An accelerated failure time regression model for illness–death data: A frailty approach

Lea Kats | Malka Gorfine

Abstract
This work presents a new model and estimation procedure for the illness–death survival data where the hazard functions follow accelerated failure time (AFT) models. A shared frailty variate induces positive dependence among failure times of a subject for handling the unobserved dependency between the nonterminal and the terminal failure times given the observed covariates. The motivation behind the proposed modeling approach is to leverage the well-known interpretability advantage of AFT models with respect to the observed covariates, while also benefiting from the simple and intuitive interpretation of the hazard functions. A semiparametric maximum likelihood estimation procedure is developed via a kernel smoothed-aided expectation-maximization algorithm, and variances are estimated by weighted bootstrap. We consider existing frailty-based illness–death models and place particular emphasis on highlighting the contribution of our current research. The breast cancer data of the Rotterdam tumor bank are analyzed using the proposed as well as existing illness–death models. The results are contrasted and evaluated based on a new graphical goodness-of-fit procedure. Simulation results and data analysis nicely demonstrate the practical utility of the shared frailty variate with the AFT regression model under the illness–death framework.

KEYWORDS
goodness of fit, illness–death model, kernel method, semicompeting risks, shared frailty

1 | INTRODUCTION

The accelerated failure time (AFT) model (Kalbfleisch & Prentice, 2002) is a well-known alternative to the popular Cox proportional hazards (PH) model (Cox, 1972). The major advantage of AFT over Cox PH is that the former is more intuitively interpretable (Wei, 1972; Cox, 1997 among others). This work focuses on the illness–death model, which is specified by AFT models for the involved transitions. In illness–death model subjects start at State 0 (e.g., healthy) and then move to State 2 (e.g., death) directly, or transit first to State 1 (e.g., the age at diagnosis of the disease under study) and then to State 2. (See Figure 1).

Xu et al. (2010) suggested an illness–death model with three Cox-based hazard functions. One of their major contributions was the inclusion of a shared gamma-frailty
variate, which acts multiplicatively on each of the hazard functions, with the aim of incorporating unobserved dependence between the time to disease diagnosis and time to death. Lee et al. (2015) adopted the model of Xu et al. (2010), but replaced their semiparametric maximum likelihood estimation procedure with a semiparametric Bayesian estimation approach. Jiang and Haneuse (2017) developed a class of transformation illness–death models that permit a nonparametric specification of the frailty distribution, but to ensure identifiability, their model is restricted to parametric transformation and error distribution. Recently, for a simpler interpretation, Gorfine et al. (2021) proposed a frailty-based illness–death model with Cox-type marginalized hazards that also accommodates delayed entry.

Various estimation procedures have been developed for shared-frailty AFT models for clustered data (without competing or semicompeting risks). Pan (2001) considered clustered survival data with an AFT model and a gamma frailty to characterize the unobserved dependence among cluster members. It is assumed that the shared frailty acts multiplicatively on the hazard function of the error term. Zhang and Peng (2007), Xu and Zhang (2010), Johnson and Strawderman (2012), and Liu et al. (2013) adopted Pan’s model and provided various estimation methods. All these AFT estimation procedures for clustered data are not directly applicable to our illness–death setting, due to the differences in the likelihood functions, as explained in Section 2.3.

Unlike Cox-type models with illness–death framework, AFT models are not well developed. The only work which provides an AFT frailty-based model with illness–death setting is that of Lee et al. (2017). Therein, in contrast to Pan (2001), they used an additive frailty variate in the log-transformed failure time model, and their parametric and semiparametric estimation methods are based on a Bayesian approach. Table 1 summarizes the available frailty-based Cox or AFT models for the illness–death setting.

The current work fills the gap and provides a gamma-frailty illness–death AFT model where the frailty acts multiplicatively on the hazards of the error terms, in the spirit of Pan (2001). We extended the estimation approaches of Zeng and Lin (2007) and Liu et al. (2013) and developed semiparametric maximum likelihood estimators (MLEs) based on a kernel-smoothing technique combined with an expectation-maximization (EM) algorithm. Conceptual differences between our model and that of Lee et al. (2017) will be demonstrated in Section 2.2.

The proposed model and estimation method along with existing methods were applied to the Rotterdam tumor bank of 1546 breast cancer patients, who had node-positive disease and underwent a tumor removal surgery between the years 1978–1993. In this example, date at tumor removal surgery is the entrance time to State 0; date at relapse and date at death are the respective entry times to states 1 and 2. Prognostic variables are age at primary surgery, menopausal status, tumor size, tumor grade, number of positive lymph nodes, levels of estrogen and progesterone receptors in the initial biopsy, hormonal therapy, and chemotherapy. For a comparison of the various models, we extended the goodness-of-fit procedure of Li et al. (2021) to any illness–death model. The results of our proposed goodness-of-fit visualizing procedure nicely demonstrate the utility of the proposed model and estimation procedure.

The remainder of this article is organized as follows. Section 2 describes the proposed gamma frailty-based AFT illness–death regression model, and the estimation method for the regression coefficients, the hazard functions, and the parameter of the gamma-frailty distribution. The illness–death goodness-of-fit procedure is provided in Section 3. In Section 4, we report the results of an extensive simulation study. Section 5 summarizes the analyses of the breast cancer data from the Rotterdam tumor bank, while comparing the proposed AFT approach and various existing AFT and Cox models, with and without frailty. A concluding discussion is provided in Section 6.

## 2 THE MODEL AND METHODS

### 2.1 The proposed multiplicative frailty-based model

Assume a sample of n independent observations. Let $T_{1i}$ and $T_{2i}$ be the times to the nonterminal and the terminal events, respectively, of subject $i$, $i = 1, ..., n$. Let $X_i$ be a time-independent vector of covariates. The illness–death model (Figure 1) is defined by
Illness–death Cox and AFT models, methods and software availability.

| Authors                  | Model                                      | Estimation procedure   | Software                           |
|--------------------------|--------------------------------------------|------------------------|-----------------------------------|
| Xu et al. (2010)         | Cox, gamma frailty, semiparametric         | Semi-parametric MLE    | None                              |
| Lee et al. (2015)        | Cox, gamma frailty, semiparametric         | Bayesian               | R package SemicompRisks           |
| Jiang and Haneuse (2017) | Transformation model, known transformation function, nonparametric frailty at the price of the known error distribution | Semiparametric efficient score | None                              |
| Lee et al. (2017)        | AFT, additive normal frailty, parametric and semiparametric | Bayesian               | R package SemicompRisks           |
| Gorfine et al. (2021)    | Cox, marginalized gamma frailty, semiparametric | Pseudo-likelihood approach | https://github.com/nirkeret/Frailty-LTRC |
| Current work             | AFT, multiplicative gamma frailty, semiparametric | Semi-parametric MLE    | https://github.com/LeaKats/semicompAFT |

\[
\log(T_{1i}) = \beta_{01}^T X_i + \varepsilon_{01i}, \quad T_{1i} > 0, \quad \text{given subject } i \text{ is free of disease}, \tag{1}
\]

\[
\log(T_{2i}) = \beta_{02}^T X_i + \varepsilon_{02i}, \quad T_{2i} > 0, \quad \text{given subject } i \text{ was diagnosed at age } T_{1i} = t_1, \tag{2}
\]

\[
\log(T_{2i}) = \beta_{12}^T X_i + \varepsilon_{12i}, \quad T_{2i} > t_1 > 0, \quad \text{given subject } i \text{ was diagnosed at age } T_{1i} = t_1, \tag{3}
\]

where \( \beta_{jk}, jk \in \{01, 02, 12\} \), is a vector of regression coefficients of transition \( jk \), and \( \varepsilon_{jk} \) are random errors with an unspecified distribution. The fact that \( X_i \) is shared by the three models does not necessitate the use of identical covariates in these models, given that the regression coefficient vectors \( \beta_{jk}, jk \in \{01, 02, 12\} \), are dependent on the transitions. As such, one can exclude a specific covariate by setting its corresponding coefficient to 0. Given that subject \( i \) was diagnosed at age \( T_{1i} = t_1 \), the support of \( T_{2i} \) is restricted by \( t_1 \), so the conditional distribution of \( T_{2i} \) is truncated by \( t_1 \). Model (3) above does not include age at diagnosis, \( t_1 \), as an additional covariate, but instead, the dependence between \( T_{1i} \) and \( T_{2i} \) is incorporated via a shared-frailty model. Given the frailty variate of subject \( i \), denoted by \( \gamma_i \), it is assumed that the respective conditional baseline hazard functions of \( \exp(\varepsilon_{jk}) \), \( jk \in \{01, 02, 12\} \), are given by

\[
\lambda_{01}^o(t | \gamma_i) = \gamma_i h_{01}^o(t), \quad t > 0, \tag{4}
\]

\[
\lambda_{02}^o(t | \gamma_i) = \gamma_i h_{02}^o(t), \quad t > 0, \tag{5}
\]

\[
\lambda_{12}^o(t | \gamma_i) = \gamma_i h_{12}^o(t), \quad t > t_1 > 0, \tag{6}
\]

where each \( h_{jk}^o(\cdot) \) is an unspecified baseline hazard function of \( \exp(\varepsilon_{jk}) \) and \( \gamma_i \) is an unobservable nonnegative random effect, taken to be independent of \( X_i \). It is assumed that \( \gamma_i \) are gamma distributed with mean 1, unknown variance \( \sigma > 0 \), and thus with density \( f(\gamma; \sigma) = \frac{1}{\sigma} e^{-\gamma/\sigma} / \Gamma(\sigma^{-1}) \). We also assume that \( \varepsilon_{01i}, \varepsilon_{02i}, \) and \( \varepsilon_{12i} \) are independent given \( (X_i, \gamma_i) \).

