# Supplementary Materials for "Multistate models with nested frailty for lifetime analysis: application to bone marrow transplantation recovery patients"

Jonathan K. J. Vasquez[1,2]; Katy C. Molina[1,2]; Vera D. Tomazella[2]; Carlos A. Diniz[2] and Adriano K. Suzuki[1]

[1] Institute of Mathematics and Computer Sciences, University of São Paulo, São Carlos, SP, Brazil.
[2] Department of Statistics, Federal University of São Carlos, São Carlos, SP, Brazil

## 1 Inference

In this section, we describe the inferential procedure for the model parameters based on the parametric and semi parametric approach.

### 1.1 Parametric estimation

In frailty models it is very common to use the Gamma distribution for random effects, due to its algebraic facility. We considered the Gamma distribution for random effects to obtain the marginal log-likelihood of the multistate model with nested frailty.

#### 1.1.1 Multistate model with Gamma frailty

The marginal log-likelihood function (Eq. 5 in the paper) depends on the distribution of $V_h$ and the Laplace transform of $W_{qh}$ and its derivative. In this case, let's assume the Gamma distributions with mean equal to 1 and variances $\theta_V$ and $\theta_q$ for the variables $V_h$ and $W_{qh}$, we have

$$f_V \overset{iid}{\sim} \text{Gamma}\left(\frac{1}{\theta_V}, \frac{1}{\theta_V}\right)$$
$$f_W \overset{iid}{\sim} \text{Gamma}\left(\frac{1}{\theta_q}, \frac{1}{\theta_q}\right),$$

where

$$f_V(v_h;\theta_V) = \frac{\left(\frac{1}{\theta_V}\right)^{1/\theta_V}}{\Gamma\left(\frac{1}{\theta_V}\right)} v_h^{1/\theta_V - 1} \exp(-v_h/\theta_V)$$
$$f_W(w_{qh};\theta_q) = \frac{\left(\frac{1}{\theta_q}\right)^{1/\theta_q}}{\Gamma\left(\frac{1}{\theta_q}\right)} w_{qh}^{1/\theta_q - 1} \exp(-w_{qh}/\theta_q).$$

With respect to the distribution of $W_{qh}$, its Laplace transform and its $d_{qh}$-th derivative of the Laplace transform at the point $k$ in the transition $q$ are, respectively,

$$\mathcal{L}_q(k) = \mathbb{E}[\exp(-W_{qh}k)] = (1+\theta_q k)^{-1/\theta_q}$$
$$\mathcal{L}_q^{d_{qh}}(k) = (-1)^{d_{qh}}\mathbb{E}\left[W_{qh}\exp(-W_{qh}k)\right] = \frac{(-1)^{d_{qh}}\Gamma(d_{qh}+\frac{1}{\theta_q})}{\Gamma\left(\frac{1}{\theta_q}\right)\theta_q^{1/\theta_q}(\frac{1}{\theta_q}+k)^{d_{qh}+1/\theta_q}}.$$

Assuming Gamma distribution for the random terms, the log-likelihood function (Eq. 5 in the paper) is given by

$$\begin{aligned}
\ell_M(\boldsymbol{\zeta}) \;=\; & \sum_{h=1}^{H}\Bigg\{\sum_{i=1}^{n_h}\sum_{q=1}^{Q}\delta_{qhi}\log\left[\lambda_{q0}(t_{qhi})\exp\left(\boldsymbol{\beta}_q^\top\mathbf{X}_{qhi}\right)\right] \\
& + \log\int_0^\infty v_h^{d_h}\prod_{q=1}^{Q}\frac{(-1)^{d_{qh}}(-1)^{d_{qh}}\Gamma\left(d_{qh}+\frac{1}{\theta_q}\right)}{\Gamma(\frac{1}{\theta_q})\theta_q^{1/\theta_q}\left[\frac{1}{\theta_q}+v_h\sum_{i=1}^{n_h}\Lambda_{q0}(t_{qhi})\exp\left(\boldsymbol{\beta}_q^\top\mathbf{X}_{qhi}\right)\right]^{d_{qh}+1/\theta_q}} \\
& \times\frac{(\frac{1}{\theta_V})^{1/\theta_V}}{\Gamma(\frac{1}{\theta_V})}v_h^{1/\theta_V-1}\exp(-v_h/\theta_V)dv_h \\
& - \log\int_0^\infty\prod_{q=1}^{Q}\left[1+\theta_q v_h\sum_{i=1}^{n_h}\Lambda_{q0}(\tau_{qhi})\exp\left(\boldsymbol{\beta}_q^\top\mathbf{X}_{qhi}\right)\right]^{-1/\theta_q} \\
& \times\frac{(\frac{1}{\theta_V})^{1/\theta_V}}{\Gamma(\frac{1}{\theta_V})}v_h^{1/\theta_V-1}\exp(-v_h/\theta_V)dv_h\Bigg\}
\end{aligned}$$

