form of the baseline hazard.

3.4 Hierarchical Likelihood (HL) Algorithm (R package: `frailtyHL`)

The `frailtyHL` package created by Ha et al. [26] implements the hierarchical-likelihood (HL) procedures for fitting semi-parametric frailty models with

non-parametric baseline hazards. The package fits shared or multilevel frailty models for correlated survival data. The lognormal or gamma distributions can be adopted as the frailty distribution, corresponding to the normal or log-gamma distributions for the log frailties. The results of estimates of fixed effects, random effects, and variance components as well as their standard errors are provided. In addition, it provides a statistical test for the variance components of frailties and three AIC criteria for the model selection. However, the package does not provide the interval estimation of frailty.

Lee, Y. & Nelder, J.A. [35] proposed the use of hierarchical likelihood for fitting the model with random effects. The hierarchical likelihood consists of data, parameters and unobserved random effects. The Laplace approximation is used to calculate the integration over the random-effect distribution. Thus, the h-likelihood can be used directly for inference on random effects.

For the observe y_{ij} and the censoring indicator is δ_{ij} , the h-likelihood [35] for a frailty model is defined by

$$hl(\beta, \theta, u; y, \delta, \hat{h}_0) = l_0(\beta; y, \delta, z, \hat{h}_0) + l_1(\theta; u), \quad (13)$$

where l_0 is the sum of conditional log densities for (y, δ) given the random effect $u = (\log z_1, \dots, \log z_g)$; then it follows:

$$\begin{aligned} l_0(\beta; y, \delta, u, \hat{h}_0) &= \sum_{ij} \log f(y_{ij}, \delta_{ij} \mid \beta, u_i, \hat{h}_0) \\ &= \sum_{ij} \delta_{ij} \left\{ \log \hat{h}_0(y_{ij}) + (\beta^T x_{ij} + u_i) \right\} \\ &\quad - \sum_{ij} \left\{ \hat{H}_0(y_{ij}) \exp(\beta^T x_{ij} + u_i) \right\} \end{aligned} \quad (14)$$

l_1 is the sum of log densities for random effects u with parameter θ , which is defined by

$$l_1(\theta; u) = \sum_i \log f_U(u_i \mid \theta). \quad (15)$$

The Breslow estimator in equation (6) is employed to estimate the baseline hazard function \hat{h}_0 . From the equation (3), the penalized partial likelihood [7, 11] is defined as,

$$\begin{aligned} l_{ppt}(\beta, u, \theta; y, \delta) &= \sum_{ij} \delta_{ij} \left\{ (\beta^T x_{ij} + u_i) - \log \left[\sum_{(q,w) \in R(y_{ij})} \exp(x_{qw}\beta + u_q) \right] \right\} \\ &\quad + \sum_i \log f_U(u_i \mid \theta). \end{aligned} \quad (16)$$

The papers [18, 36] showed that $hl(\theta, \beta)$ is equal to the $l_{ppt}(\beta, u, \theta; y, \delta)$ plus a constant,

$$hl(\beta, u, \theta) = l_{ppt}(\beta, u, \theta; y, \delta) + \sum_{(q,w) \in R(y_{ij})} d_{qw} \left\{ \log \hat{h}_0(y_{qw}) - 1 \right\}, \quad (17)$$

where $\sum_{(q,w) \in R(y_{ij})} d_{qw} \left\{ \log \hat{h}_0(y_{qw}) - 1 \right\}$ is a constant and d_{qw} is the number of element in the risk set $R(y_{qw})$. Accordingly, given the frailty variance θ , the hierarchical likelihood methods for estimating β can be obtained by maximizing the profile marginal likelihood after eliminating $H_0(t)$. The Laplace approximation can be used when the marginal likelihood is hard to obtain. Given $\hat{\beta}$ and \hat{u} , the maximum adjusted profile hierarchical likelihood for estimating the frailty variance θ can be obtained. We iterate these steps until convergence. The estimates of the standard errors can be computed [35].

3.5 Pseudo Full Likelihood (PFL) Algorithm (R package: frailtySurv)

The package `frailtySurv` [25] can be used for simulating and fitting semi-parametric shared frailty models. It can be applied to various frailty distributions, including gamma, log-normal, inverse Gaussian and power variance functions via the pseudo-full likelihood (PFL) approach. The estimators of the parameters are consistent and asymptotically normally distributed. Consequently, statistical inference, including hypothesis testing and constructing confidence intervals, can be conducted based on the normal distribution with the results from this package. Only right-censoring with clustered failures dataset is supported by `frailtySurv`. The PFL algorithm will be described as follows.

The PFL [16, 17] is a new method that can handle various parametric frailty distributions with finite moments. A simple univariate numerical integration can deal with non-conjugate frailty distributions. The cumulative hazard function is estimated via a non-iterative procedure. Other properties follow the consistency and asymptotic normality of the parameter estimators and a direct, consistent covariance estimator. It is easy to compute and implement. From the study of Gorfine et al. [17], the results for fitting the shared frailty model are very similar to the EM-based method.