Based on Equations (1)–(6), the conditional hazard functions of the three transitions, given \( (X_i, \gamma_i) \), can be written as

\[
\lambda_{0k}(t | X_i, \gamma_i) = \lim_{\Delta \to 0} \frac{1}{\Delta} \Pr(t \leq T_{ki} < t + \Delta | T_{1i} \geq t, T_{2i} \geq t, X_i, \gamma_i)
\]

\[
= \gamma_i h_{0k}^o \left( t e^{-\beta_{0k}^T X_i} \right) e^{-\beta_{0k}^T X_i}, \quad t > 0, \quad k = 1, 2, \tag{7}
\]

\[
\lambda_{12}(t | X_i, \gamma_i) = \lim_{\Delta \to 0} \frac{1}{\Delta} \Pr(t \leq T_{2i} < t + \Delta | T_{1i} = t_1, T_{2i} \geq t, X_i, \gamma_i)
\]

\[
= \gamma_i h_{12}^o \left( t e^{-\beta_{12}^T X_i} \right) e^{-\beta_{12}^T X_i}, \quad t > t_1 > 0. \tag{7}
\]

For details, see Section S1 of the Web Supplementary Material (WSM).
2.2 Comparison with the additive frailty-based model

Lee et al. (2017) proposed to model the times of the events directly via the following AFT model specification:

\[
\log(T_{1i}) = \beta^TX_i + \gamma_i + \xi_{01}, \quad T_{1i} > 0,
\]

\[
\log(T_{2i}) = \beta^TX_i + \gamma_i + \xi_{02},
\]

\(T_{2i} > 0\), given subject \(i\) is free of disease,

\[
\log(T_{2i}) = \beta^TX_i + \gamma_i + \xi_{12},
\]

\(T_{2i} > t_1 > 0\), given subject \(i\) was diagnosed at age \(T_{1i} = t_1\),

where \(\beta_{jk}, jk \in \{01,02,12\}\), are vectors of transition-specific regression coefficients. The errors \(\xi_{jki}\) are transition-specific random variables with unspecified distributions in the semiparametric setting, or with a normal distribution in the parametric setting. Also, \(\gamma_i\), \(i = 1, ..., n\), are the unobserved normally distributed frailty variates with mean zero and variance \(\sigma^2\) and are assumed to be independent of \(X_i\).

In the above additive frailty approach, the observed covariates \(X_i\) and the unobservable \(\gamma_i\) are included in the models in a similar fashion. In contrast, the popular multiplicative frailty approach of Pan (2001) in the context of clustered data and in Section 2.1 above separates the observed covariates \(X_i\) and the unobservable component \(\gamma_i\). The observed covariates directly affect time to event, and the unobserved component affects the hazard functions of the error terms.

Assume, for example, that \(\exp(\xi_{01})\) and \(\xi_{02}\) are normally distributed with mean \(\mu_0\) and variance \(\sigma^2_{0i}\). Then, the respective conditional hazard functions of the multiplicative and additive models are

\[
\lambda_{01}(t|X_i, \gamma_i) = \frac{\gamma_i e^{-\beta^TX_i}}{\omega_{01}} \phi\left(\frac{\log t - \mu_{01} - \beta_i^TX_i - \gamma_i}{\omega_{01}}\right)
\]

\[
\left\{1 - \Phi\left(\frac{\log t - \mu_{01} - \beta_i^TX_i - \gamma_i}{\omega_{01}}\right)\right\}^{-1}
\]

and

\[
\lambda_{01}(t|X_i, \gamma_i) = \frac{1}{\omega_{01}} \phi\left(\frac{\log t - \mu_{01} - \beta_i^TX_i - \gamma_i}{\omega_{01}}\right)
\]

\[
\left\{1 - \Phi\left(\frac{\log t - \mu_{01} - \beta_i^TX_i - \gamma_i}{\omega_{01}}\right)\right\}^{-1},
\]

where \(\phi(\cdot)\) and \(\Phi(\cdot)\) are the density and cumulative distribution function of the standard normal distribution, respectively. Evidently, \(\lambda_{01}\) admits a simpler interpretation than that of \(\lambda_{01}\) in terms of the unobserved frailty effect. Figure 2 displays \(\lambda_{01}(t|X_i, \gamma_i)\) as a function of \(t\) for \(X_i = 0\) and various combinations of \((\gamma_i, \mu_{01}, \omega_{01})\). This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

The top-left (respectively, bottom) plot in Figure 2 shows that \(\lambda_{01}\) decreases (respectively, increases) as a function of \(\gamma_i\) for any given value of \(t\). The top-right plot indicates that \(\lambda_{01}\) could be a non-monotone function of \(\gamma_i\) for some values of \(t\). In contrast, \(\lambda_{01}\) of the multiplicative frailty model is a monotonic increasing function of the frailty variate \(\gamma_i\) for any error distribution (see Equation 7), and thus the multiplicative-frailty model admits a simpler interpretation for the unobserved frailty effect.

2.3 The proposed estimation method

Our goal is to estimate the unknown set of parameters of the illness–death model \(\Omega = \{\beta, \rho, \sigma\}\). Let \(C_i\) denote the right censoring time of subject \(i\), \(i = 1, ..., n\). Then, the observed data consists of \(n\) independent observations \(\Theta_i = \{V_i, W_i, \delta_{1i}, \delta_{2i}, \delta_{3i}, X_i\}\), where \(V_i = \min(T_{1i}, T_{2i}, C_i)\), \(\delta_{1i} = I(T_{1i} \leq \min(T_{2i}, C_i))\), \(\delta_{2i} = I(T_{2i} \leq \min(T_{1i}, C_i))\), \(\delta_{3i} = \delta_{1i}I(T_{2i} \leq C_i)\). Here \(V_i\) refers to the first observed time, \(\delta_{1i}\) and \(\delta_{3i}\) indicate whether the first observed time was age at disease diagnosis \((T_{1i})\), age at death \((T_{2i})\), or age at censoring; \(W_i\) is age at death or age at censoring after diagnosis, and \(\delta_{3i}\) indicates whether death was observed after diagnosis. It is assumed that the censoring and the failure times are conditionally independent and noninformative, given \((X_i, \gamma_i)\), and observations are identically distributed.

