1 JOINT MODELING OF ORR&TTE ENDPOINTS

1.1 Model specification

Assume the observed data were collected from \( n \) patients. We suppose that there are \( K \) unique response states for the ORR endpoints. Let \( R_i \) denotes the response for patient \( i \). Given the covariates \( X_i \), the conditional probability for \( R_i \) is defined as

\[
\begin{cases}
\pi_k(x_i) = P(R_i = k | X_i) = \frac{\exp \{ g_k(x_i) \}}{1 + \sum_{l=1}^{K} \exp \{ g_l(x_i) \}} \\
g_k(x_i) = \ln \frac{P(R_i = k | X_i)}{P(R_i = 0 | X_i)} = x_i' \beta_R - \gamma_k \\
g_0(x_i) = 0
\end{cases}
\]

where \( R_i = 0 \) is set to be the anchoring group. Due to the limited number of clinically available CR patients in practice, this paper solely considers the case of \( K = 3 \), i.e., \( R_2, R_1, R_0 \) denoting for CR&PR, SD and PD, respectively.

Let \( T^*_i \) denotes the underlying time to event for patient \( i \). For the usual right censoring mechanism, we can observe \( T_i = \min(T^*_i, C_i) \), where \( C_i \) denoting the censoring time. Meanwhile, we define the event indicator as \( \delta_i = I(T^*_i \leq C_i) \), where \( I(\cdot) \) is the indicator function that takes the value 1 if the event is observed, and 0 otherwise. Given the covariates \( X_i \), the hazard function for the \( i \)th patient at time \( t \) is defined as

\[
\lambda(t; X_i) = \lim_{\Delta t \to 0} \frac{\Pr (t \leq T^*_i < t + \Delta t | T^*_i \geq t, X_i)}{\Delta t}
\]

The observations are consist of \( D = \{ R_i, T_i, \delta_i, X_i \}, i = 1, \ldots, n \). The random effects are employed to demonstrate the links between various clinical outcomes. Specifically, we will follow the Generalized Linear Mixed Model (GLMM) framework to introduce random effects into the multinomial logistic regression model and proportional hazards function.

\[
\begin{cases}
g_1(x_{ij}) = x_i' \beta_{R1} + \gamma_1 + U_{i1} \\
g_2(x_i) = x_i' \beta_{R2} + \gamma_2 + U_{i2} \\
\lambda(t; x_i) = \lambda_0(t) \exp(x_i' \beta_3 + V_i)
\end{cases}
\]

where \( g_0(x_i) = 0 \) denotes the reference group, \( x_i \) denotes the vector of covariates, \( \beta_k = (\beta'_{Rk}, \gamma_k) \), \( k = 1 \) or 2, denotes the corresponding logistic regression coefficients for response \( k \), and \( \beta_3 \) denotes the risk factors effects on hazard function. A positive value of coefficients in \( \beta_{Rk} \) implies a higher probability of response \( k \) associated with the corresponding covariate, while a positive value of coefficients in \( \beta_3 \) implies a higher risk of failure associated with the risk factor. \( U_{i1}, U_{i2} \) and \( V_i, i = 1, \ldots, n \), denote random effects of cluster \( i \) on the respective endpoint, which are assumed to follow the multivariate normal distribution \( N(0, \Sigma) \) with
Let \( u_1 = (U_{11}, ..., U_{1n})', u_2 = (U_{21}, ..., U_{2n})', v = (V_1, ..., V_n)', \) and \( a = (u_1', u_2', v') \). Then the vector \( a \) follows the normal distribution with mean zero and covariance matrix \( \Omega = \Sigma \otimes I_n \), where \( I_n \) is a \( n \times n \) unit matrix, and \( \otimes \) denotes the Kronecker product of two matrices. We further define the linear predictors as

\[
\begin{align*}
\eta_{1i} &= x_i' \cdot \beta_{R1} + \gamma_1 + U_{1i} \\
\eta_{2i} &= x_i' \cdot \beta_{R2} + \gamma_2 + U_{2i} \\
\eta_{3i} &= x_i' \cdot \beta_3 + V_i
\end{align*}
\]

Following the GLMM formulation, the random effects are set to be conditionally fixed, then the best linear unbiased prediction (BLUP) type log-likelihood can be written as the sum of following three terms:

\[
\begin{align*}
\ell_R &= \sum_i \left[ \sum_{k=1}^{2} R_{ki} \eta_{ki} - \log \left\{ 1 + \sum_{k=1}^{2} \exp(\eta_{ki}) \right\} \right] \\
\ell_S &= \sum_i \log \left\{ \lambda_0(t) \exp \eta_{3i} \right\} \delta_i \times \exp \left\{ -\Lambda_0(t) \exp \eta_{3i} \right\} \\
\ell_r &= \frac{1}{2} \left[ \log \left\{ (2\pi)^K n |\Omega| \right\} + a' \Omega^{-1} a \right]
\end{align*}
\]

where \( \lambda_0(t) \) is called baseline hazard function, and \( \Lambda_0(t) = \int_0^t \lambda_0(s) ds \) denotes the corresponding cumulative baseline hazard. To simplify the estimation procedure, the baseline hazard can be eliminated in the conditional partial likelihood function \( \ell_S \). Following the marginal likelihood argument, we arrange the \( n \) distinct failure times in an increasing order by \( t_1 < t_2 < \cdots < t_n \). And the log-likelihood with random effects being conditionally fixed, \( \ell_S \), can be rewritten as

\[
\ell_S = \sum_{i=1}^{n} \log \frac{\exp \eta_{3i}}{\sum_{s \in R(t_i)} \exp \eta_{3s}}
\]

where \( R(t) = \{ j : T_j \geq t \} \) is the risk set at \( t^- \). So the BLUP type log-likelihood is formulated as

\[
\begin{align*}
\ell_f &= \ell_R + \ell_S \\
\ell_{blup} &= \ell_f + \ell_r
\end{align*}
\]