$$\begin{aligned}
= \; & \sum_{h=1}^{H}\Bigg\{\sum_{i=1}^{n_h}\sum_{q=1}^{Q}\delta_{qhi}\log\left[\lambda_{q0}(t_{qhi})\exp\left(\boldsymbol{\beta}_q^\top\mathbf{X}_{qhi}\right)\right] \\
& + \log\int_0^\infty\frac{v_h^{d_h+1/\theta_V-1}\exp(-v_h/\theta_V)}{\prod_{q=1}^{Q}\left[\frac{1}{\theta_q}+v_h\sum_{i=1}^{n_h}\Lambda_{q0}(t_{qhi})\exp\left(\boldsymbol{\beta}_q^\top\mathbf{X}_{qhi}\right)\right]^{d_{qh}+1/\theta_q}}dv_h \\
& - \log\int_0^\infty\frac{v_h^{1/\theta_V-1}\exp(-v_h/\theta_V)}{\prod_{q=1}^{Q}\left[\theta_q v_h\sum_{i=1}^{n_h}\Lambda_{q0}(\tau_{qhi})\exp\left(\boldsymbol{\beta}_q^\top\mathbf{X}_{qhi}\right)+1\right]^{1/\theta_q}}dv_h \\
& + \sum_{q=1}^{Q}\log\left[\frac{\Gamma\left(d_{qh}+\frac{1}{\theta_q}\right)}{\Gamma(\frac{1}{\theta_q})\theta_q^{1/\theta_q}}\right]\Bigg\}. \qquad (1)
\end{aligned}$$

With distributions already defined for the frailty variables, the marginal likelihood function (1) can be maximized. For this, we assume a distribution for the $Q$ basic risk functions, and maximize $\ell_M(\boldsymbol{\zeta})$ to estimate the parameter vector $\boldsymbol{\zeta} = (\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\beta})'$. The maximum likelihood estimates (MLEs) of the parameters are obtained by maximizing the numerically log-likelihood functions. The optim routine in the R software package was employed for numerical maximization.

# 2 Aplication

## 2.1 Description of the data set

According to Klein and Moeschberger, the patient's recovery process is based on two intermediate events, which can occur before two final events (death and recurrence). Intermediate events are the possibility of developing AGvHD, which usually occurs within the first 100 days after transplantation, and the recovery of the platelet count to a self-sustaining level that is greater than or equal to $40\times10^9/1$ ( being called platelet recovery). Immediately after transplantation, patients have a drop in platelet counts and are free of AGvHD. However, at some point in the process, they may have AGvHD or have their platelets recovered, which can change their prognosis. These events can occur in any order, and they can happen one after the other and, in the sequence, patients can die or have the disease relapse. Patients can also die or have the disease relapse without experiencing any of these intermediate events.

We considered a multicenter study of patients prepared for transplantation with a conditioned free radiation regimen. In total, 137 patients were treated at one of four hospitals: 76 at Ohio State University Hospitals (OSU) in Colombo; 17 at Hahnemann University (HU) in Philadelphia; 23 at St. Vincent's Hospital (SVH) in Sydney; and 21 at Alfred Hospital (AH) in Melbourne. For more details on this study, see Copelan et al. .

The data set is described as follows: **patient**: Patient identification, **hospital**: hospital in which the patient underwent treatment (1: OSU; 2: HU; 3: SVH; 4: AH), **t1**: time (in days) to death or last follow-up, **t2**: time (in days) of disease-free survival (time to relapse, time to death, or last follow-up), **d1**: death indicator (1: yes; 0: no), **d2**: recurrence indicator (1: yes; 0: no), **ta**: time (in days) for the development of AGvHD, **da**: AGvHD indicator (1: yes; 0: no), **tp**: time (in days) for platelet recovery, **dp**: platelet recovery indicator (1: platelets returned to normal; 0: platelets did not return to normal), $\mathbf{X_1}$: patient's age (in years) at the time of transplantation, $\mathbf{X_2}$: donor age (in years).

If a patient develops AGvHD after the platelet recovery count, that state will be represented as AGvHD/Pla.R. In Table 1 we have the total number of events for each type of transition. Therefore, according to the results presented in Table 1, of the 137 post-transplant patients, 116 had their platelet count recovered (transition T→ Pla.R), 8 developed acute graft versus host disease (transition T→AGvHD), 3 had disease relapse without experiencing intermediate events ( transition T→R) and 10 died without also experiencing the intermediate events (transition T→D). Of the 116 patients who had their platelet count recovered after transplantation, 18 of them developed AGvHD (transition Pla.R→AGvHD), 20 died without developing AGvHD (transition Pla.R→D) and most of them (34) had a recurrence of the disease (transition Pla.R→ R).

