Journal Title
XX(X):2–39

DOI: 10.1177/ToBeAssigned
www.sagepub.com/

Survival analysis with delayed entry in selected families with application to human longevity

M. Rodríguez-Girondo[1], J. Deelen[2], P.E. Slagboom[2] and J.J. Houwing-Duistermaat[1,3]

Abstract

In the field of aging research, family-based sampling study designs are commonly used to study the lifespans of long-lived family members. However, the specific sampling procedure should be carefully taken into account in order to avoid biases. This work is motivated by the Leiden Longevity Study (LLS), a family-based cohort of long-lived siblings. Families were invited to participate in the study if at least two siblings were 'long-lived', where 'long-lived' meant being older than 89 years for men or older than 91 years for women. As a result, more than 400 families were included in the study and followed for around 10 years. For estimation of marker-specific survival probabilities and correlations among life times of family members, delayed entry due to outcome-dependent sampling mechanisms has to be taken into account. We consider shared frailty models to model left-truncated correlated survival data. The treatment of left truncation in shared frailty models is still an open issue and the literature on this topic is scarce. We show that the current approaches provide, in general biased estimates and we propose a new method to tackle this selection problem by applying a correction on the likelihood estimation by means of inverse probability weighting at the family level.

Prepared using sagej.cls

Keywords

Survival analysis; shared-frailty model; delayed entry; family data, inverse probability weighting

1 Introduction

Family-based cohort studies are frequently used in epidemiology in order to investigate traits which aggregate within families. In the field of aging research, human longevity has shown to cluster within families (1–5) and this has motivated numerous family-based sampling study designs based on the selection of long-lived (according to a set of predefined criteria) family members (e.g. siblings) from a reference population of interest. The study of their survival times provides insights into the factors affecting survival in old individuals, marker-specific survival probabilities, and the level of lifespan correlation within families. However, the specific sampling procedure should be carefully taken into account in the statistical analysis of the resulting data in order to avoid biases that may lead to wrong conclusions. In general, given that the selection of participants is based on age criteria, left truncation by death plays an important role when studying longevity or extreme survival. Challenges in this framework are to deal with the delayed entry resulting from the sampling mechanism, to take into account correlation between family members, and to deal with the interplay between them.

This work is motivated by the Leiden Longevity Study (6, 7), a family-based cohort of long-lived siblings together with their offspring and the partners thereof. The goal of the recruitment strategy was to enrich for genetic variants involved in aging. Families were invited to participate in the study if at least two siblings were 'long-lived', where

[1] Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, The Netherlands.

[2]Department of Molecular Epidemiology, Leiden University Medical Center, The Netherlands

[3]Department of Statistics, University of Leeds, United Kingdom

Corresponding author:

Mar Rodríguez-Girondo. Department of Medical Statistics and Bioinformatics, Leiden University Medical Center. Post Zone S5-P, PO Box 9600, 2300 RC Leiden, The Netherlands.
Email: M.Rodriguez_Girondo@lumc.nl

Prepared using sagej.cls *[Version: 2015/06/09 v1.01]*

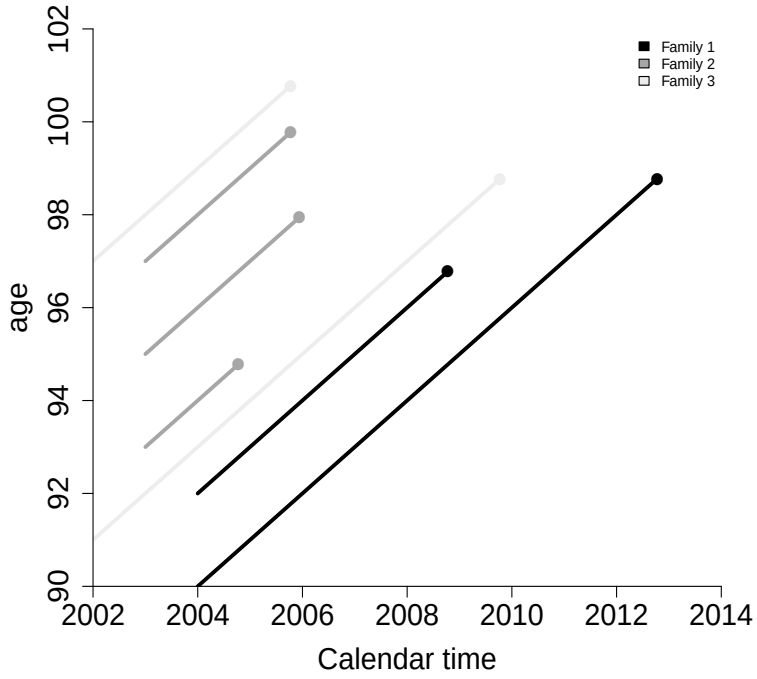

Figure 1. An example of three included families in the Leiden Longevity Study.

'long-lived' meant being older than 89 years for men or older than 91 years for women. As a result, more than 400 families were included in the study between 2002 and 2006 and followed for around 10 years. Note that similar designs are also used by other studies, such as the European study GEHA (Genetics of Healthy Aging, (8)) and the international (U.S.and Denmark) LLFS (Long Life Family Study, (9)).

The Lexis diagram displayed in Figure 1 illustrates the selection procedure in the Leiden Longevity Study (LLS). The aim of our data analysis is two fold. On the one hand, we are interested in estimating the effect of (genetic) markers affecting survival in the elderly and their corresponding marker-specific survival rates, by using the 'long-lived' siblings of the LLS. On the other hand, we are interested in estimating the level of familial correlation of lifespan in the subpopulation of long-lived.

We adopt a conditional approach by considering shared frailty models (10–12) to model correlated survival data, using age as time scale. The frailty variance represents

the within-family aggregation of the studied survival times (see, for example, Section 4.1. in (11)), and hence, its correct estimation is of primary interest in aging research (1, 5, 13). Additionally, the prediction of marker-specific survival rates is a relevant topic in the field of longevity (14). This can also be addressed from a frailty model perspective, for which the unbiased estimation of both the marker effect and baseline hazard is required.

Since the inclusion criterion relies on being alive at a certain enrollment period, individuals are only observed if their age at death is greater than certain age at entry (determined by the enrollment mechanism). This leads to the presence of left-truncated survival times due to delayed entry.

The treatment of left truncation in shared frailty models is still an open issue and the literature on this topic is scarce. Left truncation was already considered by Nielsen et al. (15) in their seminal paper on frailty models. Namely, these authors studied the correlation between the lifespans of parents, biological and adopted children. Left truncation due to delayed entry is handled by adapting the at-risk indicators in this example. Later, Jensen et al. (16) and Rondeau et al. (17) independently, proposed an alternative approach which accounts for left truncation at the frailty distribution level. Recently, van der Berg and Drepper (18) revisited the problem and proposed the same likelihood as Jensen et al. (16), and Rondeau et al. (17) for the specific case where each cluster contains two units and both are observed (but under delayed entry). Also, in the field of recurrent events, both approaches have been discussed (20).

In this paper, we revisit the former approaches for dealing with left truncation in family data in order to provide clear guidelines about their assumptions and their appropriateness according to the data at hand. Specifically, we will discuss two selection mechanisms, across and within families, which influence the configuration of the observed sample. On the one hand, left-truncated cluster survival data can be regarded as a problem of frailty-dependent (non-random) selection of families, as it has been previously recognized (16, 20). On the other hand, the presence of left -truncation due to delayed entry induces a within-family selection phenomenon which has been less studied so far. We illustrate the different impact of both selection mechanisms according to the size of the family and different selection criteria and we show that the current approaches provide, in general biased estimates due to the assumption of non-informative selection of individuals within families. We propose a new method to

tackle this selection problem by proposing a correction on the likelihood estimation by means of inverse probability weighting at the family level.

The rest of the paper is organized as follows. In Section 2, we introduce notation and we establish a general framework for different sampling schemes resulting on left-truncated clustered survival data. Shared frailty models are revisited in Section 3. In Section 4, we present existing and new methods for dealing with left-truncation in shared frailty models. An intensive simulation study is presented in Section 5, while in Section 6 the methods are applied to the LLS. Main conclusions and a final discussion follow in Section 7.

2 Notation and problem description

Let the observations be given by $(B_{ij}, V_{ij}, T_{ij}, \delta_{ij}, \mathbf{x}_{ij})$, where $i = 1, \ldots, n$ index all the studied families, $j = 1, \ldots, m_i$ the observed siblings from the i-th family, B_{ij} is the date of birth, V_{ij} is the date of enrollment in the study, T_{ij} is the date of death or end of follow-up, δ_{ij} is the non-censoring indicator and $\mathbf{x}_{ij} = (x_{1ij}, \ldots, x_{qij})'$ a vector of q individual-specific covariates which may affect survival. We focus on $t_{ij} = T_{ij} - B_{ij}$, the potentially right-censored survival time given in age scale. Since the survival times and covariates of the included individuals are only observed if they are alive at certain specific date V_{ij} determined by the enrollment process, individuals are observed only if their age at death (t_{ij}) is greater than certain age at enrollment in the study, defined as $t_{0ij} = V_{ij} - B_{ij}$.

This type of data is the result of outcome-dependent sampling schemes and hence, the specific sampling mechanism can not be neglected in the models (see (19, 20) and references therein). Denote by $A_i = I\left\{\left[\sum_j^{n_i} I(t_{ij} > t_{0ij} \geq c_0)\right] \geq K\right\}$ the inclusion indicator for family i according to the pre-defined study design (c_0, K and n_i deterministic). $A_i = 1$ if family i is included in the study, i.e. $A_i = 1$ if at least K alive siblings were older than a predefined value $c_0 \geq 0$ by the recruitment period. In general, c_0 represents the age at the origin of the follow-up time and may be common for all the individuals, as in the GEHA project ($c_0 = 90$) or covariate-specific, as it is the case in the Leiden Longevity Study, where gender-specific entry criteria were considered ($c_0 = 89$ for men, $c_0 = 91$ for women). n_i is the size of the family i (including those siblings not included due to death previous to recruitment but with

$t_{ij} \geq c_0$). Note that n_i is known because the genealogical information regarding birth and death dates is available for the complete family, i.e. also for those members deceased before recruitment started, and for which none of the covariates $\mathbf{x}$ are available. Finally, denote by M_i the random variable referring to the number of included individuals from family i, $M_i = m_i$, and define the sampling event for family as $\Omega_i = \{t_{i1} \geq t_{0i1}, \ldots, t_{im_i} \geq t_{0im_i}, M_i = m_i, n_i\}$.

In the LLS, the members of a given family are selected at the same timepoint ($V_{ij} = V_i,\ j = 1, \ldots, m_i$) but at different ages (due to their different birth dates), which provokes different entry ages in the sample across individuals. Specifically, let us consider the recruitment period given in chronological time by a discrete process of dates $[\tau_1, \ldots \tau_Q]$ and suppose that an arbitrary family i is invited to participate at time τ_q given that it verifies $[A_i = 1|\tau_q, c_0]$. This means that those members who are too young to be included at τ_q are not recovered in a posterior sampling time point and that $V_{ij} = V_i = \tau_q, i = 1, \ldots, n_i$. Under such sampling scheme, $t_{0ij} = \tau_q - B_{ij} \geq c_0$, where B_{ij} is the date of birth of sibling j from family i.

Note that the former general definition of A_i covers a large number of outcome-dependent sampling scenes, all affected by delayed entry, for example, when considering age as time scale. On the one hand, one may consider the selection of a given family only if all its members are observed, we refer to this situation as 'fully observed' families, where $m_i = n_i$ by design. This means that a family is included if and only if all its eligible members (i.e., those n_i with $t_{ij} > c_0$) are alive at the recruitment timepoint V_{ij}, so that we can follow them all, even if their entry times differ. Such sampling schemes are typically used in twin studies (1, 5) and imply that M_i is deterministic. On the other hand, less restrictive selection schemes, where families are partially observed (M_i is random and $m_i \leq n_i$), are common in epidemiological studies. Family-based studies relying on arbitrary number of siblings, as it is the case of the LLS, select siblings if at least two of the total number of the sibship are alive at the recruitment period ($K = 2$). Also, the dynamic sampling framework considered in Jensen et al. (16) to study family aggregation of childhood mortality implies different number of selected individuals per family, without fixing any minimum number of individuals per cluster, i.e. $K = 1$.

Given that the inclusion of families under left-truncation is driven through the inclusion of (some of) their members, we can treat the unobserved family members

as missing data. Set $R_{ij} = I\{t_{ij} > t_{0ij} \geq c_0\} = 1$ if the member j of family i is included and 0 otherwise, i.e., $R_{ij} = 0$ if $t_{0ij} > t_{ij} \geq c_0$. Let $\mathbf{R}_i = (R_{i1}, \ldots, R_{in_i})$ be the vector of non-missingness indicator of family i. $\mathbf{R}_i$ is always observed as long as n_i (the family size at sampling time) is known. Consequently, we can redefine the number of observed siblings of family i as $M_i = \sum_{j=1}^{n_i} R_{ij}$.

In the next Section, we provide a general methodological overview to deal with left-truncated frailty models, paying special attention to the impact in the inference of the different selection procedures according to the specific choices of K and the resulting patterns of $\mathbf{R}_i$.