**Remark:** When the response degrades to a dichotomous indicator (clinically, the tumor status is frequently reduced to response or non-response to simplify the study), the multinomial logistic regression in the proposed model degenerates to a binary logistic regression.
To estimate the corresponding regression coefficient vectors $\beta_1$, $\beta_2$ and the risk effect $\beta_3$, the conditionally fixed random effects $(u_1', u_2', v')$ and the variance component parameters $(\theta_1, \theta_2, \theta_3, \rho_1, \rho_2, \rho_3)$, the estimation procedure can be performed under GLMM framework. The use of GLMM can avoid the need for intractable high-dimensional integration involved in the marginal likelihood methods or time-consuming Monte Carlo approximations in the E-step for maximum likelihood estimation. The original work of McGilchrist extended the BLUP method for linear mixed models (LMMs) to a GLMM setting (with additional random effect terms in the linear predictor) through the residual maximum likelihood (REML) estimation procedures, which involves an iterative numerical procedure:

\[
\begin{pmatrix}
\hat{\beta}_1 \\
\hat{\beta}_2 \\
\hat{\beta}_3 \\
\hat{u}_1 \\
\hat{u}_2 \\
\hat{v}
\end{pmatrix} =
\begin{pmatrix}
\beta_{10} \\
\beta_{20} \\
\beta_{30} \\
u_{10} \\
u_{20} \\
v_{0}
\end{pmatrix}
+ I_0^{-1}
\begin{Bmatrix}
X' R 0 0 \\
0 X' R 0 \\
0 0 X' \\
0 Z' 0 \\
0 0 Z'
\end{Bmatrix}
\begin{bmatrix}
\partial \ell_f / \partial \eta_1 \\
\partial \ell_f / \partial \eta_2 \\
\partial \ell_f / \partial \eta_3
\end{bmatrix}
- \frac{1}{\theta_1 \theta_2 \theta_3 \left[ 1 - \rho_1^2 - \rho_2^2 - \rho_3^2 + 2\rho_1 \rho_2 \rho_3 \right]}
\begin{pmatrix}
0 \\
0 \\
0
\end{pmatrix}
\times
\begin{pmatrix}
\rho_1 \rho_2 \rho_3 \\
\rho_1 \rho_2 \rho_3 \\
\rho_1 \rho_2 \rho_3
\end{pmatrix}
\begin{pmatrix}
u_{10} \left( 1 - \rho_3^2 \right) \theta_2 \theta_3 - u_{20} \left( \rho_1 - \rho_2 \rho_3 \right) \theta_3 \sqrt{\theta_1 \theta_2} - v_0 \left( \rho_2 - \rho_1 \rho_3 \right) \theta_2 \sqrt{\theta_1 \theta_3} - u_{10} \left( \rho_1 - \rho_2 \rho_3 \right) \theta_3 \sqrt{\theta_1 \theta_2} + u_{20} \left( 1 - \rho_2^2 \right) \theta_1 \theta_3 - v_0 \left( \rho_3 - \rho_1 \rho_2 \right) \theta_1 \sqrt{\theta_2 \theta_3} - u_{10} \left( \rho_2 - \rho_1 \rho_3 \right) \theta_2 \sqrt{\theta_1 \theta_3} - u_{20} \left( \rho_3 - \rho_1 \rho_2 \right) \theta_1 \sqrt{\theta_2 \theta_3} + v_0 \left( 1 - \rho_2^2 \right) \theta_1 \theta_2
\end{pmatrix}
\]
where $\beta_{10}$, $\beta_{20}$, $\beta_{30}$, $u_{10}$, $u_{20}$, $v_0$ are initial values of $\beta_1$, $\beta_2$, $\beta_3$, $u_1$, $u_2$, $v$ respectively, and are replaced by their updated estimates in each iteration, for given values of $\theta_1$, $\theta_2$, $\theta_3$, $\rho_1$, $\rho_2$, $\rho_3$. $X_R = (X, 1)$ indicates the design matrix with added intercepts $I$. $f$ is the negative second derivative of log-likelihood $\ell_{blup}$ with respect to $\beta_1$, $\beta_2$, $\beta_3$, $u_1$, $u_2$, $v$. Specifically,
I = \begin{bmatrix}
X_R' & 0 & 0 \\
0 & X_R' & 0 \\
Z' & 0 & 0 \\
0 & Z' & 0 \\
0 & 0 & Z'
\end{bmatrix}
\begin{bmatrix}
-\partial^2 \ell_f & -\partial^2 \ell_f & -\partial^2 \ell_f \\
\partial \eta_1 \partial \eta_1 & \partial \eta_1 \partial \eta_2 & \partial \eta_1 \partial \eta_3 \\
-\partial^2 \ell_f & -\partial^2 \ell_f & -\partial^2 \ell_f \\
\partial \eta_2 \partial \eta_1 & \partial \eta_2 \partial \eta_2 & \partial \eta_2 \partial \eta_3 \\
-\partial^2 \ell_f & -\partial^2 \ell_f & -\partial^2 \ell_f \\
\partial \eta_3 \partial \eta_1 & \partial \eta_3 \partial \eta_2 & \partial \eta_3 \partial \eta_3
\end{bmatrix}
\begin{bmatrix}
X_R' & 0 & 0 & Z' & 0 & 0 \\
0 & X_R' & 0 & 0 & Z' & 0 \\
0 & 0 & X' & 0 & 0 & Z
\end{bmatrix}
+ \frac{1}{\theta_1 \theta_2 \theta_3 (1 - \rho_1^2 - \rho_2^2 - \rho_3^2 + 2 \rho_1 \rho_2 \rho_3)}
\begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & (1 - \rho_3^2) \theta_2 \theta_3 & (1 - \rho_2 \rho_3) \theta_3 \theta_1 \theta_2 & - (\rho_2 - \rho_1 \rho_3) \theta_2 \theta_1 \theta_3 \\
0 & T & (1 - \rho_3^2) \theta_1 \theta_3 & - (\rho_3 - \rho_1 \rho_2) \theta_1 \theta_2 \theta_3 \\
0 & T & 0 & (1 - \rho_1^2) \theta_1 \theta_2
\end{bmatrix}

where the mark $T$ implies that the matrix is symmetric.

