Shared parameter models under random effects misspecification

Shared parameter models under random-effects misspecification

By DIMITRIS RIZOPOULOS, GEERT VERBEKE

Biostatistical Centre, Catholic University of Leuven, Kapucijnenvoer 35, B–3000 Leuven, Belgium

dimitris.rizopoulos@med.kuleuven.be geert.verbeke@med.kuleuven.be

and GEERT MOLENBERGHS

Center for Statistics, Hasselt University, Agoralaan 1, B–3590 Diepenbeek, Belgium

geert.molenberghs@uhasselt.be

Summary

A common objective in longitudinal studies is the investigation of the association structure between a longitudinal response process and the time to an event of interest. An attractive paradigm for the joint modelling of longitudinal and survival processes is the shared parameter framework where a set of random-effects is assumed to induce their interdependence. In this work, we propose an alternative parameterization for shared parameter models and investigate the effect of misspecifying the random-effects distribution in the parameter estimates and their standard errors.

Some key words: Copula functions; Joint modelling; Sandwich variance estimator.

1. Introduction

In follow-up studies, it is common that each subject provides both a sequence of longitudinal response measurements as well as the time to an event of interest. In such studies, the

main scientific interest may focus on three distinct aspects, i.e., on either the longitudinal process in which the event occurrence causes informative dropout, on the survival process in which the longitudinal measurements are considered as a time-dependent covariate measured with error, or on the association structure between the two processes. Typical examples in this setting include HIV studies, in which longitudinal measurements of CD4 cell counts or the estimated viral load are predictive for the time to onset of clinical AIDS or death, as well as kidney disease studies where longitudinal glomerular filtration rate measurements are predictive for the time to kidney failure.

Shared parameter models (SPMs) (Wu & Carroll, 1988; Wulfsohn & Tsiatis, 1997; Tsiatis & Davidian, 2004) offer an appealing framework for the joint modelling of survival and longitudinal processes. In particular, in SPMs it is assumed that a latent process, expressed by a set of time-invariant random-effects, induces the dependence between the two explicitly observed processes. These random-effects are usually assumed to be normally distributed, even though this choice is not made on the grounds of computational simplicity. Some authors have questioned the Gaussian assumption, in the sense that the resulting inferences can be sensitive to assumptions not easily verifiable from the available data (see e.g., discussion to Scharfstein et al., 1999). To this end, some approaches have been proposed that either relax the distributional assumptions (Song et al., 2002) or make no parametric assumptions at all (Tsiatis & Davidian, 2001) about the random-effects distribution. However, the main empirical result from these approaches is that the parameter estimates are rather robust to random-effects misspecification. Huang et al. (2006) have explored a similar behaviour in structural measurement error models.

In this paper, we consider the SPMs framework and formally investigate the effect of misspecifying the random-effects distribution using two possible parameterizations. In par-

ticular, we show that, as the number of repeated longitudinal measurements per individual grows, the effect of random-effects misspecification vanishes for certain parameters. The intuitive justification for this claim is based on two arguments. First, as the number of repeated measurements per individual increases, the dominating part in the SPMs factorization is the longitudinal measurement model and thus any erroneous assumption about the random-effects distribution is alleviated. Second, as it will be shown, SPMs assume in general a restrictive association structure for the joint distribution of the two processes and this partially explains robustness with respect to the random-effects distribution. Two types of random-effects structure parameterizations are considered, namely either a common set or different sets of random-effects for the two processes. For the second type, we propose a copula representation of the random-effects distribution, allowing for different types of dependence structure between the underlying measurement and survival processes, thus enabling sensitivity analysis regarding the association structure.

The remainder of the paper is organized as follows. In § 2 we present the shared parameter model factorization, discuss some of its features, and show the two possible parameterizations. In § 3, we formally investigate the effect of random-effects misspecification as a function of the number of repeated measurements per individual. In § 4 we describe the results of a simulation study and § 5 considers a real data application.

2. Shared Parameter Models Framework

2·1 *Model Specification*

Let T_i^* denote the true event time for the ith subject and consider a random sample of n subjects ($i = 1, \ldots, n$). Letting C_i denote the underlying potential censoring for subject i, one observes $T_i = \min(T_i^*, C_i)$ and $\delta_i = I(T_i^* \leq C_i)$, where $I(\cdot)$ is the indicator function. Moreover, let $y_i(t_{ij})$ denote the longitudinal measurement for subject i taken at time t_{ij},

$j = 1, \ldots, n_i$. Clearly, $y_i(t_{ij})$ is observed whenever $t_{ij} \leq T_i$, and generally $y_i(T_i)$ is not available. Let $\mathcal{Y}_i = \{y_i(t_{ij}), j = 1, \ldots, n_i\}$ denote the observed longitudinal process for the ith subject. Finally, let b_i represent time-independent random-effects that underly both the longitudinal measurement and survival processes. Under this setting, the shared parameter model is defined as follows:

$$p(\mathcal{Y}_i, T_i; \theta) = \int p(\mathcal{Y}_i \mid b_i; \theta_y)\, p(T_i, \delta_i \mid b_i; \theta_t)\, p(b_i; \theta_b)\, db_i, \tag{1}$$

where $\theta^\top = (\theta_y^\top, \theta_t^\top, \theta_b^\top)$ is the vector containing the parameters of each one of the sub-models, with $\top$ denoting the transpose, and $p(\cdot)$ denoting the appropriate probability density functions. Here $p(T_i, \delta_i \mid b_i; \theta_t) = p_{T_i^*}(T_i \mid b_i; \theta_t)^{\delta_i} S_{T_i^*}(T_i \mid b_i; \theta_t)^{1-\delta_i}$, i.e., it equals either the density for the true event times or the survival function for censored observations. Moreover, we assume that, conditionally on b_i, the longitudinal measurements $\mathcal{Y}_i$ are independent, that is

$$p(\mathcal{Y}_i \mid b_i; \theta_y) = \prod_j p(y_i(t_{ij}) \mid b_i; \theta_y). \tag{2}$$

An implicit assumption in factorization (1) is that both the censoring and the visiting processes are noninformative, i.e., independent of b_i, and can be ignored in the modelling procedure. Although such an assumption might be questionable in certain situations, we adhere to it here and revisit it in § 6.

SPMs are built under the so-called conditional independence assumption, where the survival and longitudinal processes are assumed independent given the random-effects b_i, implying that all association is induced by the random-effects. It is customary to assume b_i to follow a normal distribution, even though this does not usually lead to a tractable form for the integral in (1) and hence numerical integration remains a requirement to evaluate the associated likelihood. According to (1), distributional assumptions for the random-effects allegedly play an important role in the SPM's factorization since the b_i's link the two processes

of interest. However, empirical results (Wang & Taylor, 2001; Song et al., 2002; Tsiatis & Davidian, 2004) show that misspecification of the random-effects distribution does not have a great impact on the parameter estimates, except for extreme cases such as discrete distributions. We investigate this phenomenon in more detail in § 3.

2·2 *Two Parameterizations*

The usual SPMs assume that the longitudinal and event processes share a common set of random-effects. In particular, the conditional sub-models for $\mathcal{Y}_i$ and T_i have the form

$$\begin{aligned} \mathcal{Y}_i \mid b_i \sim N(\eta_{yi}, \sigma_y^2 \mathrm{I}_{n_i}) \quad ; \quad \log T_i \mid b_i \sim \mathcal{P} \quad \text{with} \quad E(\log T_i \mid b_i) = \eta_{ti} \\ \eta_{yi} = X_{yi}\beta + Z_{yi}b_i \quad ; \quad \eta_{ti} = x_{ti}^\top \gamma + (Z_{yi}b_i)^\top \alpha, \end{aligned} \tag{3}$$

where I_{n_i} denotes the n_i-dimensional identity matrix, $\mathcal{P}$ denotes a parametric distribution (e.g., extreme value, normal, logistic or other), X_{yi} and Z_{yi} are known fixed- and random-effects design matrices, respectively, for the longitudinal process, β is a vector of unknown fixed-effects parameters, σ_y^2 is the error variance, x_{ti} is a vector of covariates for the event process with an associated coefficient vector γ, and α denotes a vector of association parameters linking the survival process with the random-effects structure of the measurement process. If $\alpha = 0$, then the two processes are unrelated, implying that joint modelling is not required under the posited model.