In the shared frailty model, we further assume that the observed data consists of (y, δ, x) . The proposed approach can estimate the regression coefficient vector β , the frailty distribution's parameter θ , and the non-parametric cumulative baseline hazard H_0 . Let τ be the end of the observation period. The full

likelihood [16, 17] can be defined as

$$L(\beta, \theta, H_0) = \prod_{i=1}^g \prod_{j=1}^{n_i} \left\{ h_0(y_{ij}) \exp(\beta^T x_{ij}) \right\}^{\delta_{ij}} \prod_{i=1}^g (-1)^{N_{i\bullet}(\tau)} \mathcal{L}^{(N_{i\bullet})} \{H_{i\bullet}(\tau)\}, \quad (18)$$

where $N_{ij}(t) = \delta_{ij} I(y_{ij} \leq t)$, $N_{i\bullet}(t) = \sum_{j=1}^{n_i} N_{ij}(t)$, $H_{ij}(t) = H_0(\min\{y_{ij}, t\}) \exp(\beta^T x_{ij})$, $H_{i\bullet}(t) = \sum_{j=1}^{n_i} H_{ij}(t)$, \mathcal{L} is the Laplace transform of the frailty distribution and $\mathcal{L}^{(m)}$, $m = 1, 2, \dots$ are the m th derivatives of \mathcal{L} . Note that the m th derivatives of the Laplace transform evaluated at $H_{i\bullet}(\tau)$ equals to $(-1)^{N_{i\bullet}(\tau)} \int z^{N_{i\bullet}(\tau)} \exp\{-z H_{i\bullet}(\tau)\} f(z) dz$. The log-likelihood [16, 17] equals to

$$l(h_0, \theta, \beta) = \sum_{i=1}^g \sum_{j=1}^{n_i} \left\{ \delta_{ij} \log\{h_0(y_{ij}) \exp(\beta^T x_{ij})\} \right\} + \sum_{i=1}^g \log \mathcal{L}^{(N_{i\bullet})} \{H_{i\bullet}(\tau)\}. \quad (19)$$

An estimate of H_0 is required in the log-likelihood function to obtain estimates of β and θ . In the initialization step, a standard Cox model is fitted to obtain an initial estimate of β , then θ is initiated by solving the score equation under within-cluster independence or very weak dependency [25]. For given these two initial values, H_0 is estimated via the Breslow estimator with jumps at the ordered observed failure times τ_v , $v = 1, \dots, r$. The detailed step of estimating the baseline hazard function is referred to Gorfine et al. [17]. Then, \hat{H}_0 is substituted into the log-likelihood function. The estimators of $\hat{\beta}$ and $\hat{\theta}$ can be obtained by maximizing the log-likelihood function. Iterate these steps until convergence.

3.6 Maximization Penalized Likelihood (MPL)

Algorithm (R package: frailtypack)

The **frailtypack** package [28] allows fitting Cox models and four types of frailty models (shared, nested, joint, additive). The function **frailtyPenal** fits the shared frailty model by using the maximum penalized likelihood (MPL) method with the splines to estimate the baseline hazard. As a result, due to the use of splines with a specified number of knots for modelling the baseline hazard, this approach can be considered a parametric model. According to the reference manual, the baseline hazard can also be modelled using a piecewise constant function or a Weibull function. However, it is important to note that the default function for the baseline hazard in the **frailtyPenal** is splines. Right-censored or left-truncated data can be fitted in this package. The arguments are the terms including the fixed effect, the cluster variable, and the data set. In addition, there are two arguments in the formula that need to be specified: **n.knots** (4 up to 20) and **kappal** (smoothing parameter). In our simulation, we employed the splines function with 15 knots and set the value of **kappal** to

1. Additionally, we utilized the argument cross-validation, a logical value that must be set to ‘true’. The description of the MPL algorithm is as follows.

The MPL estimation [21, 22] can be applied to the nonparametric estimation of a continuous hazard function in a shared frailty model. This approach is based on the penalized full likelihood, which is opposed to the penalized partial likelihood. We assume that the frailty effects are distributed from a gamma distribution with mean 1 and variance θ . For the observe y , δ , and the truncation times τ , the full marginal log-likelihood for the shared gamma frailty model has an analytical formulation [15]

$$\begin{aligned} l(\beta, \theta, h_0) = \sum_{i=1}^g \left\{ \left[\sum_{j=1}^{n_i} \delta_{ij} \log h_0(y_{ij}) \right] - \left(\frac{1}{\theta} + m_i \right) \log \left[1 + \theta \sum_{j=1}^{n_i} H_0(y_{ij}) \right] \right. \\ \left. + \frac{1}{\theta} \log \left[1 + \theta \sum_{j=1}^{n_i} H_0(\tau_{ij}) \right] + I(m_i \neq 0) \sum_{k=1}^{m_i} \log \left(1 + \theta(m_i - k) \right) \right\} \end{aligned} \quad (20)$$

where the number of recurrent events is $m_i = \sum_{j=1}^{n_i} \delta_{ij}$.

The penalized log-likelihood function for the shared gamma frailty model [21, 22] follows

$$pl(\beta, \theta, h_0) = l(\beta, \theta, h_0) - k \int_0^\infty h_0''(t)^2 dt \quad (21)$$

where k is a positive smoothing parameter that controls the trade-off between the data fit and the smoothness of the functions. The smoothing parameter needs to be a fixed value, and the estimators of β and θ can be obtained via the maximization of the penalized likelihood. The robust Marquardt algorithm [37] is used to estimate parameters, which is a combination of a Newton Raphson algorithm and the steepest descent algorithm. The estimator of the baseline hazard function $h_0(\cdot)$ can be approximated based on Cubic M-splines with Q knots [38]. The splines, the regression coefficients, and the variance of the frailty term are initialized to 0.1 in the shared frailty model. The model can be fit firstly, then adjusted Cox model to give new initial values for the splines and the regression coefficients.

Table 1 presents a concise overview of the six estimation methods that we introduce above, specifically focusing on their distinctions in terms of the primary R function, fitting algorithm, likelihood, the form of baseline hazard, and the method for handling frailty terms, frailty distribution, censoring type and data type. This table provides a clear visualization of the similarities and differences among the various methods, allowing for comparisons and assessments.

Table 1 An overview of six estimation methods for fitting shared frailty models in terms of the primary R function, fitting algorithm, likelihood, the form of baseline hazard, and the method for handling frailty terms, frailty distribution, censoring type and data type. PPL = penalized partial likelihood, MML = maximum marginal likelihood, EM = expectation maximization, PFL = pseudo full likelihood, HL = h-likelihood, MPL = maximization penalized log-likelihood.