Then, the likelihood function for \(\Omega\) is proportional to

\[
L(\Omega) = \prod_{i=1}^{n} L_i(\Omega) d\gamma_i,
\]

where

\[
L_i(\Omega) = \left\{\frac{\gamma_i e^{-\beta^TX_i}}{\omega_{01}} \phi\left(\frac{\log t - \mu_{01} - \beta_i^TX_i - \gamma_i}{\omega_{01}}\right)\right\}^{\delta_{1i}}
\]

\[
\left\{1 - \Phi\left(\frac{\log t - \mu_{01} - \beta_i^TX_i - \gamma_i}{\omega_{01}}\right)\right\}^{-1}
\]

and

\[
\lambda_{01}(t|X_i, \gamma_i) = \frac{\gamma_i e^{-\beta^TX_i}}{\omega_{01}} \phi\left(\frac{\log t - \mu_{01} - \beta_i^TX_i - \gamma_i}{\omega_{01}}\right)
\]

\[
\left\{1 - \Phi\left(\frac{\log t - \mu_{01} - \beta_i^TX_i - \gamma_i}{\omega_{01}}\right)\right\}^{-1},
\]

and \(\frac{\gamma_i e^{-\beta^TX_i}}{\omega_{01}} \phi\left(\frac{\log t - \mu_{01} - \beta_i^TX_i - \gamma_i}{\omega_{01}}\right)\)

for obtaining semiparametric maximum likelihood estimators (Dempster et al., 1977). It can be verified (see S3
of WSM) that the conditional expectation of the complete log-likelihood given the observed data $\mathcal{O} = \{O_i, i = 1, \ldots, n\}$ and the parameters’ values at the $m$th step, $\hat{\Omega}^{(m)}$, equals

$$E\left(l(\sigma) | \mathcal{O}, \hat{\Omega}^{(m)} \right) + E\left(l(\beta_0, h_0) | \mathcal{O}, \hat{\Omega}^{(m)} \right) + E\left(l(\beta_{12}, h_{12}) | \mathcal{O}, \hat{\Omega}^{(m)} \right), \quad (8)$$

where

$$E\left(l(\sigma) | \mathcal{O}, \hat{\Omega}^{(m)} \right) = \frac{1}{n} \sum_{i=1}^{n} \left( D_i + \frac{1}{\hat{\sigma}} \right)_{1i}^{(m)} - \frac{1}{\hat{\sigma}} \log \sigma - \log \Gamma \left( \frac{1}{\hat{\sigma}} \right), \quad (9)$$

$$E\left(l(\beta_{01}, h_{01}) | \mathcal{O}, \hat{\Omega}^{(m)} \right) = \frac{1}{n} \sum_{i=1}^{n} \left[ \delta_{1i} \left( \log h_{01} (V_i e^{-\beta_{01} T_i X_i}) - \beta_{01} T_i X_i \right) \right] + \hat{H}^{\alpha(m)}_{01} \left( W_i e^{-\beta_{01} T_i X_i} \right)$$

$$E\left(l(\beta_{12}, h_{12}) | \mathcal{O}, \hat{\Omega}^{(m)} \right) = \frac{1}{n} \sum_{i=1}^{n} \left[ \delta_{1i} \left( \log h_{12} (W_i e^{-\beta_{12} T_i X_i}) - \beta_{12} T_i X_i \right) \right] + \hat{H}^{\alpha(m)}_{12} \left( W_i e^{-\beta_{12} T_i X_i} \right)$$

Figure 2 The additive frailty-based AFT model of Lee et al. (2017): conditional hazard of transition from a healthy state to disease diagnosis. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.
and finally \( D_i = \sum_{k=1}^{3} \delta_{ki} \), \( \Gamma(x) \) is the Gamma function and \( \Psi(x) = \Gamma'(x)/\Gamma(x) \) is the digamma function.

The M-step consists of the maximization of the expression in Equation (8). While maximizing \( E(l(\sigma)|\Theta, \hat{\Omega}^{(m)}) \) as a function of \( \sigma \) can be done by gradient-based optimization algorithms, maximizing the other three expectations cannot be done directly with respect to \( (\hat{\beta}_{jk}, h_{jk}^{\circ}) \), \( jk \in \{01, 02, 12\} \), due to a very nonsmooth estimator of the cumulative hazard functions; see Zeng and Lin (2007) for more details under the standard univariate AFT model. Therefore, we aimed to find a smooth alternative and to this aim we extended the kernel-smoothing approach of Zeng and Lin (2007) and Liu et al. (2013) to accommodate our semicompeting risks setting.

We start with a simple case of piecewise-constant hazard functions

\[
\hat{H}^{\circ}_{jk}(t) = \sum_{i=1}^{J_{jk,n}} c_{jk,i} I(t_{jk,i-1} \leq t < t_{jk,i}),
\]

where \( 0 = t_{jk,0} < t_{jk,1} < \cdots < t_{jk,J_{jk,n}} = M_{jk}, \ jk \in \{01, 02, 12\} \), are equally spaced, \( M_{0k} \) are the respective upper bounds for \( V_i \exp(-\hat{\beta}_{jk}^{T}X_i) \), \( k \in \{1, 2\} \), and \( M_{12} \) is the upper bound for \( W_i \exp(-\hat{\beta}_{12}^{T}X_i) \). Then, the cumulative baseline hazard functions are

\[
\tilde{H}^{\circ}_{jk}(t) = \sum_{i=1}^{J_{jk,n}} c_{jk,i}(t - t_{jk,i-1})I(t_{jk,i-1} \leq t < t_{jk,i})
\]

\[
+ \left( M_{jk}/J_{jk,n} \right) \sum_{i=1}^{J_{jk,n}} c_{jk,i} I(t \geq t_{jk,i}).
\]

The functions \( \tilde{H}^{\circ}_{jk} \) and \( \tilde{H}^{\circ}_{jk} \) are plugged in \( E(l(\beta_{jk}, h_{jk}^{\circ})|\Theta, \hat{\Omega}^{(m)}) \) and by maximizing the resulting expression with respect to \( c_{jk,i}, \ l = 1, \ldots, J_{jk,n}, \ jk \in \{01, 02, 12\} \), for a given \( \beta_{jk} \), we are left with a closed-form estimator of \( \hat{c}_{jk,i}^{(m)}, \ c_{jk,i}^{(m)} \). Plugging \( \hat{c}_{jk,i}^{(m)} \) in \( E(l(\beta_{jk}, h_{jk}^{\circ})|\Theta, \hat{\Omega}^{(m)}) \) provides an approximated profile-likelihood function of \( \beta_{jk} \). However, even these profile-likelihood functions are not smooth and have multiple local maxima, and thus an additional smoothing step is required. To this end, it can be shown that each of the profile likelihoods converges to a limit function of \( \hat{\beta}_{jk} \), \( jk \in \{01, 02, 12\} \), as \( n \to \infty \), \( J_{jk,n} \to \infty \) and \( J_{jk,n}/n \to 0 \). Then, for a given kernel function \( K \) with bandwidths \( a_{jk,n} \), the estimators of \( \beta_{jk} \), \( jk \in \{01, 02, 12\} \), are obtained by maximizing a smoothed approximation of the limit function. In the illness–death setting, the age of death after disease diagnosis is truncated by the age at diagnosis. Therefore, the kernel-smoothing approach is adopted to accommodate left truncation. Given the estimators of the regression coefficients, the proposed smoothing approach also yields estimators of the baseline hazard functions. Details of the above summary can be found in Section S4 of the WSM. Here we provide the resulting estimation procedure.

Define \( R^{\circ}_{jk}(\hat{\beta}) = \log V_i - \hat{\beta}^{T}X_i \), \( R^{W}_{jk}(\hat{\beta}) = \log W_i - \hat{\beta}^{T}X_i \). Then, \( \hat{\beta}_{0k} \) is estimated by maximization of \( I^{\circ}_{0k}(\hat{\theta}_{0k}) \), \( k = 1, 2 \), where

\[
I^{\circ}_{0k}(\hat{\theta}_{0k}) = - \frac{1}{n} \sum_{i=1}^{n} \delta_{ki} \log V_i
\]

\[
+ \frac{1}{n} \sum_{i=1}^{n} \delta_{ki} \log \left\{ \frac{1}{n a_{0k,n}} \sum_{j=1}^{J_{0k,n}} \delta_{kj} K \left( \frac{R^{\circ}_{jk}(\hat{\theta}_{0k}) - R^{\circ}_{0k,n}(\hat{\theta}_{0k})}{a_{0k,n}} \right) \right\}
\]

\[
- \frac{1}{n} \sum_{i=1}^{n} \delta_{ki} \log \left\{ \frac{1}{n} \sum_{j=1}^{J_{0k,n}} \delta_{kj} K(s)ds \right\},
\]

(10)

\( \beta_{12} \) is estimated by maximization of

\[
I^{\circ}_{12}(\hat{\theta}_{12}) = - \frac{1}{n} \sum_{i=1}^{n} \delta_{ki} \log W_i
\]

\[
+ \frac{1}{n} \sum_{i=1}^{n} \delta_{ki} \log \left\{ \frac{1}{n a_{12,n}} \sum_{j=1}^{J_{12,n}} \delta_{kj} K \left( \frac{R^{W}_{jk}(\hat{\theta}_{12}) - R^{W}_{12,n}(\hat{\theta}_{12})}{a_{12,n}} \right) \right\}
\]

\[
- \frac{1}{n} \sum_{i=1}^{n} \delta_{ki} \log \left\{ \frac{1}{n} \sum_{j=1}^{J_{12,n}} \delta_{kj} K(s)ds \right\}
\]