[ii] To calculate the asymptotic standard errors for the regression parameter estimates $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$, we denote

$$
I^{-1} = \begin{pmatrix}
T_{11} & T_{12} & T_{13} & T_{14} \\
T_{21} & T_{22} & T_{23} & T_{24} \\
T_{31} & T_{32} & T_{33} & T_{34} \\
T_{41} & T_{42} & T_{43} & T_{44}
\end{pmatrix}
$$

where the block diagonal matrices $I^{-1}$ can be partitioned conformably into $\beta_1 \mid \beta_2 \mid \beta_3 \mid a$. The corresponding block diagonal matrices $T_{11}, T_{22}$ and $T_{33}$ can provide the asymptotic variance for $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$, respectively. Following the GLMM method, given $\hat{a} = \begin{pmatrix} \hat{u}_1' \hat{u}_2' \hat{v}' \end{pmatrix}$ and the block diagonal matrix $T_{44}$, the REML estimates of $\theta_1, \theta_2, \theta_3, \rho_1, \rho_2, \rho_3$ can be obtained by solving the following nonlinear equations:

$$
\begin{align*}
\text{tr} \Omega^{-1} \frac{\partial \Omega}{\partial \theta_m} + \text{tr} \left( T_{44} + \hat{a} \hat{a}' \right) \frac{\partial \Omega^{-1}}{\partial \theta_m} &= 0 \\
\text{tr} \Omega^{-1} \frac{\partial \Omega}{\partial \rho_m} + \text{tr} \left( T_{44} + \hat{a} \hat{a}' \right) \frac{\partial \Omega^{-1}}{\partial \rho_m} &= 0
\end{align*}
$$

(S7)

where $m = 1, 2, 3$.

[iii] Repeat steps [i] and [ii] until convergence criteria are fulfilled.

As the above procedure involves manipulations of large matrices, the computational complexity and cost of the proposed method are extremely high. For example, update based on eq(S6) in Step [i] appears in every iteration and involves the computation of matrices $I$ and $I^{-1}$, and the time for calculating the standard errors in Step [ii] is also high. But it is needed only once after the simplified REML estimators have been obtained. Furthermore, predictions of the endpoints can be obtained using the empirical Bayes (EB) method to estimate the point estimates and the variances of the random effects.
Following the developments in McGilchrist (1993), simplified version of the above REML estimators for the variance component parameters and their asymptotic standard errors can be obtained.

\[ \eta_k' = (\eta_{k1}, \ldots, \eta_{kn}), \quad (k = 1, 2) \]

\[ w_i = \exp \eta_{3i}, \quad a_i = \delta_i / \sum_{j=i}^n w_j, \]

\[ b_i = \sum_{j=1}^i a_j, \quad d = (\delta_i, \ldots, \delta_n) \]

\[ W = \text{diag} \{w_1, \ldots, w_n\}, \quad A = \text{diag} \{a_1, \ldots, a_n\}, \quad B = \text{diag} \{b_1, \ldots, b_n\} \]

We can rewrite

\[ \frac{\partial l_f}{\partial \eta_k} - \frac{\partial^2 l_f}{\partial \eta_k \partial \eta'_k} - \frac{\partial^2 l_f}{\partial \eta_1 \partial \eta_2} - \frac{\partial^2 l_f}{\partial \eta_3 \partial \eta_3} \]

as

\[ \frac{\partial l_f}{\partial \eta_k} = \left( R_{k1} - \frac{\exp \eta_{11} + \exp \eta_{21}}{1 + \exp \eta_{11} + \exp \eta_{21}}, \ldots, R_{kn} - \frac{\exp \eta_{kn} + \exp \eta_{2n}}{1 + \exp \eta_{kn} + \exp \eta_{2n}} \right)' \]

where \( R_{ki} = I (R_i = k), \quad k = 1, 2, \ldots \)

\[ - \frac{\partial^2 l_f}{\partial \eta_k \partial \eta'_k} = \text{diag} \left\{ \frac{\exp \eta_k + \exp (\eta_1 + \eta_2)}{(1 + \exp \eta_1 + \exp \eta_2)^2} \right\} \]

\[ - \frac{\partial^2 l_f}{\partial \eta_1 \partial \eta_2} = \text{diag} \left\{ -\frac{\exp (\eta_1 + \eta_2)}{(1 + \exp \eta_1 + \exp \eta_2)^2} \right\} \]

\[ \frac{\partial l_f}{\partial \eta_3} = d - WMA1 \]

\[ - \frac{\partial^2 l_f}{\partial \eta_3 \partial \eta_3} = WB - WMA^2 M'W \]

where \( 1' = (1, \ldots, 1) \).