3 Shared frailty models revisited

We consider shared frailty proportional hazard models for the analysis of clustered survival data (10–12):

$$\lambda_{ij}(t) = u_i \lambda_0(t, \boldsymbol{\gamma}) \exp(\boldsymbol{\beta}\mathbf{x}_{ij}), \quad i = 1, \ldots n, \quad j = 1, \ldots, n_i, \tag{1}$$

where $\lambda_0(t, \boldsymbol{\gamma})$ refers to the baseline hazard, $\boldsymbol{\beta}$ are the regression coefficients corresponding to the vector of covariates $\mathbf{x}$ and the term $u > 0$ refers to an unobserved random effect (frailty) shared by the members of the same family. The baseline hazard $\lambda_0(t, \boldsymbol{\gamma})$ is specified in terms of the vector of parameters $\boldsymbol{\gamma}$. If $\boldsymbol{\gamma}$ is infinitive, the baseline hazard is completely unspecified and it corresponds to a frailty Cox model, otherwise, when $\boldsymbol{\gamma}$ is a finite-dimensional vector, we refer to parametric frailty models. The unobserved heterogeneity shared within families accounts for genetic or (early life) environmental factors common to members of a given sample and it is assumed to follow certain parametric distribution G in the population. In this paper, we assume that u follows a gamma distribution. Gamma frailties have been broadly used because of their attractive mathematical properties, given that the dependence induced by the frailty can be expressed in terms of their Laplace transforms, which allows the derivation of closed-form likelihoods when assuming a parametric baseline hazard λ_0, i.e. when $\boldsymbol{\gamma}$ is a finite-dimensional vector. Otherwise, if $\boldsymbol{\gamma}$ is infinite-dimensional, EM algorithms (15) or penalized likelihood approaches (17) have been proposed to fit model (1). See Duchateau and Janssen (11) for a recent review on frailty distributions and discussion on existing estimation procedures for frailty models.

Due to identifiability reasons, we assume that $u \sim \Gamma(1/\theta, 1/\theta)$, which ensures that $E(u) = 1$ and $var(u) = \theta$.

Inference of gamma shared frailty models has received a lot of attention in past decades and it is well established (see Cortiñas et al. (21) for a review). In general, without left truncation, and assuming that conditional on u, right-censoring is non-informative, the marginal likelihood contribution of family i is given by:

$$L_i = E_u[f_c(t_1, \ldots, t_{n_i}|\mathbf{x_i}, u_i)] = \int_u \prod_{j=1}^{n_i} [u\lambda_0(t_{ij}, \gamma)\exp(\boldsymbol{\beta}\mathbf{x}_{ij})]^{\delta_{ij}} \exp[-u\Lambda_{ij}(t_{ij})]dG(u), \tag{2}$$

where f_c refers to the conditional probability density function, and $\Lambda_{ij}(t) = \Lambda_0(t)\exp(\boldsymbol{\beta}\mathbf{x}_{ij})$, $\Lambda_0(t) = \int_0^t \lambda_0(s, \boldsymbol{\gamma})ds$ is the cumulative hazard.

Recalling the Laplace transform derivatives of u as $\mathcal{L}^{(r)}(s) = (-1)^r E_u\left[u^r \exp(-us)\right]$, and denoting by D_i the number of uncensored observations of family i, we can rewrite the likelihood contribution of a family i as:

$$L_i = \prod_{j=1}^{n_i} [\lambda_0(t_{ij}, \gamma)\exp(\boldsymbol{\beta}\mathbf{x}_{ij})]^{\delta_{ij}} \, (-1)^{D_i} \mathcal{L}^{(D_i)}\left[\textstyle\sum_{j=1}^{n_i} \Lambda(t_{ij})\right] \tag{3}$$

and the parameters of interest $(\boldsymbol{\gamma}, \boldsymbol{\beta}, \theta)$ are estimated by maximizing the log-likelihood obtained from equation (3), namely solving:

$$\max_{\boldsymbol{\gamma},\boldsymbol{\beta},\theta} \ell(\boldsymbol{\gamma}, \boldsymbol{\beta}, \theta) = \max_{\boldsymbol{\gamma},\boldsymbol{\beta},\theta} \sum_{i=1}^{n} \ell_i(\gamma, \beta, \theta) = \max_{\gamma,\beta,\theta} \sum_{i=1}^{n} \log L_i(\boldsymbol{\gamma}, \boldsymbol{\beta}, \theta) \tag{4}$$

4 Shared frailty models with delayed entry

So far, the inference of left-truncated gamma frailty models has been approached from two different points of view. The fundamental difference among them relies on the specification of the frailty distribution when constructing the marginal likelihood.

Note that the frailty distribution G corresponds to the distribution of the frailty values at the population level, which corresponds with the origin time c_0 of the study. However, given that frailer individuals die first, the outcome-dependent selection related to left-truncated provokes that families with larger values of u_i are less likely to be observed. As a result, the frailty distribution in the population of survivors at a given time $t > c_0$ differs from the original one given by G. Specifically, the mean of the frailty distribution becomes smaller as the stronger individuals remain (those with

smaller values of u). At the same time, the variance also becomes smaller since the remaining individuals at risk are more alike. A nice property of the gamma distribution is that the conditional frailty density at time $t > c_0$ is still a gamma density (this property is commonly referred as updating).

Following the notation in Jensen et al. (16), we refer to the two existing approaches as 'naïve' (15) and 'updated' (16–18, 20, 22).

4.1 'Naïve' approach

The first approximation to adapt the likelihood expression given by equation (2) to the presence of delayed entry relies on the same rationale which is standard in the context of survival analysis for left-truncated independent observations. Specifically, delayed entry is handled by adapting the risk sets at the level of the likelihood contribution of a given individual, i.e. replacing $I\left(s \leq t_{ij}\right)$ by $I\left(t_{0ij} < s \leq t_{ij}\right)$ in the definition of the cumulative hazard. Consequently, to account for delayed entry, $\Lambda_0(t_{ij}) = \int_0^{t_{ij}} \lambda_0(s, \boldsymbol{\gamma}) ds = \int_0^{\infty} I(s \leq t_{ij}) \lambda(s, \boldsymbol{\gamma}) ds$ is replaced by $\Lambda_0(t_{ij}) = \int_{t_{0ij}}^{t_{ij}} \lambda(s, \boldsymbol{\gamma}) ds = \int_0^{\infty} I(t_{0ij} < s \leq t_{ij}) \lambda(s, \boldsymbol{\gamma}) ds$ in equations (2) and (3). As a result, provided that truncation is independent from survival of each unit, the resulting likelihood contribution of family i with m_i observed individuals is given by:

$$
\begin{aligned}
&L_i^N = E_u[f_c(t_{i1}, \ldots, t_{im_i} | t_{i1} > t_{0i1}, \ldots, t_{im_i} > t_{0im_i}, \mathbf{x}_i, u_i)] = \\
&\int_0^{\infty} \prod_{j=1}^{m_i} \frac{f_c(t_{ij} | x_{ij}, u_i)}{S_c(t_{0ij} | x_{ij}, u_i)} dG(u_i) = \\
&\int_u \prod_{j=1}^{m_i} \left[u \lambda_0(t_{ij}, \gamma) \exp(\boldsymbol{\beta} \mathbf{x}_{ij})\right]^{\delta_{ij}} \exp\left\{-u \left[\Lambda_{ij}(t_{ij}) - \Lambda_{ij}(t_{0ij})\right]\right\} dG(u)
\end{aligned}
\tag{5}
$$

where f_c and S_c refer to the conditional probability density and survival functions, respectively. The second equality in expression (5) implies that the frailty distribution is not affected by the selection process induced by the delayed entry of the individuals within families, i.e. expression (5) assumes that $G(u_i) = G(u_i | t_{i1} > t_{0i1}, \ldots, t_{im_i} > t_{0im_i})$, as it has been pointed out by Jensen et al. (16) and van den Berg and Drepper (18). However, such an assumption is unrealistic, since, as mentioned, in general, lower values of u will be over-represented when increasing age at entry, as a direct consequence of the fact that frailer individuals (those with higher values of u) die first

and hence, the probability of surviving until their corresponding entry time is lower for them. Hence, the estimates $(\widehat{\gamma}, \widehat{\boldsymbol{\beta}}, \widehat{\theta})$ resulting from expression (5) will be, in general, inconsistent. However, the size of the bias differs according to the level of discrepancy between $G(u_i)$ and $G(u_i|t_{i1} > t_{0i1}, \ldots, t_{im_i} > t_{0im_i})$ in the data at hand. In general, under common frailty distribution and random truncation patterns, one would expect that the level of bias depends on the size of the families in the underlying population. This is due to the fact that even if the size of the family is non-informative with regard to the survival, it affects the distribution of the frailty term in the selected families (23, 24). To illustrate this, suppose that the members of a given family i share a fixed truncation point t_0, then the conditional selection probabilities at the family level can be written as $P(A_i = 1|n_i, t_0, u_i) = \sum_{j=K}^{n_i} \binom{n_i}{j} S_j(t_0)^j \left[1 - S_j(t_0)\right]^{n_i - j}$, where $S_j(t_0) = S(t_0|u_i)$ is the conditional survival at the entry time t_0 for a given member j of family i (free of the particular value of n_i). Note that among families with similar frailty term u_i, larger families are more likely to be included. Moreover, the under-represented higher values of the frailty distribution are more likely to be observed under delayed-entry when belonging to larger families, which potentially would entail $G(u) \approx G(u|t > t_0)$. The practical impact of this issue is empirically evaluated by means of Simulations in the next Section.

4.2 'Updated' approach

An alternative strategy for dealing with left-truncation in the shared-gamma frailty model relies on writing the likelihood as follows (17, 22):

$$L_i^{UP} = \frac{E_u[f_c(t_{i1}, \ldots, t_{im_i}|\mathbf{x}_i, u_i)]}{E_u[S_c(t_{0i1}, \ldots, t_{0im_i}|\mathbf{x}_i, u_i)]} \tag{6}$$

By using the gamma distribution properties, the numerator can be expressed in terms of D_i- derivative of the Laplace transform of u, taking the form of equation (3), while the denominator can be written as $E_u[S_c(t_{0i1}, \ldots, t_{0im_i}|\mathbf{x}_i, u_i)] = \mathcal{L}\left[\sum_{j=1}^{m_i} \Lambda(t_{0ij})\right]$

The former equation (6) can be rewritten as:

$$L_i^{UP} = \frac{E_u[f_c(t_{i1}, \ldots, t_{im_i}|\mathbf{x}_i, u_i)]}{E_u[S_c(t_{0i1}, \ldots, t_{0im_i}|\mathbf{x}_i, u_i)]} = \frac{\int_0^\infty \prod_{j=1}^{m_i} f(t_{ij}|t_{ij} > t_{0ij}, x_{ij}, u_i) P(t_{ij} > t_{0ij}|x_{ij}, u_i) dG(u_i)}{\int_0^\infty \prod_{j=1}^{m_i} P(t_{ij} > t_{0ij}|x_{ij}, u_i) dG(u_i)} \tag{7}$$

Note that by applying Bayes's theorem, we obtain

$$G(u_i|t_{i1} > t_{0i1}, \ldots, t_{im_i} > t_{0im_i}) = \frac{\prod_{j=1}^{m_i} P(t_{ij} > t_{0ij}|x_{ij}, u_i)G(u_i)}{\int_0^\infty \prod_{j=1}^{m_i} P(t_{ij} > t_{0ij}|x_{ij}, u_i)dG(u_i)} \quad (8)$$

and hence, equation (7) is equivalent to:

$$L_i^{UP} = \int_0^\infty \prod_{j=1}^{m_i} f_c(t_{ij}|t_{ij} > t_{0ij}, x_{ij}, u_i)dG(u_i|t_{i1} > t_{0i1}, \ldots, t_{im_i} > t_{0im_i}) \quad (9)$$

Equation (9) explicitly shows the updating nature of this approach. In contrast to the 'naïve' approach, the conditional density of the observed units within a family is averaged over the conditional frailty distribution given the entry times of the family members. This allows to tackle the first selection process (across families) mentioned introduced in Section 1, by adapting the level of dependency within the observed families to the informative selection process. Instead of assuming mean one frailties to all the selected families (as in the naïve approach), in the updated approach the mean of the frailty depends on the number and timing of the observed events for each family.

This approach is the state of art method for dealing with left-truncated correlated survival data in frailty models, however, it still relies on a strong assumption in order to provide valid estimates, namely that $m_i = n_i$ fixed. This can be observed by rewriting the likelihood contribution of family i in terms of the random variable M_i and considering the whole sampling event $\Omega_i = \{t_{i1} \geq t_{0i1}, \ldots, t_{i1} \geq t_{0i1}, M_i = m_i, n_i\}$:

$$L_i = \int_u f_c(t_{i1}, \ldots, t_{im_i}|\mathbf{x}_i, u_i, \Omega_i)dG(u_i|\Omega_i) = \\ \frac{\int_0^\infty \prod_{j=1}^{m_i} f(t_{ij}|t_{ij} \geq t_{0ij}, x_{ij}, u_i)P(t_{ij} \geq t_{0ij}|x_{ij}, u_i)P(M_i = m_i|u_i, n_i)dG(u_i)}{\int_0^\infty \prod_{j=1}^{m_i} P(t_{ij} \geq t_{0ij}|x_{ij}, u_i)P(M_i = m_i|u_i, n_i)dG(u_i)} \quad (10)$$

Note that equation (10) reduces to equation (7) if and only if $P(M_i = m_i|u_i, n_i)$ is assumed to be independent of u_i. In that case, the term $P(M_i = m_i)$ can be moved outside the integrals in both numerator and denominator of equation (10) and it cancels out. However, such assumption requires to consider M_i deterministic and reformulate

the sampling event for family i as $\Omega_i = \{t_{i1} \geq t_{0i1}, \ldots, t_{im_i} \geq t_{0im_i}\}$. This holds in the very special case in which $m_i = n_i$ by design ('all in/all out' selection procedure) and consequently equation (7) leads to correct estimates in that case.