An implicit feature of parameterization (3) is that it assumes perfect linear correlation between the latent structures of the two processes, since the same random-effects are shared. This could be regarded as a rather restrictive assumption that may not be desirable, especially in settings in which the association structure between the measurement and event processes is of interest. Therefore, we propose a more flexible parameterization that considers two separate sets of random effects for the two processes, linking them using a copula

function. Copulas (Nelsen, 1999) are multivariate cumulative distribution functions with uniform marginals, which provide a natural approach to construct joint distributions and explore dependence. The consideration of two separate random-effects is in the spirit of the approach proposed by Henderson et al. (2000) who postulate a bivariate Gaussian latent process shared by the two processes. In particular, we assume that

$$\eta_{yi} = X_{yi}\beta + Z_{yi}b_{yi} \quad \text{and} \quad \eta_{ti} = x_{ti}^{\top}\gamma + b_{ti}, \tag{4}$$

$$p(b_{yi}, b_{ti}) = c\{F_y(b_{yi}), F_t(b_{ti}); \alpha\}\; p(b_{yi})\; p(b_{ti}), \tag{5}$$

where b_{yi} are random-effects for the measurement process and b_{ti} is a frailty term for the survival process. The frailty term is assumed to represent an unobserved covariate explaining heterogeneity (Keiding et al., 1997). For the joint density of $\{b_{yi}, b_{ti}\}$ given in (5) we assume a copula representation, where $c(\cdot)$ denotes the density of a copula function $C(\cdot)$, and $F_y(\cdot)$ and $F_t(\cdot)$ are the marginal cumulative distributions functions for b_{yi} and b_{ti}, respectively. In the case of multivariate b_{yi}, we assume that the copula behind $F_y(\cdot)$ is directly compatible with $C(\cdot)$ (Nelsen, 1999, pp. 85–86). It is important to note that under, (4), the association parameter is a parameter of the random-effects model and specifically of the copula function, in contrast to (3) where α is a parameter of the event process model. The main advantage of parameterization (4) is the flexibility in considering different dependence structures between the two processes by using different copula functions while keeping all other aspects of the model fixed. For instance, under the usual normality assumption for b_i, parameterization (3) is a special case of (4) with $C(\cdot)$ being the Gaussian copula with a restricted correlation matrix assuming $\text{corr}(b_{yi}, b_{ti}) = \pm 1$ depending on the sign of α under (3), and Gaussian marginals $F_y(\cdot)$ and $F_t(\cdot)$. In this example, $b_{ti} = \alpha b_{yi}$, that is, α^2 is merely a rescaling factor for the variance of b_{yi}.

However, even though the latter parameterization offers increased flexibility for the as-

sociation structure between the two processes, we should note that SPMs, in general, imply a restrictive representation of the marginal joint distribution $\{\mathcal{Y}_i, T_i\}$. To see this, consider the following simple but instructive example. Assume no censoring and moreover that all processes involved, namely $\mathcal{Y}_i \mid b_{yi}$, $\log T_i \mid b_{ti}$ and $\{b_{yi}, b_{ti}\}$ follow a normal distribution. Then the covariance for the marginal distribution $\{\mathcal{Y}_i, \log T_i\}$ is of the form $V = \tilde{Z} D \tilde{Z}^\top + \Sigma$, where $\tilde{Z} = \text{diag}\{Z_y, 1\}$, D is the covariance matrix for the joint distribution of $\{b_{yi}, b_{ti}\}$, and Σ is the residual covariance matrix for the joint distribution of $\{\mathcal{Y}_i, \log T_i \mid b_{yi}, b_{ti}\} = \{\mathcal{Y}_i \mid b_{yi}\} \cdot \{\log T_i \mid b_{ti}\}$. Clearly, V is of a specific form assuming positive correlation and not a general variance-covariance matrix.

3. Random-Effects Misspecification

In this section, we investigate the effect of misspecifying the random-effects distribution in parameter estimates and standard errors under the SPMs framework. Unless explicitly stated, we will denote by b_i the set of random-effects under both parameterizations (3) and (4); in the latter case $b_i^\top = (b_{yi}^\top, b_{ti})$. In particular, we assume that the true random-effects probability density function is $p(b_i)$, whereas the fitted one is $f(b_i; \theta_b)$, where both $p(b_i)$ and $f(b_i; \theta_b)$ are absolutely continuous. Moreover, we assume that there is no $\theta_b \in \Theta_b$, such that $f(b_i; \theta_b) = p(b_i)$, where Θ_b is the parameter space of θ_b. Finally, the conditional models for the longitudinal measurement and event processes, $p(\mathcal{Y}_i \mid b_i; \theta_y)$ and $p(T_i, \delta_i \mid b_i; \theta_t)$, respectively, are assumed correctly specified.

3·1 *Parameter estimates*

We will distinguish between two sets of parameters, namely $\theta_{yt}^\top = (\theta_y^\top, \theta_t^\top)$ and θ_b. The effect of using $f(b_i; \theta_b)$ instead of the true $p(b_i)$ in the parameter estimates is described in the following theorem.

THEOREM 1. *For fixed sample size n and as the number n_i of repeated measurements per individual in the longitudinal process $\mathcal{Y}_i$ increases, then the maximum likelihood estimator $\tilde{\theta}_{yt}$ under $f(b_i;\theta_b)$ converges to the maximum likelihood estimator $\hat{\theta}_{yt}$ under the correct model $p(b_i)$.*

The proof can be found in the Appendix. The key argument behind Theorem 1 lies in the fact that, as n_i grows the longitudinal measurement model $p(\mathcal{Y}_i \mid b_i;\theta_y)$ becomes the dominating part in the posterior distribution of the random-effects $p(b_i \mid \mathcal{Y}_i, T_i, \delta_i;\theta)$, implying that the choice between $p(b_i)$ or $f(b_i;\theta_b)$ is of minimal importance. However, the above theorem does not hold for θ_b, and in this case the effect of misspecification will be more prominent. According to White (1982), the maximum likelihood estimator $\tilde{\theta}_b$ will converge in probability to the value θ_b^o that minimizes the Kullback-Leibler distance $\mathcal{D}(p : f;\theta_b) = \int\int p(\mathcal{Y},T,\delta)\ \log\{p(\mathcal{Y},T,\delta)/f(\mathcal{Y},T,\delta;\theta_b)\}\ d\mathcal{Y}dT$.

Two remarks based on the above theorem are worth making. First, in many clinical examples the main interest lies in the degree of the association between the longitudinal measurements and the survival process. As we noted in § 2·2, in the common parameterization (3) that assumes perfect correlation, the association parameter α is, in fact, a parameter of the survival model or a parameter of the longitudinal model, if (3) was written as $\eta_{yi} = X_{yi}\beta + Z_{yi}b_i^*$, $\eta_{ti} = x_{ti}^\top\gamma + b_i$, with $b_i^* = \alpha \cdot b_i$. Thus, under Theorem 1, α will be minimally affected by misspecification of the random-effects distribution, which explains the empirical results reported by other authors (Wang & Taylor, 2001; Song et al., 2002; Tsiatis & Davidian, 2004). However, under parameterization (4), α is a parameter of the copula function, which is a part of the random-effects model. Thus, even for large n_i, we may observe some sensitivity in the estimation of α under different choices for $C(\cdot)$. Second, a straightforward extension of Theorem 1 shows that θ_y will be unbiasedly estimated, even if the event process model is

misspecified. This has a direct impact in the missing data context where SPMs are also used to correct for nonignorable dropout (Follmann & Wu, 1995). In particular, if the informative censoring mechanism producing the missing data in the longitudinal process is described by a SPM, then the effect of misspecifying both the survival and the random-effects model will be minimal as the number of repeated longitudinal measurements per individual increases.