Package	Function	Algorithm	Likelihood	Form of H_0	Handling z	Frailty distribution	Censoring	Data
survival	coxph	PPL	Partial likelihood	Step function	Penalization	Gamma, Log-normal, t	Right, interval, Left	Clustered failures, Recurrent events
parfm	parfm	MML	Parametric full likelihood	Parametric distributions	Laplace transform	Gamma, Log-normal, Positive Stable, Inverse Gaussian	Right	Clustered failures, Left truncation
frailtyEM	emfrail	EM	Partial likelihood	Step function	EM Algorithm	Gamma, Positive Stable, Inverse Gaussian, Compound Poisson, Power Variance Function	Right	Clustered failures, Recurrent events, Left truncation
frailtySurv	fitfrail	PFL	Full likelihood	Step function	Laplace transform	Gamma, Log-normal, Inverse Gaussian, Power Variance Function	Right	Clustered failures
frailtyHL	frailtyHL	HL	Partial likelihood	Step function	Laplace approximation	Gamma, Log-normal	Right	Clustered failures
frailtypack	frailtyPenal	MPL	Full likelihood	Spline	Integration	Gamma, Log-normal	Right	Clustered failures, Recurrent events, Left truncation, Correlated structure

4 Simulations and Results

We conducted simulation studies to investigate the performances of the estimation methods implemented in packages for fitting the shared frailty models. We generated the true failure time from a Weibull regression model with shape parameter ($\alpha = 3$) and scale parameter ($\lambda = 0.007$) [29]. More specifically $t_{ij} = \{-\log(\kappa_{ij})/[\lambda \exp(\beta_1 x_{ij}^{(1)} + \beta_2 x_{ij}^{(2)} + \beta_3 x_{ij}^{(3)} + z_i)]\}^{(1/\alpha)}$, where $i = \{1, \dots, g\}$, $j = \{1, \dots, n_i\}$ and κ_{ij} was simulated from Uniform[0, 1]. The censoring time C_i was simulated from an exponential distribution, $\exp(\eta)$, where η was set to obtain three different censoring rates (c): 20%, 50%, and 80%, respectively. Three covariates were generated including $x_{ij}^{(1)}$ from a Uniform[0, 1], $x_{ij}^{(2)}$ from a Normal(0, 1), and $x_{ij}^{(3)}$ from a Bern(0.25). We set true regression parameters for the three covariates as $\beta_1 = 1$, $\beta_2 = -1$, $\beta_3 = 0.5$, respectively. The frailty term was generated from a gamma distribution with a variance of 0.5. All the parameter settings are consistent with the previous study[29]. We considered fitting a shared frailty gamma model assuming $h_{ij}(t_{ij}) = z_i \exp(\beta_1 x_{ij}^{(1)} + \beta_2 x_{ij}^{(2)} + \beta_3 x_{ij}^{(3)}) h_0(t_{ij})$ as a true model. Via investigating if the performances of estimation methods depend on sample size, we simulated datasets with varying sample sizes n ranging from 100 to 800. For a sample of size 100, the observations were grouped into 10 clusters of 10 observations. For a sample of size 400, the observations were grouped into 10 clusters of 40 observations or 40 clusters of 10 observations. For a sample of size 800, the observations were grouped into 10 clusters of 80 observations or 80 clusters of 10 observations. All considered estimation methods available in R packages were applied to the same simulated dataset in each scenario. Using 1000 datasets generated under each scenario, we examined the precision

of the parameter estimates in terms of bias and standard errors of the estimated parameters, as well as the coverage probability (CP) of the estimated parameters. We also investigated the performance of the packages in terms of convergence rate and average computing time under each simulation scenario. We will use the package's name to represent each estimation method in our comparative study. The simulation studies were conducted with equal cluster sizes. The results for the simulation with slightly unequal cluster sizes are similar to those with equal cluster sizes.

4.1 Performance of Parameter Estimation

Figure 1 presents the boxplots of the estimated regression coefficients and the frailty variances over 1000 repeated samples when the sample size n was 100 and 400. The results indicate that all the packages considered in this paper performed similarly in estimating the regression coefficients. However, it should be noted that **frailtypack** slightly overestimated β_2 for the sample size of 100. Not surprisingly, as the censoring rate increases, the estimated regression coefficients are subject to more variability.

The frailty variance was underestimated with mostly negative biases by almost all the packages considered in this paper. The underestimation is shown by the boxplots in the bottom two rows of Figure 1, and the values of biases are given in Tables S1 - S6 in the Supplementary Materials. This underestimation was more prominent when there were 10 clusters of 10 observations and 10 clusters of 40 observations, compared to 40 clusters of 10 observations. The underestimation was even observed in the settings without censoring. The maximum likelihood variance estimator in linear mixed models tends to underestimate the true variance. This discrepancy arises because high values of random effects are penalized. The underestimation of the frailty variance in the shared frailty model may share a similar reason with the underestimation observed in fitting linear mixed models. The underestimation might be one of the reasons that the **survival** package does not provide a standard error for the estimated frailty variance. Figure 1 also suggests that a lower number of clusters results in higher variability in estimating the frailty variance.

Most importantly, Figure 1 shows that the distribution of the estimated frailty variance is highly positively-skewed, which was previously reported in [39, 40]. To enhance the visualization of the estimated frailty variance, the boxplots of the $\log(\hat{\theta})$ are also presented in the bottom row of Figure 1. These plots show that the **frailtypack** package sometimes gives extremely small estimates for the frailty variance. The skewness is particularly severe when the sample size is small such as 100 or when the censoring rate is high such as 80%. This skewness explains the poor coverage of confidence intervals for θ , which we will show details in Section 4.2. Due to this observation, we were inspired to construct a confidence interval for θ based on the distribution of $\log(\hat{\theta})$ rather than that of $\hat{\theta}$, as given below in equation (22).