Adapting from Lai and Yau (2008), the derivatives for \( \Omega \) and \( \Omega^{-1} \), which are now applied to the multi-endpoint joint model, are given as follows:

\[ \Omega^{-1} = \frac{1}{\theta_1 \theta_2 \theta_3 J} \times \begin{pmatrix}
(1 - \rho_3^2) \theta_2 \theta_3 I_n & - (\rho_1 - \rho_2 \rho_3) \theta_3 \sqrt{\theta_1} \theta_2 I_n & - (\rho_2 - \rho_1 \rho_3) \theta_2 \sqrt{\theta_1 \theta_3} I_n \\
T & (1 - \rho_2^2) \theta_1 \theta_3 I_n & - (\rho_3 - \rho_1 \rho_2) \theta_1 \sqrt{\theta_2} \theta_3 I_n \\
T & T & (1 - \rho_1^2) \theta_1 \theta_2 I_n
\end{pmatrix} \]

\[ \frac{\partial \Omega^{-1}}{\partial \theta_1} = \frac{1}{2 \theta_1^2 \theta_2 \theta_3 J} \times \begin{pmatrix}
-2 (1 - \rho_3^2) \theta_2 \theta_3 & (\rho_1 - \rho_2 \rho_3) \theta_3 \sqrt{\theta_1} \theta_2 & (\rho_2 - \rho_1 \rho_3) \theta_2 \sqrt{\theta_1 \theta_3} \\
T & 0 & 0 \\
T & T & 0
\end{pmatrix} \]
Then the estimating equations for the variance component parameters can be rewritten as:

\[
\frac{\partial \Omega^{-1}}{\partial \theta_2} = \frac{1}{2 \theta_1^2 \theta_2 J} \times \begin{pmatrix}
0 & (\rho_1 - \rho_2 \rho_3) \theta_3 \sqrt{\theta_1 \theta_2} & (\rho_3 - \rho_1 \rho_2) \theta_1 \sqrt{\theta_2 \theta_3} \\
T & -2 (1 - \rho_2^2) \theta_1 \theta_3 & 0 \\
0 & 0 & (\rho_3 - \rho_1 \rho_3) \theta_1 \sqrt{\theta_2 \theta_3}
\end{pmatrix}
\]

\[
\frac{\partial \Omega^{-1}}{\partial \theta_3} = \frac{1}{2 \theta_1^2 \theta_2 J} \times \begin{pmatrix}
0 & 0 & (\rho_3 - \rho_1 \rho_3) \theta_1 \sqrt{\theta_2 \theta_3} \\
T & -2 (1 - \rho_1^2) \theta_1 \theta_2 & 0 \\
0 & 0 & (\rho_2 - \rho_1 \rho_3) \theta_2 \sqrt{\theta_1 \theta_3}
\end{pmatrix}
\]

\[
\frac{\partial \Omega^{-1}}{\partial \rho_1} = \frac{1}{\theta_1 \theta_2 \theta_3 J^2} \times \begin{pmatrix}
2 (\rho_1 - \rho_2 \rho_3) (1 - \rho_2^2) \theta_2 \theta_3 & \left\{ J + 2 (\rho_1 - \rho_2 \rho_3)^2 \right\} \theta_3 \sqrt{\theta_1 \theta_2} & \left\{ \rho_3 (1 + \rho_1^2 + \rho_2^2 - \rho_3^2) - 2 \rho_1 \rho_2 \right\} \theta_2 \sqrt{\theta_1 \theta_3} \\
T & 2 (\rho_1 - \rho_2 \rho_3) (1 - \rho_2^2) \theta_1 \theta_3 & \left\{ \rho_2 (1 + \rho_1^2 - \rho_2^2 - \rho_3^2) - 2 \rho_1 \rho_3 \right\} \theta_1 \sqrt{\theta_2 \theta_3} + 2 \left\{ \rho_1 (\rho_2^2 + \rho_3^2) - \rho_2 \rho_3 (\rho_1^2 + 1) \right\} \theta_1 \theta_2 \\
T & 2 (\rho_2 \rho_1 \rho_3 + 1) \theta_1 \theta_3 & \left\{ J + 2 (\rho_2 - \rho_1 \rho_3)^2 \right\} \theta_2 \sqrt{\theta_1 \theta_3} - \left\{ \rho_1 (1 - \rho_1^2 + \rho_2^2 + \rho_3^2) - 2 \rho_2 \rho_3 \right\} \theta_1 \sqrt{\theta_2 \theta_3} + 2 \left( \rho_2 - \rho_1 \rho_3 \right) (1 - \rho_2^2) \theta_1 \theta_2
\end{pmatrix}
\]

\[
\frac{\partial \Omega^{-1}}{\partial \rho_2} = \frac{1}{\theta_1 \theta_2 \theta_3 J^2} \times \begin{pmatrix}
2 (\rho_2 - \rho_1 \rho_3) (1 - \rho_2^2) \theta_2 \theta_3 & \left\{ \rho_3 (1 + \rho_1^2 + \rho_2^2 - \rho_3^2) - 2 \rho_1 \rho_2 \right\} \theta_3 \sqrt{\theta_1 \theta_2} & \left\{ J + 2 (\rho_2 - \rho_1 \rho_3)^2 \right\} \theta_2 \sqrt{\theta_1 \theta_3} \\
T & 2 \left\{ \rho_2 (\rho_1^2 + \rho_3^2) - \rho_1 \rho_3 (\rho_2^2 + 1) \right\} \theta_1 \theta_3 & \left\{ \rho_1 (1 - \rho_1^2 + \rho_2^2 + \rho_3^2) - 2 \rho_2 \rho_3 \right\} \theta_1 \sqrt{\theta_2 \theta_3} + 2 \left( \rho_2 - \rho_1 \rho_3 \right) (1 - \rho_2^2) \theta_1 \theta_2 \\
T & 2 (\rho_3 - \rho_1 \rho_2) (1 - \rho_2^2) \theta_1 \theta_3 & \left\{ J + 2 (\rho_2 - \rho_1 \rho_3)^2 \right\} \theta_2 \sqrt{\theta_1 \theta_3} - \left\{ \rho_1 (1 - \rho_1^2 + \rho_2^2 + \rho_3^2) - 2 \rho_2 \rho_3 \right\} \theta_1 \sqrt{\theta_2 \theta_3} + 2 \left( \rho_2 - \rho_1 \rho_3 \right) (1 - \rho_2^2) \theta_1 \theta_2
\end{pmatrix}
\]