(11)

and given \( \hat{\theta}^{(m)}_{jk} \), \( jk \in \{01, 02, 12\} \), the baseline hazard functions are estimated by

\[
\hat{h}^{(m)}_{0k}(t) = \frac{1}{n a_{0k,n}} \sum_{i=1}^{n} \delta_{ki} K \left( \frac{R^{\circ}_{jk}(\hat{\theta}^{(m)}_{0k}) - \log t}{a_{0k,n}} \right) K(s)ds
\]

\[
= \frac{1}{n a_{12,n}} \sum_{i=1}^{n} \delta_{ki} K \left( \frac{R^{W}_{jk}(\hat{\theta}^{(m)}_{12}) - \log t}{a_{12,n}} \right) K(s)ds
\]

(12)

\[
\hat{h}^{(m)}_{12}(t) = \frac{1}{n a_{12,n}} \sum_{i=1}^{n} \delta_{ki} K \left( \frac{R^{W}_{jk}(\hat{\theta}^{(m)}_{12}) - \log t}{a_{12,n}} \right) K(s)ds
\]

(13)

and \( \hat{\beta}^{(m)}_{jk}(t) = \int_{0}^{t} \hat{h}^{(m)}_{jk}(s)ds \). The following is a summary of our EM-based estimation algorithm:

**Step 0 (Initial values):** Set \( m = 0 \) and \( \varepsilon^{(m)}_{1i} = 1 \), \( i = 1, \ldots, n \). (See also Section 2.3.1.)

**Step 1 (E-step):** Set \( m = m + 1 \) and get \( \varepsilon^{(m)}_{1i} \) and \( \varepsilon^{(m)}_{2i} \), \( i = 1, \ldots, n \).

**Step 2 (M-step):** Obtain \( \hat{\theta}^{(m)} \) by maximizing Equation (9), and get
\( \hat{\beta}_{jk}^{(m)} \) by maximizing Equations (10) and (11). Obtain \( \hat{\gamma}_{jk}^{(m)}(t) \) and \( \hat{H}_{jk}^{(m)}(t) \), \( jk \in \{01, 02, 12\} \).

**Step 3:** Repeat Steps 1 and 2 until convergence is reached.

The asymptotic results of Liu et al. (2013, Theorem 1) for clustered data can be extended to establish the asymptotic properties of the proposed illness–death model’s estimators. Assume \( \sigma^*, \beta_{jk}^*, H_{jk}^* \) are the true parameter values, and let \( \Xi = (\sigma, \beta_{01}^T, \beta_{02}^T, \beta_{12}^T)^T \). It can be shown that under the regularity conditions listed in Section S6 of WSM, as \( n a_{jk,n} \to \infty \), \( na_{jk,n}^4 \to 0 \), and \( n \to \infty \), \( \sup_{t \in [0, \tau]} |\hat{\gamma}_{jk}(t) - \gamma_{jk}^*(t)| \to 0 \), \( jk \in \{01, 02, 12\} \), \( \hat{\gamma} \to \Xi \) almost surely, and \( n^{1/2}(\hat{\gamma} - \gamma^*) \) converges to a mean-zero multivariate normal distribution.

### 2.3.1 Initial values

Based on a comprehensive simulation study, our code starts with naive estimates \( \hat{\beta}_{jk}^{(0)} \) based on the rank-based estimates of the R package aftgee (Chiou et al., 2014), \( \hat{\gamma}_{jk}^{(0)}(\cdot) \) are based on Equations (12) and (13) but without frailty, and \( \hat{H}_{jk}^{(0)}(\cdot) \) by adaptive quadrature (e.g., integrate function of R), \( jk \in \{01, 02, 21\} \). Finally, \( \hat{\gamma}^{(0)}(\cdot) \) is set to be between 2 and 5. Given these initial values, the code applies the above EM algorithm but only with respect to \{\( \beta_{01}, \beta_{02}, H_{01}^0, H_{02}^0, H_{12}^0, \sigma \}\} and takes the resulting estimates as the initial values for the above EM algorithm.

### 2.3.2 Bandwidth selection

For the bandwidth parameters \( a_{jk,n} \), we recommend a modified version of the optimal bandwidths of Jones (1990) and Jones and Sheather (1991), in the spirit of Liu et al. (2013). The smoothed profile-likelihood function involves the kernel density for uncensored subjects and the cumulative kernel for all subjects. The recommended bandwidth for the kernel density is \( \xi \hat{\tau}_{jk}(8 \sqrt{2}/3)^{1/3} n_{jk}^{-1/5} \), \( jk \in \{01, 02, 12\} \), where \( \hat{\tau}_{jk} \) is the sample standard deviation of \( \log(T_{jk}) \) among the subjects with observed event time of transition \( j \to k \), denoted by \( n_{jk} \). The recommended bandwidth for the cumulative kernel is \( \xi \hat{\gamma}_{jk}^{4/3} n_{jk}^{1/3} \), where \( \hat{\gamma}_{jk} \) is the sample standard deviation of \( \log(V_j) \) for \( j = 0 \) and \( \log(W_j) \) for \( j = 1 \) among all subjects in state \( j \), denoted by \( n_j \), \( j = 0, 1 \). Based on our extensive simulation study, \( \xi = 0.5 \) is recommended for \( \hat{\beta}_{jk} \), and \( \xi = 0.01 \) for \( H_{jk}^0 \).

### 2.3.3 Variance estimation

In highly censored data, standard bootstrap could produce a low number of distinct event times, which often causes convergence failure. Alternatively, the weighted bootstrap approach can be used (Kosorok et al., 2004). At each bootstrap step, a random weight sampled from the standard exponential distribution is assigned for each observation, and the estimators are derived by the respective weighted log-likelihood functions or weighted sums (see Section S5 of WSM for details).

### 2.3.4 The proposed model and estimation method without frailty

The proposed estimation method can also be implemented under the model described by Equations (1)–(6), where \( y_i \equiv 1 \) for all \( i = 1, \ldots, n \), that is, with no frailty effect. Namely, AFT models with competing risks, while model (3) is an AFT model adjusted for left truncated data, since age at death after disease diagnosis is truncated by the age at diagnosis. Assuming that the functional form of age at diagnosis that affects time to death after diagnosis is known, and frailty is unnecessary, including age at diagnosis as one of the covariates in model (3) is advisable. In such a scenario, the estimation process is simplified to two steps: estimating the vectors of regression coefficients and the baseline hazard function estimators, without requiring an iterative process (see Section S7 of WSM).

## 3 VISUALIZING GOODNESS OF FIT

A goodness-of-fit (GOF) procedure aimed to evaluate how closely observed data mirror expected data under the assumed model. Recently, Li et al. (2021) proposed a GOF method for an arbitrary univariate survival model with right censored data which is based on randomized survival probabilities (RSP). Their key idea is to replace the survival probability of a censored failure time with a uniform random number between 0 and the survival probability at the censored time. They showed that RSPs always uniformly distributed on the (0,1) interval, under the true model. Then, graphical methods for comparing the distribution of the RSPs with the standard uniform distribution could be used to detect a lack of model fit. In contrast, the
TABLE 2  Simulation results with frailty: means of estimates, empirical standard deviations (SD), bootstrap standard errors (SE), and empirical coverage rates.

| σ  | True values | $\hat{\sigma}$ | $\beta_{01,1}$ | $\beta_{01,2}$ | $\beta_{02,1}$ | $\beta_{02,2}$ | $\beta_{12,1}$ | $\beta_{12,2}$ | $\beta_{12,3}$ |
|----|-------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 2  | Mean        | 1.94           | 1.01           | 0.50           | 0.12           | 0.13           | 0.10           | 0.11           | 0.12           |
|    | Empirical SD| 0.25           | 0.09           | 0.12           | 0.08           | 0.07           | 0.08           | 0.07           | 0.09           |
|    | Bootstrap SE| 0.27           | 0.10           | 0.10           | 0.12           | 0.11           | 0.11           | 0.12           | 0.12           |
|    | CR          | 0.97           | 0.96           | 0.93           | 0.93           | 0.97           | 0.97           | 0.96           | 0.97           |
| 1  | Mean        | 0.94           | 1.02           | 0.50           | 1.00           | 0.50           | 0.51           | 0.51           | 1.05           |
|    | Empirical SD| 0.16           | 0.07           | 0.08           | 0.07           | 0.08           | 0.08           | 0.09           | 0.10           |
|    | Bootstrap SE| 0.16           | 0.09           | 0.09           | 0.09           | 0.09           | 0.10           | 0.11           | 0.11           |
|    | CR          | 0.94           | 0.97           | 0.96           | 0.96           | 0.96           | 0.94           | 0.97           | 0.93           |
| 0.5| Mean        | 0.47           | 1.00           | 0.53           | 1.02           | 0.99           | 0.52           | 0.52           | 1.02           |
|    | Empirical SD| 0.11           | 0.06           | 0.07           | 0.07           | 0.07           | 0.07           | 0.08           | 0.11           |
|    | Bootstrap SE| 0.11           | 0.07           | 0.08           | 0.08           | 0.08           | 0.09           | 0.09           | 0.10           |
|    | CR          | 0.96           | 0.97           | 0.96           | 0.97           | 0.97           | 0.97           | 0.97           | 0.95           |