3·2 *Standard Errors*

As we argued in the previous section, the MLE $\tilde{\theta}_{yt}$ under the misspecified random-effects model converges to the MLE $\hat{\theta}_{yt}$ under the correct random-effects model, which implies that misspecification does not affect consistency. However, the effect of misspecifying the random-effects distribution will be more prominent in the estimation of standard errors of $\tilde{\theta}_{yt}$. This becomes more transparent by examining the form of the minus inverse Hessian matrix under the SPM (1), which we would have used as a consistent estimator of the asymptotic inverse Fisher Information matrix if misspecification was ignored. In particular, adopting the notation introduced in Appendix A1 and for $k, k' = y, t, b$ we set $H_{kk'} = n^{-1} \sum_i \partial L_i^f(\tilde{\theta}_k)/\partial \theta_{k'}$ to denote the corresponding block of the Hessian matrix H, where

$$\frac{\partial L_i^f(\tilde{\theta}_k)}{\partial \theta_{k'}} = \begin{cases} E_f\left\{\partial h(\cdot;\tilde{\theta}_k)/\partial \theta_{k'}\right\} + E_f\left[h(\cdot;\tilde{\theta}_k)\left\{h(\cdot;\tilde{\theta}_k) - L_i^f(\tilde{\theta}_k)\right\}^\top\right], & k' = k \\ E_f\left[h(\cdot;\tilde{\theta}_k)\left\{h(\cdot;\tilde{\theta}_{k'}) - L_i^f(\tilde{\theta}_{k'})\right\}^\top\right], & k' \neq k \end{cases} \tag{6}$$

with $E_f\{\cdot\}$ denoting the expectation with respect to the posterior distribution $f(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta)$. Following this notation and assuming that H^{-1} exists, the asymptotic variance of $\tilde{\theta}_{yt}$ under standard likelihood methods has the form $\widehat{\text{var}(\tilde{\theta}_{yt})} = -(H_{yt} - H_{yt,b}H_{bb}^{-1}H_{yt,b}^\top)^{-1}$, where $H_{yt,b}^\top = \{H_{yb}^\top, H_{tb}^\top\}$. The second part in the parenthesis is clearly affected by misspecification; to see this we focus on the H_{yb} block of $H_{yt,b}$, with the results for H_{tb} and H_{bb} following

similarly. For H_{yb}, (6) can be rewritten as

$$H_{yb} = n^{-1} \sum_{i=1}^{n} E_f \left[\left\{ \sum_{j=1}^{n_i} \frac{\partial}{\partial \theta_y} \log p(y_i(t_{ij}) | b_i; \tilde{\theta}_y) \right\} \left\{ \frac{\partial}{\partial \theta_b} \log f(b_i; \tilde{\theta}_b) \right\}^\top \right] - \left\{ L_i^f(\tilde{\theta}_y) \right\} \left\{ L_i^f(\tilde{\theta}_b) \right\}^\top .$$

If we let n_i grow, then $E_f\{\} \to E_p\{\}$, i.e., in the corresponding expectations the true posterior is used. However, note that both parts of H_{yb} still depend on the misspecified random-effects model, since $L_i^p(\theta_b) = \int \{\partial \log f(b_i; \theta_b)/\partial \theta_b\}\, p(b_i \mid \mathcal{Y}_i, T_i; \theta)\, db_i$, which will result in some bias in the standard error estimates.

Following standard maximum likelihood theory under misspecification (White, 1982), the asymptotic covariance matrix for $\tilde{\theta}$ is $\text{var}(\tilde{\theta}) = K^{-1} D K^{-1}$, where $K = E\{-H(\theta_{yt}^*, \theta_b^o)\}$, $D = E\{L^f(\theta_{yt}^*, \theta_b^o) L^f(\theta_{yt}^*, \theta_b^o)^\top\}$, and the expectations are taken with respect to the true distribution $p(\mathcal{Y}, T, \delta; \theta^*)$. Using the sandwich variance estimator as a consistent estimator for this covariance matrix we obtain,

$$\widetilde{\text{var}(\tilde{\theta}_{yt})} = \mathcal{A}\mathcal{X}\mathcal{A} + 2\mathcal{B}\mathcal{Z}^\top \mathcal{A} + \mathcal{B}\mathcal{W}\mathcal{B}^\top, \tag{7}$$

where $\mathcal{A} = \widehat{\text{var}(\tilde{\theta}_{yt})} = -(H_{yt} - H_{yt,b} H_{bb}^{-1} H_{yt,b}^\top)^{-1}$, $\mathcal{X} = n^{-1} \sum_i L_i^f(\tilde{\theta}_{yt}) L_i^f(\tilde{\theta}_{yt})^\top$, $\mathcal{B} = -H_{yt}^{-1} H_{yt,b} (H_b - H_{yt,b}^\top H_{yt}^{-1} H_{yt,b})^{-1}$, $\mathcal{Z} = n^{-1} \sum_i L_i^f(\tilde{\theta}_{yt}) L_i^f(\tilde{\theta}_b)^\top$, and $\mathcal{W} = n^{-1} \sum_i L_i^f(\tilde{\theta}_b) L_i^f(\tilde{\theta}_b)^\top$. Note that as, $n \to \infty$ and if the correct model $p(b_i)$ had been used, then $\|\widetilde{\text{var}(\tilde{\theta}_{yt})} - \widehat{\text{var}(\tilde{\theta}_{yt})}\| \to 0$. Straightforward algebra then implies that $\text{diag}\{\mathcal{A}\mathcal{X} + 2\mathcal{B}\mathcal{Z}^\top + \mathcal{B}\mathcal{W}\mathcal{B}^\top \mathcal{A}^{-1}\} - 1$ quantifies the extra variance for $\tilde{\theta}_{yt}$ owing to misspecification.

4. A Simulation Study

4·1 *Study Set-up*

A small simulation study was performed to empirically corroborate the arguments unfolded in § 3. Since the one random-effect case has been extensively studied in the literature (e.g., by Song et al., 2002), here we investigate the effect of misspecifying the random-effects

distribution in parameter estimates and standard errors for the two random-effects case (5). The study considers a two-group comparison with $n = 200$. In particular, for the linear mixed-effects model in (3), we assume measurement error variance $\sigma_y^2 = 0{\cdot}5^2$, and linear predictor given by $\eta_{yi} = (\beta_0 + b_{yi}) + \beta_1 \mathcal{T}_i + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \beta_4 \mathcal{T}_i t_{ij} + \beta_5 \mathcal{T}_i t_{ij}^2$, where b_{yi} denotes a random-intercepts term, $\mathcal{T}_i$ is the binary treatment indicator and $(\beta_0, \ldots, \beta_5)^\top = (1, 0, 1{\cdot}5, 2{\cdot}5, -0{\cdot}5, -1)$. For the survival model in (3), we assume that $\mathcal{P}$ follows the extreme value distribution with scale parameter $\sigma_t = 0{\cdot}5$, and linear predictor given by $\eta_{ti} = (\gamma_0 + b_{ti}) + \gamma_1 \mathcal{T}_i$, where b_{ti} is a frailty term and $(\gamma_0, \gamma_1)^\top = (2, 1{\cdot}5)$. The censoring mechanism follows an exponential distribution with mean 20, resulting in about 50% censoring and the visiting times t_{ij} are random. For the random-effects model in (4) the following three scenarios are considered: (i) a bimodal mixture distribution $0{\cdot}45 \times N((-2, -2{\cdot}1)^\top, \Sigma) + 0{\cdot}55 \times N((1{\cdot}636, 1{\cdot}718)^\top, \Sigma)$, with $\Sigma = \text{vech}(1{\cdot}5^2, 1^2, 0{\cdot}5)$ (where in $\text{vech}(s_1^2, s_2^2, \rho)$, s_1^2 and s_2^2 denote the two variances and ρ the correlation of the covariance matrix Σ); (ii) a unimodal skewed mixture distribution $0{\cdot}7 \times N((1{\cdot}3, 0{\cdot}9)^\top, \Sigma) + 0{\cdot}3 \times N((-3{\cdot}033, 2{\cdot}1)^\top, \Sigma)$, with $\Sigma = \text{vech}(1{\cdot}6^2, 1{\cdot}7^2, 0{\cdot}7)$; and (iii) a normal distribution $N(0, \Sigma)$, with $\Sigma = \text{vech}(2{\cdot}5^2, 2{\cdot}2^2, 0{\cdot}82)$. The parameter values have been chosen such that the variances σ_{by}^2 and σ_{bt}^2, and the degree of association (in terms of Kendall's τ) for the random-effects are of the same magnitude for all three scenarios. For n_i, two cases are considered, namely the large-n_i case where $\max_i\{n_i\} = 15$ with 10 measurements per subject on average, and the small-n_i case where $\max_i\{n_i\} = 4$ with 2·5 measurements per subject on average. Finally, for each scenario and for each n_i-case, 100 data-sets are simulated.