However, the application of the updated approach in samples with partially observed families ($\mathbf{R}_i$ contains at least one zero-entry) leads to biased results since the missing data mechanism associated to the random variable M_i depends, in general, on the frailty term u_i. Under the common assumption of independence between the left-truncation and survival times (the age at entry is independent of the lifetime) and independence between covariates and the frailty term (the covariates are evenly distributed across the population), the frailty term determines the level of within-family selection. Specifically, given two arbitrary families of the same size n_i with different frailty terms and assuming that the recruitment ages of their members has common support, the family with larger u_i is likely to have smaller m_i due to the effect of the frailty on the lifetime of an individual.

We now propose a new method based on the correction of expression (7) which relies on inverse probability weighting for dealing with missing data.

4.3 New proposal: 'weighted' approach

As previously discussed, the updated method is only valid for fully observed families. However, in many applications, our sample consists of a mixture of both, fully and partially observed families. One way to deal with this situation is to remove all partially observed families from the sample, i.e. to consider only the families for which $\mathbf{R}_i = \mathbf{1}$, but this may lead to discarding a substantial part of the data at hand, with an evident cost in efficiency.

Alternatively, we propose to correct the updated method proposed in the Subsection 4.2. accounting for the non-observed individuals in each family by means of inverse probability weighting (IPW).

The general idea of IPW is to weigh the contributions of the observed units in the estimation by the inverse of the probability of being observed. Denote such probabilities by π_i for an arbitrary unit i. If π_i is consistently estimated then the estimation relying on the pseudo-population resulting from weighing each i observation with $1/\pi_i$ provides consistent estimates (see (25–27) and references therein).

In our case, this principle may be applied at the family level to weigh the likelihood contribution of family i given by equation (7), which is actually correctly specified in absence of within-family missing data (i.e., it coincides with equation (10) when $m_i = n_i, i = 1, \ldots, n$).

The general idea of our method is to weigh each family i contributing with $M_i = m_i$ observed individuals with the probability π_i of having observed exactly m_i members. Recall the inclusion indicator for family i $A_i = I\left\{\left[\sum_j^{n_i} I(t_{ij} > t_{0ij} \geq c_0)\right] \geq K\right\}$ introduced in Section 2. We can use the more restrictive inclusion indicator of family i, defined in terms of the m_i observed members out of n_i as $A_i' = I(\Omega_i) = A_i \times I(M_i = m_i, n_i)$ and the maximization problem given by expression (4) can be reformulated in terms of $\pi = (\pi_1, \ldots, \pi_n)$ as follows:

$$\max_{\boldsymbol{\gamma},\boldsymbol{\beta},\theta} \ell(\boldsymbol{\gamma}, \boldsymbol{\beta}, \theta) = \max_{\boldsymbol{\gamma},\boldsymbol{\beta},\theta} \sum_{i=1}^{n} \ell_i(\gamma, \beta, \theta) = \max_{\boldsymbol{\gamma},\boldsymbol{\beta},\theta} \sum_{i=1}^{n} \frac{A_i'}{\pi_i} \log L_i^{UP}(\gamma, \beta, \theta) \qquad (11)$$

In this way, the new pseudo-population resembles a sample of fully-observed families and expression (11) provides consistent estimates under correct specification of the vector of weights $\pi = (\pi_1, \ldots, \pi_n)$. Our approach is motivated by Molenberghs et al. (28) who investigated inverse probability weighting in the context of partially observed longitudinal data. The proof of consistency of the estimator derived from expression (11) in Appendix B shows that their results still apply here.

Since L_i^{UP}, the base of our proposed weighted estimating procedure given by expression (11), is conditioned to $A_i = 1$, we consider an extra weight to account for incomplete selection also conditioned to $A_i = 1$. Specifically, we define $\pi_i = P(M_i = m_i | A_i = 1, n_i)$, $i = 1, \ldots, n$.

Note that the impact of the unobserved frailty term u_i on the selection of family i is captured in L_i^{UP} and the denominator in expressions (6) and (7) can be regarded as an estimate of $P(A_i = 1|u_i)$, which 'weighs' the numerator to correct for the informative selection of families induced by left-truncation.

Assume, without loss of generality, that the first m_i family members are observed while the remaining $n_i - m_i$ are missing, $(t_{i1}, \ldots, t_{im_i}, t_{im_i+1}, \ldots, t_{n_i}) = (\mathbf{t_i}^{obs}, \mathbf{t_i}^{miss})$. As stated before, the missing data mechanism is informative of the family-specific frailty term u, but note that it is missing completely at random (MCAR) given u, i.e., within families, provided usual assumptions such as independence

between random left-truncation and survival times and that covariates are evenly distributed across and within families. This implies that we can assume that the individuals within a family are exchangeable, in the sense that the distribution of any sub-vector of $(t_{i1}, \ldots, t_{in_i})$ is the same than that corresponding to any other sub-vector of equal length, i.e. for a given family i, $E(\mathbf{t_i^{miss}}|u_i) = E(\mathbf{t_i^{obs}}|u_i)$.

In practice, given the conditional MCAR nature of the missingness mechanism within each selected family, we propose to estimate $\pi_i = P(M_i = m_i|A_i = 1, n_i)$ according to a completely at random selection of m_i individuals from the total n_i members of the selected family i. Specifically, we propose the following IPW weights:

$$\pi_i = \binom{n_i}{m_i} \left(\frac{1}{n_i}\right)^{m_i} \left(1 - \frac{1}{n_i}\right)^{n_i - m_i}.$$

4.4 Software implementation

For practical application of the presented methods, we created an user friendly R function, `LTfrailty` which is available from the authors upon request. The function requires the user to introduce clustered survival data set in their standard presentation consisting of the observed survival times, censoring indicator, cluster identifier, and vector of covariates. Additionally, for the weighted approach the cluster size must be provided by the user. The updated approach is implemented in our function using the `parfm` R package (29). The three implemented methods rely on a Weibull specification for the baseline hazard and the same optimization algorithm is used in order to maximize the log-likelihood. Namely, the `optim() R` function was employed, based on a quasi-Newton method (option `method="L-BFGS-B"`). An alternative existing implementation of the updated approach with gamma frailty is the `frailtypack` R package by (30). Moreover, `coxme` uses the naïve approach under a lognormal frailty distribution specification (31, 32).

5 Simulation study

5.1 Simulation setup

A simulation study was conducted to assess the performance of the new method based on weighted pseudo-likelihood and to compare it with the two existing approaches, naïve and updated, in different controlled scenarios intended to mimic relevant

situations in practice. We generated 1000 Monte Carlo trials based on the following theoretical model:

$$\lambda_{ij}(t) = u_i \lambda_0 \exp(1.5 x_{ij}), \quad i = 1, \dots n, \quad j = 1, \dots, n_i$$

where t is the observed survival time, $\lambda_0 = 1$ represents the constant baseline hazard and x is a binary risk factor ($\beta = 1.5$). The latent frailty term u_i shared for the n_i members of a given family i is drawn from a gamma-distribution with mean 1 and variance θ ($\theta = 0.1$, 0.5 and 2 were considered). In order to check the impact of the cluster size on the performance of the three analyzed methods, we compare the results corresponding to populations composed of 'small' clusters ($n_i = 2$), large 'clusters' ($n_i = 8$) and a mixture of both. Left-truncation times were drawn from a uniform distribution with support [0,4] ($C \sim U[0, 4]$). We assumed a 50% of truncated observations and no right-censoring. In terms of sample size, we considered three different situations ($n = 400$, 800 and 1600 clusters).

In our basic simulation setting observations were removed from the analysis if their truncation time was larger than their survival time ($K = 1$ according to definition of A_i of Section). Additionally, we considered the selection criterion used in the LLS ($K = 2$). Note that this corresponds with $K = n_i$ (complete families selection) when considering families of size 2. The complete family selection based on populations containing large families ($n_i = 8$) is omitted. We considered a Weibull specification for the baseline hazard ($\boldsymbol{\gamma} = (\lambda_0, \rho_0)$) which enables to derive close-form expressions for the maximum likelihood estimates for the three studied methods. The explicit expressions for the naïve approach (given by expression (5)), and updated (given by expression (7)) corresponding to the Weibull hazard specification are given as Supplemental material in Appendix A. Standard errors of the parameters were also estimated. For the naïve and updated methods, they were computed as the square roots of the diagonal elements of the observed hessian matrix. For the weighted approach, robust estimates of the standard errors were obtained using a sandwich estimator (see Tsiatis (25) for technical details and Rondeau et al. (17) for application in frailty models). Coverage rates of the 95% confidence intervals for each method are also reported.

5.2 Simulation results

For each of the studied scenarios we provide results on mean estimated relative bias (defined as the difference between the simulated mean and true parameter value divided by the true value), empirical standard deviation and mean square error (MSE) across the 1000 Monte Carlo trials of the variance parameter of the frailty term, θ, the covariate effect, β, baseline parameter λ_0 ($\rho_0 = 1$ is efficiently estimated by the three methods, data not shown), and resulting population survival estimates at $t = 1$. Population hazards and survival probabilities can be estimated from conditional models such as frailty models (11). Specifically, the population survival at time t derived from gamma-frailty models may be expressed as $S_p(t) = \{1 - \theta\Lambda(t)\}^{\frac{-1}{\theta}}$.

In Tables 1-4 we report the performance of the three methods to fit frailty models in presence of left-truncation. Specifically, for each of the simulated scenarios, the estimated relative bias (reBias), standard deviation (SD) and mean square error (MSE) of the variance parameter of the frailty term θ are summarized in Table 1, while the same summary measures for the estimation of the baseline hazard λ_0 are reported in Table 2. Table 3 contains the results regarding the covariate effect β and Table 4 summarizes the performance of the three studied methods in the estimation of $S_p(1)$. Results for the basic selection setting ($K = 1$) are presented in the top part of the tables and the bottom parts show results for $K = 2$ (families are included if at least 2 members are observed). Estimated standard errors and coverage probabilities are given as supplemental material in Appendix C.

Roughly speaking, the overall difference of the naïve and updated approaches mainly depends on the size of the families (n_i), and the level of familial correlation (θ). According to the results presented in the top part of Table 1 ($K = 1$), we observe that the updated method clearly underestimates the frailty variance, and the bias increases with the size of the frailty variance and the size of the clusters, i.e. the bias tends to become more severe in scenarios where the number of selected members by family is variable and, in general, smaller than n_i. This issue appears to be huge in the situation in which the relying population is composed of large families ($n_i = 8$) and $\theta = 2$, where the estimated relative bias of the updated method is larger than 50%. Still, when the population of reference consists of families of small size ($n_i = 2$) and $\theta = 2$, the relative bias is noticeable (around 14%). Note that the bias is systematic since it does

Table 1. Relative bias (reBias), standard deviation (SD) and mean square error (MSE) for $\widehat{\theta}$ along 1000 trials for several family sizes (n_i), selection schemes ($K = k$: family is included if at least k members are observed) and number of families (n). 50% of left-truncated observations.