Cumulative baseline hazard functions

| σ  | t       | $H_{01}^\sigma(t)$ | $H_{02}^\sigma(t)$ | $H_{12}^\sigma(t)$ |
|----|---------|--------------------|--------------------|--------------------|
| 2  | 0.10    | 0.01               | 0.01               | 0.01               |
|    | 0.20    | 0.01               | 0.02               | 0.02               |
|    | 0.30    | 0.01               | 0.03               | 0.03               |
|    | 0.40    | 0.01               | 0.04               | 0.04               |
|    | 0.50    | 0.01               | 0.05               | 0.05               |
|    | 0.60    | 0.01               | 0.06               | 0.06               |
|    | 0.70    | 0.01               | 0.07               | 0.07               |
|    | 0.80    | 0.01               | 0.08               | 0.08               |
|    | 0.90    | 0.01               | 0.09               | 0.09               |
|    | 1.00    | 0.01               | 0.10               | 0.10               |
| 1  | 0.10    | 0.01               | 0.04               | 0.09               |
|    | 0.20    | 0.01               | 0.05               | 0.05               |
|    | 0.30    | 0.01               | 0.06               | 0.06               |
|    | 0.40    | 0.01               | 0.07               | 0.07               |
|    | 0.50    | 0.01               | 0.08               | 0.08               |
|    | 0.60    | 0.01               | 0.09               | 0.09               |
|    | 0.70    | 0.01               | 0.10               | 0.10               |
|    | 0.80    | 0.01               | 0.11               | 0.11               |
|    | 0.90    | 0.01               | 0.12               | 0.12               |
|    | 1.00    | 0.01               | 0.13               | 0.13               |

(Continues)
distributions of well-known residuals (e.g., Cox–Snell) under the true model are complicated due to censoring and cannot be characterized clearly with a known distribution, since their distribution depends on the censoring distribution. Hence, there is a lack of reference distributions for conducting the GOF procedure, and the most widely used diagnostic tool is to apply the Kaplan–Meier estimator on the residual; see Li et al. (2021) and references therein. We extend the RSP approach to the illness–death model.

In our setting of the frailty-based AFT illness–death model, the marginal survival functions should be used since the frailties are unobserved. The illness–death model will be examined by two sets of RSPs: (i) the probability of remaining at State 0; and (ii) the probability of remaining at State 1 among those who are diagnosed with the disease. In particular (for a detailed derivation, see Section S8 of WSM),

$$S_{01}^M(t|X_i) = \Pr(T_{1i} > t, T_{2i} > t|X_i)$$

where the superscript M denotes the marginal distribution with respect to the frailty variate. Clearly, in the absence of frailty (i.e., $\gamma_i \equiv 1$ for all $i = 1, ..., n$) we get

$$S_{01}^M(t|X_i) = \exp \left\{-H_{01}^0(te^{-\beta_1 T_{01} X_i}) - H_{02}^0(te^{-\beta_2 T_{02} X_i}) \right\}$$

and

$$S_{12}^M(t|t_1, X_i) = \exp \left\{-H_{12}^0(te^{-\beta_1 T_{12} X_i}) + H_{12}^0(t_1 e^{-\beta_2 T_{12} X_i}) \right\}, \quad t > t_1.$$ 

Now we are in a position to define the RSPs:

$$S_{01}^M(V_i, \delta_{1i}, \delta_{2i}, U_{1i}, X_i) = (\delta_{1i} + \delta_{2i})S_{01}^M(V_i|X_i) + (1 - \delta_{1i} - \delta_{2i})U_{1i}s_{01}^M(V_i|X_i),$$

and

$$S_{12}^M(W_i, V_i, \delta_{12i}, U_{2i}, X_i) = \delta_{3i}s_{12}^M(W_i|V_i, X_i) + (1 - \delta_{3i})U_{2i}s_{12}^M(W_i|V_i, X_i),$$

where $U_{1i}$ and $U_{2i}$ are independent random samples from the standard uniform distribution $U(0,1)$. It can be shown that if the censoring and failure times are independent, given the observed covariates and the frailty variate, the distributions of $S_{01}^M(V_i, \delta_{1i}, \delta_{2i}, U_{1i}, X_i)$ and $S_{12}^M(W_i, V_i, \delta_{12i}, U_{2i}, X_i)$ are uniform over $(0,1)$. (See Section S9 of WSM for a detailed proof.)

Finally, a visualized GOF procedure for any illness–death model could be accomplished by comparing the histograms of the estimates $S_{01}^M(V_i, \delta_{1i}, \delta_{2i}, U_{1i}, X_i)$ and

| Cumulative baseline hazard functions | $\sigma$ | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 | 1.00 |
|-------------------------------------|---------|------|------|------|------|------|------|------|------|------|------|
| $H_{01}^0(t)$                       | 0.01    | 0.04 | 0.09 | 0.16 | 0.25 | 0.36 | 0.49 | 0.64 | 0.81 | 1.00 |      |
| Mean                               | 0.01    | 0.04 | 0.09 | 0.16 | 0.25 | 0.37 | 0.50 | 0.66 | 0.83 | 1.02 |      |
| Empirical SD                       | 0.00    | 0.01 | 0.01 | 0.02 | 0.03 | 0.04 | 0.06 | 0.08 | 0.11 | 0.14 |      |
| Bootstrap SE                       | 0.00    | 0.01 | 0.01 | 0.02 | 0.03 | 0.05 | 0.07 | 0.09 | 0.12 | 0.15 |      |
| CR                                 | 0.92    | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| $H_{02}^0(t)$                       | 0.02    | 0.06 | 0.14 | 0.24 | 0.38 | 0.54 | 0.74 | 0.96 | 1.22 | 1.50 |      |
| Mean                               | 0.02    | 0.06 | 0.14 | 0.24 | 0.38 | 0.54 | 0.74 | 0.97 | 1.22 | 1.50 |      |
| Empirical SD                       | 0.00    | 0.01 | 0.02 | 0.03 | 0.04 | 0.06 | 0.09 | 0.12 | 0.16 | 0.22 |      |
| Bootstrap SE                       | 0.00    | 0.01 | 0.02 | 0.03 | 0.04 | 0.06 | 0.09 | 0.12 | 0.17 | 0.22 |      |
| CR                                 | 0.95    | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| $H_{11}^0(t)$                       | 0.01    | 0.04 | 0.09 | 0.16 | 0.25 | 0.36 | 0.49 | 0.64 | 0.81 | 1.00 |      |
| Mean                               | 0.01    | 0.04 | 0.09 | 0.16 | 0.26 | 0.38 | 0.51 | 0.67 | 0.84 | 1.04 |      |
| Empirical SD                       | 0.00    | 0.01 | 0.02 | 0.03 | 0.04 | 0.06 | 0.07 | 0.10 | 0.12 | 0.14 |      |
| Bootstrap SE                       | 0.00    | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.08 | 0.10 | 0.12 |      |
| CR                                 | 0.95    | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |

0 Abbreviations: CR, confidence interval; SD, standard deviation; SE, standard error.
Simulation results of model misspecification where frailty is ignored: means of estimates, empirical standard deviations, and empirical coverage rates.

| $\sigma$ | $\beta_{01,1}$ | $\beta_{01,2}$ | $\beta_{02,1}$ | $\beta_{02,2}$ | $\beta_{12,1}$ | $\beta_{12,2}$ | $\beta_{12,3}$ |
|---------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 2       | True values    | Mean           | Empirical SD   | CR             |
|         | 1.27           | 1.15           | 1.27           | 0.25           | 0.32           | 0.90           | 0.94           | 0.45           |
|         | 0.10           | 0.15           | 0.15           | 0.32           | 0.31           | 0.94           | 0.45           | 0.50           |
|         | 0.25           | 0.82           | 0.53           | 0.90           | 0.94           |                |                |                |
| 1       | Mean           | 1.19           | 1.08           | 1.17           | 0.37           | 0.46           | 1.63           |                |
|         | 0.09           | 0.10           | 0.09           | 0.16           | 0.16           | 0.19           | 0.36           |                |
|         | 0.48           | 0.87           | 0.56           | 0.86           | 0.93           | 0.08           |                |                |
| 0.5     | Mean           | 1.09           | 1.06           | 1.10           | 0.44           | 0.47           | 1.33           |                |
|         | 0.07           | 0.08           | 0.08           | 0.11           | 0.12           | 0.14           |                |                |
|         | 0.70           | 0.94           | 0.89           | 0.72           | 0.91           | 0.96           | 0.37           |                |