where \( J = 1 - \rho_1^2 - \rho_2^2 - \rho_3^2 + 2 \rho_1 \rho_2 \rho_3 \). Let \( J_1 = \begin{pmatrix} I_n & 0 & 0 \\ 0 & I_n & 0 \\ 0 & 0 & I_n \end{pmatrix} \), \( J_2 = \begin{pmatrix} 0 & I_n & 0 \\ I_n & 0 & 0 \\ 0 & 0 & I_n \end{pmatrix} \), \( J_3 = \begin{pmatrix} 0 & 0 & I_n \\ I_n & 0 & 0 \\ 0 & 0 & I_n \end{pmatrix} \), and

\[
\begin{align*}
J_4 &= \begin{pmatrix} 0 & 0 & 0 \\ I_n & 0 & 0 \\ 0 & I_n & 0 \end{pmatrix}, \quad J_5 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & I_n \\ 0 & I_n & 0 \end{pmatrix}, \quad J_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & I_n \\ 0 & I_n & 0 \end{pmatrix},
\end{align*}
\]

and

\[
\begin{align*}
L_1 &= \text{tr} \cdot J_1 \left( T_{44} + \hat{a} \hat{a}' \right), \quad L_2 = \text{tr} \cdot J_2 \left( T_{44} + \hat{a} \hat{a}' \right) / 2, \quad L_3 = \text{tr} \cdot J_3 \left( T_{44} + \hat{a} \hat{a}' \right) / 2, \\
L_4 &= \text{tr} \cdot J_4 \left( T_{44} + \hat{a} \hat{a}' \right), \quad L_5 = \text{tr} \cdot J_5 \left( T_{44} + \hat{a} \hat{a}' \right) / 2, \quad L_6 = \text{tr} \cdot J_6 \left( T_{44} + \hat{a} \hat{a}' \right)
\end{align*}
\]

Then the estimating equations for the variance component parameters can be rewritten as:

\[
f_1 = \frac{n}{\theta_1} + \frac{1}{\theta_1^2 \theta_2 \theta_3 J} \left\{ -(1 - \rho_2^2) \theta_2 \theta_3 L_1 + (\rho_1 - \rho_2 \rho_3) \theta_3 \sqrt{\theta_1 \theta_2} L_2 + (\rho_2 - \rho_1 \rho_3) \theta_2 \sqrt{\theta_1 \theta_3} L_3 \right\} \quad \text{(S8)}
\]

\[
f_2 = \frac{n}{\theta_2} + \frac{1}{\theta_1^2 \theta_2^2 \theta_3 J} \left\{ -(1 - \rho_2^2) \theta_1 \theta_3 L_4 + (\rho_1 - \rho_2 \rho_3) \theta_3 \sqrt{\theta_1 \theta_2} L_2 + (\rho_1 - \rho_1 \rho_2) \theta_1 \sqrt{\theta_2 \theta_3} L_5 \right\} \quad \text{(S9)}
\]
\[ f_3 = \frac{n}{\theta_3} + \frac{1}{\theta_1 \theta_2 \theta_3^2 J} \left\{ -(1 - \rho_1^2) \theta_1 \theta_2 L_6 + (\rho_2 - \rho_1 \rho_3) \theta_2 \sqrt{\theta_1 \theta_3 L_3} + (\rho_3 - \rho_1 \rho_2) \theta_1 \sqrt{\theta_2 \theta_3 L_5} \right\} \]  
(S10)

\[ f_4 = -\frac{2(\rho_1 - \rho_2 \rho_3)}{J} n + \frac{1}{\theta_1 \theta_2 \theta_3 J^2} \times 
\left[ 2(\rho_1 - \rho_2 \rho_3)(1 - \rho_1^2) \theta_2 \theta_3 L_1 - 2 \left\{ J + 2 (\rho_1 - \rho_2 \rho_3)^2 \right\} \theta_3 \sqrt{\theta_1 \theta_2 L_2} 
+ 2 \left\{ \rho_3(1 + \rho_1^2 + \rho_2^2 - \rho_3^2) - 2 \rho_1 \rho_2 \right\} \theta_2 \sqrt{\theta_1 \theta_3 L_3} 
+ 2(\rho_1 - \rho_2 \rho_3)(1 - \rho_2^2) \theta_1 \theta_3 L_4 
+ 2 \left\{ \rho_2(1 + \rho_1^2 - \rho_2^2 + \rho_3^2) - 2 \rho_1 \rho_3 \right\} \theta_1 \sqrt{\theta_2 \theta_3 L_5} 
+ 2 \left\{ \rho_1(\rho_2^2 + \rho_3^2) - \rho_2 \rho_3 (\rho_1^2 + 1) \right\} \theta_1 \theta_2 L_6 \right] \]  
(S11)