Selection	n	n_i	θ	Naïve			Updated			Weighted		
K				reBias	SD	MSE	reBias	SD	MSE	reBias	SD	MSE
	400	8	0.1	-0.030	0.022	0.000	-0.082	0.018	0.000	-0.125	0.091	0.008
	800	8	0.1	-0.040	0.016	0.000	-0.080	0.013	0.000	-0.093	0.069	0.005
	1600	8	0.1	-0.040	0.011	0.000	-0.070	0.009	0.001	-0.035	0.049	0.002
	400	8	0.5	-0.040	0.049	0.003	-0.314	0.028	0.025	-0.050	0.167	0.029
	800	8	0.5	-0.040	0.033	0.002	-0.312	0.020	0.025	-0.018	0.113	0.013
	1600	8	0.5	-0.034	0.025	0.001	-0.310	0.014	0.024	-0.012	0.079	0.006
	400	8	2	-0.031	0.134	0.022	-0.508	0.067	1.038	-0.021	0.249	0.064
	800	8	2	-0.029	0.093	0.012	-0.507	0.048	1.031	-0.012	0.178	0.032
	1600	8	2	-0.027	0.069	0.008	-0.509	0.034	1.037	-0.010	0.123	0.015
$K = 1$	400	2	0.1	-0.131	0.060	0.004	-0.061	0.056	0.003	-0.056	0.057	0.003
	800	2	0.1	-0.170	0.045	0.002	-0.032	0.041	0.002	-0.046	0.042	0.002
	1600	2	0.1	-0.122	0.031	0.001	-0.024	0.029	0.001	-0.013	0.029	< 0.001
	400	2	0.5	-0.132	0.093	0.013	-0.056	0.085	0.008	-0.041	0.094	0.009
	800	2	0.5	-0.128	0.068	0.009	-0.064	0.060	0.005	-0.033	0.065	0.004
	1600	2	0.5	-0.122	0.046	0.006	-0.062	0.041	0.003	-0.038	0.044	0.002
	400	2	2	-0.135	0.200	0.113	-0.138	0.171	0.105	-0.078	0.195	0.062
	800	2	2	-0.139	0.143	0.098	-0.143	0.119	0.096	-0.082	0.137	0.045
	1600	2	2	-0.133	0.106	0.082	-0.140	0.085	0.086	-0.085	0.095	0.038
	400	8	0.1	-0.052	0.022	0.004	-0.082	0.018	< 0.001	-0.127	0.095	0.009
	800	8	0.1	-0.041	0.016	0.003	-0.083	0.012	< 0.001	-0.081	0.067	0.005
	1600	8	0.1	-0.030	0.011	< 0.001	-0.073	0.008	< 0.001	-0.068	0.048	0.002
	400	8	0.5	-0.034	0.049	0.003	-0.313	0.028	0.025	-0.062	0.166	0.029
	800	8	0.5	-0.030	0.036	0.001	-0.312	0.020	0.025	-0.015	0.114	0.013
	1600	8	0.5	-0.032	0.025	0.001	-0.311	0.014	0.024	-0.014	0.077	0.006
	400	8	2	-0.023	0.137	0.021	-0.508	0.068	1.034	-0.017	0.255	0.066
	800	8	2	-0.023	0.097	0.012	-0.507	0.048	1.030	-0.010	0.182	0.033
	1600	8	2	-0.024	0.068	0.007	-0.508	0.034	1.033	-0.011	0.127	0.017
$K = 2$	400	2	0.1	-0.100	0.073	0.005	-0.022	0.068	0.005	-0.049	0.060	0.004
	800	2	0.1	-0.100	0.053	0.003	-0.021	0.048	0.002	-0.005	0.044	0.002
	1600	2	0.1	-0.092	0.039	0.002	-0.005	0.034	0.001	0.001	0.031	0.001
	400	2	0.5	0.026	0.123	0.015	-0.004	0.105	0.011	-0.042	0.092	0.009
	800	2	0.5	0.022	0.088	0.008	0.002	0.074	0.005	-0.039	0.061	0.004
	1600	2	0.5	0.024	0.062	0.004	-0.006	0.051	0.003	-0.040	0.043	0.002
	400	2	2	0.078	0.300	0.115	-0.007	0.226	0.051	0.012	0.182	0.034
	800	2	2	0.070	0.210	0.064	< 0.001	0.162	0.026	0.003	0.163	0.026
	1600	2	2	0.072	0.139	0.040	-0.003	0.114	0.013	-0.001	0.114	0.013

not vanish by increasing the sample size. As expected, this problem is solved when considering complete family selection framework, where $m_i = n_i$, as reflected in the bottom part of Table 1 for the situation with $K = 2$ and $n_i = 2$, with values of relative bias inferior to 5%.

In contrast, the naïve method performs reasonably well regarding the estimation of θ in the studied scenarios where $n_i = 8$ (relative bias $< 5\%$). The reason behind this good performance of the naïve method in such settings has been explained in

Subsection 4.1., and comes from the fact that the wrong assumption of equal frailty distribution at $c_0 = 0$ and conditioned to the truncation times approximately holds if the family is large enough. The performance of the naïve method is worse for small families ($n_i = 2$), providing too low estimates of θ with relative bias of around 13% for all the studied values of θ when $K = 1$ (note however, that its performance is still similar to the updated approach). The bottom part of Table 1 reflects the limitations of the naïve method to deal with situations with highly-selected families ($K = 2$, $n_i = 2$, $\theta = 2$), in which we observe a slight overestimate of the frailty variance. As for the updated approach the observed bias does not vanish by increasing the sample size.

The results regarding the estimation of the constant baseline hazard $\lambda_0 = 1$ for the basic setting with $K = 1$ are displayed in the top part of Table 2. We observe that the updated approach systematically overestimates the baseline hazard, while the naïve method underestimates it, consistently across all the considered sample sizes. In both cases, the worst performance scenarios coincide with the worst results in terms of estimation of θ. The updated approach presents relative biases greater than 100% for large cluster size combined with large θ situations ($reBias = 1.175$ for $n_i = 8$, $\theta = 2$, $reBias = 0.339$ for $n_i = 2$, $\theta = 2$), and in general the relative bias is grater than 5% in all the studied situations. The naïve approach presents the worst estimates of the baseline hazard for $n_i = 2$ and $\theta = 2$, however the performance is in general better than for the updated approach (maximum relative bias is smaller than 30%). In terms of variance, the updated approach also provides worse results than the naïve method. Increasing the selection level ($K = 2$, bottom part of Table 2) does not affect the performance of the updated method (no improvement in the estimation of λ_0 is observed), but the naïve method clearly becomes worse, reaching relative bias levels around 50% when applied to highly selected small families ($K = 2$, $n_i = 2$, $\theta = 2$).

The estimation of β (results shown in Table 3) with the updated method is satisfactory, so it seems that even in the cases where the estimation of the baseline hazard and frailty variance are biased, the relative difference among the two groups defined by x is well estimated. In agreement with the results from Tables 1 and 2, the naïve method presents satisfactory results for the large family case ($n_i = 8$). However, its performance with small families is clearly unsatisfactory (the relative bias on the estimate of β is around 10% with $n_i = 2$, $\theta = 2$, for both $K = 1$ and $K = 2$).

Table 2. Relative bias (reBias), standard deviation (SD) and mean square error (MSE) for $\widehat{\lambda}_0$ along 1000 trials for several family sizes (n_i), selection schemes ($K = k$: family is included if at least k members are observed) and number of families (n). 50% of left-truncated observations.

Selection	n	n_i	θ	Naïve			Updated			Weighted		
K				reBias	SD	MSE	reBias	SD	MSE	reBias	SD	MSE
	400	8	0.1	-0.010	0.039	0.002	0.052	0.046	0.005	0.030	0.378	0.144
	800	8	0.1	-0.013	0.027	0.001	0.053	0.035	0.004	0.021	0.238	0.057
	1600	8	0.1	-0.011	0.020	< 0.001	0.053	0.024	0.003	0.006	0.146	0.021
	400	8	0.5	-0.030	0.051	0.004	0.274	0.077	0.081	0.089	0.667	0.452
	800	8	0.5	-0.028	0.038	0.002	0.271	0.055	0.077	0.037	0.365	0.135
	1600	8	0.5	-0.027	0.026	0.001	0.271	0.039	0.075	0.041	0.249	0.064
	400	8	2	-0.039	0.084	0.009	1.183	0.211	1.444	0.472	2.083	4.561
	800	8	2	-0.036	0.060	0.005	1.180	0.140	1.411	0.219	0.860	0.787
	1600	8	2	-0.039	0.042	0.003	1.175	0.096	1.389	0.101	0.516	0.277
$K = 1$	400	2	0.1	-0.017	0.078	0.006	0.011	0.092	0.009	0.006	0.090	0.008
	800	2	0.1	-0.017	0.057	0.004	0.011	0.061	0.004	0.002	0.064	0.004
	1600	2	0.1	-0.019	0.038	0.002	0.009	0.044	0.002	0.008	0.046	0.002
	400	2	0.5	-0.088	0.089	0.016	0.054	0.120	0.017	0.043	0.121	0.017
	800	2	0.5	-0.084	0.058	0.010	0.058	0.081	0.010	0.041	0.084	0.009
	1600	2	0.5	-0.082	0.044	0.009	0.058	0.058	0.007	0.037	0.059	0.005
	400	2	2	-0.225	0.097	0.060	0.343	0.212	0.163	0.232	0.209	0.098
	800	2	2	-0.225	0.068	0.055	0.340	0.149	0.138	0.225	0.148	0.072
	1600	2	2	-0.229	0.048	0.055	0.339	0.108	0.126	0.215	0.102	0.057
	400	8	0.1	-0.009	0.041	0.002	0.054	0.047	0.005	0.018	0.415	0.172
	800	8	0.1	-0.011	0.029	0.001	0.052	0.034	0.004	0.022	0.231	0.054
	1600	8	0.1	-0.012	0.019	0.001	0.054	0.023	0.003	< 0.001	0.151	0.023
	400	8	0.5	-0.029	0.052	0.004	0.269	0.074	0.078	0.067	0.631	0.402
	800	8	0.5	-0.028	0.036	0.002	0.273	0.055	0.077	0.033	0.362	0.132
	1600	8	0.5	-0.030	0.026	0.002	0.271	0.037	0.075	0.029	0.231	0.054
	400	8	2	-0.046	0.085	0.009	1.180	0.193	1.430	0.486	2.199	5.073
	800	8	2	-0.047	0.060	0.006	1.183	0.138	1.417	0.167	0.851	0.752
	1600	8	2	-0.047	0.041	0.004	1.183	0.097	1.409	0.147	0.577	0.353
$K = 2$	400	2	0.1	-0.029	0.102	0.011	0.008	0.089	0.008	0.008	0.092	0.008
	800	2	0.1	-0.031	0.073	0.006	0.009	0.061	0.004	0.003	0.067	0.005
	1600	2	0.1	-0.031	0.053	0.004	0.010	0.043	0.002	0.007	0.045	0.002
	400	2	0.5	-0.153	0.105	0.034	0.056	0.119	0.017	0.040	0.119	0.016
	800	2	0.5	-0.147	0.074	0.027	0.053	0.081	0.009	0.038	0.084	0.008
	1600	2	0.5	-0.148	0.051	0.025	0.053	0.053	0.006	0.036	0.060	0.005
	400	2	2	-0.461	0.093	0.222	0.029	0.219	0.049	0.013	0.209	0.044
	800	2	2	-0.470	0.060	0.224	0.007	0.150	0.022	0.008	0.149	0.022
	1600	2	2	-0.473	0.043	0.225	0.006	0.102	0.012	0.008	0.102	0.011

With regard to the new method based on weights, its performance is less affected for the family size n_i and K and it outperforms the existing methods in terms of relative bias in the estimation of θ and λ_0 in a number of situations. Moreover, for $K = 1$, $n_i = 2$ and $\theta = 2$ the new method is the preferable strategy with regard to the estimation of θ (minimum MSE for all the studied sample sizes). In general, provided that the sample size is large enough the new method presents better results for the estimation of the frailty variance θ than the existing methods (relative bias in the

Table 3. Relative bias (reBias), standard deviation (SD) and mean square error (MSE) for $\widehat{\beta}$ along 1000 trials for several family sizes (n_i), selection schemes ($K = k$: family is included if at least k members are observed) and number of families (n). 50% of left-truncated observations.

Selection	n	n_i	θ	Naïve			Updated			Weighted		
K				reBias	SD	MSE	reBias	SD	MSE	reBias	SD	MSE
	400	8	0.1	-0.004	0.059	0.004	< 0.001	0.059	0.003	0.058	0.409	0.175
	800	8	0.1	-0.002	0.042	0.002	< 0.001	0.0425	0.002	0.017	0.267	0.071
	1600	8	0.1	-0.003	0.030	0.001	< 0.001	0.032	0.001	0.005	0.184	0.034
	400	8	0.5	-0.009	0.063	0.004	0.003	0.060	0.004	0.015	0.263	0.070
	800	8	0.5	-0.011	0.044	0.002	< 0.001	0.043	0.002	0.015	0.189	0.036
	1600	8	0.5	-0.011	0.030	0.001	0.001	0.030	0.001	0.004	0.128	0.016
	400	8	2	-0.017	0.056	0.004	0.009	0.056	0.003	0.007	0.114	0.013
	800	8	2	-0.017	0.040	0.002	0.011	0.039	0.002	0.001	0.081	0.007
	1600	8	2	-0.015	0.029	0.001	0.011	0.029	0.001	0.002	0.057	0.003
$K = 1$	400	2	0.1	-0.003	0.132	0.017	0.001	0.127	0.016	0.003	0.134	0.018
	800	2	0.1	-0.009	0.091	0.009	-0.002	0.091	0.008	0.004	0.097	0.009
	1600	2	0.1	-0.007	0.062	0.004	0.002	0.066	0.004	-0.002	0.070	0.005
	400	2	0.5	-0.047	0.146	0.026	0.011	0.152	0.023	0.008	0.156	0.024
	800	2	0.5	-0.046	0.102	0.015	0.003	0.103	0.011	0.004	0.108	0.012
	1600	2	0.5	-0.048	0.069	0.010	0.001	0.076	0.006	0.002	0.081	0.007
	400	2	2	-0.118	0.157	0.056	0.002	0.162	0.026	0.010	0.176	0.031
	800	2	2	-0.118	0.103	0.042	0.009	0.115	0.013	0.009	0.117	0.014
	1600	2	2	-0.119	0.079	0.038	0.007	0.085	0.007	0.003	0.088	0.007
	400	8	0.1	-0.004	0.059	0.002	0.003	0.060	0.004	0.045	0.391	0.158
	800	8	0.1	-0.004	0.042	0.001	-0.002	0.042	0.002	0.021	0.268	0.073
	1600	8	0.1	-0.004	0.031	< 0.001	-0.000	0.030	0.001	0.007	0.179	0.032
	400	8	0.5	-0.010	0.059	0.004	0.003	0.061	0.004	0.015	0.265	0.071
	800	8	0.5	-0.013	0.041	0.002	0.002	0.044	0.002	0.012	0.176	0.031
	1600	8	0.5	-0.012	0.031	0.001	< 0.001	0.031	0.001	0.009	0.125	0.016
	400	8	2	-0.015	0.059	0.004	0.008	0.060	0.004	0.004	0.112	0.013
	800	8	2	-0.015	0.043	0.002	0.008	0.042	0.002	0.003	0.080	0.006
	1600	8	2	-0.017	0.030	0.002	0.006	0.029	0.001	0.002	0.058	0.003
$K = 2$	400	2	0.1	< 0.001	0.166	0.028	0.008	0.163	0.027	-0.004	0.137	0.019
	800	2	0.1	-0.002	0.117	0.014	< 0.001	0.120	0.014	0.002	0.095	0.009
	1600	2	0.1	-0.007	0.085	0.007	0.003	0.082	0.007	0.003	0.067	0.005
	400	2	0.5	-0.027	0.183	0.035	0.004	0.179	0.032	0.004	0.154	0.024
	800	2	0.5	-0.038	0.133	0.021	0.003	0.129	0.017	0.002	0.108	0.012
	1600	2	0.5	-0.039	0.089	0.011	< 0.001	0.092	0.008	0.001	0.077	0.006
	400	2	2	-0.101	0.176	0.054	0.005	0.185	0.034	0.012	0.182	0.034
	800	2	2	-0.099	0.126	0.038	0.002	0.126	0.016	0.006	0.129	0.017
	1600	2	2	-0.101	0.089	0.031	0.001	0.093	0.009	0.003	0.085	0.007

estimate of θ is lower than 10%). For small sample size ($n = 400$) and small frailty variance ($\theta = 0.1$), we detect a slight underestimation due to lack of information. However, the performance of the new method improves with larger samples (this does not happen with the existing methods for which the bias do not vanish by increasing the sample size).