4·2 *Analyses Models*

For each simulated data-set four joint models are fitted. In particular, for the longitudinal $p(\mathcal{Y}_i \mid b_{yi}; \theta_y)$ and the survival $p(T_i, \delta_i \mid b_{ti}; \theta_t)$ processes the correct models are assumed,

whereas for the random-effects model four copulas are considered, namely the Frank, Gumbel, Gaussian and Student's-t (df = 4) copulas, with normal marginals. Thus, under scenarios (i) and (ii) all fitted models are misspecified, whereas for scenario (iii) the normal copula random-effects model corresponds to the true joint model we simulated from. Furthermore, the quality of the model based standard errors $\widehat{\text{var}(\tilde{\theta})}$ (i.e., assuming the random-effects distribution had been correctly specified) and the sandwich estimator standard errors $\widetilde{\text{var}(\tilde{\theta})}$ is compared to the empirical standard errors obtained by $\left\{\sum_{m=1}^{M}(\hat{\theta}_m - \overline{\theta})^2/(M-1)\right\}^{1/2}$, where $\hat{\theta}_m$ denotes the MLEs in the mth simulated data-set, $\overline{\theta} = \sum_{m=1}^{M}\hat{\theta}_m/M$, and $M = 100$. The models are fitted using an EM algorithm in which the random-effects are treated as missing values; more details can be found in Appendix A2. All computations have been performed in R (R Development Core Team, 2006).

4·3 *Results*

Tables 1, 2 and 3 present the simulation study results.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

For all scenarios we observe that the parameter estimates for the longitudinal and survival models are rather robust to random-effects misspecification. On the contrary, the parameter estimates for random-effects model, and especially the estimates of the association parameter, show greater sensitivity regarding the choice of the copula $C(\cdot)$. Furthermore, the small-n_i case yielded relatively more sensitive results for the parameter estimates, which is in accordance with Theorem 1. An interesting feature is that the Gaussian copula performed rather well under misspecification. This feature can be explained by the concept

of local dependence introduced by Holland & Wang (1987). The local dependence function equals $\partial^2 \log p(b_{yi}, b_{ti})/\partial b_{yi}\partial b_{ti}$ and is used to quantify dependence when both the degree and the direction of the dependence is different in different regions of the plane (Jones, 1996). A numerical comparison between the values of the local dependence function of the true random-effects densities under scenarios (i) and (ii), and the corresponding values of the assumed copulas, reveals that the Gaussian copula is on average closer to the true densities than the other copulas. Finally, regarding the estimation of standard errors, we observe that the average of sandwich estimates is closer to the empirical standard errors than the corresponding model-based ones, with exception scenario (iii) under the Gaussian copula, where the model-based standard errors, as expected, show good behaviour.

5. Application

In this section, we present the analysis of data coming from a longitudinal study on patients who received a kidney transplant. The main scientific focus lies in the time a patient can maintain the new graft. In this case, a good marker for the kidneys' performance is the level of serum creatinine in blood. However, due to the fact that the observed levels of this marker are directly influenced by a person's muscle activity, the glomerular filtration rate (GFR) is typically used, which is an inverse function of serum creatinine correcting also for sex, weight, and age.

During the 10 year follow-up period GFR measurements are regularly taken and our aim here is to explore the association structure between longitudinal GFR measurements and the time to graft failure. Out of the 432 patients, 91 (21·1%) experienced the event; moreover, patients made on average 72 visits (standard deviation 22·4 visits), resulting in a total of 31,062 records. Based on descriptive measures and plots we adopted the following models for the two processes. For the longitudinal process a linear random-intercepts model

is assumed with fixed-effects quadratic time trends for the first 6 months, followed by linear time trends for the remaining follow-up period. For the survival process we include the age, weight, and sex as main effects, and a frailty term related to the random-intercept term of the measurement model.

To investigate the influence of parametric assumptions on the size of the association between the two processes, we performed a sensitivity analysis under different copula functions and assuming normal marginals for the joint distribution of the random-effects, and different survival distributions. In particular, we considered the Frank, Gumbel, Gaussian, and Student's-t (df = 4) copulas, and the Weibull, log-normal and log-logistic as survival distributions. The estimates of Kendall's τ for each scenario are presented in Table 4.

[Table 4 about here.]

For the entire analysis we observed similar results as in § 4. In particular, the main effects for both the linear mixed and survival models were minimally affected by different assumptions regarding the random-effects, whereas the degree of the association between the two processes was influenced to a much larger extent by the choice of the copula function. The results suggest a moderate positive association between the underlying latent processes, ranging from 0·56 to 0·86. However, note that this is far from the perfect correlation that the common parameterization (3) assumes.

6. Concluding Remarks

In this paper, we investigated the effect of misspecifying the random-effects distribution under the shared parameter model framework. In particular, we showed that, as the number n_i of repeated longitudinal measurements per individual increases, the effect of misspecification becomes minimal for certain parameter estimates. However, estimation of the standard errors

under the misspecified model will generally be affected, and thus the use of the sandwich estimator is recommended. How large n_i has to be depends on the type of information for b_i that is included in $\mathcal{Y}_i$. In particular, we expect that for linear mixed models, smaller values of n_i will suffice as opposed to generalized or nonlinear mixed models. The reason is that in the former case $\log p(\mathcal{Y}_i \mid b_i; \theta_y)$ will be quadratic in b_i, which implies that convergence of $p(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta)$ to a normal distribution will be faster. In addition, note that Theorem 1 requires all subjects to have a relatively large number of repeated measurements. Thus, in cases in which some groups of subjects have very few measurements (e.g., many subjects dropout too early in the study), choosing the correct random-effects distribution will be important. Moreover, our results are based on the assumption that both $p(b_i)$ and $f(b_i; \theta_b)$ are continuous densities, excluding the case in which the true random-effects distribution is discrete, with few support points. In this setting we would expect that the robustness for $\tilde{\theta}_{yt}$ is seriously affected.

Moreover, the formulation of the SPM presented in § 2 assumed a noninformative visiting process, which enabled an easier likelihood construction. However, in cases where such an assumption is erroneous, ignoring the visiting process may considerably influence results since each subject will have n_i measurement occasions leading to a multivariate model. Thus, the posterior distribution of the random-effects will then heavily depend on both the longitudinal and visit process models.

Finally, we have assumed that the parameter space of the survival model is of finite dimension. This excludes the commonly used semiparametric framework in which the baseline hazard is left unspecified. Extensions of the results presented here for this case are under consideration.

