Supplementary Document for the Manuscript entitled "Joint Modelling for Organ Transplantation Outcomes for Patients with Diabetes and the End-Stage Renal Disease"

1 Monte Carlo EM algorithm

1.1 M-step

To make the notation short, let $E^{(t)}(g(\boldsymbol{\beta}_i)) = E[g(\boldsymbol{\beta}_i)|t, \boldsymbol{w}(t), S, \delta, \boldsymbol{Z}, Y(t), \Theta^{(t)}]$ be the conditional log likelihood based on the current estimate $\Theta^{(t)}$ for any function $g(\boldsymbol{\beta}_i)$. The MLE of $\boldsymbol{b}$, $\boldsymbol{B}$, $\boldsymbol{\alpha}$, and σ^2 can be written as

$$
\begin{aligned}
\hat{\boldsymbol{b}} &= \sum_{i=1}^{n} E^{(t)}(\boldsymbol{\beta}_i) \\
\hat{\boldsymbol{B}} &= \sum_{i=1}^{n} E^{(t)}(\boldsymbol{\beta}_i - \hat{\boldsymbol{b}})(\boldsymbol{\beta}_i - \hat{\boldsymbol{b}}) \\
\hat{\boldsymbol{\alpha}} &= \sum_{i=1}^{n} E^{(t)}((\boldsymbol{Z}^T\boldsymbol{Z})^{-}\boldsymbol{Z}^T(\boldsymbol{Y}_i - \boldsymbol{\beta}_i^T\boldsymbol{\xi}(\boldsymbol{t}_i))) \\
\hat{\sigma}^2 &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{E^{(t)}(Y_{ij} - \hat{\boldsymbol{\alpha}}^T\boldsymbol{Z} - \boldsymbol{\beta}_i^T\boldsymbol{\xi}(t_{ij}))^2}{\sum_{i=1}^{n} m_i},
\end{aligned}
$$

where $\boldsymbol{Y}_i = (Y_{i1}, \cdots, Y_{im_i})^T$, and $\boldsymbol{t}_i = (t_{i1}, \cdots, t_{im_i})^T$.

We estimate the baseline hazard function λ_0 by a step-function. Let $T_1, \cdots, T_H$ be all observed event times, then the baseline failure time is

$$\Phi(T_h, Z_i, \boldsymbol{w}_i(t_h), \boldsymbol{\beta}_i, \boldsymbol{\gamma}) = \int_0^{T_h} exp[\boldsymbol{\gamma}_1^T\boldsymbol{Z}_i + \boldsymbol{\gamma}_2^T\boldsymbol{\beta}_i + \boldsymbol{w}_i(s|\boldsymbol{\gamma}_3)]ds,$$

where $h = 1, \cdots, H$. Let $\mu_h = \Phi(t_h, Z_i, \boldsymbol{w}_i(t_h), \boldsymbol{\beta}_i, \boldsymbol{\gamma})$, we estimate μ_h by plugging in the current estimate of $\boldsymbol{\beta}_i$ and $\boldsymbol{\gamma}_i^T = (\boldsymbol{\gamma}_1^T, \boldsymbol{\gamma}_2^T, \boldsymbol{\gamma}_3^T)$. We get $0 = \hat{\mu}_{(0)} \leq \hat{\mu}_{(1)} \leq \cdots \leq \hat{\mu}_{(H)}$ by ordering these estimate in the data. Then the baseline function can be specified as $\lambda_0(\mu) = \sum_{h=1}^{H} C_h 1_{\{\hat{\mu}_{(h-1)} < \mu \leq \hat{\mu}_{(h)}\}}$. Now let the derivative of $E^{(t)}(l(\Theta))$ w.r.t C_h be equal to zero, then we obtain the maximum likelihood estimate for C_h:

$$\hat{C}_h = \frac{\sum_{i=1}^n E_i^{(t)}[\delta_i 1_{\hat{\mu}_{(h-1)} < \mu_i \le \hat{\mu}_{(h)}}]}{\sum_{i=1}^n E_i^{(t)}[\{\hat{\mu}_{(h)} - \hat{\mu}_{(h-1)}\} 1_{\{\hat{\mu}_{(h)} \le \mu_i\}}]}.$$

If we insert the baseline hazard function $\hat{\lambda}_0(\mu)$ into the conditional log likelihood, then we have

$$\begin{aligned} Q(\Theta|\Theta^{(t)}) &= \sum_{i=1}^n E^{(t)}\Big[\delta_i \log\{\sum_{h=1}^H \hat{C}_h 1_{\{\hat{\mu}_{(h-1)} < \mu_i \le \hat{\mu}_{(h)}\}}\} + \delta_i(\boldsymbol{\gamma}_1^T \boldsymbol{Z} + \boldsymbol{\gamma}_2^T \beta_i + \\ &\quad \boldsymbol{w}(t|\boldsymbol{\gamma}_3)) - \sum_{h=1}^H \hat{C}_h \{\hat{\mu}_{(h)} - \hat{\mu}_{(h-1)}\} 1_{\{\hat{\mu}_{(h)} \le \mu_i\}} + \\ &\quad \sum_{j=1}^{m_i} \log f(Y_{ij}|\boldsymbol{\beta}_i, \boldsymbol{\alpha}, \sigma^2) + \log f(\boldsymbol{\beta}_i|\boldsymbol{b}, \boldsymbol{B})\Big]. \end{aligned}$$

After we have obtained the estimate for the parameters $\boldsymbol{b}$, $\boldsymbol{B}$, $\boldsymbol{\alpha}$, $\hat{\sigma^2}$, and the baseline hazard function $\hat{\lambda}_0(t)$, the last parameter to estimate is $\boldsymbol{\gamma}$. The estimate for $\boldsymbol{\gamma}$ has no closed-form. So we use the numeric optimization algorithm such as *optim*() in R to estimate $\boldsymbol{\gamma}$ in the M-step.