Similar results were obtained with regard to the estimation of λ_0 and β with the new method. In both cases, we observe a slight overestimation of the true parameters

and large variance when $n = 400$. Even so, the relative bias in the estimate of β is always lower than 10% with the new method and its performance clearly improves when increasing the sample size. For the scenarios with $n_i = 2$ and $\theta = 2$, we observe overestimation of the baseline hazard (even for $n = 1600$) with the new method, but it still outperforms in terms of MSE both updated method (for both $K = 1$, $K = 2$) and naïve method ($K = 2$).

Table 4 shows the performance of the three methods in terms of marginal survival estimates at time $t = 1$ (mean survival time). The estimation of the population survival based on frailty models summarizes the interplay between frailty variance, baseline hazard and covariate effect. We observe that both updated and new methods underestimate the survival probability at $t = 1$ but the new method presents, in general, lower relative bias and MSE than the updated method. The naïve method shows overestimation of the population survival (especially for $n_i = 2$), comparable in magnitude to the performance of the new method for $K = 1$, but it clearly performs poorly for $K = 2$ and $n_i = 2$. Overall, the new method provides the best results in terms of estimation of $S_p(1)$ across the studied scenarios.

With regard to the estimation of the standard errors (Tables S1-S3 in Appendix C), in general terms, the mean estimates are close to the Monte Carlo estimates of the standard deviation of the parameters of interest for the three approaches. As a result, when the estimation is unbiased, the coverage probabilities are close to 0.95. We find an exception in the estimation of β by the naïve approach. The standard errors are systematically overestimated and the resulting coverage probabilities are too large. Finally, as one could expect due to the extra quantity estimated in the weighted approach, the sandwich estimator standard errors tend to provide larger estimates than those provided by the naïve and updated approaches.

Table 4. Relative bias (reBias), standard deviation (SD) and mean square error (MSE) for absolute survival at time 1 along 1000 trials for several family sizes (n_i), selection schemes ($K = k$: family is included if at least k members are observed) and number of families (n). 50% of left-truncated observations.

Selection	n	n_i	θ	Naïve			Updated			Weighted		
K				reBias	SD	MSE	reBias	SD	MSE	reBias	SD	MSE
	400	8	0.1	0.034	0.004	< 0.001	-0.185	0.003	< 0.001	0.085	0.024	< 0.001
	800	8	0.1	0.031	0.003	< 0.001	-0.192	0.002	< 0.001	0.067	0.018	< 0.001
	1600	8	0.1	0.029	0.002	< 0.001	-0.193	0.002	< 0.001	0.051	0.011	< 0.001
	400	8	0.5	0.034	0.009	< 0.001	-0.552	0.005	0.003	0.149	0.064	0.004
	800	8	0.5	0.031	0.007	< 0.001	-0.551	0.003	0.003	0.044	0.040	0.002
	1600	8	0.5	0.035	0.005	< 0.001	-0.552	0.003	0.003	0.004	0.030	< 0.001
	400	8	2	0.012	0.017	< 0.001	-0.716	0.010	0.052	0.070	0.123	0.016
	800	8	2	0.011	0.012	< 0.001	-0.717	0.007	0.052	0.023	0.089	0.008
	1600	8	2	0.012	0.009	< 0.001	-0.717	0.005	0.052	0.005	0.061	0.004
$K = 1$	400	2	0.1	0.017	0.007	< 0.001	-0.029	0.007	< 0.001	-0.024	0.007	< 0.001
	800	2	0.1	0.030	0.005	< 0.001	-0.029	0.005	< 0.001	-0.028	0.005	< 0.001
	1600	2	0.1	0.034	0.004	< 0.001	-0.026	0.003	< 0.001	-0.009	0.004	< 0.001
	400	2	0.5	0.108	0.016	< 0.001	-0.132	0.013	< 0.001	-0.092	0.015	< 0.001
	800	2	0.5	0.116	0.011	0.002	-0.133	0.009	0.002	-0.087	0.010	< 0.001
	1600	2	0.5	0.119	0.008	< 0.001	-0.130	0.007	0.002	-0.084	0.008	< 0.001
	400	2	2	0.118	0.023	0.002	-0.228	0.022	0.006	-0.146	0.023	0.003
	800	2	2	0.118	0.016	0.002	-0.230	0.015	0.006	-0.149	0.016	0.003
	1600	2	2	0.120	0.011	0.002	-0.230	0.011	0.005	-0.147	0.012	0.002
	400	8	0.1	0.029	0.004	< 0.001	-0.196	0.003	< 0.001	0.158	0.025	0.001
	800	8	0.1	0.029	0.003	< 0.001	-0.186	0.002	< 0.001	0.033	0.017	< 0.001
	1600	8	0.1	0.036	0.002	< 0.001	-0.194	0.001	< 0.001	0.041	0.011	< 0.001
	400	8	0.5	0.034	0.010	< 0.001	-0.550	0.005	0.003	0.145	0.066	0.004
	800	8	0.5	0.038	0.007	< 0.001	-0.551	0.004	0.003	0.050	0.038	0.001
	1600	8	0.5	0.036	0.005	< 0.001	-0.551	0.002	0.003	-0.008	0.027	0.001
	400	8	2	0.016	0.017	< 0.001	-0.717	0.010	0.052	0.087	0.129	0.017
	800	8	2	0.014	0.013	< 0.001	-0.718	0.007	0.052	0.041	0.089	0.008
	1600	8	2	0.014	0.009	< 0.001	-0.718	0.005	0.052	-0.007	0.063	0.004
$K = 2$	400	2	0.1	0.085	0.010	< 0.001	-0.004	0.009	< 0.001	0.011	0.007	< 0.001
	800	2	0.1	0.090	0.007	< 0.001	-0.008	0.005	< 0.001	-0.018	0.005	< 0.001
	1600	2	0.1	0.099	0.005	< 0.001	-0.007	0.003	< 0.001	-0.025	0.003	< 0.001
	400	2	0.5	0.357	0.024	0.002	-0.134	0.013	< 0.001	-0.082	0.015	< 0.001
	800	2	0.5	0.374	0.016	0.002	-0.136	0.009	< 0.001	-0.084	0.010	< 0.001
	1600	2	0.5	0.370	0.012	0.001	-0.134	0.006	< 0.001	-0.084	0.007	< 0.001
	400	2	2	0.440	0.029	0.020	-0.231	0.021	0.006	-0.149	0.023	0.003
	800	2	2	0.447	0.022	0.020	0.002	0.021	< 0.001	-0.143	0.016	0.002
	1600	2	2	0.446	0.015	0.020	0.001	0.011	< 0.001	-0.149	0.012	0.002

6 Application to the Leiden Longevity Study

To illustrate the performance of the three methods introduced in Section 4, and discussed in Section 5, we analyzed data from the LLS, introduced in Section 1. The sample contained 404 families with at least two long-lived members , which corresponded to 915 individuals. Most of the sample consisted of pairs of siblings (309 families contributed with 2 members. i.e. 76% of the studied families), but 84 families (21 %) contributed with 3 members, 10 families contributed with 4 members,

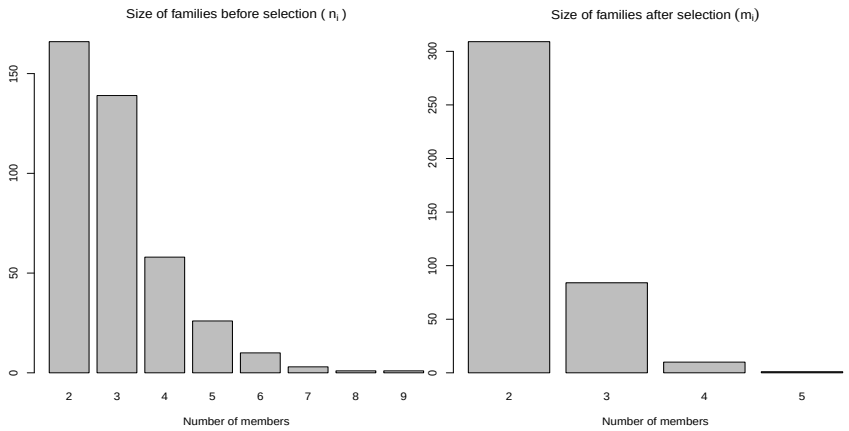

Figure 2. Left: Size of the families at age c_0 ($c_0 = 89$ for men; $c_0 = 91$ for women), i.e., number of members of each family that reached c_0 even if no included in the study. Right: Observed size of the families, i.e., number of siblings with available covariate information.

and 1 family contributed with 5 members. The median age at inclusion for men was 91 years (range: 89-102) and 94 for women (range: 91-103), resulting in a truncation rate which was around 80% for both genders. 7% of the participants were alive by the end of follow-up (February 2014), being the median age of death 95 years (range: 89-106) for men and 98 years (range: 91-108) for women. The genealogical information, i.e. the birth and deceased dates, of the complete sibship of the included families was recovered and used for the calculation of the complete family size at the beginning of follow-up (n_i). As explained in Section 2, we considered the family members whose lifespan was longer than the gender-specific minimum age of entry c_0 (89 for men, 91 for women). Due to the retrospective nature of the sampling, for the siblings with $t > c_0$ but death before sampling it was not possible to determine any covariate, and they were treated as missing data. Moreover, we excluded from the calculations of n_i all those family members too young to determine if their lifespan was longer than the corresponding c_0,i.e., all those family members with $t < c_0$ and $\delta_{ij} = 0$ and all those members died before c_0, i.e., all those family members with $t < c_0$ and $\delta_{ij} = 1$. The size of the families before and after recruitment are presented in Figure 2.

We considered three different models, each fitted with each of the three methods: naïve, updated and weighted. On the one hand, we, consider a null model, without presence of covariates, to specifically focus on the level of familial correlation between lifespans in elderly populations. On the other hand, we separately, evaluated the effect of two binary genetic markers, the indicator of being a carrier of the APOE-ϵ2 and APOE-ϵ4 allele, respectively. In a recent meta-analyses of GWAS studies, Deelen et al. (33) reported a protective effect of APOE-ϵ2 allele in order to survive until old age and evidence of APOE-ϵ4 as a risk factor, while other studies did not find significant effects (34). Note that the design and size of the sample differs among studies, so the quantification of the effect of these variants in terms of hazard ratios for the population of extreme survivors is still not clear.

For six individuals (from six different included families), the information on APOE-ϵ2 and APOE-ϵ4 was missing. We considered that this lack of information was completely at random, i.e. independent of any observed and unobserved variables related to the survival process. Therefore, we removed those cases from the sample and the final effective sample size was 909 individuals, 20% of them were carrier of the APOE-ϵ2 and 17% carried the APOE-ϵ4 allele.

As in the Simulation Study, we considered a Weibull specification for the baseline hazard, and maximum likelihood estimates were derived in terms of the expressions detailed in Appendix A. Estimates of the baseline hazard parameters $\gamma = (\lambda_0, \rho_0)$, θ and β and their respective standard errors are reported in Table 5. As in the Simulation Study, standard errors for the naïve and updated methods were computed as the square roots of the diagonal elements of the observed hessian matrix while for the weighted approach, robust estimates were obtained using a sandwich estimator. Frailty-based estimates of the Kendall's tau between lifespans of family members ($\tau = \frac{\widehat{\theta}}{\widehat{\theta}+2}$) are also reported.

From the results in upper part of Table 5, referring to the null model, we observe that the largest estimated frailty variance is provided by the weighted approach ($\widehat{\theta} = 0.079$), followed by the naïve method ($\widehat{\theta} = 0.069$), while the updated approach provides a lower level of the within family aggregation ($\widehat{\theta} = 0.060$). On the other hand, while the naïve and updated approach provide similar estimates for the baseline hazard at time t given by $\lambda_0 t^{\rho_0}$, the weighted approach provides slightly higher estimates for the baseline hazard in the null model. However, the impact of theses differences is

Table 5. Application to the Leiden Longevity Study (LLS). For each method and model specification estimates of frailty variance ($\widehat{\theta}$) and effect of each genetic marker ($\widehat{\beta}$) and their standard errors are provided.