Appendix

Technical Details

A1. Proof of Theorem 1

Let $p(\mathcal{Y}, T, \delta; \theta)$ and $f(\mathcal{Y}, T, \delta; \theta)$ denote the marginal densities under the correctly specified $p(b)$ and the misspecified $f(b; \theta_b)$ random-effects distributions, respectively. First, we work under parameterization (3) with a common random-effect b_i for the two processes. We make the following assumptions: (i) both $p(\mathcal{Y}, T, \delta; \theta)$ and $f(\mathcal{Y}, T, \delta; \theta)$ are well-defined densities under the usual regularity conditions (Cox & Hinkley, 1974, pp. 281); (ii) for fixed n we define the log-likelihood functions $\ell_n^p(\theta) = n^{-1} \sum_{i=1}^n \log p(\mathcal{Y}_i, T_i, \delta_i; \theta)$, $\ell_n^f(\theta) = n^{-1} \sum_{i=1}^n \log f(\mathcal{Y}_i, T_i, \delta_i; \theta)$, and in addition, we assume that $\ell_n^p(\theta)$ and $\ell_n^f(\theta)$ have unique maxima at $\hat{\theta}_{yt} \in \Theta_{yt}$ and $\tilde{\theta}_{yt} \in \Theta_{yt}$, respectively, with θ_y and θ_t having disjoint parameter spaces, i.e., $\Theta_{yt} = \Theta_y \times \Theta_t$; (iii) for the score vectors $L_n^p(\theta) = \partial \ell_n^p(\theta)/\partial\theta$, and $L_n^f(\theta) = \partial \ell_n^f(\theta)/\partial\theta$ we assume that the required conditions hold which allow differentiation to be taken inside the integral sign; (iv) finally, we assume that both $\log p(b_i)$ and $\log f(b_i; \theta_b)$ are bounded and smooth functions of b_i around the neighborhood of the mode $\hat{b}_i$ of $\log p(\mathcal{Y}_i \mid b_i; \theta_y) = \sum_{j=1}^{n_i} \log p(y_i(t_{ij}) \mid b_i; \theta_y)$.

Next we note that, under assumption (iii), the score vector takes the form

$$\begin{aligned} L_n^p(\theta) &= \sum_{i=1}^n \frac{\partial}{\partial\theta} \log \int p(\mathcal{Y}_i \mid b_i; \theta_y)\, p(T_i, \delta_i \mid b_i; \theta_t)\, p(b_i)\, db_i \\ &= \sum_{i=1}^n \int h(\cdot; \theta)\, p(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta)\, db_i, \end{aligned} \tag{A.1}$$

where $h(\cdot; \theta)$ denotes the corresponding score vector of each of the sub-models (e.g., for the measurement process, $L_n^p(\theta_y)$ requires $h(\cdot; \theta) = \partial \log p(\mathcal{Y}_i \mid b_i; \theta_y)/\partial\theta_y$). Analogously, the

misspecified score vector takes the form

$$L_n^f(\theta) = \sum_{i=1}^{n} \int h(\cdot;\theta)\ f(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta)\ db_i. \tag{A.2}$$

(A.2) differs from (A.1) in that $f(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta)$ is the posterior under $f(b_i; \theta_b)$, but also that for $L_n^f(\theta_b)$, $h(\cdot;\theta) = \partial \log f(b_i;\theta_b)/\partial\theta_b$. For fixed n, the score vectors are functions of the number of repeated measurements n_i. Following we assume that for all i, $n_i \to \infty$. Then, under assumptions (i) and (iv) both posterior distributions $p_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta)$ and $f_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta)$ will concentrate around the neighborhood of the mode $\hat{b}_i$ of the correctly specified longitudinal model $\log p(\mathcal{Y}_i \mid b_i; \theta_y) = \sum_{j=1}^{n_i} \log p(y_i(t_{ij}) \mid b_i; \theta_y)$ (Cox & Hinkley, 1974, pp. 399–400), which implies that, as $n_i \to \infty$, $|f_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta) - p_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta)| \to 0$. Based on this result, we have that for every $\epsilon > 0$ there exists an integer m such that for all $n_i \geq m$ and for all $\theta_{yt} \in \Theta_{yt}$ we get

$$\begin{aligned}
&\|L_{n_i}^f(\theta_{yt}) - L_{n_i}^p(\theta_{yt})\| = \\
&= \|\sum_i \int h(\cdot;\theta_{yt}) \left\{ f_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta) - p_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta) \right\}\ db_i\| \\
&\leq \sum_i \int \|h(\cdot;\theta_{yt}) \left\{ f_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta) - p_{n_i}(b_i \| \mathcal{Y}_i, T_i, \delta_i; \theta) \right\}\| \ db_i \\
&\leq \sum_i \int \|h(\cdot;\theta_{yt})\| \cdot \mid f_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta) - p_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta) \mid \ db_i < \epsilon,
\end{aligned}$$

where $\|\cdot\|$ denotes the Euclidean vector norm. The last statement combined with the application of the mean value theorem to either $L_{n_i}^f(\theta_{yt})$ or $L_{n_i}^p(\theta_{yt})$ implies that $\|\tilde{\theta}_{yt} - \hat{\theta}_{yt}\| \to 0$.

Under the two random-effects parameterization (4), the arguments raised above can be adapted accordingly to show that $\tilde{\theta}_y$ will converge to $\hat{\theta}_y$. However, for $\tilde{\theta}_t$ we have that

$$\begin{aligned}
L_n^f(\theta_t) &= \sum_{i=1}^{n} \iint \frac{\partial}{\partial\theta_t} \log p(T_i, \delta_i \mid b_{ti}; \theta_t)\ f(b_{yi}, b_{ti} \mid \mathcal{Y}_i, T_i, \delta_i; \theta)\ db_{yi} db_{ti} \\
&= \sum_{i=1}^{n} \int \frac{\partial}{\partial\theta_t} \log p(T_i, \delta_i \mid b_{ti}; \theta_t)\ f(b_{ti} \mid \mathcal{Y}_i, T_i, \delta_i; \theta)\ db_{ti},
\end{aligned}$$

where $f(b_{ti} \mid \mathcal{Y}_i, T_i, \delta_i; \theta) \propto f(\mathcal{Y}_i \mid b_{ti}; \theta_y, \theta_b) p(T_i, \delta_i \mid b_{ti}; \theta_t) f(b_{ti}; \theta_b)$, with $f(\mathcal{Y}_i \mid b_{ti}; \theta_y, \theta_b) = \int p(\mathcal{Y}_i \mid b_{yi}; \theta_y) f(b_{yi} \mid b_{ti}; \theta_b) db_{yi}$. Heuristically, as long as $f(b_{yi} \mid b_{ti}; \theta_b) \neq f(b_{yi}; \theta_b)$, then as n_i increases the $p(\mathcal{Y}_i \mid b_{yi}; \theta_y)$ part of $f(\mathcal{Y}_i \mid b_{ti}; \theta_y, \theta_b)$ provides increasing information regarding b_{ti}, and this information becomes greater as the association between b_{yi} and b_{ti} gets stronger. Formally, note that for $\tilde{\theta}_t$ to converge to $\hat{\theta}_t$ as $n_i \to \infty$, we need $|f_{n_i}(b_{ti} \mid \mathcal{Y}_i, T_i, \delta_i; \theta) - p_{n_i}(b_{ti} \mid \mathcal{Y}_i, T_i, \delta_i; \theta)| \to 0$. This would be the case as along as $f_{n_i}(b_{ti} \mid \mathcal{Y}_i, T_i, \delta_i; \theta)$ is concentrated around the neighborhood of the mode $\hat{b}_{ti}$ of $f_{n_i}(\mathcal{Y}_i \mid b_{ti}; \theta_y, \theta_b)$. To ensure this, we moreover adopt the regularity conditions of Heyde & Johnstone (1979), under which both $f_{n_i}(b_{ti} \mid \mathcal{Y}_i, T_i, \delta_i; \theta)$ and $p_{n_i}(b_{ti} \mid \mathcal{Y}_i, T_i, \delta_i; \theta)$ will converge to the same normal distribution with mean $\hat{b}_{ti}$, even though conditional independence does not hold in this case, i.e, $f(\mathcal{Y}_i \mid b_{ti}; \theta_y, \theta_b) \neq \prod_{j=1}^{n_i} f(y_i(t_{ij}) \mid b_{ti}; \theta_y, \theta_b)$.