Cumulative baseline hazard functions

| $\sigma$ | $t$ | $0.10$ | $0.20$ | $0.30$ | $0.40$ | $0.50$ | $0.60$ | $0.70$ | $0.80$ | $0.90$ | $1.00$ |
|---------|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 2       | $H_{01}^0(t)$ | Mean  | 0.01   | 0.04   | 0.09   | 0.16   | 0.25   | 0.36   | 0.49   | 0.64   | 0.81   | 1.00   |
|         | Mean  | 0.62   | 0.10   | 0.14   | 0.19   | 0.23   | 0.28   | 0.32   | 0.36   | 0.40   | 0.46   | 0.51   |
|         | 0.96   | 0.73   | 0.31   | 0.06   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   |
| 1       | $H_{02}^0(t)$ | Mean  | 0.02   | 0.06   | 0.14   | 0.24   | 0.38   | 0.54   | 0.74   | 0.96   | 1.22   | 1.50   |
|         | Mean  | 0.22   | 0.11   | 0.17   | 0.23   | 0.29   | 0.35   | 0.40   | 0.46   | 0.51   | 0.56   | 0.64   |
|         | 0.00   | 0.01   | 0.02   | 0.02   | 0.03   | 0.03   | 0.04   | 0.04   | 0.04   | 0.04   | 0.04   | 0.04   |
| 0.5     | $H_{12}^0(t)$ | Mean  | 0.19   | 0.44   | 0.66   | 0.88   | 1.07   | 1.26   | 1.43   | 1.58   | 1.73   | 1.89   |
|         | Mean  | 0.88   | 0.57   | 0.29   | 0.19   | 0.09   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   |
|         | 0.90   | 0.86   | 0.82   | 0.75   | 0.72   | 0.69   | 0.68   | 0.68   | 0.72   | 0.77   |                |                |

(Continues)
of $\hat{S}^M_2(W_i, V_i, \delta_{3i}, U_{2i}, X_i)$ with the expected values under the standard uniform distribution. This procedure will be demonstrated in Section 5.

4 | SIMULATION STUDY

4.1 | Simulation setup

To demonstrate the finite-sample properties of the proposed estimation method, we conducted an extensive simulation study. Failure times were generated from models (1)–(3) with $X_i = (X_{1i}, X_{2i}, X_{3i}, X_{4i})^T$, $\beta_{01} = (1, 0.5, 0, 0)^T$, $\beta_{02} = (0, 1, 1, 0)^T$, $\beta_{12} = (0.5, 0.5, 0, 1)^T$, and a sample size of $n = 1000$. The baseline hazard functions of $\exp(\xi_{jk}(t))$, $j,k \in \{01, 02, 12\}$, were $h_{01}^o(t) = 2t$, $h_{02}^o(t) = 3t$, and $h_{12}^o(t) = 2t$, and $X_{1i}, X_{2i}, X_{3i}, X_{4i}$ were sampled independently such that $X_{1i}, X_{2i}, X_{3i} \sim \text{Uniform}(-1,1)$ and $X_{2i} \sim \text{Bernoulli}(0.5)$. Frailty variates $\gamma_i$ were generated from a gamma distribution with various dependence magnitudes $\sigma = 0.5, 1$, and 2. Failure times, $T_1$ and $T_2$, were generated by solving $U = \exp(\gamma H_{12}^o (X_i e^{-\beta T X_i}))$, $k \in \{1, 2\}$, for $T$, where $U$ is uniformly distributed over (0,1). For those diagnosed with the disease (i.e., $T_1 < T_2$), new values of $T_2$ were generated from the appropriate left-truncated distribution at $T_1$, by solving $U = \exp(-\gamma [H_{12}^o (X_i e^{-\beta T X_i}) - H_{12}^o (T_1 e^{-\beta T X_i})])$ for $T$ and a new random $U$. Censoring times were sampled from $U(0,15)$, such that for $\sigma = 2$ about 16% of observations were censored prior to disease diagnosis or death; and among those diagnosed, about 13% were censored before death. The corresponding censoring rates were 9% and 10% for $\sigma = 1$, and 7% and 8% for $\sigma = 0.5$. The analysis in this work uses the Gaussian kernel with bandwidths values according to Subsection 2.3.2. A range of values for $\xi$ were studied (see Section S10 of WSM). We set $\hat{\sigma}^{[0]} = 2$. Finally, the convergence criteria were $\max_{1 \leq q \leq p}\frac{1}{n_{jk}} \sum_{i=1}^{n_{jk}} [\hat{H}^{(m+1)}_{jk}(f_{jk_i}) - \hat{H}^{(m)}_{jk}(f_{jk_i})] < 0.0001$, $\frac{1}{n_{jk}} \sum_{i=1}^{n_{jk}} [\hat{H}^{(m+1)}_{jk}(f_{jk_i}) - \hat{H}^{(m)}_{jk}(f_{jk_i})] < 0.0001$, $j,k \in \{01, 02, 12\}$, and $|\hat{\sigma}^{(m+1)} - \hat{\sigma}^{(m)}| < 0.0001$, where $p_{jk}$ is the number of components in $\beta_{jk}$, $\hat{\beta}_{jkq}$ denotes the $q$th component of $\beta_{jk}$, $n_{jk}$ is the number of subjects relevant to the transition $j \rightarrow k$, $f_{jk_i} = V_i e^{-\beta_{jk X_i}}$, $j,k \in \{01, 02\}$, and $f_{12i} = W_i e^{-\beta_{12 X_i}}$.

4.2 | Simulation results

Table 2 presents the performance of the proposed estimation method with frailty. They show the empirical mean, empirical standard deviations (SDs), estimated standard errors (SEs), and the empirical coverage rate of 95% Wald-type confidence interval (CR) of the dependence parameter, the regression coefficients, and the cumulative baseline hazard functions at selected time points. Results are based on 100 repetitions. Table 2 indicates that the proposed approach performs well in terms of bias and coverage rates. Moreover, the empirical SDa and the estimated SEs are reasonably close.

Table 3 presents simulation results of model misspecification. The data were generated with a gamma frailty effect, but we applied our estimation procedure without the frailty effect provided in Subsection 2.3.4. Clearly, for high dependence, ignoring the frailty leads to biased estimates and poor coverage rates. For example, with $\sigma = 2$ and true coefficients $\beta_{01} = (1, 0.5)^T$, $\beta_{02} = (1, 1)^T$ and $\beta_{12} = (0.5, 0.5, 0, 1)^T$, the respective mean estimates were $\hat{\beta}_{01} = (1.27, 0.36)^T$, $\hat{\beta}_{02} = (1.15, 1.27)^T$, and $\hat{\beta}_{12} = (0.25, 0.41, 2.09)^T$, with extremely poor coverage rates. In Section S10 of the WSM it is demonstrated that when the true model is free of frailty, our approach proposed in Subsection 2.3.4 performs very well in terms of bias and coverage rate.

Additional simulation results with a smaller sample size of $n = 500$, higher censoring rates, and censoring distribution that depends on the covariates, are included in Section S10 of WSM. The proposed estimation procedure performs very well in terms of bias and coverage rate under these settings as well.