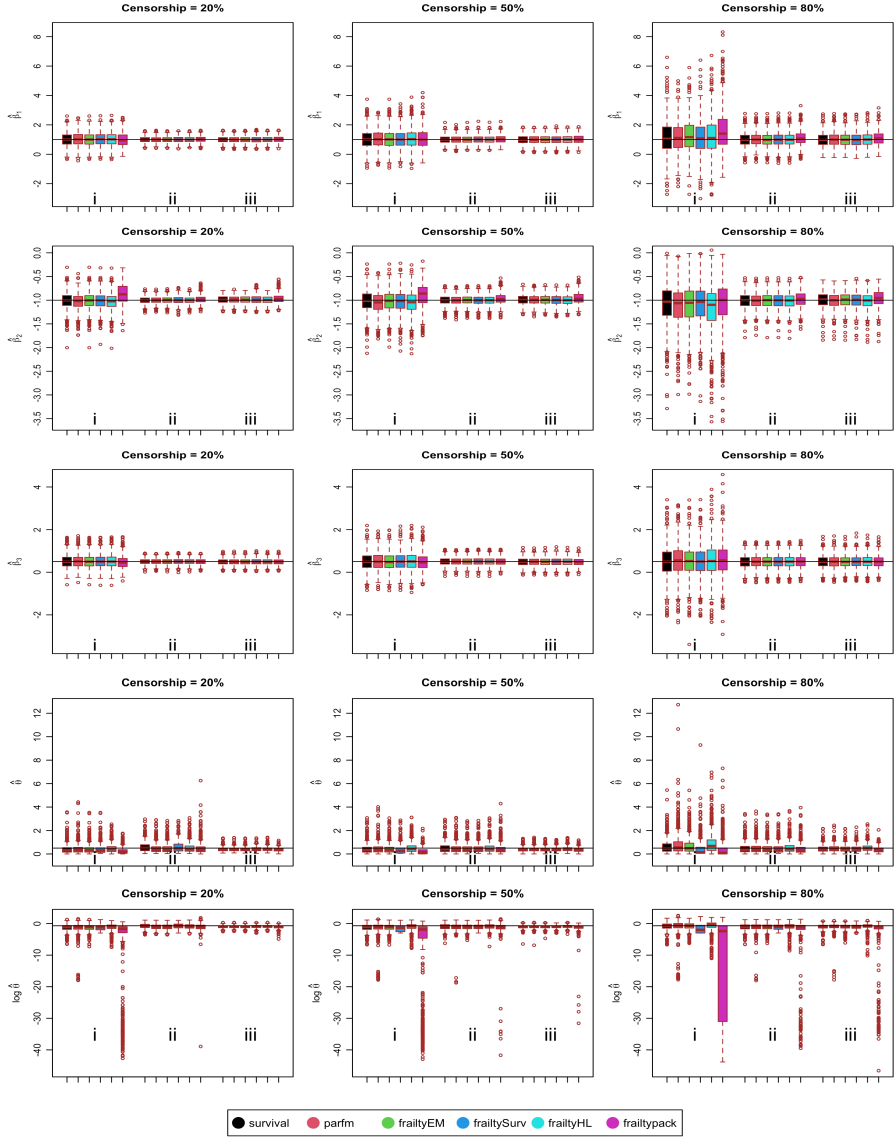


Fig. 1 Boxplots of estimated regression coefficients and frailty variances over 1000 samples simulated from the true model. True values of the parameters are indicated as horizontal lines. The first, second and third columns correspond to censoring rates of 20%, 50% and 80%, respectively. In each panel, the left group corresponds to scenario (i) with 10 clusters of 10 observations; the middle group corresponds to scenario (ii) with 10 clusters of 40 observations; the right group corresponds to scenario (iii) with 40 clusters of 10 observations.