Finally, to make the above results probabilistic in nature and to ensure that $\tilde{\theta}_{yt}$ converges in probability to the true parameter vector θ^*_{yt}, we assume that $n \to \infty$ and that $\hat{\theta}_{yt}$ is a consistent estimator for θ^*_{yt}.

A2. EM Steps

The maximum likelihood estimates for the parameter vector θ are obtained using an EM algorithm, where b_{yi} and b_{ti} are treated as missing data. We assume the following sub-models for the processes involved in the specification of the SPM:

$$\mathcal{Y}_i = X_{yi}\beta + Z_{yi}b_{yi} + \varepsilon_{yi} \quad \text{and} \quad \log T_i = x_{ti}^\top \gamma + b_{ti} + \sigma_t^{-1}\varepsilon_{ti},$$

where $\varepsilon_{yi} \sim N_{n_i}(0, V_i = \sigma_y^2 Q_i)$ with Q_i being a correlation matrix with an associated parameter vector κ, $\varepsilon_{ti} \sim \mathcal{P}$ where $\mathcal{P}$ denotes an appropriate distribution function with corresponding survival function S and density function p, and σ_t is a scale parameter (Kalbfleisch & Prentice, 2002, Ch. 3). Finally, the joint density of $\{b_{yi}, b_{ti}\}$ follows (5), with copulas belonging to either the Archimedean or elliptical classes and Gaussian marginals.

For the E-step we set $\ddot{A}$ to denote $E\{A(b_{yi}, b_{ti}) \mid \mathcal{Y}_i, T_i, \delta_i; \theta\}$, where the required integrals are approximated using a Gauss-Hermite quadrature rule. For the parameters with no closed-form solutions, we set $\ell(\cdot)$ to denote the score vector of the complete data log-likelihood. The expected value $\ddot{\ell}(\cdot)$ of $\ell(\cdot)$, with respect to $p(b_{yi}, b_{ti} \mid \mathcal{Y}_i, T_i, \delta_i; \theta)$, is used to numerically maximize the expected value of the complete data log-likelihood, based on a quasi-Newton algorithm. In particular, the following expressions define the M-step.

Longitudinal measurement model:

$$\begin{aligned}
\beta &= \left\{\sum_{i=1}^{n} X_{yi}^{\top} V_i^{-1} X_{yi}\right\}^{-1} \left\{\sum_{i=1}^{n} X_{yi}^{\top} V_i^{-1} (y_i - Z_{yi}\ddot{b}_{yi})\right\}, \\
\sigma_y^2 &= \frac{1}{N}\sum_{i=1}^{n} \mu_{yi}^{\top} Q_i^{-1}(\mu_{yi} - 2Z_{yi}\ddot{b}_{yi}) + \mathrm{tr}(Z_{yi}^{\top} Q_i^{-1} Z_{yi} \ddot{v b}_{yi}) + \ddot{b}_{yi}^{\top} Z_{yi}^{\top} Q_i^{-1} Z_{yi} \ddot{b}_{yi}, \\
\ddot{\ell}(\kappa) &= \frac{1}{2}\sum_{i=1}^{n} \mathrm{tr}(-Q_i^{-1} W_i) + \mu_{yi}^{\top} K_i(\mu_{yi} - 2Z_{yi}\ddot{b}_{yi}) + \mathrm{tr}(M_i \ddot{v b}_{yi}) + \ddot{b}_{yi}^{\top} M_i \ddot{b}_{yi},
\end{aligned}$$

where $N = \sum_{i=1}^{n} n_i$, $\mu_{yi} = y_i - X_{yi}\beta$, $\ddot{b}_{yi} = E(b_{yi} \mid \mathcal{Y}_i, T_i; \theta)$, $\ddot{v b}_{yi} = \mathrm{var}(b_{yi} \mid \mathcal{Y}_i, T_i; \theta)$, $W_i = \partial Q_i / \partial \kappa$, $K_i = Q_i^{-1} W_i Q_i^{-1}$, $M_i = Z_{yi}^{\top} K_i Z_{yi}$.

Event process model:

$$\ell(\gamma) = \sigma_t^{-1} \sum_{i=1}^{n} x_{ti} a_i \quad \text{and} \quad \ell(\sigma_t) = \sigma_t^{-1} \sum_{i=1}^{n} \omega_i a_i - \delta_i,$$

where $a_i = -\delta_i\{\partial \log p(\omega_i)/\partial \omega_i\} - (1-\delta_i)\{\partial \log S(\omega_i)/\partial \omega_i\}$, and $\omega_i = (\log T_i - x_{ti}^{\top}\gamma - b_{ti})/\sigma_t$.

Random-effects model: We distinguish between the following cases. First, the Gaussian copula combined with normal marginals results in a multivariate normal distribution with known derivatives for the variance components. Second, the Student's-t copula involves the inverse distribution function of the Student's-t distribution and thus numerical derivatives are used. Finally, for Archimedean copulas, $\ell(\alpha)$ is derived for each particular copula separately, whereas for the parameters θ_{by} and θ_{bt} of the marginal models for b_{yi} and b_{ti}, the following general formula is used

$$\ell(\theta_{by}) = \sum_{i=1}^{n} \left[\left\{\frac{g^{(3)}(C(u_i, v_i))}{g^{(2)}(C(u_i, v_i))} - 3\frac{g^{(2)}(C(u_i, v_i))}{g^{(1)}(C(u_i, v_i))}\right\} c_u(v_i) + \frac{g^{(2)}(u_i)}{g^{(1)}(u_i)}\right] \frac{\partial u}{\partial \theta_{by}} + \frac{\partial \log p(b_{yi}; \theta_{by})}{\partial \theta_{by}},$$

where $g(\cdot)$ is the generator function of the archimedean copula with $g^{(l)}(\cdot)$ denoting its lth derivative, $c_u(v) = \partial C(u,v)/\partial u$, u and v are the distribution functions of the marginal Gaussian distributions for b_{yi} and b_{ti}, respectively, and $\ell(\theta_{bt})$ is derived analogously.

References

Cox, D. & Hinkley, D. (1974). *Theoretical Statistics*. Chapman & Hall, London.

Follmann, D. & Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51**, 151–168.

Henderson, R., Diggle, P. & Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 465–480.

Heyde, C. & Johnstone, I. (1979). On asymptotic posterior normality for stochastic processes. *Journal of the Royal Statistical Society, Series B* **41**, 184–189.

Holland, P. & Wang, Y.-J. (1987). Dependence function for continuous bivariate densities. *Comm. Statist., Theory Methods* **16**, 863–876.

Huang, X., Stefanski, L. & Davidian, M. (2006). Latent-model robustness in structural measurement error models. *Biometrika* **93**, 53–64.

Jones, M. (1996). The local dependence function. *Biometrika* **83**, 899–904.

Kalbfleisch, J. & Prentice, R. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, New York, 2nd edition.

Keiding, N., Andersen, P. K. & Klein, J. (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine* **16**, 215–224.

Nelsen, R. (1999). *An Introduction to Copulas*. Springer-Verlag, New York.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Scharfstein, D., Rotnitzky, A. & Robins, J. (1999). Adjusting for nonignorable dropout using semiparametric non-response models (with discussion). *Journal of the American Statistical Association* **94**, 1096–1120.

Song, X., Davidian, M. & Tsiatis, A. (2002). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* **58**, 742–753.

Tsiatis, A. & Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* **88**, 447–458.

Tsiatis, A. & Davidian, M. (2004). An overview of joint modeling of longitudinal and time-to-event data. *Statistica Sinica* **14**, 793–818.

Wang, Y. & Taylor, J. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association* **96**, 895–905.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1–26.

Wu, M. & Carroll, R. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175–188.

Wulfsohn, M. & Tsiatis, A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330–339.