5 | ROTTERDAM TUMOR BANK DATA

5.1 | Data and models

We analyzed the Rotterdam tumor bank of 1546 breast cancer patients, who had node-positive disease and underwent a tumor removal surgery between the years 1978 and 1993; the dataset is available in the survival R package (Therneau, 2021). $T_1$ is the time from surgery to relapse, and $T_2$ is the time from surgery or relapse to death. Of
TABLE 4 Rotterdam Tumor Bank Data: Estimates (Est) or posterior medians (PM), standard errors (SE), exponent of estimated regression coefficients (exp), p-values, and Holm's adjusted p-values for the frequentist approach and credible intervals at credibility level 0.05 for the hazard-ratio parameters under the Bayesian approach. Bold results are significant at 0.05 based on Holm's adjusted p-value.

| Proposed model (zeta=65) | Marginalized Cox (Gorfine et al. 2021) | Conditional Cox (Lee et al. 2015) |
|--------------------------|----------------------------------------|----------------------------------|
|                          | Est (SE) exp p-value Holm                | Est (SE) exp p-value Holm PM (SE) exp Credible interval |
| **σ**                    | 2.18 (0.73) - 0.003 0.058                | 2.52 (0.54) - 0.000 0.000 1.47 (0.23) - (1.046,1.956) |
| **Transition: surgery → relapse** |                                      |                                  |
| Age at surgery (divided by 10) | 0.14 (0.06) 1.15 0.012 0.185 | -0.15 (0.06) 0.86 0.014 0.262 | -0.22 (0.08) 0.80 (0.685,0.918) |
| log of lymph nodes        | -0.40 (0.05) 0.67 0.000 0.000 | 0.42 (0.04) 1.53 0.000 0.000 | 0.71 (0.07) 2.03 (1.795,2.326) |
| log of estrogen+1         | 0.07 (0.03) 1.07 0.030 0.390 | -0.03 (0.02) 0.97 0.186 1.000 | -0.10 (0.04) 0.90 (0.839,0.964) |
| log of progesterone+1     | 0.09 (0.02) 1.09 0.000 0.005 | -0.04 (0.02) 0.96 0.065 1.000 | -0.11 (0.03) 0.90 (0.845,0.958) |
| Postmenopausal (vs. premenopausal) | -0.34 (0.15) 0.71 0.023 0.328 | 0.13 (0.13) 1.14 0.296 1.000 | 0.34 (0.19) 1.40 (0.980,2.081) |
| Tumor size (ref < 20 mm)  |                                      |                                  |
| 20–50 mm                  | -0.32 (0.09) 0.73 0.001 0.015 | 0.20 (0.07) 1.22 0.006 0.116 | 0.40 (0.12) 1.49 (1.180,1.882) |
| > 50 mm                   | -0.49 (0.11) 0.61 0.000 0.000 | 0.38 (0.11) 1.46 0.001 0.020 | 0.79 (0.16) 2.19 (1.625,3.007) |
| Hormone therapy           | 0.60 (0.13) 1.83 0.000 0.000 | -0.38 (0.08) 0.68 0.000 0.000 | -0.88 (0.15) 0.41 (0.310,0.541) |
| Chemotherapy              | 0.49 (0.11) 1.64 0.000 0.000 | -0.37 (0.11) 0.69 0.001 0.023 | -0.79 (0.16) 0.46 (0.329,0.615) |
| Postmenopausal (vs. premenopausal) | -0.25 (0.09) 0.78 0.004 0.081 | 0.21 (0.08) 1.23 0.008 0.155 | 0.44 (0.13) 1.56 (1.216,1.986) |
| **Transition: surgery → death** |                                      |                                  |
| Age at surgery (divided by 10) | -0.43 (0.14) 0.65 0.002 0.051 | 1.32 (0.37) 3.74 0.000 0.009 | 1.43 (0.18) 4.20 (2.987,5.923) |
| log of lymph nodes        | -0.14 (0.08) 0.87 0.091 1.000 | 0.13 (0.12) 1.14 0.298 1.000 | 0.44 (0.15) 1.54 (1.163,2.092) |
| log of estrogen+1         | 0.04 (0.04) 1.04 0.287 1.000 | -0.01 (0.06) 0.99 0.816 1.000 | -0.11 (0.08) 0.89 (0.765,1.040) |
| log of progesterone+1     | 0.01 (0.04) 1.01 0.827 1.000 | 0.08 (0.06) 1.08 0.205 1.000 | 0.01 (0.07) 1.01 (0.884,1.163) |
| Postmenopausal (vs. premenopausal) | -0.15 (0.34) 0.86 0.647 1.000 | -0.30 (0.50) 0.74 0.554 1.000 | -0.35 (0.70) 0.70 (0.179,2.997) |
| Tumor size (ref < 20 mm)  |                                      |                                  |
| 20–50 mm                  | -0.13 (0.15) 0.88 0.376 1.000 | -0.16 (0.25) 0.85 0.526 1.000 | -0.04 (0.28) 0.96 (0.554,1.653) |
| > 50 mm                   | -0.19 (0.18) 0.82 0.275 1.000 | 0.15 (0.31) 1.16 0.634 1.000 | 0.58 (0.35) 1.79 (0.933,3.488) |
| Hormone therapy           | 0.41 (0.18) 1.51 0.019 0.290 | -0.21 (0.25) 0.81 0.389 1.000 | -0.69 (0.29) 0.50 (0.275,0.851) |
| Chemotherapy              | 1.13 (0.30) 3.09 0.000 0.005 | -0.22 (0.81) 0.81 0.789 1.000 | -0.78 (0.63) 0.46 (0.130,1.531) |
| Tumor grade 3 (vs. 2)     | -0.06 (0.13) 0.94 0.641 1.000 | -0.01 (0.28) 0.99 0.961 1.000 | 0.21 (0.27) 1.23 (0.750,2.148) |

(Continues)
| Transition: relapse → death | Proposed model (zeta=65) | Marginalized Cox (Gorfine et al. 2021) | Conditional Cox (Lee et al. 2015) |
|-----------------------------|--------------------------|--------------------------------------|----------------------------------|
|                             | Est (SE) | exp | p-value | Holm | Est (SE) | exp | p-value | Holm | PM (SE) | exp | Credible interval |
| **Age at surgery (divided by 10)** | 0.00 (0.07) | 1.00 | 0.956 | 1.000 | 0.03 (0.08) | 1.03 | 0.700 | 1.000 | 0.08 (0.07) | 1.08 | (0.931,1.232) |
| **log of lymph nodes** | −0.25 (0.07) | 0.78 | 0.000 | **0.010** | 0.25 (0.05) | 1.28 | 0.000 | **0.000** | 0.38 (0.07) | 1.47 | (1.271,1.687) |
| **log of estrogen+1** | 0.04 (0.05) | 1.04 | 0.341 | 1.000 | −0.03 (0.02) | 0.97 | 0.193 | 1.000 | −0.10 (0.04) | 0.90 | (0.838,0.973) |
| **log of progesterone+1** | 0.13 (0.04) | 1.14 | 0.001 | **0.021** | −0.08 (0.02) | 0.92 | 0.000 | **0.003** | −0.19 (0.04) | 0.83 | (0.771,0.884) |
| **Postmenopausal (vs. premenopausal)** | −0.21 (0.17) | 0.81 | 0.203 | 1.000 | −0.05 (0.13) | 0.95 | 0.731 | 1.000 | 0.04 (0.20) | 1.04 | (0.705,1.527) |
| **Tumor size (ref. < 20 mm)** | | | | | | | | | | | |
| **20–50 mm** | −0.37 (0.14) | 0.69 | 0.008 | 0.131 | 0.23 (0.07) | 1.26 | 0.001 | **0.024** | 0.46 (0.14) | 1.58 | (1.234,2.112) |
| **> 50 mm** | −0.52 (0.17) | 0.60 | 0.002 | **0.044** | 0.40 (0.10) | 1.49 | 0.000 | **0.002** | 0.67 (0.18) | 1.96 | (1.405,2.764) |
| **Hormone therapy** | 0.39 (0.14) | 1.48 | 0.005 | 0.090 | −0.18 (0.09) | 0.84 | 0.037 | 0.633 | −0.48 (0.16) | 0.62 | (0.452,0.835) |
| **Chemotherapy** | 0.23 (0.18) | 1.25 | 0.205 | 1.000 | −0.16 (0.13) | 0.85 | 0.227 | 1.000 | −0.18 (0.17) | 0.84 | (0.604,1.179) |
| **Tumor grade 3 (vs. 2)** | −0.26 (0.13) | 0.77 | 0.047 | 0.569 | 0.21 (0.09) | 1.23 | 0.024 | 0.440 | 0.43 (0.14) | 1.54 | (1.177,2.034) |
FIGURE 3  Goodness of fit plots for the illness-death models. Histograms of $\tilde{S}_{\text{M}}(V_i|X_i)$ (left of (A), (C), (E)), $\tilde{S}_{12}^M(W_i,V_i|X_i)$ (left of (B), (D), (F)), and $\tilde{S}_{12}^M(W_i,V_i,\delta_i, U_i, X_i)$ (right of (B), (D), (F)). The dashed lines are the expected values under the uniform distribution.