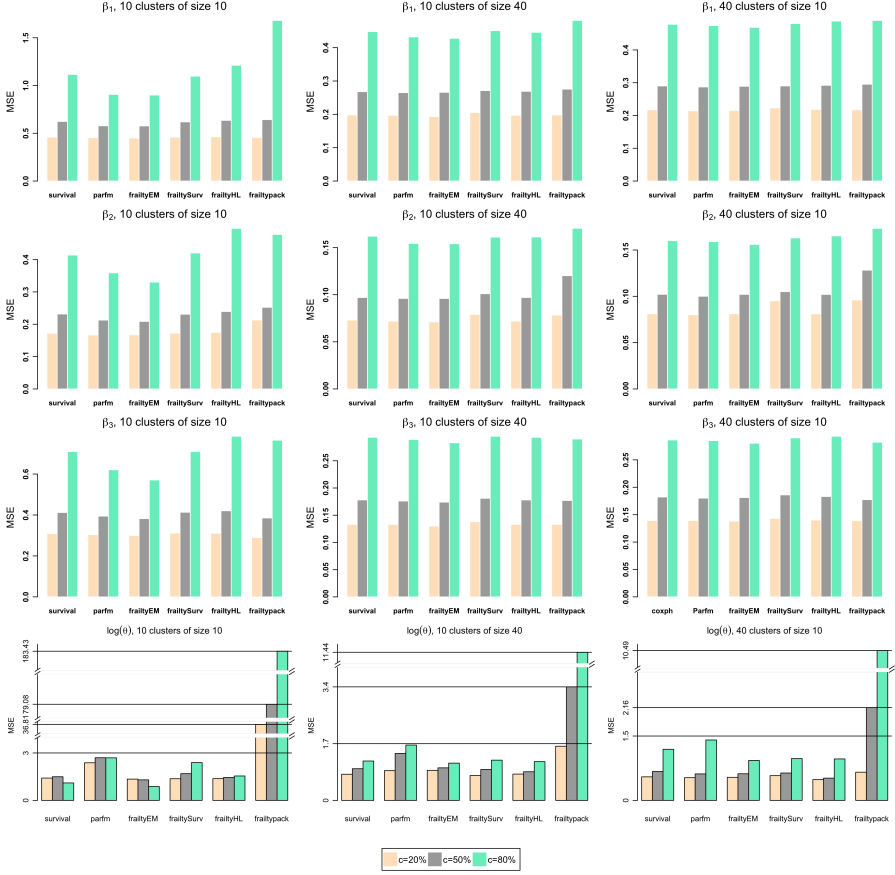


Fig. 2 Comparison of the MSEs of estimated regression coefficients and the log variance of the frailty terms. The first, second, third and fourth rows correspond to the results for β_1 , β_2 , β_3 and θ , respectively. In each panel, the yellow, gray and green bars correspond to censoring rates of 20%, 50% and 80%, respectively.

Figure 2 displays the MSEs of the estimators for all the parameters in the scenario of 10 clusters of 10 observations (left panels), 10 clusters of 40 observations (middle panels) and 40 clusters of 10 observations (right panels). We see that the MSEs of the estimated regression coefficients increase as the censoring rate increases. The **frailtypack** had slightly larger MSEs for β_2 than other packages. Not surprisingly, the MSE of the estimated frailty variance increases for most cases as the censoring percentage increases. Most importantly, we noticed that the MSEs of the estimated log frailty variance given by the **frailtypack** package are significantly larger compared to the other packages, especially in the case of the sample size of 100.

4.2 Coverage Probability (CP)

For all the packages considered in this paper, 95% confidence intervals (CI) of regression coefficients are constructed based on a normal approximation of the distribution of $\hat{\beta}$, i.e., $\hat{\beta} \pm 1.96 \times \text{SE}(\hat{\beta})$. Figure S1 in the Supplementary Materials displays the coverage probabilities (CPs) of the 95% confidence intervals (CIs) of the three regression coefficients in different scenarios. The left panels correspond to the scenario of 10 clusters of 10 observations, the middle panels depict the scenario of 10 clusters of 40 observations, and the right panels showcase 40 clusters of 10 observations. The CPs of the 95% CIs for most of the packages were very close to 95%. However, in the scenario with 10 clusters of 10 observations, the CIs for β_1 and β_2 by **frailtypack** have slightly lower CPs than the nominal level. Similarly, in the case of 10 clusters of 40 observations, the CIs for β_2 by **frailtypack** and the CIs for β_3 by **frailtySurv** have slightly lower CPs than the nominal level 95%. The detailed results are displayed in Table S7 in the Supplementary Materials.

For the frailty variance, most packages construct CIs based on a normal approximation for the distribution of $\hat{\theta}$ as follows

$$\text{CI}^{(1)} : \hat{\theta} \pm 1.96 \times \widehat{\text{SE}}(\hat{\theta}).$$

We call this type of interval as $\text{CI}^{(1)}$. The first row in Figure 3 clearly showed that the CPs for $\text{CI}^{(1)}$ failed to attain the nominal level of 95%. Moreover, in the scenario of 10 clusters of 10 observations with an 80% censoring rate, the CPs for $\text{CI}^{(1)}$ given by the **parfm**, **frailtyEM**, and **frailtySurv** exceeded the 95% nominal level. These results are not surprising, since the distribution of $\hat{\theta}$ is highly skewed as shown by Figure 1. Better CI may be constructed with a normal approximation for the distribution of $\log(\hat{\theta})$, which is more symmetric as shown by Figure 1. Based on this reasoning, we construct a 95% CI for $\log(\theta)$ and then exponentiate the boundaries for obtaining 95% CI for θ , given as follows:

$$\text{CI}^{(2)} : \exp(\log(\hat{\theta}) \pm 1.96 \times \widehat{\text{SE}}(\log(\hat{\theta}))). \quad (22)$$

We call this type of interval by $\text{CI}^{(2)}$. Most packages do not provide an estimate of $\text{SE}(\log(\hat{\theta}))$ directly. However, we can estimate it with an estimate of $\text{SE}(\hat{\theta})$ by applying the general formula for computing the Fisher information of transformed parameters (or the Delta method for calculating the asymptotic variance of a transformed random variable) [41–43], given as follows:

$$\widehat{\text{SE}}(\log(\hat{\theta})) = \frac{1}{\hat{\theta}} \widehat{\text{SE}}(\hat{\theta}). \quad (23)$$

To make this paper self-contained, we provide a brief derivation of equation (23) in the appended Section A.

As shown by Figure 3, the confidence intervals computed with $\text{CI}^{(2)}$ have consistently better coverage probabilities (closer to 0.95) than those computed with $\text{CI}^{(1)}$ across most packages. Interestingly, as the censoring rate increases,

the CPs of both $CI^{(1)}$ and $CI^{(2)}$ for θ became closer to 95%. This is likely because a higher censoring rate results in a larger estimate of the standard error of $\hat{\theta}$ or $\log(\hat{\theta})$, which then gives a wider confidence interval. This finding was previously reported by Balan et al. [40]. In addition, the confidence intervals for θ by all the packages considered in this paper had lower CPs in the scenario with 10 clusters of 40 observations compared to the scenario with 40 clusters of 10 observations; this is presumably because that the distribution of $\hat{\theta}$ (or $\log(\hat{\theta})$) is closer to a normal when the number of clusters is larger.