\[ f_5 = -\frac{2(\rho_2 - \rho_1 \rho_3)}{J} n + \frac{1}{\theta_1 \theta_2 \theta_3 J^2} \times 
\left[ 2(\rho_2 - \rho_1 \rho_3)(1 - \rho_2^2) \theta_3 \theta_4 L_1 
+ 2 \left\{ \rho_3(1 + \rho_1^2 + \rho_2^2 - \rho_3^2) - 2 \rho_1 \rho_2 \right\} \theta_3 \sqrt{\theta_1 \theta_2 L_2} 
- 2 \left\{ J + 2(\rho_2 - \rho_1 \rho_3)^2 \right\} \theta_2 \sqrt{\theta_1 \theta_3 L_3} 
+ 2 \left\{ \rho_2(\rho_1^2 + \rho_3^2 - \rho_1 \rho_3 (\rho_2^2 + 1)) \right\} \theta_1 \theta_3 L_4 
+ 2 \left\{ \rho_1(1 - \rho_1^2 + \rho_2^2 + \rho_3^2) - 2 \rho_2 \rho_3 \right\} \theta_1 \sqrt{\theta_2 \theta_3 L_5} 
+ 2(\rho_2 - \rho_1 \rho_3)(1 - \rho_1^2) \theta_1 \theta_2 L_6 \right] \]  
(S12)

\[ f_6 = -\frac{2(\rho_3 - \rho_1 \rho_2)}{J} n + \frac{1}{\theta_1 \theta_2 \theta_3 J^2} \times 
\left[ 2 \left\{ \rho_3(\rho_1^2 + \rho_2^2 - \rho_1 \rho_2 (\rho_3^2 + 1)) \right\} \theta_2 \theta_3 L_1 
+ 2 \left\{ \rho_2(1 + \rho_1^2 - \rho_2^2 + \rho_3^2) - 2 \rho_1 \rho_3 \right\} \theta_3 \sqrt{\theta_1 \theta_2 L_2} 
+ 2 \left\{ \rho_1(1 - \rho_1^2 + \rho_2^2 + \rho_3^2) - 2 \rho_2 \rho_3 \right\} \theta_2 \sqrt{\theta_1 \theta_3 L_3} 
+ 2(\rho_3 - \rho_1 \rho_2)(1 - \rho_2^2) \theta_1 \theta_3 L_4 
- 2 \left\{ J + 2(\rho_3 - \rho_1 \rho_2)^2 \right\} \theta_1 \sqrt{\theta_2 \theta_3 L_5} 
+ 2(\rho_3 - \rho_1 \rho_2)(1 - \rho_1^2) \theta_1 \theta_2 L_6 \right] \]  
(S13)

After simplifications, we can obtain the analytical solution of the REML estimations for the variance component parameters

\[
\hat{\theta}_1 = L_1/n \quad \text{(S14)}
\]

\[
\hat{\theta}_2 = L_4/n \quad \text{(S15)}
\]

\[
\hat{\theta}_3 = L_6/n \quad \text{(S16)}
\]

\[
\hat{\rho}_1 = L_2/\sqrt{L_1 L_4} \quad \text{(S17)}
\]

\[
\hat{\rho}_2 = L_3/\sqrt{L_1 L_6} \quad \text{(S18)}
\]

\[
\hat{\rho}_3 = L_5/\sqrt{L_4 L_6} \quad \text{(S19)}
\]

To compute the asymptotic variances for the variance component parameter estimates, let
Then, following the arguments in McGilchrist and Yau (1995), the covariance matrix of the estimators in the variance component is given by

\[
\var \begin{pmatrix}
\hat{\theta}_1 \\
\hat{\theta}_2 \\
\hat{\theta}_3 \\
\hat{\rho}_1 \\
\hat{\rho}_2 \\
\hat{\rho}_3 \\
\end{pmatrix} = 2 \begin{pmatrix}
a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\
* & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\
* & * & a_{33} & a_{34} & a_{35} & a_{36} \\
* & * & * & a_{44} & a_{45} & a_{46} \\
* & * & * & * & a_{55} & a_{56} \\
* & * & * & * & * & a_{66} \\
\end{pmatrix}^{-1}
\]

\[
a_{11} = \text{tr} (K_1 - K_2)^2 \\
a_{12} = \text{tr} (K_1 K_3 + K_2 K_4 - 2K_1 K_4) \\
a_{13} = \text{tr} (K_1 K_5 + K_2 K_6 - 2K_1 K_6) \\
a_{14} = \text{tr} (K_1 K_7 + K_2 K_8 - 2K_1 K_8) \\
a_{15} = \text{tr} (K_1 K_9 + K_2 K_{10} - 2K_1 K_{10}) \\
a_{16} = \text{tr} (K_1 K_{11} + K_2 K_{12} - 2K_1 K_{12}) \\
a_{22} = \text{tr} (K_3 - K_4)^2 \\
a_{23} = \text{tr} (K_3 K_5 + K_4 K_6 - 2K_3 K_6) \\
a_{24} = \text{tr} (K_3 K_7 + K_4 K_8 - 2K_3 K_8) \\
a_{25} = \text{tr} (K_3 K_9 + K_4 K_{10} - 2K_3 K_{10}) \\
a_{26} = \text{tr} (K_3 K_{11} + K_4 K_{12} - 2K_3 K_{12}) \\
a_{33} = \text{tr} (K_5 - K_6)^2 \\
a_{34} = \text{tr} (K_5 K_7 + K_6 K_8 - 2K_5 K_8) \\
a_{35} = \text{tr} (K_5 K_9 + K_6 K_{10} - 2K_5 K_{10}) \\
a_{36} = \text{tr} (K_5 K_{11} + K_6 K_{12} - 2K_5 K_{12}) \\
a_{44} = \text{tr} (K_7 - K_8)^2 \\
a_{45} = \text{tr} (K_7 K_9 + K_8 K_{10} - 2K_7 K_{10}) \\
a_{46} = \text{tr} (K_7 K_{11} + K_8 K_{12} - 2K_7 K_{12}) \\
a_{55} = \text{tr} (K_1 - K_{10})^2 \\
a_{56} = \text{tr} (K_9 K_{11} + K_{10} K_{12} - 2K_9 K_{12}) \\
a_{66} = \text{tr} (K_{11} - K_{12})^2
\]