The 1546 patients, 924 showed a relapse of the disease (63%), 106 died without evidence of relapse (7%), and 771 patients died after a relapse (79% of the patients who showed a relapse of the cancer). The prognostic variables are age at the primary surgery (in years), menopausal status (0 = premenopausal, 1 = postmenopausal), tumor size ($\leq 20$, 20–50, and $> 50$ mm), tumor grade (2 = moderately differentiated, 3 = poorly differentiated), number of positive lymph nodes, estrogen, and progesterone receptors in the initial biopsy (fmol/L), having hormone therapy (0 = no, 1 = yes), and application of chemotherapy (0 = no, 1 = yes).

The following methods were used: the proposed AFT model and estimation procedure with gamma frailty (SEs are based on 500 bootstrap samples and the initial value of $\sigma$ was set to 2), the gamma-frailty Cox model of Lee et al. (2015), and the marginalized gamma-frailty of Gorfine et al. (2021). The AFT additive-frailty model of Lee et al. (2017) is implemented in the R package SemicompRisks (Alvares et al., 2019) with sojourn time $T_2 - T_1$ when death occurs.
after the disease. Applying it to the current data resulted in convergence failure (the potential scale reduction factors that should be less than 1.05 are much larger for most of the parameters). We hypothesize that the convergence failure is due to the use of the sojourn time, which could be negatively correlated with time from surgery to relapse, while a gamma frailty model assumes a positive dependence (see Section S11 of WSM for more details). Additionally, the data were analyzed with the proposed AFT model without frailty and with a Cox illness–death model (see Section S11 of WSM for more details).

5.2 | Results

Table 4 presents the estimates of the frailty-based methods with Holm-adjusted p-values correction for multiplicity (Holm, 1979) for the frequentist approaches. Hereafter, a result is considered significant based on the adjusted p-value and a significance level of 0.05. The proposed model suggests that when holding the other covariates constant, higher age, higher progesterone level, having chemotherapy, and having hormone therapy, each goes with a longer time to relapse after surgery. Also, a higher number of positive lymph nodes, a larger tumor size, and poorly differentiated tumor are related to earlier relapse after the surgery. Hormonal treatment and chemotherapy after surgery are associated with a longer time to subsequent relapse.

In the transition from surgery to death, a higher age and a higher number of positive lymph nodes are related to earlier death after surgery. For the transition from relapse to death, the proposed AFT model implies that a lower number of positive lymph nodes and a higher progesterone level are associated with longer postsurgery survival time, given that the patient experienced relapse. Finally, the proposed AFT model indicates a strong dependence between time to relapse and time to death. The marginalized Cox model also indicates a high level of dependence between relapse and death times, while the conditional Cox model shows a somewhat lower dependence. The directions of the covariates’ effect under the Cox models are similar to each other and those of the proposed AFT model, but inference results based on these three models are somewhat different.

GOF assessment is done by a visual inspection of the histograms of the unmodified estimated survival probabilities and the RSPs histograms, shown in Figure 3. As expected, the unmodified histograms are far from that of a uniform distribution. However, based on the RSPs histogram, it is evident that the proposed AFT model fits well the data and is a better fit to the data, especially in comparison with the marginalized Cox model. Figure S1 in WSM displays the GOF plots of the models that lack frailty. The plots suggest that the models lacking frailty exhibit a poorer fit to the data compared to the models that incorporate frailty.

6 | DISCUSSION

This work makes a dual contribution: First, it offers a new estimation method and a semiparametric model for frailty-based AFT regression in the illness–death framework. Second, it proposes an exploratory technique for assessing the suitability of any illness–death model.

The proposed model allows for covariates and handles potential residual dependency between nonterminal and terminal failure times via a shared frailty. The estimation method is applicable with or without frailty, and simulation results demonstrate the estimators’ good performance in terms of bias and variance. The motivation behind the proposed model is to leverage the interpretability advantage of AFT models concerning observed covariates, while also benefiting from the simple and intuitive interpretation of hazard functions. In contrast, as demonstrated in Section 2.2, interpreting the hazard functions of the model proposed by Lee et al. (2017) could be challenging.

The proposed model and methods can be extended to other types of multi-state models, for example, with multiple nonterminal events and a vector of random effects (frailties) capturing multiple levels of dependence among the event. Additional work is required to extend the proposed estimation method to the case of time-dependent covariates.

ACKNOWLEDGMENTS

The work was supported by the Israel Science Foundation grant number 767/21 and by a grant from the Tel Aviv University Center for AI and Data Science (TAD).

DATA AVAILABILITY STATEMENT

The data that support the findings in this paper are available from the survival R package at https://cran.r-project.org/package=survival (Therneau, 2021).

OPEN RESEARCH BADGES

This article has earned Open Data and Open Materials badges. Data and materials are available as supporting material.

ORCID

Lea Kats https://orcid.org/0000-0002-6623-4497

REFERENCES

Alvares, D., Haneuse, S., Lee, C., & Lee, K.H. (2019) Semicompriks: an R package for the analysis of independent and clustered semi-competing risks data. The R Journal, 11(1), 376.
Chiou, S.H., Kang, S. & Yan, J. (2014) Fitting accelerated failure time models in routine survival analysis with R package afte. Journal of Statistical Software, 61(11), 1–23.

Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B, 39(1), 1–22.

Gorfine, M., Keret, N., Ben Arie, A., Zucker, D. & Hsu, L. (2021) Marginalized frailty-based illness-death model: Application to the UK-Biobank survival data. Journal of the American Statistical Association, 116(535), 1155-1167.

Holm, S. (1979) A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6, 65–70.

Jiang, F. & Haneuse, S. (2017) A semi-parametric transformation frailty model for semi-competing risks survival data. Scandinavian Journal of Statistics, 44(1), 112–129.

Johnson, L.M. & Strawderman, R.L. (2012) A smoothing expectation and substitution algorithm for the semiparametric accelerated failure time frailty model. Statistics in Medicine, 31(21), 2335–2358.

Jones, M. & Sheather, S. (1991) Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. Statistics & Probability Letters, 11(6), 511–514.

Jones, M.C. (1990) The performance of kernel density functions in kernel distribution function estimation. Statistics & Probability Letters, 9(2), 129–132.

Kalbfleisch, J.D. & Prentice, R.L. (2002) The statistical analysis of failure time data, Wiley Series in Probability and Statistics, volume 360. Hoboken, NJ: John Wiley & Sons.

Kosorok, M.R., Lee, B.L., Fine, J.P. (2004) Robust inference for univariate proportional hazards frailty regression models. The Annals of Statistics, 32(4), 1448–1491.

Lee, K.H., Haneuse, S., Schrag, D. & Dominici, F. (2015) Bayesian semi-parametric analysis of semi-competing risks data: investigating hospital readmission after a pancreatic cancer diagnosis. Journal of the Royal Statistical Society. Series C, Applied statistics, 64(2), 253.

Lee, K.H., Rondeau, V. & Haneuse, S. (2017) Accelerated failure time models for semi-competing risks data in the presence of complex censoring. Biometrics, 73(4), 1401–1412.

Li, L., Wu, T. & Feng, C. (2021) Model diagnostics for censored regression via randomized survival probabilities. Statistics in Medicine, 40(6), 1482–1497.

Liu, B., Lu, W. & Zhang, J. (2013) Kernel smoothed profile likelihood estimation in the accelerated failure time frailty model for clustered survival data. Biometrika, 100(3), 741–755.

Pan, W. (2001) Using frailties in the accelerated failure time model. Lifetime Data Analysis, 7(1), 55–64.

Therneau, T.M. (2021) A package for survival analysis in R. R package version 3.2-13.

Xu, J., Kalbfleisch, J.D. & Tai, B. (2010) Statistical analysis of illness–death processes and semicompeting risks data. Biometrics, 66(3), 716–725.

Xu, L. & Zhang, J. (2010) An EM-like algorithm for the semiparametric accelerated failure time gamma frailty model. Computational Statistics & Data Analysis, 54(6), 1467–1474.

Zeng, D. & Lin, D. (2007) Efficient estimation for the accelerated failure time model. Journal of the American Statistical Association, 102(480), 1387–1396.

Zhang, J. & Peng, Y. (2007) An alternative estimation method for the accelerated failure time frailty model. Computational Statistics & Data Analysis, 51(9), 4413–4423.

SUPPORTING INFORMATION
Web Appendices, tables, and figures referenced in Sections 2, 3, 4, and 5 are available with this paper at the Biometrics website on Wiley Online Library. The R code for performing the simulations and data analysis in this paper is available at https://github.com/LeaKats/semicompAFT and it also would be posted with the main paper at the Biometrics website on Wiley Online Library.

How to cite this article: Kats, L. & Gorfine, M. (2023) An accelerated failure time regression model for illness–death data: A frailty approach. Biometrics, 79, 3066–3081.
https://doi.org/10.1111/biom.13880