Table 1. *Parameter estimates and standard errors under the bimodal scenario (i). The top part contains the results for the large-n_i case (i.e., $\max_i\{n_i\} = 15$), whereas the bottom part contains the results for the small-n_i case (i.e., $\max_i\{n_i\} = 4$). The 'Std. Err.' columns contain the empirical/average sandwich/average model-based standard errors, respectively.*

		Frank		Gumbel		Gaussian		Student's-t	
	True	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
β_0	1	0·979	0·141/0·111/0·081	0·966	0·146/0·122/0·089	0·982	0·145/0·136/0·098	0·980	0·146/0·133/0·093
β_1	0	0·038	0·215/0·191/0·174	0·038	0·206/0·183/0·107	0·029	0·201/0·194/0·155	0·024	0·216/0·191/0·166
β_2	1·5	1·502	0·022/0·023/0·014	1·503	0·022/0·023/0·018	1·502	0·022/0·023/0·013	1·502	0·022/0·024/0·014
β_3	2·5	2·499	0·002/0·002/0·001	2·500	0·002/0·002/0·001	2·500	0·002/0·002/0·001	2·500	0·002/0·002/0·001
β_4	-0·5	-0·502	0·031/0·030/0·020	-0·503	0·031/0·030/0·022	-0·503	0·031/0·030/0·019	-0·502	0·031/0·031/0·018
β_5	-1	-1·000	0·002/0·002/0·001	-0·999	0·002/0·002/0·001	-1·001	0·002/0·002/0·001	-1·000	0·002/0·002/0·001
σ_y	0·5	0·498	0·013/0·010/0·008	0·497	0·015/0·017/0·009	0·499	0·013/0·016/0·009	0·498	0·014/0·013/0·009
γ_0	2	1·978	0·210/0·198/0·188	1·992	0·209/0·199/0·172	2·005	0·201/0·191/0·172	1·982	0·201/0·203/0·176
γ_1	1·5	1·496	0·272/0·264/0·167	1·510	0·260/0·242/0·188	1·492	0·269/0·272/0·196	1·492	0·290/0·273/0·198
σ_t	0·5	0·418	0·199/0·188/0·173	0·610	0·308/0·313/0·266	0·429	0·218/0·190/0·150	0·669	0·210/0·207/0·183
τ	0·62	0·619	0·075/0·065/0·036	0·724	0·117/0·108/0·098	0·634	0·067/0·059/0·042	0·693	0·098/0·100/0·068
σ_{by}	2·35	2·305	0·104/0·108/0·025	2·319	0·140/0·129/0·099	2·349	0·105/0·110/0·078	2·372	0·116/0·117/0·073
σ_{bt}	2·15	2·172	0·143/0·145/0·136	2·141	0·199/0·172/0·145	2·205	0·164/0·146/0·101	2·121	0·197/0·192/0·151
β_0	1	0·952	0·195/0·136/0·134	0·905	0·259/0·130/0·131	0·964	0·210/0·192/0·104	0·941	0·214/0.183/0.120
β_1	0	0·083	0·290/0·186/0·170	0·021	0·339/0·257/0·190	0·060	0·309/0·287/0·193	0·053	0·308/0.259/0.189
β_2	1·5	1·505	0·029/0·029/0·006	1·508	0·029/0·029/0·011	1·506	0·029/0·029/0·009	1·505	0·029/0.029/0.006
β_3	2·5	2·500	0·002/0·003/0·001	2·500	0·003/0·003/0·001	2·500	0·002/0·001/0·003	2·500	0·003/0.002/0.001
β_4	-0·5	-0·505	0·035/0·038/0·008	-0·507	0·035/0·038/0·012	-0·506	0·035/0·038/0·012	-0·506	0·035/0.038/0.008
β_5	-1	-1·000	0·003/0·003/0·001	-0·999	0·003/0·003/0·001	-1·000	0·003/0·003/0·003	-1·000	0·003/0.003/0.001
σ_y	0·5	0·497	0·023/0·015/0·225	0·498	0·023/0·023/0·015	0·497	0·022/0·019/0·015	0·497	0·022/0.015/0.008
γ_0	2	2·032	0·215/0·139/0·061	1·989	0·277/0·156/0·087	2·059	0·251/0·185/0·097	2·036	0·256/0.196/0.090
γ_1	1·5	1·511	0·321/0·201/0·080	1·479	0·378/0·219/0·091	1·495	0·348/0·219/0·091	1·473	0·354/0.344/0.272
σ_t	0·5	0·327	0·128/0·095/0·033	0·580	0·358/0·124/0·092	0·411	0·203/0·174/0·122	0·615	0·285/0.254/0.174
τ	0·62	0·663	0·045/0·039/0·013	0·758	0·131/0·107/0·067	0·660	0·045/0·056/0·012	0·727	0·106/0.095/0.046
σ_{by}	2·35	2·294	0·107/0·114/0·037	2·266	0·215/0·196/0·116	2·338	0·108/0·120/0·033	2·408	0·222/0.201/0.171
σ_{bt}	2·15	2·298	0·154/0·162/0·025	2·191	0·239/0·210/0·173	2·279	0·179/0·171/0·055	2·256	0·295/0.265/0.191

Table 2. *Parameter estimates and standard errors under the skewed scenario (ii). The top part contains the results for the large-n_i case (i.e., $\max_i\{n_i\} = 15$), whereas the bottom part contains the results for the small-n_i case (i.e., $\max_i\{n_i\} = 4$). The 'Std. Err.' columns contain the empirical/average sandwich/average model-based standard errors, respectively.*

		Frank		Gumbel		Gaussian		Student's-t	
	True	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
β_0	1	0·992	0·169/0·151/0·142	0·993	0·171/0·166/0·156	0·995	0·173/0·161/0·155	0·995	0·177/0·167/0·118
β_1	0	-0·014	0·256/0·210/0·199	-0·009	0·260/0·231/0·220	-0·024	0·269/0·272/0·251	-0·022	0·269/0·256/0·193
β_2	1·5	1·501	0·022/0·021/0·017	1·502	0·022/0·022/0·011	1·501	0·022/0·021/0·016	1·502	0·022/0·022/0·011
β_3	2·5	2·499	0·002/0·002/0·001	2·500	0·002/0·002/0·001	2·500	0·002/0·002/0·001	2·500	0·002/0·002/0·001
β_4	-0·5	-0·499	0·029/0·028/0·019	-0·500	0·029/0·028/0·015	-0·500	0·029/0·028/0·015	-0·501	0·030/0·028/0·013
β_5	-1	-1·000	0·002/0·002/0·001	-1·000	0·002/0·002/0·001	-1·000	0·002/0·001/0·001	-1·000	0·002/0·002/0·001
σ_y	0·5	0·504	0·013/0·010/0·007	0·506	0·021/0·018/0·010	0·506	0·021/0·019/0·010	0·506	0·021/0·018/0·012
γ_0	2	1·947	0·226/0·235/0·195	1·980	0·222/0·207/0·174	2·028	0·238/0·219/0·178	2·004	0·222/0·209/0·188
γ_1	1·5	1·502	0·326/0·308/0·268	1·518	0·330/0·336/0·273	1·506	0·352/0·326/0·296	1·501	0·374/0·354/0·308
σ_t	0·5	0·582	0·260/0·269/0·203	0·574	0·282/0·257/0·199	0·471	0·226/0·203/0·177	0·589	0·244/0·223/0·197
τ	0·63	0·654	0·103/0·101/0·085	0·684	0·093/0·100/0·077	0·626	0·060/0·055/0·033	0·661	0·082/0·065/0·051
σ_{by}	2·56	2·549	0·109/0·107/0·072	2·523	0·121/0·131/0·067	2·527	0·108/0·118/0·083	2·530	0·107/0·117/0·088
σ_{bt}	2·19	2·105	0·218/0·221/0·179	2·110	0·209/0·196/0·149	2·159	0·205/0·193/0·127	2·092	0·222/0·202/0·176
β_0	1	0·981	0·272/0·213/0·190	1·027	0·249/0·189/0·154	1·022	0·251/0·199/0·178	1·004	0·260/0·213/0·143
β_1	0	0·006	0·314/0·295/0·269	-0·006	0·327/0·295/0·191	0·003	0·329/0·310/0·299	0·018	0·337/0·298/0·234
β_2	1·5	1·502	0·028/0·029/0·006	1·502	0·028/0·029/0·007	1·502	0·028/0·029/0·006	1·502	0·028/0·029/0·007
β_3	2·5	2·500	0·003/0·003/0·001	2·500	0·003/0·003/0·001	2·500	0·003/0·003/0·001	2·500	0·003/0·003/0·001
β_4	-0·5	-0·504	0·036/0·038/0·008	-0·504	0·036/0·038/0·008	-0·504	0·036/0·038/0·007	-0·504	0·036/0·038/0·008
β_5	-1	-1·000	0·003/0·003/0·001	-0·999	0·003/0·003/0·001	-1·000	0·003/0·003/0·001	-1·000	0·003/0·003/0·001
σ_y	0·5	0·496	0·021/0·015/0·009	0·496	0·021/0·015/0·008	0·496	0·021/0·015/0·009	0·496	0·021/0·017/0·009
γ_0	2	1·937	0·233/0·191/0·112	1·996	0·204/0·196/0·101	2·022	0·228/0·184/0·088	1·983	0·225/0·197/0·123
γ_1	1·5	1·539	0·320/0·299/0·111	1·511	0·317/0·297/0·133	1·544	0·332/0·310/0·185	1·557	0·352/0·294/0·193
σ_t	0·5	0·552	0·290/0·244/0·096	0·535	0·271/0·178/0·088	0·431	0·203/0·163/0·082	0·628	0·246/0·191/0·091
τ	0·63	0·687	0·105/0·089/0·021	0·709	0·081/0·068/0·014	0·649	0·048/0·035/0·012	0·672	0·081/0·063/0·013
σ_{by}	2·56	2·570	0·138/0·117/0·057	2·543	0·134/0·131/0·040	2·539	0·123/0·130/0·034	2·545	0·124/0·130/0·041
σ_{bt}	2·19	2·180	0·213/0·179/0·085	2·159	0·190/0·160/0·024	2·203	0·177/0·173/0·076	2·098	0·202/0·196/0·088