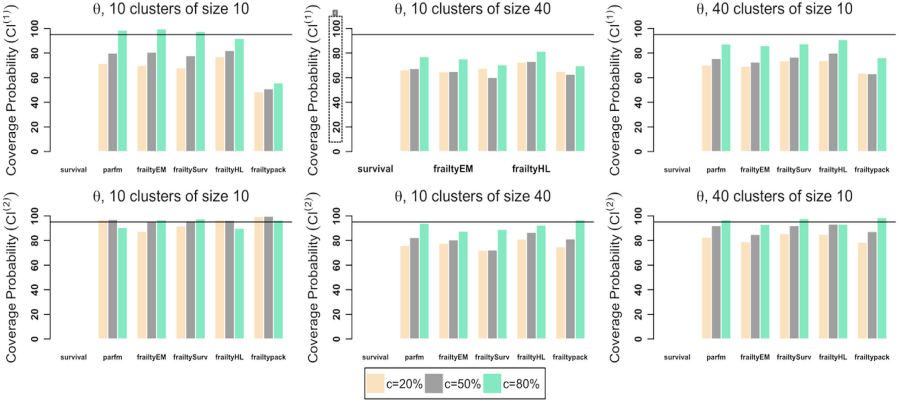


Fig. 3 Coverage probabilities of 95% confidence intervals for the frailty variance. The black horizontal line indicates the 95% nominal level. The first and second rows correspond to $CI^{(1)}$ and $CI^{(2)}$ as described in section 4.2 respectively.

4.3 Convergence Rate

Table S8 in the Supplementary Materials presents the results of the convergence rates. The **survival**, **frailtySurv**, and **frailtypack** packages had convergence rates over 97% in all scenarios. When the sample size is small with a large censoring rate, the **frailtyHL**, **parfm** and **frailtyEM** packages had relatively lower convergence rates compared to others. In the scenario with a low sample size such as 100 and a high censoring rate such as 80%, the **parfm** and **frailtyEM** packages had significantly lower convergence rates (66.3% and 57.2%) than other packages.

4.4 Computing Time

Table 2 reports the average computing time for fitting the shared frailty model using the packages under each simulation scenario. The package **survival** is the fastest one, followed by **frailtyEM** and **frailtypack**, and **parfm**, **frailtySurv** and **frailtyHL**. In general, a larger number of clusters or a larger cluster size results in more computing time for most packages, except for the **frailtyEM** package.

Table 2 Average computing time (in seconds) of the packages under each simulation scenario. “100c” represents the censoring rate, and “Obs” represents the cluster size.

<i>n</i>	Clusters	Obs	100c	survival	parfm	frailtyEM	frailtySurv	frailtyHL	frailtypack
100	10	10	20	0.013	3.551	0.486	0.248	0.463	0.451
400	40	10	20	0.030	10.693	1.319	14.579	19.694	1.135
400	10	40	20	0.020	6.581	2.151	9.881	5.044	1.097
800	80	10	20	0.059	17.854	2.310	164.660	123.478	2.222
800	10	80	20	0.032	15.908	6.512	124.758	44.476	2.239
100	10	10	50	0.014	3.859	0.388	0.233	0.553	0.409
400	40	10	50	0.020	7.281	0.763	13.860	11.572	1.022
400	10	40	50	0.020	7.281	0.763	13.860	11.572	1.022
800	80	10	50	0.059	24.918	2.068	230.582	160.960	2.450
800	10	80	50	0.023	10.093	3.450	125.591	19.418	1.656
100	10	10	80	0.011	3.180	0.278	0.184	0.643	0.266
400	40	10	80	0.021	9.575	0.478	13.555	22.614	0.860
400	10	40	80	0.017	6.551	0.611	9.659	7.316	0.902
800	80	10	80	0.032	16.223	0.754	162.441	141.127	1.394
800	10	80	80	0.020	9.521	1.412	124.499	18.479	1.235

5 Conclusions and Discussions

In this paper, all the packages considered for fitting the shared frailty models gave very similar and unbiased parameter estimates for the fixed-effect regression coefficients, regardless of the sample size, cluster sizes and censoring rates. However, there are differences in the performance of estimating the frailty variance between the considered packages. All the packages considered in this paper consistently underestimated the frailty variance. This is because that the values of the u_i ’s are shrunk towards 0. As θ is the variance of the u_i , the estimated θ tends to be smaller than its real value. This finding was previously reported by other researchers; see [29, 44]. Most importantly, we have also observed that the distribution of estimated frailty variance is highly positively-skewed. Our results also showed that a larger number of clusters can lead to higher precision in estimating the frailty variance. The CPs of the 95% CIs of the regression coefficients given most packages are very close to the nominal 95%. Most packages had convergence rates over 97% in all scenarios, except for the **parfm** and **frailtyEM** packages in the scenario with a small sample size ($n=100$) and large censor rate (80%). The computing times of **survival**, **frailtyEM** and **frailtypack** packages for all scenarios are within 0.1 minutes; the **parfm** takes no more than 0.5 minutes. However, the computing times of **frailtySurv** and **frailtyHL** packages need two to three minutes under the sample size $n=800$.

Considering the severe skewness of the distribution of $\hat{\theta}$, we have suggested a new type of confidence interval for the frailty variance θ , based on a normal approximation for the distribution of $\log(\hat{\theta})$. Our empirical results show that the proposed confidence interval’s coverage probabilities are much closer to the nominal level than those of the confidence intervals based on a normal approximation for the distribution of $\hat{\theta}$. We recommend that the authors of the considered R packages in this paper implement this new type of confidence interval.