$Model$		Naïve	Updated	Weigthed
Null	$\widehat{\theta}$ (s.e.)	0.069 (0.041)	0.060 (0.036)	0.079 (0.053)
	$\widehat{\lambda_0}$ (s.e.)	0.040 (0.193)	0.041 (0.008)	0.070 (0.338)
	$\widehat{\rho_0}$ (s.e.)	1.722 (0.048)	1.742 (0.090)	1.496 (0.104)
APOE-ϵ2	$\widehat{\theta}$ (s.e.)	0.065 (0.040)	0.056 (0.035)	0.075 (0.052)
	$\widehat{\beta}$ (s.e.)	-0.246 (0.094)	-0.250 (0.097)	-0.415 (0.185)
	$\widehat{\lambda_0}$ (s.e.)	0.041 (0.192)	0.042 (0.008)	0.075 (0.341)
	$\widehat{\rho_0}$ (s.e.)	1.735 (0.048)	1.750 (0.090)	1.506 (0.103)
APOE-ϵ4	$\widehat{\theta}$ (s.e.)	0.063 (0.040)	0.055 (0.036)	0.061 (0.049)
	$\widehat{\beta}$ (s.e.)	0.183 (0.097)	0.184 (0.099)	0.202 (0.209)
	$\widehat{\lambda_0}$ (s.e.)	0.039 (0.194)	0.040 (0.080)	0.067 (0.349)
	$\widehat{\rho_0}$ (s.e.)	1.721 (0.048)	1.738 (0.090)	1.483 (0.103)

small with regard to both the the estimation of population survival (e.g. $S_p(5) \approx 0.47$ for the weighted approach while $S_p(5) \approx 0.52$ for naïve and updated approaches) and the estimates of within-family lifetimes correlation. Specifically, the corresponding estimated Kendall's tau between the lifespans of members of the same family is $\tau = 0.038$ according to the weighted method, while the naïve and updated methods provide $\tau = 0.033$ and $\tau = 0.029$, respectively.

With regard to the model including the APOE-ϵ2 as covariate (middle part of Table 5), the findings with respect to the frailty variance remain the same than in the null model with the weighted approach providing the largest within-family correlations estimate and the updated approach the lowest. With regard to the estimates of the effect of the APOE-ϵ2, the three methods provide a significant (at a 5% level) protective effect in favor to extreme survival for the carriers of this allele. It is noteworthy that the estimated effect and its corresponding standard deviation provided by the weighted approach ($\widehat{\beta} = -0.415$, $s.e. = 0.185$) is notably larger than those provided by the other methods. This result resemble the simulated scenarios with 400 simulated families, 'large' families ($n_i = 8$) and small frailty variance ($\theta = 0.1$), which may suggest a

slight overestimate of the effect of the APOE-$\epsilon 2$ by the weighted method, as we observed in comparable simulated scenarios. On the other hand, the baseline hazard estimated by the weighted approach is slightly larger than the estimation corresponding to the naïve and updated methods, both provide very close estimates of λ_0 and ρ_0. As in the null cases, the difference in terms of marginal survival after five years of follow-up are are very small (the three methods provide $S_p(5) \approx 0.58$).

The results from the bottom part of Table 5 suggests that even if the three methods identify the APOE-$\epsilon 4$ allele as inversely associated to extreme survival in the elderly. However, its adverse effect is of less magnitude (and not statistically significant at 5% level) than the protective effect of APOE-$\epsilon 2$.

According to these results, we conclude that level of familial correlation in the population of long-lived seems to be low and that the allele APOE-$\epsilon 2$ presents a protective effect for extreme survival. The identification of APOE-$\epsilon 4$ as a risk factor acting against survival in our target population of long-lived remains unclear. As the level of within-family correlation seems to be low, the differences among methods are, overall, small. The sample size is a limitation of the Leiden Longevity Study, especially for the application of the weighted approach which seems require larger sample sizes to provide valid estimates when the frailty term is small and the sample consists of clusters containing more than two members.

7 Discussion

In this paper, we have revisited the problem of inference of frailty models with left-truncated and clustered survival data. Our methodological research was motivated by epidemiological questions from the framework of aging research. Namely, we are interested in the study of extreme survival based on family-based cohorts of siblings, such as the GEHA (Genetics of Healthy Ageing) project, or in particular, the LLS. In this context, dealing with left-truncation by death due to retrospective sampling may play an important role.

The first of the analyzed methods to deal with these type of data, the naïve approach, handles left-truncation by adapting risk sets at individual level. However, the outcome-dependent selection related to left-truncation provokes that families with larger values of the frailty term are less likely to be observed. This issue is ignored

by the naïve approach. Alternatively, the second of the revisited methods, the updated approach, takes into account delayed-entry at the frailty distribution level, and hence, the frailty-dependent selection of families. However, it relies on a complete-family observation assumption, i.e., all the members of each family are observed, even if not from the origin of the follow-up time. To overcome the limitations of the existing approximations, we have proposed an inverse probability correction based on the updated method. Specifically, we have proposed family-based weights to account for the within-family selection process, in such a way that the resulting weighted sample satisfies the assumptions of the updated approach. The weights calculation relies on the original family size (n_i) and the assumption of completely at random missing data at each family. The new method is interesting since it is conceptually simple and it can be easily implemented. It only requires the computation of the weights for each family and to conduct a weighted regression based on existing methods.

According to our results, the naïve approach outperforms the updated approximation when the underlying population is composed of large clusters, while the updated approach seems to be appropriated in complete-family designs (e.g.: twin studies) or in situations where the underlying target population is composed of small families. Interestingly, the updated approach provides unbiased estimates of the regression coefficient in all the studied situations, which indicates that it is an appropriated method when the interest specifically relies on estimating the conditional effect of a given marker. However, this is at the cost of introducing bias in the estimation of the baseline hazard and the frailty variance, which may have a big impact in the estimates of marker-specific survival, within-family correlation and risk prediction. The new method may outperform the existing approaches, provided that the sample size is large enough and, specially, when the level of within-family correlation is large. As a limitation, we have observed that the new method provides biased estimates of the covariate effects when the sample does not provide enough information to correctly estimate the weights, i.e., when applied to relatively small samples (< 800 families) in combination with low within-family correlation. This may be improved by using external information based on population mortality tables. In Tsonaka et al. (35), a penalty term based on the disease prevalence is introduced in the context of maximum likelihood estimation in logistic regression with selected families. Following the same idea in the context of frailty models, we could incorporate a penalty to guarantee a

given value of overall survival. As mentioned in Subsection 5.2., from the frailty model, we can estimate the population survival at a given time t as $S_p(t) = \{1 - \theta\Lambda(t)\}^{\frac{-1}{\theta}}$. Since the population survival is often available in population-based registries, one could introduce a penalization term over the difference between the estimated and the registry-based values. This is left as future research.

Our application to the Leiden Longevity Study suggests an underestimate of the level of within family correlation with the updated method, which appears to be corrected by the new method based on weights, and to a lesser extent, by the naïve method. Overall, it seems that level of the within-family correlation is low in the LLS (~ 0.08). The three methods lead to similar conclusions with regard to the conditional effects of the two studied genetic markers. However, the methods do not agree on the size of the protective effect of the APOE-ϵ2 allele. A large sample size is required to get more insights in this issue.

Both in our Simulation Study and the real data analysis, we have considered a parametric formulation for the baseline hazard, mainly for mathematical convenience which eases the practical implementation of the studied methods. The extension to more flexible settings of the new approach is left as future research. Also beyond the scope of this paper, the estimation of standard errors for the proposed weighted approach needs further research. As noted by (27), the widely used sandwich adjustment used here may be anti-conservative, given that the variability in the estimate of the weights is ignored. Alternative approaches under the sandwich principle, as those suggested by Seaman and White (27) should be investigated. Alternatively, a family-based bootstrapping approach may be adopted, but it is not appropriated in the case of the LLS, due to sample size limitations.

We have considered frailty models, which seems a natural choice in our context, given that we are explicitly interested in the within-family correlation of lifespans. Note that marginal approaches may be also of interest but they provide different interpretations of the estimated parameters in survival analysis. Problems due to informative selection discussed in this paper may also affect the results of marginal approaches, so extensions of the current weighted approach in such direction are currently under investigation.

Finally, we would like to emphasize the importance of analyzing the sampling mechanism that resulted in the left-truncated clustered survival data at hand, in order to choose a proper method to deal with it.

Acknowledgements

Work supported by Grant MIMOmics of the European Union's Seventh Framework Programme (FP7-Health-F5-2012) number 305280.

Appendix

A. Weibull gamma shared-frailty model

Recall the shared frailty proportional hazard model specification introduced in expression (1) in Section 3:

$$\lambda_{ij}(t) = u_i \lambda_0(t, \boldsymbol{\gamma}) \exp(\boldsymbol{\beta}\mathbf{x}_{ij}), \quad i = 1, \dots n, \quad j = 1, \dots, n_i,$$

Assume a Weibull distribution for the baseline hazard with parameters $\boldsymbol{\gamma} = (\lambda_0, \rho_0)$, $\lambda_0, \rho_0 > 0$:

$$\lambda_0(t, \lambda_0, \rho_0) = \lambda_0 \rho_0 t^{\rho_0 - 1}, \quad \Lambda_0(t, \lambda_0, \rho_0) = \lambda_0 t^{\rho_0}$$

then, the likelihood contribution of cluster i under a Weibull shared-frailty specification can be rewritten as:

$$L_i = \prod_{j=1}^{n_i} \left[\lambda_{ij}(t_{ij})\right]^{\delta_{ij}} (-1)^{D_i} \mathcal{L}^{(D_i)} \left[\sum_{j=1}^{n_i} \Lambda(t_{ij})\right] =$$

$$\prod_{j=1}^{n_i} \left[\lambda_0 \rho_0 t_{ij}^{\rho_0 - 1} exp(\boldsymbol{\beta}\mathbf{x}_{ij})\right]^{\delta_{ij}} (-1)^{D_i} \mathcal{L}^{(D_i)} \left[\sum_{j=1}^{n_i} \lambda_0 t_{ij}^{\rho_0} exp(\boldsymbol{\beta}\mathbf{x}_{ij})\right]$$

Assuming that the frailty term u follows a gamma distribution G with mean 1 and variance θ ($\theta > 0$) ($u \sim G = \Gamma(1/\theta, 1/\theta)$), which density is given by:

$$g(u) = \frac{\theta^{-\frac{1}{\theta}} u^{\frac{1}{\theta} - 1} exp(-u/\theta)}{\Gamma(1/\theta)},$$

the corresponding Laplace transform derivatives are given by:

$$\mathcal{L}^{(r)}(s) = (-1)^r E_u\left[u^r \exp(-us)\right] = (-1)^r (1+\theta s)^{-r}\left[\prod_{q=0}^{r-1}(1+q\theta)\right](1+\theta s)^{-1/\theta}$$

As a result, a Weibull gamma shared-frailty specification allows for the following explicit expression of the log-likelihood for cluster i $\ell_i(\lambda_0, \rho_0, \boldsymbol{\beta}, \theta) = \log L_i(\lambda_0, \rho_0, \boldsymbol{\beta}, \theta)$ for the no left-truncated case:

$$\ell_i(\lambda_0, \rho_0, \boldsymbol{\beta}, \theta) = \left\{\sum_{j=1}^{n_i} \delta_{ij}\left[\log(\lambda_0\rho_0 t_{ij}^{\rho_0-1}) + \boldsymbol{\beta}\boldsymbol{x}_{ij}\right]\right\}$$
$$-\left(D_i + \frac{1}{\theta}\right)\log\left\{1+\theta\left[\sum_{j=1}^{n_i}\lambda_0 t_{ij}^{\rho_0} exp(\boldsymbol{\beta}\mathbf{x}_{ij})\right]\right\} + \sum_{q=0}^{D_i-1}\log(1+q\theta)$$

The naïve approach for left-truncated Weibull gamma shared-frailty presented in Subsection 4.1. provides the following expression of the log-likelihood for cluster i $\ell_i^N(\lambda_0, \rho_0, \boldsymbol{\beta}, \theta) = \log L_i^N(\lambda_0, \rho_0, \boldsymbol{\beta}, \theta)$:

$$\ell_i^N(\lambda_0, \rho_0, \boldsymbol{\beta}, \theta) = \left\{\sum_{j=1}^{n_i} \delta_{ij}\left[\log(\lambda_0\rho_0 t_{ij}^{\rho_0-1}) + \boldsymbol{\beta}\boldsymbol{x}_{ij}\right]\right\}$$
$$-\left(D_i + \frac{1}{\theta}\right)\log\left\{1+\theta\left[\sum_{j=1}^{n_i}\lambda_0 t_{ij}^{\rho_0} exp(\boldsymbol{\beta}\mathbf{x}_{ij}) - \sum_{j=1}^{n_i}\lambda_0 t_{0ij}^{\rho_0} exp(\boldsymbol{\beta}\mathbf{x}_{ij})\right]\right\} + \sum_{q=0}^{D_i-1}\log(1+q\theta)$$

The updated approach for left-truncated Weibull gamma shared-frailty presented in Subsection 4.2. provides the following expression of the log-likelihood for cluster i $\ell_i^{UP}(\lambda_0, \rho_0, \boldsymbol{\beta}, \theta) = \log L_i^{UP}(\lambda_0, \rho_0, \boldsymbol{\beta}, \theta)$:

$$\ell_i^{UP}(\lambda_0, \rho_0, \boldsymbol{\beta}, \theta) = \left\{\sum_{j=1}^{n_i} \delta_{ij}\left[\log(\lambda_0\rho_0 t_{ij}^{\rho_0-1}) + \boldsymbol{\beta}\boldsymbol{x}_{ij}\right]\right\}$$
$$-\left(D_i + \frac{1}{\theta}\right)\log\left\{1+\theta\left[\sum_{j=1}^{n_i}\lambda_0 t_{ij}^{\rho_0} exp(\boldsymbol{\beta}\mathbf{x}_{ij})\right]\right\} +$$
$$\sum_{q=0}^{D_i-1}\log(1+q\theta) + \frac{1}{\theta}\log\left\{1+\theta\left[\sum_{j=1}^{n_i}\lambda_0 t_{0ij}^{\rho_0} exp(\boldsymbol{\beta}\mathbf{x}_{ij})\right]\right\}$$

B. Robustness of the weighted approach

Theorem:

If $\pi = (\pi_1, \ldots, \pi_n)$ *are correctly specified, then* $\max_{\boldsymbol{\gamma},\boldsymbol{\beta},\theta} \sum_{i=1}^{n} \frac{A'_i}{\pi_i} \log L_i^{UP}(\gamma, \beta, \theta)$ *produces consistent estimates of* $(\boldsymbol{\gamma}, \boldsymbol{\beta}, \theta)$.