### 1.3 Measurement errors

To reducing the biasing effect caused by measurement errors and obtain a more robust TMB threshold, we integrated the widely applicable corrected-score with the joint model, resulting in approximately consistent estimators based on the observed data. The covariate \(X_i = \{Z_i, TMB_i\}^T\), where \(Z_i\) denotes a vector of covariates, e.g., age, gender, treatment indicator, cancer stage. \(TMB_i = TMB_i^* + \epsilon_i\). \(TMB_i^*\) is the true TMB value which cannot be observed. Score function is the first order derivative of the likelihood \(\Psi = \left[\frac{\partial l}{\partial \eta_1}, \frac{\partial l}{\partial \eta_2}, \frac{\partial l}{\partial \eta_3}\right]\). A corrected score is a function \(\Psi^*_c\) of the observed data having the important property that

\[
E \{\Psi^*_c (R_i, T_i, \delta_i, Z_i, TMB_i^*; \Theta) | R_i, T_i, \delta_i, Z_i, TMB_i \} = \Psi (R_i, T_i, \delta_i, Z_i, TMB_i; \Theta)
\]

(S20)
which is conditionally unbiased for the true-data score function according to the property of conditional expectation, 
\( E \{ \psi_c^* (R_i, T_i, \delta_i, Z_i, TB_i; \Theta) \} = 0 \). The corrected scores provide an approach to reducing bias incurred by covariate measurement error.

Based on the correct-score, \( \psi^*_c = \left[ \left( \frac{\partial L}{\partial \beta_{R1i}} \right)_c, \left( \frac{\partial L}{\partial \beta_{R2i}} \right)_c, \left( \frac{\partial L}{\partial \beta_T} \right)_c \right] \), specifically:

\[
\begin{align*}
\eta_{1i} &= Z_{i}^T \cdot \beta_{R1i} + TB_{j,i} \cdot \beta_{R1m} + U_{i1} \\
\eta_{2i} &= Z_{i}^T \cdot \beta_{R2i} + TB_{j,i} \cdot \beta_{R2m} + U_{i2} \\
\left( \frac{\partial L}{\partial \beta_{R1i}} \right)_c &= \left( \frac{\partial L}{\partial \eta_i} \right)_c \cdot X_i = \left[ J^{-1} \sum_{j=1}^{J} \frac{\partial L}{\partial (\eta_i)}^T \right] \cdot X_i \\
\left( \frac{\partial L}{\partial \beta_{R2i}} \right)_c &= \left( \frac{\partial L}{\partial \eta_i} \right)_c \cdot X_i = \left[ J^{-1} \sum_{j=1}^{J} \frac{\partial L}{\partial (\eta_i)}^T \right] \cdot X_i \\
\left( \frac{\partial L}{\partial \beta_T} \right)_c &= \sum_{i} \delta_i \left[ X_i - \sum_{(j, T_j \geq T_i)} X_j \exp \left( X_j^T \beta_T + V_j \right) \right] / \sum_{(j, T_j \geq T_i)} \exp \left( X_j^T \beta_T + V_j \right) + \left( \begin{array}{c} 0 \\ \sigma^2 \beta_{Tm} \end{array} \right)
\end{align*}
\]

where the complex variate \( \tilde{TB}_{j,i} = TB_{j,i} + \sqrt{-1} \xi_{j,i} \), and \( \xi_{j,i} \) is a normal random vector with zero mean and variance \( \sigma_x \). Since the logistic function does not satisfy the smoothness condition required by the corrected-score theory, it has no intuitively exact corrected score but obtains an approximation by means of complex variable extrapolation. Additionally,

\[
\left( \frac{\partial L}{\partial \beta_T} \right)_c = \sum_{i} \delta_i \left[ X_i - \sum_{(j, T_j \geq T_i)} X_j \exp \left( X_j^T \beta_T + V_j \right) \right] / \sum_{(j, T_j \geq T_i)} \exp \left( X_j^T \beta_T + V_j \right) + \left( \begin{array}{c} 0 \\ \sigma^2 \beta_{Tm} \end{array} \right)
\]

Solving the equation \( \psi_c^* (R_i, T_i, \delta_i, Z_i, TB_i; \Theta) = 0 \), the approximately consistent estimators \( \tilde{\theta}, \tilde{u}_i \) and \( \tilde{v}_i \) can be derived.

## 2 SIMULATIONS

### 2.1 Simulation study without considering TMB measurement errors

In order to assess the proposed estimation procedure’s performance, we conducted a series of simulation studies whose primary objective was to assess the fixed effect estimates and the variance of the random effects. Data are simulated in an oncology trial context, with random effects correlated among clusters’ multiple endpoints. In the simulations, we assume 200 patients, i.e., \( i = 1, ..., 200 \). For each cluster \( i \), we generate the random effects \( U_{i1}, U_{i2}, \) and \( V_i \) from a multivariate normal distribution with zero mean, \( \theta_1 = \theta_2 = \theta_3 = 1.0 \), and general correlations \( \rho_1, \rho_2, \rho_3 \). We consider three distinct tumor response states CR&PR (\( k = 0 \) is being setting as the reference), SD (\( k = 1 \)) and PD (\( k = 2 \)). The response data is generated based on the logistic probabilities for multiple classifications, \( \pi_0 = 1 - (\pi_1 + \pi_2) \), \( \pi_1 = \exp \eta_1 / (1 + \exp \eta_1 + \exp \eta_2) \), \( \pi_2 = \exp \eta_2 / (1 + \exp \eta_1 + \exp \eta_2) \). The event time for the patient is generated from the probability density function \( \lambda_0(t) = \exp \eta_{3i} S_0(t)^{\exp \eta_{3j}} \), where the baseline hazard \( \lambda_0(t) \) is assumed to follow the exponential distribution with a mean equal to 500, and \( S_0(t) = \exp(- \int_0^t \lambda_0(s) ds) \) is the baseline survival function. Censoring time \( C \) is generated from the uniform distribution \( U(50, 500) \).