Table 1: Transition frequencies.

| From/For | T | Pla.R | AGvHD | AGvHD/Pla.R | R | D |
|---|---|---|---|---|---|---|
| T | - | 116 | 8 | - | 3 | 10 |
| Pla.R | - | - | - | 18 | 34 | 20 |
| AGvHD | - | - | - | - | 0 | 7 |
| AGvHD/Pla.R | - | - | - | - | 3 | 6 |
| R | - | - | - | - | - | - |
| D | - | - | - | - | - | - |

The regression parameter estimates and 95% confidence intervals using the MMNF are shown in the Table 2. are found

Table 2: Regression parameter estimates and 95% confidence intervals for the multistate models with nested frailty.

| Parameters | LL | exp($\beta$) | UL | Parameters | LL | exp($\beta$) | UL |
|---|---|---|---|---|---|---|---|
| $\beta_{1.1}$ | 0.9960 | 1.0280 | 1.0600 | $\beta_{2.1}$ | 0.9590 | 0.9854 | 1.0120 |
| $\beta_{1.2}$ | 0.9741 | 1.1020 | 1.2470 | $\beta_{2.2}$ | 0.9047 | 1.0020 | 1.1090 |
| $\beta_{1.3}$ | 0.8894 | 1.1040 | 1.3700 | $\beta_{2.3}$ | 0.8259 | 0.9814 | 1.1660 |
| $\beta_{1.4}$ | 0.8490 | 0.9474 | 1.0570 | $\beta_{2.4}$ | 1.0030 | 1.1020 | 1.2100 |
| $\beta_{1.5}$ | 0.9458 | 1.0180 | 1.0960 | $\beta_{2.5}$ | 0.9411 | 1.0180 | 1.1010 |
| $\beta_{1.6}$ | 0.9898 | 1.0440 | 1.1000 | $\beta_{2.6}$ | 0,9255 | 0.9764 | 1.0300 |
| $\beta_{1.7}$ | 0.9500 | 1.0210 | 1.0980 | $\beta_{2.7}$ | 0.9022 | 0.9697 | 1.0420 |
| $\beta_{1.8}$ | 0.7697 | 0.9364 | 1.1390 | $\beta_{2.8}$ | 0.9184 | 1.0830 | 1.2760 |
| $\beta_{1.9}$ | 0.6535 | 0.8401 | 1.0800 | $\beta_{2.9}$ | 0.8345 | 1.1740 | 1.6520 |
| $\beta_{1.10}$ | 0.9454 | 1.0470 | 1.1600 | $\beta_{2.10}$ | 0.7869 | 0.9640 | 1.1810 |

Finally, we analized each transition intensities in detail for the multistate models with nested frailty (MMNF), so that

**Transition 1 (T→Pla. R):** For a one-unit increment in age, his risk of having his platelet count recovered soon after transplantation (transition 1) increases by 2.76%. Therefore, the older the patient, the greater the risk of recovering their platelet count. Already, for each additional year in the donor's age, the patient's risk of recovering their platelets decreases by 1.46% (1-0.9854).

**Transition 2 (T→AGvHD):** Age also has a positive effect on this transition, increasing by 10.23% the patient's risk of developing AGvHD soon after transplantation. For each additional year in donor age, the patient's risk of developing AGvHD increases by 0.17%.

**Transition 3 (T→R):** For each year of the patient's age, the risk of having a recurrence of the disease, without having their platelet count recovered and having developed AGvHD, increases by 10.37%. For each year of the donor's age, the patient's risk of having the disease relapse, without experiencing the intermediate events, decreases by 1.86%.

**Transition 4 (T→D):** For each year of patient age, the risk of death after transplantation without experiencing the intermediate events decreases by 5.26%. For each year of age of the donor, the risk of this transition increases by 10.15%.

**Transition 5 (Pla. R→AGvHD/Pla. R):** For each year of patient age, the risk of a patient, who after transplantation had platelet recovery and then developed AGvHD, increases by 1.82%. Donor age also has a positive effect on this transition, increasing its risk by 1.77% for each year of donor age.

**Transition 6 (Pla. R→R):** The risk of a patient whose platelet count recovered after transplantation and then having the disease relapse increases by 4.36% for each year of their age. However, the risk of this transition decreases by 2.36% for each year of age of the donor.

**Transition 7 (Pla. R→D):** The risk of a patient who had their platelet count recovered after transplantation and then dying increases by 2.14% for each year of their age. The risk of this transition decreases by 3.03% for each year of age of the donor.

**Transition 8 (AGvHD→D):** For each year of patient age, the risk of developing AGvHD, given that he had platelet count recovery soon after transplantation, decreases by 6.36%. The risk of

transition increases by 8.27% for each year of age of the donor.

**Transition 9 (AGvHD/Pla. R→R):** A patient's risk of developing AGvHD soon after transplantation decreases by about 16% for each year of age. The risk of transition also increases for each year of age of the donor, with an increase of 17.42%.

**Transition 10 (AGvHD/Pla. R→D):** A patient's risk of relapse, given that he developed AGvHD after his platelets recovered, increases by 4.74% for each year of his age. However, the risk for this transition decreases by 3.6% for each year of age of the donor.