Proof:

Recall the indicator $A'_i = I(\Omega_i) = A_i \times I(M_i = m_i, n_i)$ and the maximization problem resulting from considering the updated approach for dealing with left-truncated gamma-frailty models, introduced in Subsection 4.2. If (and only if) the assumption of complete families holds ($m_i = n_i$, $A'_i = A_i$), solving $\max_{\boldsymbol{\gamma},\boldsymbol{\beta},\theta} \ell^{UP}(\boldsymbol{\gamma}, \boldsymbol{\beta}, \theta) = \max_{\boldsymbol{\gamma},\boldsymbol{\beta},\theta} \sum_{i=1}^{n} A_i \ell_i^{UP}(\gamma, \beta, \theta) = \max_{\gamma,\beta,\theta} \sum_{i=1}^{n} A_i \log L_i^{UP}(\boldsymbol{\gamma}, \boldsymbol{\beta}, \theta)$ provides consistent estimates of $(\gamma, \boldsymbol{\beta}, \theta)$.

Consider the score vector of the first derivative of the log-likelihood given by L^{UP} in expression (7), $U = \sum_{i=1}^{n} A_i U_i(\lambda_0, \rho_0, \boldsymbol{\beta}, \theta | \mathbf{t_{0}}_i, \mathbf{t}_i, \mathbf{x}_i, \boldsymbol{\delta}_i, M_i = n_i) = \sum_{i=1}^{n} A_i U_i(\mathbf{t}_i) = \sum_{i=1}^{n} A_i U_i(t_{i1}, \ldots, t_{in_i})$; $U_i(\mathbf{t}) = \frac{\partial}{\partial(\lambda_0, \rho_0, \boldsymbol{\beta}, \theta)} \log L_i^{UP}(\gamma, \beta, \theta)$. Under $m_i = n_i$ (M_i non-random):

$$E_{\mathbf{t}} \left[\sum_{i=1}^{n} A_i U_i(\mathbf{t_i}) \right] = 0 \tag{S1}$$

Consider now the general situation with M_i random (in general, $m_i < n_i$) and recall the division of the vector of complete survival times of family i in terms of an observed and a missing subvectors: $\mathbf{t}_i = (\mathbf{t_i^{obs}}, \mathbf{t_i^{miss}})$. Note that $\mathbf{t}_i$ is a member of the sampling event Ω_i. Since the missing data procedure is MCAR within clusters, $E(t_{ij}) = E(t_{ij}^{miss})$ and $E_{\mathbf{t_i^{miss}}|\mathbf{t}_i^{obs}} [U_i(\mathbf{t}_i)] = U_i(\mathbf{t_i^{obs}})$. Assume that the inverse probability weights $\pi = (\pi_1, \ldots, \pi_n)$ are correctly specified (i.e. $P(M_i = m_i | n_i, A_i = 1) = \pi_i$) and that they are bounded away from zero. Consider the weighted score vector $\frac{I(M_i = m_i, n_i)}{\pi_i} A_i U_i(\mathbf{t_i^{obs}})$. Consistency of the new method follows from its expectation being 0:

$$E_{\mathbf{t}}\left\{\sum_{i=1}^{n}\frac{I(M_i=m_i,n_i)}{\pi_i}A_iU_i(\mathbf{t}_i^{\mathbf{obs}})\right\}=E_{\mathbf{t}}\left\{\sum_{i=1}^{n}E_{M_i|\mathbf{t_i}}\left[\frac{I(M_i=m_i,n_i)}{\pi_i}A_iE_{\mathbf{t_i^{miss}}|\mathbf{t_i}^{obs}}U_i(\mathbf{t}_i)\right]\right\}=$$

$$E_{\mathbf{t}}\left\{\sum_{i=1}^{n}\left[\frac{E_{M_i|\mathbf{t_i}}(I(M_i=m_i,n_i))}{\pi_i}A_iE_{\mathbf{t_i^{miss}}|\mathbf{t}_i^{obs}}U_i(\mathbf{t}_i)\right]\right\}=$$

$$E_{\mathbf{t}}\left\{\sum_{i=1}^{n}\left[\frac{P(M_i=m_i|n_i,A_i=1)}{\pi_i}A_iE_{\mathbf{t_i^{miss}}|\mathbf{t}_i^{obs}}U_i(\mathbf{t}_i)\right]\right\}=$$

$$\sum_{i=1}^{n}A_iE_{\mathbf{t}_i}\left\{E_{\mathbf{t_i^{miss}}|\mathbf{t}_i^{obs}}\left[U_i(\mathbf{t_i})\right]\right\}=E_{\mathbf{t}}\left[\sum_{i=1}^{n}A_iU_i(\mathbf{t_i})\right]=0,$$

where the last equality follows from expression (S1).

C. Simulation study. Standard errors and coverage probabilities

Table S1. Relative bias (reBias), mean standard error (s.e.) and coverage probabilities (Coverage) for $\widehat{\theta}$ along 1000 trials for several family sizes (n_i), selection schemes ($K = k$: family is included if at least k members are observed) and number of families (n). 50% of left-truncated observations.

Selection	n	n_i	θ	Naïve			Updated			Weighted		
K				reBias	s.e.	Coverage	reBias	s.e.	Coverage	reBias	s.e.	Coverage
	400	8	0.1	-0.040	0.022	0.942	-0.082	0.012	0.904	-0.125	0.060	0.639
	800	8	0.1	-0.040	0.016	0.939	-0.080	0.012	0.887	-0.093	0.053	0.789
	1600	8	0.1	-0.040	0.011	0.929	-0.070	0.009	0.849	-0.035	0.043	0.863
	400	8	0.5	-0.040	0.048	0.905	-0.314	0.028	0.000	-0.050	0.133	0.818
	800	8	0.5	-0.040	0.034	0.902	-0.312	0.020	0.000	-0.018	0.101	0.886
	1600	8	0.5	-0.034	0.024	0.877	-0.310	0.014	0.000	-0.012	0.074	0.935
	400	8	2	-0.031	0.135	0.906	-0.508	0.057	0.000	-0.021	0.251	0.929
	800	8	2	-0.029	0.096	0.898	-0.507	0.040	0.000	-0.012	0.176	0.929
$K = 1$	1600	8	2	-0.027	0.068	0.875	-0.509	0.028	0.000	-0.010	0.125	0.923
	400	2	0.1	-0.131	0.060	0.899	-0.061	0.057	0.933	-0.056	0.054	0.931
	800	2	0.1	-0.170	0.043	0.904	-0.032	0.002	0.940	-0.046	0.041	0.926
	1600	2	0.1	-0.122	0.031	0.916	-0.024	0.029	0.943	-0.013	0.029	0.939
	400	2	0.5	-0.132	0.091	0.842	-0.056	0.084	0.905	-0.041	0.088	0.919
	800	2	0.5	-0.128	0.064	0.791	-0.064	0.059	0.913	-0.033	0.063	0.918
	1600	2	0.5	-0.122	0.045	0.720	-0.062	0.042	0.863	-0.038	0.044	0.920
	400	2	2	-0.135	0.198	0.667	-0.138	0.171	0.607	-0.078	0.193	0.838
	800	2	2	-0.139	0.140	0.509	-0.143	0.120	0.343	-0.082	0.136	0.746
	1600	2	2	-0.133	0.099	0.221	-0.140	0.085	0.103	-0.085	0.096	0.600
	400	8	0.1	-0.052	0.022	0.931	-0.082	0.018	0.908	-0.127	0.060	0.636
	800	8	0.1	-0.041	0.016	0.936	-0.083	0.013	0.873	-0.081	0.056	0.813
	1600	8	0.1	-0.030	0.011	0.936	-0.073	0.009	0.868	-0.068	0.043	0.861
	400	8	0.5	-0.034	0.048	0.919	-0.313	0.028	0.001	-0.062	0.134	0.822
	800	8	0.5	-0.030	0.034	0.907	-0.312	0.020	0.000	-0.015	0.101	0.914
	1600	8	0.5	-0.032	0.024	0.866	-0.311	0.014	0.000	-0.014	0.074	0.915
	400	8	2	-0.023	0.136	0.926	-0.508	0.057	0.000	-0.017	0.247	0.935
	800	8	2	-0.023	0.096	0.885	-0.507	0.040	0.000	-0.010	0.176	0.932
	1600	8	2	-0.024	0.068	0.887	-0.508	0.028	0.000	-0.011	0.124	0.947
	400	2	0.1	-0.100	0.071	0.840	-0.022	0.067	0.910	-0.049	0.060	0.931
	800	2	0.1	-0.100	0.052	0.926	-0.021	0.048	0.922	-0.005	0.046	0.925
$K = 2$	1600	2	0.1	-0.092	0.037	0.933	-0.005	0.034	0.950	0.001	0.033	0.940
	400	2	0.5	0.026	0.123	0.936	-0.004	0.103	0.941	-0.042	0.102	0.921
	800	2	0.5	0.022	0.087	0.962	0.002	0.073	0.948	-0.039	0.072	0.914
	1600	2	0.5	0.024	0.062	0.948	-0.006	0.052	0.942	-0.040	0.051	0.918
	400	2	2	0.078	0.287	0.928	-0.007	0.231	0.937	0.012	0.183	0.949
	800	2	2	0.070	0.201	0.919	< 0.001	0.162	0.947	0.003	0.162	0.950
	1600	2	2	0.072	0.142	0.846	-0.003	0.114	0.946	-0.001	0.114	0.942

Table S2. Relative bias (reBias), mean standard error (s.e.) and coverage probabilities (Coverage) for $\widehat{\lambda}_0$ along 1000 trials for several family sizes (n_i), selection schemes ($K = k$: family is included if at least k members are observed) and number of families (n). 50% of left-truncated observations.