In summary, the best package to estimate the parameters of a frailty model is the **survival** package, which is computationally fast with a high convergence rate in almost all simulation scenarios. Furthermore, the **survival** package can effectively handle a wide variety of structures for the frailties. A drawback of

the **survival** package is that it does not provide an estimate for the standard error of estimated frailty variance. Since the performances of the EM and PPL algorithms in estimating fixed effects are similar, **frailtyEM** can be used to substitute **survival** if the standard error of the frailty variance is required in a real application. However, we do not suggest using **frailtyEM** package when the sample size is small with a large censoring rate due to its lower convergence rate. The **parfm** has a lower convergence rate as well in the scenario with a small sample size and a large censoring rate. The estimators of the regression coefficients and the frailty variance are more accurate if the baseline hazard distribution is given. In such cases, the **parfm** is a good choice if the sample size is also large. The **frailtySurv** package fits the frailty model with a wide range of frailty distributions, and the **frailtyHL** allows multilevel frailties in the frailty model. However, the **parfm**, **frailtySurv** and **frailtyHL** packages require more computing time compared to other methods, which is due to the method used for modelling the frailties. When the sample size is large, the **frailtypack** package demonstrates similar performance in parameter estimation with other packages. However, it may produce less accurate estimates when the sample size is 100. Hence, caution should be exercised when using the **frailtypack** package for datasets with small sample sizes and few clusters. To improve the performance of **frailtypack** package, the user may need to carefully choose the number of knots and other parameters for using spline methods. The number of knots may play a crucial role in estimating the baseline hazard using spline methods. The baseline hazard function might impact the estimated frailty variance. Previous research also reported that using a smaller number of knots typically helps to circumvent the problem of overestimating the fixed effect [29]. On the other hand, the **frailtypack** package offers the advantage of accommodating more complex structures for the frailties, including nested and joined frailties, as well as frailty interactions, which enable more flexible modelling options for the frailties.

A Derivation of $SE(\log(\hat{\theta}))$

Let X be a random vector (data) with the PDF $f(x | \theta)$. Let $I_1(\theta)$ denote the Fisher information of θ and $l_1(\theta; x)$ denote the log-likelihood of θ given x . Suppose we re-parameterize $\theta = \Theta(\phi)$, where $\Theta(\cdot)$ is a differentiable function. The log-likelihood function [42, 43, 45] for ϕ , $l_2(\phi; x)$, is given by:

$$l_2(\phi; x) = l_1(\Theta(\phi); x) = \log(f(x | \Theta(\phi))). \quad (24)$$

Then the derivative of l_2 is given by:

$$\frac{\partial l_2(\phi; x)}{\partial \phi} = \frac{\frac{\partial f(x|\theta)}{\partial \theta} \frac{\partial \Theta(\phi)}{\partial \phi}}{f(x | \Theta(\phi))}. \quad (25)$$

Then, it follows that the Fisher's information of ϕ , $I_2(\phi)$, is obtained as follows:

$$I_2(\phi) = E_X \left\{ \left(\frac{\frac{\partial f(X|\theta)}{\partial \theta}}{f(X|\theta(\phi))} \right)^2 \right\} \left(\frac{\partial \Theta(\phi)}{\partial \phi} \right)^2 = I_1(\theta)(\Theta'(\phi))^2, \quad (26)$$

where Θ' denotes the derivative function of Θ . Applying the above general rule (26) to $\Theta(\phi) = \exp(\phi)$ (i.e., $\phi = \log(\theta)$), we arrive at the following equation:

$$I_2(\phi) = I_1(\theta)\theta^2. \quad (27)$$

We know that $SE(\hat{\theta}) = \frac{1}{\sqrt{I_1(\hat{\theta})}}$, where $\hat{\theta}$ is the maximum likelihood estimation (MLE) of θ [42, 43, 45]. Finally, we arrive at the following relationship:

$$SE(\log(\hat{\theta})) = \frac{1}{\sqrt{I_2(\hat{\phi})}} = \frac{1}{\theta} \frac{1}{\sqrt{I_1(\theta)}} = \frac{1}{\theta} SE(\hat{\theta}). \quad (28)$$

B Supplementary Materials

Additional figures, tables, and simulation results are available via [this publisher's link](#).

References

- [1] Cox, D.R.: Regression models and life-tables. Journal of the Royal Statistical Society. Series B, Methodological **34**(2), 187–220 (1972)
- [2] Lin, D.Y., Wei, L.J., Ying, Z.: Accelerated failure time models for counting processes. Biometrika **85**(3), 605–618 (1998)
- [3] Govindarajulu, U.S., Lin, H., Lunetta, K.L., D'Agostino Sr, R.B.: Frailty models: Applications to biomedical and genetic studies. Statistics in medicine **30**(22), 2754–2764 (2011)
- [4] Balan, T.A., Putter, H.: Nonproportional hazards and unobserved heterogeneity in clustered survival data: When can we tell the difference? Statistics in medicine **38**(18), 3405–3420 (2019)
- [5] Collett, D.: Modelling Survival Data in Medical Research, 3rd edn. Texts in statistical science, (2015)
- [6] Vaupel, J.W., Manton, K.G., Stallard, E.: The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography **16**(3), 439–454 (1979)