Table 3. *Parameter estimates and standard errors under the normal scenario (iii). The top part contains the results for the large-n_i case (i.e., $\max_i\{n_i\} = 15$), whereas the bottom part contains the results for the small-n_i case (i.e., $\max_i\{n_i\} = 4$). The 'Std. Err.' columns contain the empirical/average sandwich/average model-based standard errors, respectively.*

		Frank		Gumbel		Gaussian		Student's-t	
	True	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
β_0	1	1·056	0·215/0·199/0·118	1·055	0·216/0·194/0·110	1·004	0·212/0·190/0·223	1·066	0·214/0·193/0·163
β_1	0	0·077	0·220/0·113/0·089	0·073	0·225/0·175/0·102	0·022	0·215/0·222/0·210	0·062	0·230/0·198/0·120
β_2	1·5	1·498	0·019/0·024/0·014	1·498	0·019/0·024/0·017	1·497	0·018/0·026/0·019	1·498	0·018/0·021/0·014
β_3	2·5	2·500	0·001/0·002/0·001	2·500	0·001/0·002/0·001	2·500	0·002/0·002/0·002	2·500	0·002/0·002/0·001
β_4	-0·5	-0·496	0·024/0·031/0·015	-0·496	0·024/0·028/0·013	-0·496	0·024/0·032/0·025	-0·497	0·024/0·030/0·015
β_5	-1	-1·000	0·002/0·002/0·002	-1·000	0·002/0·002/0·002	-1·000	0·002/0·003/0·002	-1·000	0·002/0·002/0·001
σ_y	0·5	0·501	0·013/0·011/0·007	0·501	0·014/0·017/0·008	0·501	0·014/0·019/0·011	0·501	0·013/0·012/0·009
γ_0	2	2·000	0·246/0·226/0·193	2·028	0·271/0·236/0·189	2·034	0·264/0·279/0·257	2·026	0·275/0·202/0·172
γ_1	1·5	1·528	0·438/0·374/0·253	1·542	0·464/0·423/0·396	1·506	0·448/0·460/0·438	1·538	0·497/0·461/0·395
σ_t	0·5	0·682	0·301/0·282/0·220	0·655	0·340/0·335/0·271	0·527	0·255/0·234/0·228	0·641	0·259/0·233/0·196
τ	0·61	0·712	0·124/0·096/0·054	0·730	0·130/0·116/0·096	0·610	0·036/0·040/0·033	0·676	0·114/0·127/0·085
σ_{by}	2·5	2·519	0·135/0·123/0·094	2·486	0·131/0·137/0·109	2·487	0·127/0·132/0·126	2·487	0·128/0·121/0·099
σ_{bt}	2·2	2·049	0·184/0·158/0·112	2·103	0·159/0·146/0·117	2·174	0·160/0·215/0·148	2·098	0·191/0·198/0·138
β_0	1	0·966	0·344/0·331/0·293	0·997	0·332/0·303/0·343	1·043	0·323/0·353/0·344	0·994	0·323/0·335/0·216
β_1	0	0·028	0·321/0·280/0·149	-0·017	0·293/0·254/0·190	0·098	0·311/0·373/0·354	0·098	0·303/0·285/0·196
β_2	1·5	1·504	0·029/0·030/0·005	1·505	0·029/0·030/0·016	1·505	0·029/0·030/0·026	1·505	0·029/0·030/0·007
β_3	2·5	2·500	0·003/0·003/0·001	2·500	0·003/0·003/0·002	2·500	0·003/0·003/0·001	2·500	0·003/0·003/0·001
β_4	-0·5	-0·506	0·037/0·038/0·007	-0·507	0·037/0·038/0·011	-0·508	0·037/0·045/0·038	-0·508	0·037/0·038/0·008
β_5	-1	-0·999	0·003/0·003/0·001	-0·999	0·003/0·003/0·001	-0·999	0·003/0·003/0·001	-0·999	0·003/0·003/0·001
σ_y	0·5	0·495	0·019/0·016/0·009	0·495	0·019/0·016/0·010	0·495	0·019/0·023/0·017	0·495	0·019/0·015/0·010
γ_0	2	1·976	0·250/0·215/0·165	2·007	0·249/0·216/0·136	2·024	0·252/0·291/0·235	1·994	0·242/0·222/0·164
γ_1	1·5	1·580	0·294/0·236/0·142	1·534	0·290/0·247/0·197	1·572	0·300/0·283/0·293	1·551	0·289/0·253/0·195
σ_t	0·5	0·494	0·263/0·236/0·123	0·502	0·280/0·258/0·213	0·424	0·197/0·250/0·207	0·559	0·245/0·246/0·183
τ	0·61	0·644	0·094/0·071/0·027	0·666	0·091/0·075/0·048	0·604	0·064/0·082/0·061	0·642	0·074/0·058/0·019
σ_{by}	2·5	2·515	0·147/0·126/0·091	2·478	0·158/0·146/0·055	2·482	0·136/0·167/0·159	2·492	0·135/0·127/0·092
σ_{bt}	2·2	2·205	0·223/0·189/0·109	2·178	0·199/0·171/0·124	2·156	0·189/0·198/0·179	2·149	0·207/0·199/0·121

Table 4. *Estimated Kendall's tau (sandwich/model-based standard errors) for the association between time to graft failure and GFR longitudinal measurements under different copulas and survival models.*

	Frank	Gumbel	Gaussian	Student's-t
Weibull	0·569 (0·091/0·062)	0·803 (0·051/0·021)	0·855 (0·022/0·011)	0·657 (0·068/0·030)
log-normal	0·564 (0·103/0·064)	0·802 (0·062/0·022)	0·629 (0·042/0·019)	0·747 (0·066/0·026)
log-logistic	0·566 (0·088/0·066)	0·802 (0·048/0·022)	0·747 (0·075/0·040)	0·591 (0·071/0·031)