Selection	n	n_i	θ	Naïve			Updated			Weighted		
K				reBias	s.e.	Coverage	reBias	s.e.	Coverage	reBias	s.e.	Coverage
	400	8	0.1	-0.010	0.040	0.943	0.052	0.047	0.817	0.030	0.250	0.816
	800	8	0.1	-0.013	0.028	0.948	0.053	0.033	0.621	0.017	0.195	0.900
	1600	8	0.1	-0.011	0.020	0.926	0.053	0.023	0.372	0.006	0.143	0.916
	400	8	0.5	-0.030	0.053	0.932	0.274	0.077	0.029	0.089	0.375	0.853
	800	8	0.5	-0.028	0.038	0.902	0.271	0.054	0.000	0.037	0.290	0.898
	1600	8	0.5	-0.027	0.027	0.798	0.271	0.038	0.000	0.041	0.212	0.928
	400	8	2	-0.039	0.087	0.928	1.183	0.218	0.000	0.472	0.637	0.812
	800	8	2	-0.036	0.062	0.907	1.180	0.154	0.000	0.219	0.503	0.855
$K = 1$	1600	8	2	-0.039	0.043	0.851	1.175	0.109	0.000	0.101	0.402	0.911
	400	2	0.1	-0.017	0.080	0.951	0.011	0.088	0.953	0.006	0.090	0.944
	800	2	0.1	-0.017	0.057	0.942	0.011	0.061	0.949	0.002	0.064	0.937
	1600	2	0.1	-0.019	0.040	0.909	0.009	0.044	0.948	0.008	0.045	0.942
	400	2	0.5	-0.088	0.095	0.885	0.054	0.115	0.947	0.043	0.114	0.930
	800	2	0.5	-0.084	0.067	0.773	0.058	0.081	0.931	0.041	0.080	0.918
	1600	2	0.5	-0.082	0.047	0.551	0.058	0.057	0.879	0.037	0.057	0.906
	400	2	2	-0.225	0.129	0.552	0.343	0.220	0.740	0.232	0.169	0.712
	800	2	2	-0.225	0.091	0.250	0.340	0.153	0.387	0.225	0.119	0.566
	1600	2	2	-0.229	0.064	0.019	0.339	0.108	0.064	0.215	0.084	0.294
	400	8	0.1	-0.009	0.040	0.944	0.054	0.047	0.796	0.018	0.255	0.815
	800	8	0.1	-0.011	0.028	0.935	0.052	0.033	0.640	0.022	0.201	0.906
	1600	8	0.1	-0.012	0.020	0.919	0.054	0.023	0.367	< 0.001	0.144	0.938
	400	8	0.5	-0.029	0.054	0.930	0.269	0.077	0.024	0.067	0.387	0.865
	800	8	0.5	-0.028	0.038	0.890	0.273	0.054	0.000	0.033	0.296	0.889
	1600	8	0.5	-0.030	0.027	0.816	0.271	0.038	0.000	0.029	0.214	0.949
	400	8	2	-0.046	0.087	0.927	1.180	0.220	0.000	0.489	0.631	0.824
	800	8	2	-0.047	0.062	0.880	1.183	0.155	0.000	0.167	0.521	0.841
$K = 2$	1600	8	2	-0.047	0.044	0.827	1.183	0.109	0.000	0.147	0.398	0.887
	400	2	0.1	-0.029	0.103	0.951	0.008	0.113	0.935	0.008	0.112	0.941
	800	2	0.1	-0.031	0.073	0.936	0.009	0.080	0.938	0.003	0.079	0.949
	1600	2	0.1	-0.031	0.052	0.909	0.010	0.056	0.943	0.007	0.056	0.952
	400	2	0.5	-0.153	0.124	0.804	0.056	0.140	0.949	0.040	0.139	0.939
	800	2	0.5	-0.147	0.087	0.628	0.053	0.099	0.943	0.038	0.098	0.911
	1600	2	0.5	-0.148	0.062	0.303	0.053	0.070	0.945	0.036	0.069	0.888
	400	2	2	-0.461	0.161	0.070	0.029	0.216	0.942	0.013	0.208	0.954
	800	2	2	-0.470	0.113	0.000	0.007	0.148	0.937	0.008	0.147	0.956
	1600	2	2	-0.473	0.080	0.000	0.006	0.103	0.939	0.008	0.103	0.949

Table S3. Relative bias (reBias), mean standard error (s.e.) and coverage probabilities (Coverage) for $\widehat{\beta}$ along 1000 trials for several family sizes (n_i), selection schemes ($K = k$: cluster is included if at least k members are observed) and sample sizes (n). 50% of left-truncated observations.

Selection	n	n_i	θ	Naïve			Updated			Weighted		
K				reBias	s.e.	Coverage	reBias	s.e.	Coverage	reBias	s.e.	Coverage
	400	8	0.1	-0.004	0.242	1.000	< 0.001	0.060	0.953	0.058	0.302	0.812
	800	8	0.1	-0.002	0.165	1.000	< 0.001	0.042	0.944	0.017	0.232	0.910
	1600	8	0.1	-0.003	0.115	1.000	< 0.001	0.030	0.949	0.005	0.170	0.936
	400	8	0.5	-0.009	0.101	0.998	0.003	0.061	0.942	0.015	0.239	0.899
	800	8	0.5	-0.011	0.071	0.996	< 0.001	0.043	0.947	0.015	0.172	0.931
	1600	8	0.5	-0.011	0.050	0.996	0.001	0.030	0.946	0.004	0.123	0.948
	400	8	2	-0.017	0.070	0.973	0.009	0.059	0.952	0.007	0.112	0.942
	800	8	2	-0.017	0.049	0.969	0.011	0.041	0.934	0.001	0.080	0.949
$K = 1$	1600	8	2	-0.015	0.035	0.945	0.011	0.029	0.901	0.002	0.056	0.958
	400	2	0.1	-0.003	4.089	0.958	0.001	0.129	0.951	0.003	0.134	0.936
	800	2	0.1	-0.009	1.131	0.984	-0.002	0.092	0.949	0.004	0.096	0.941
	1600	2	0.1	-0.007	0.562	1.000	0.002	0.065	0.947	-0.002	0.068	0.945
	400	2	0.5	-0.047	0.217	0.996	0.011	0.149	0.953	0.008	0.156	0.951
	800	2	0.5	-0.046	0.149	0.989	0.003	0.105	0.951	0.004	0.109	0.949
	1600	2	0.5	-0.048	0.104	0.989	0.001	0.074	0.943	0.002	0.077	0.947
	400	2	2	-0.118	0.115	0.622	0.002	0.164	0.945	0.010	0.170	0.957
	800	2	2	-0.118	0.081	0.437	0.009	0.116	0.951	0.009	0.119	0.962
	1600	2	2	-0.119	0.057	0.196	0.007	0.082	0.952	0.003	0.085	0.952
	400	8	0.1	-0.004	0.240	1.000	0.003	0.060	0.940	0.045	0.301	0.801
	800	8	0.1	-0.004	0.164	1.000	-0.002	0.042	0.953	0.021	0.231	0.903
	1600	8	0.1	-0.004	0.115	1.000	0.001	0.030	0.949	0.007	0.171	0.919
	400	8	0.5	-0.010	0.101	0.999	0.003	0.061	0.960	0.015	0.240	0.908
	800	8	0.5	-0.013	0.071	0.996	0.002	0.002	0.954	0.012	0.173	0.936
	1600	8	0.5	-0.012	0.050	0.993	< 0.001	0.030	0.930	0.009	0.124	0.915
	400	8	2	-0.015	0.070	0.972	0.008	0.058	0.946	0.004	0.112	0.955
	800	8	2	-0.015	0.049	0.958	0.008	0.041	0.927	0.003	0.080	0.938
	1600	8	2	-0.017	0.035	0.952	0.006	0.029	0.923	0.002	0.056	0.951
	400	2	0.1	< 0.001	4.092	0.944	0.008	0.166	0.945	-0.004	0.166	0.947
	800	2	0.1	-0.002	2.014	0.982	< 0.001	0.118	0.952	0.004	0.118	0.946
$K = 2$	1600	2	0.1	-0.007	0.661	0.997	0.003	0.083	0.947	0.003	0.083	0.947
	400	2	0.5	-0.027	0.253	0.992	0.004	0.182	0.944	0.004	0.183	0.957
	800	2	0.5	-0.038	0.175	0.988	0.003	0.129	0.940	0.002	0.129	0.948
	1600	2	0.5	-0.039	0.121	0.987	< 0.001	0.091	0.951	0.001	0.091	0.948
	400	2	2	-0.101	0.133	0.745	0.005	0.182	0.964	0.012	0.183	0.949
	800	2	2	-0.099	0.094	0.609	0.002	0.128	0.960	0.006	0.128	0.954
	1600	2	2	-0.101	0.066	0.401	0.001	0.090	0.951	0.003	0.090	0.949

References

1. Herskind, A.M, McGue, M., Holm, N.V., Sørensen, T.I.A., Harvald, B., and Vaupel, J.W. (1996). The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. *Human Genetics* **97,** 319–323.

2. Ljungquist, B., Berg, S., Lanke, J., McClearn, G.E., and Pedersen, N.L. (1998). The effect of genetic factors for longevity: a comparison of identical and fraternal twins in the Swedish Twin Registry. *The Journals of gerontology. Series A, Biological sciences and medical sciences* **53,** 441–446.

3. Gavrilova, N.S., Gavrilov, L.A., Evdokushkina,G.N., Semyonova, V.G., Gavrilova, A.L., Evdokushkina, N.N., et al. (1998). Evolution, mutations, and human longevity: European royal and noble families. *Human Biology* **70,** 799–804.

4. Kerber, R.A., O'Brien, E., Smith, K.R., and Cawthon, R.M. (2001). Familial excess longevity in Utah genealogies. *The Journals of gerontology. Series A, Biological sciences and medical sciences* **56,** 130–139.

5. Skytthe, A., Pedersen, N.L., Kaprio, J., Stazi, M.A., Hjelmborg, J.v.B., Iachine, I., Vaupel, J.W., and Christensen, K. (2003). Longevity Studies in GenomEUtwin. *Twin Research* **6,** 448–454.

6. Schoenmaker, M., de Craen, A.J.M., de Meijer, P.H.E.M., Beekman, M., Blauw, G.J., et al. (2006). Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study *Eur J of Human Gen* **14,** 79–84.

7. Houwing-Duistermaat, J.J., Callegaro, A., Beekman, M., Westendorp, R.G., Slagboom, P.E., and van Houwelingen, J.C. (2009). Weighted statistics for aggregation and linkage analysis of human longevity in selected families: The Leiden Longevity Study. *Statistics in Medicine* **28,** 140–151.

8. Skytthe, A., Valensin, S., Jeune, B., Cevenini, E., Balard, F., Beekman, M., et al. (2011). Design, recruitment, logistics, and data management of the GEHA (Genetics of Healthy Ageing) project *Experimental Gerontology* **46,** 934–945.

9. Newman, A.B., Glynn, N.W., Taylor, C.A., Sebastiani, P., Perls, T.T., Mayeux, R., et al. (2011). Health and function of participants in the Long Life Family Study: A comparison with other cohorts. *Aging* **3,** 63–76.

10. Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141–151.

11. Duchateau,L., and Janssen, P. (2008). *The frailty model*. New York: Springer.

12. Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer.

13. Vaupel, J.W., Carey, J.R., Christensen, K., Johnson, T.E., Yashin, A.I., Holm, N.V., et al. (1998). Biodemographic trajectories of longevity. *Science* **280,** 855–860.

14. Tan, Q., Jacobsen, R., Sørensen, M., Christiansen, L., Kruse, T.A., and Christensen, K. (2013). Analyzing age-specific genetic effects on human extreme age survival in cohort-based longitudinal studies. *European Journal of Human Genetics* **21,** 451–454.

15. Nielsen, G. G., Gill R. D., Andersen, P. K., and Srensen, T. I. A. (1992). A Counting Process Approach to Maximum Likelihood Estimation in Frailty Models. *Scandinavian Journal of Statistics* **19,** 25–43.

16. Jensen, H., Brookmeyer, R., Aaby, P., and Andersen, P.K. (2004). Shared Frailty Model for Left-Truncated Multivariate Survival Data. *Working paper, University of Copenhagen*.

17. Rondeau, V., Commenges, D., and Joly, P. (2003). Maximum penalized likelihood estimation in a gamma-frailty model. *Lifetime Data Analysis* **9**, 139–153.

18. van den Berg, G. J., and Drepper, B. (2012). Inference for shared frailty survival models with left truncated data. *Working paper, Institute for Labour Market Policy Evaluation (IFAU)*.

19. Lawless, J. F., and Fong, D. Y. T. (1999). State duration models in clinical and observational studies. *Statistics in Medicine* **18**, 2365–2376.

20. Kvist, K., Andersen, P. K., Angst, J., and Kessing, L. V. (2010). Event dependent sampling of recurrent events *Lifetime Data Analysis* **16**, 580–598.

21. Cortinãs Abrahantes, J., Legrand, C., Burzykowski, T., Janssen, P., Ducrocq, V., and Duchateau, L. (2007). Comparison of different estimation procedures for proportional hazards model with random effects. *Computational Statistics & Data Analysis* **51**, 3913–3930.

22. Rondeau, V., Mauguen, A., Laurent, A., Berr, C., and Helmer, C. (2015). Dynamic prediction models for clustered and interval-censored outcomes: Investigating the intra-couple correlation in the risk of dementia. *Statistical Methods in Medical Research (in press)*.

23. Williams, J.S.(1977). Assumptions for different ascertainment models in human genetics. *Biometrics* **33**, 523–527.

24. Elston, R.C., and Sobel, E. (1979). Sampling considerations in the gathering and analysis of pedigree data. *American Journal of Human Genetics* **31**, 62–69.

25. Tsiatis, A.A. (2006). *Semiparametric theory and missing data*. New York: Springer.

26. Rotnitzky, A. (2009). *Longitudinal data analysis*. New York: Springer.

27. Seaman, S.R., and White, I.R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research* **22**, 278–295.

28. Molenberghs, G., Kenward, M.G., Verbeke, G., and Birhanu, T. (2011). Pseudo-likelihood estimation for incomplete data. *Statistica Sinica* **21**, 187–206.

29. Munda, M., Rotolo, F., and Legrand, C. (2012). parfm: Parametric Frailty Models in R *Journal of Statistical Software* **51**, 1–20.

30. Rondeau, V., Mazroui, Y., and González, J.R. (2012). 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* **17**, 1–28.

31. Ripatti, S., and Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics* **56**, 1016–1022.

32. Therneau, T. M., Grambsch, P.M., and Pankratz, V.S. (2003). Penalized survival models and frailty. *Journal of Computational and Graphical Statistics* **12**, 156–175.

33. Deelen, J., Beekman, M., Uh, H.W., Helmer, Q., Kuningas, M., et al. (2011). Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell* **10**, 686–698.

34. Lindahl-Jacobsen, R., Tan Q., Mengel-From, J., Christensen, K., Nebel, A., and Christiansen, L. (2013). Effects of the APOE 2 allele on mortality and cognitive function in the oldest old. *J. Gerontol. A Biol. Sci. Med. Sci.* **68**, 389–394.

35. Tsonaka, R., de Visser, M.C.H., and Houwing-Duistermaat, J. (2013). Estimation of genetic effects in multiple cases family studies using penalized maximum likelihood methodology. *Biostatistics* **14**, 220–231.