- [7] Duchateau, L., Duchateau, L.a.: The Frailty Model. Statistics for biology and health. Springer, New York (2008)
- [8] Clayton, D.G.: A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**(1), 141–151 (1978)
- [9] Karagrigoriou, A.: Frailty models in survival analysis. *Journal of Applied Statistics* **38**(12), 2988–2989 (2011)
- [10] Hanagal, D.: Modeling survival data using frailty models. *Statistical methods in medical research* **24**(6), 936–936 (2015)
- [11] Therneau, T.M., Grambsch, P.M., Pankratz, V.S.: Penalized survival models and frailty. *Journal of computational and graphical statistics* **12**(1), 156–175 (2003)
- [12] Duchateau, L., Janssen, P.: Penalized partial likelihood for frailties and smoothing splines in time to first insemination models for dairy cows. *Biometrics* **60**(3), 608–614 (2004)
- [13] Ripatti, S., Palmgren, J.: Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics* **56**(4), 1016–1022 (2000)
- [14] Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B, Methodological* **39**(1), 1–38 (1977)
- [15] Klein, J.P.: Semiparametric estimation of random effects using the cox model based on the em algorithm. *Biometrics* **48**(3), 795–806 (1992)
- [16] Zucker, D.M., Gorfine, M., Hsu, L.: Pseudo-full likelihood estimation for prospective survival analysis with a general semiparametric shared frailty model: Asymptotic theory. *Journal of statistical planning and inference* **138**(7), 1998–2016 (2008)
- [17] Gorfine, M., Zucker, D.M., Hsu, L.: Prospective survival analysis with a general semiparametric shared frailty model: A pseudo full likelihood approach. *Biometrika* **93**(3), 735–741 (2006)
- [18] Ha, I.D., Lee, Y., Song, J.: Hierarchical likelihood approach for frailty models. *Biometrika* **88**(1), 233–233 (2001)
- [19] van den Berg, G.J., Drepper, B.: Inference for shared-frailty survival models with left-truncated data. *Econometric reviews* **35**(6), 1075–1098 (2016)
- [20] Lam, K.F., Kuk, A.Y.C.: A marginal likelihood approach to estimation in

- frailty models. *Journal of the American Statistical Association* **92**(439), 985–990 (1997)
- [21] Joly, P., Commenges, D., Letenneur, L.: A penalized likelihood approach for arbitrarily censored and truncated data: Application to age-specific incidence of dementia. *Biometrics* **54**(1), 185–194 (1998)
- [22] Rondeau, V., Commenges, D., Joly, P.: Maximum penalized likelihood estimation in a gamma-frailty model. *Lifetime data analysis* **9**(2), 139–153 (2003)
- [23] Therneau, T.M.: A Package for Survival Analysis in R. (2022). R package version 3.3-1. <https://CRAN.R-project.org/package=survival>
- [24] Balan, T.A., Putter, H.: frailtyEM : An R package for estimating semi-parametric shared frailty models. *Journal of statistical software* **90**(7), 1–29 (2019)
- [25] Monaco, J.V., Gorfine, M., Hsu, L.: General semiparametric shared frailty model: Estimation and simulation with frailtysurv. *Journal of statistical software* **86**(4), 1–42 (2018)
- [26] Ha, D. Il, Noh, M., Lee, Y.: frailtyHL: A package for fitting frailty models with H-likelihood. *The R journal* **4**(2), 28 (2012)
- [27] Munda, M., Rotolo, F., Legrand, C.: parfm : Parametric frailty models in R. *Journal of statistical software* **51**(11) (2012)
- [28] Rondeau, V., Mazroui, Y., Gonzalez, J.R.: frailtypack : An R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal of statistical software* **47**(4) (2012)
- [29] Hirsch, K., Wienke, A.: Software for semiparametric shared gamma and log-normal frailty models: An overview. *Computer methods and programs in biomedicine* **107**(3), 582–597 (2011)
- [30] Henderson, R.: Analysis of multivariate survival data. *Statistics in Medicine* **20**(16), 2533–2534 (2001)
- [31] Hougaard, P.: Frailty models for survival data. *Lifetime data analysis* **1**(3), 255–273 (1995)
- [32] Johnson, N.L., Kotz, S.: Distributions in statistics: Continuous univariate distributions. *Advances in Mathematics* **26**(3), 327–327 (1977)
- [33] McGilchrist, C.A.: Reml estimation for survival models with frailty. *Biometrics* **49**(1), 221–225 (1993)

- [34] Lin, D.Y.: On the breslow estimator. *Lifetime data analysis* **13**(4), 471–480 (2007)
- [35] Lee, Y., Nelder, J.A.: Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B, Methodological* **58**(4), 619–678 (1996)
- [36] Ha, I.D., Noh, M., Lee, Y.: Bias reduction of likelihood estimators in semiparametric frailty models. *Scandinavian journal of statistics* **37**(2), 307–320 (2010)
- [37] Marquardt, D.W.: An algorithm for least-squares estimation of non-linear parameters. *Journal of the Society for Industrial and Applied Mathematics* **11**(2), 431–441 (1963)
- [38] Ramsay, J.O.: Monotone regression splines in action. *Statistical science* **3**(4), 425–441 (1988)
- [39] McCulloch, C.E., Neuhaus, J.M.: Misspecifying the shape of a random effects distribution: Why getting it wrong may not matter. *Statistical science* **26**(3), 388–402 (2011)
- [40] Balan, T.A., Jonker, M.A., Johannesma, P.C., Putter, H.: Ascertainment correction in frailty models for recurrent events data. *Statistics in medicine* **35**(23), 4183–4201 (2016)
- [41] Gatenby, R.A., Frieden, B.R.: The critical roles of information and nonequilibrium thermodynamics in evolution of living systems. *Bulletin of mathematical biology* **75**(4), 589–601 (2013)
- [42] Devore, J.L.: *Modern Mathematical Statistics with Applications*, 2nd edn. Springer Texts in Statistics, (2012)
- [43] Knight, K.: *Mathematical Statistics*. Chapman & Hall/CRC Press, (2000)
- [44] Petersen, L., Sørensen, T.I.A., Nielsen, G.G., Andersen, P.K.: Inference methods for correlated left truncated lifetimes: parent and offspring relations in an adoption study. *Lifetime data analysis* **12**(1), 5–20 (2006)
- [45] Casella, G.C., Berger, R.L.: *Statistical Inference*, 2nd edn. Thomson Learning, (2002)