Furthermore, we set \( \beta_{R1} = 1.6, \gamma_1 = -0.2, \beta_{R2} = 1.0, \gamma_1 = -0.8, \beta_T = -1.2, \theta_1 = \theta_2 = \theta_3 = 1.0 \). The correlation \( \rho_m (m = 1, 2, 3) \) is chosen from -0.9, -0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8 and 0.9, in order to evaluate the performance of the estimators with different correlation levels. In the negative correlation scenario simulations, only two of the three values for \( \rho_m \) can be set to negative, given the positive definite limitation of the covariance matrix of the multivariate Gaussian distribution. With the specified parameters, 1000 replications are simulated in each setting. Results of the simulations are presented in Table [S1](#). We report the fitted value, average bias, SE and SD for each parameter, where SE and SD are defined as the average of the standard error of the estimates and the standard error of the estimates over 1000 simulations, respectively.

The results in Table [S1](#) indicate that the regression parameter estimates in three function sections perform fairly well under various multivariate correlation settings, given that the hazard estimation is performed using a semi-parametric method. The
biases in estimating the variance component parameters $\theta_1, \theta_2, \theta_3, \rho_1, \rho_2, \rho_3$ are likewise relatively minimal. In general, the results support the suggested multi-endpoint joint model with correlated random effects being applicable. When comparing SE and SD, the accuracy of the reported standard errors is generally acceptable, though slightly underestimated when $\rho_m$ approaches the limiting values of $\pm 0.9$. In particular, the sample standard errors SD for 1000 correlation estimates are minimal due to the limited one-way volatility on correlation estimates at boundaries (i.e., the correlation estimates can only be larger than -1 or smaller than 1). However, the estimation method for standard error still assumes two-way variation. As a result, when $\rho_m$ is close to the limits, bootstrap estimates for the standard error of the correlation parameter could be employed.

Table S1: Bias and standard error of estimators for joint model with multivariate random effects.

| Parameter | True value | Fitted value | Average Bias | SE   | SD   |
|-----------|------------|--------------|--------------|------|------|
| Simulation 1                      |            |              |              |      |      |
| $\beta_{R1}$ | 1.6       | 1.574        | 0.026        | 0.642| 0.646|
| $\gamma_1$    | -0.2      | -0.207       | 0.007        | 0.421| 0.423|
| $\beta_{R2}$ | 1.0       | 0.946        | 0.054        | 0.775| 0.809|
| $\gamma_2$    | -0.8      | -0.806       | 0.006        | 0.486| 0.495|
| $\beta_T$     | -1.2      | -1.190       | 0.010        | 0.484| 0.481|
| $\theta_1$    | 1.0       | 1.059        | 0.059        | 0.553| 0.546|
| $\theta_2$    | 1.0       | 1.114        | 0.114        | 0.673| 0.638|
| $\theta_3$    | 1.0       | 0.986        | 0.014        | 0.450| 0.448|
| $\rho_1$      | 0.9       | 0.859        | 0.041        | 0.208| 0.152|
| $\rho_2$      | 0.9       | 0.851        | 0.049        | 0.251| 0.153|
| $\rho_3$      | 0.9       | 0.834        | 0.066        | 0.308| 0.172|
| Simulation 2                      |            |              |              |      |      |
| $\beta_{R1}$ | 1.6       | 1.543        | 0.057        | 0.642| 0.621|
| $\gamma_1$    | -0.2      | -0.184       | 0.016        | 0.416| 0.394|
| $\beta_{R2}$ | 1         | 0.971        | 0.029        | 0.772| 0.795|
| $\gamma_2$    | -0.8      | -0.801       | 0.001        | 0.482| 0.492|
| $\beta_T$     | -1.2      | -1.212       | 0.012        | 0.487| 0.478|
| $\theta_1$    | 1         | 0.987        | 0.013        | 0.530| 0.556|
| $\theta_2$    | 1         | 1.042        | 0.042        | 0.648| 0.581|
| $\theta_3$    | 1         | 0.976        | 0.024        | 0.447| 0.449|
| $\rho_1$      | 0.8       | 0.767        | 0.033        | 0.264| 0.213|
| $\rho_2$      | 0.8       | 0.763        | 0.037        | 0.287| 0.197|
| $\rho_3$      | 0.8       | 0.725        | 0.075        | 0.342| 0.233|
| Simulation 3                      |            |              |              |      |      |
| $\beta_{R1}$ | 1.6       | 1.560        | 0.040        | 0.650| 0.671|
| $\gamma_1$    | -0.2      | -0.187       | 0.013        | 0.421| 0.408|
| $\beta_{R2}$ | 1         | 0.998        | 0.002        | 0.769| 0.767|
| $\gamma_2$    | -0.8      | -0.776       | 0.024        | 0.478| 0.478|
| $\beta_T$     | -1.2      | -1.211       | 0.011        | 0.484| 0.505|
| $\theta_1$    | 1         | 1.016        | 0.016        | 0.544| 0.585|
| $\theta_2$    | 1         | 1.005        | 0.005        | 0.634| 0.599|
| $\theta_3$    | 1         | 0.926        | 0.074        | 0.429| 0.450|
| $\rho_1$      | 0.6       | 0.593        | 0.007        | 0.349| 0.305|
| $\rho_2$      | 0.6       | 0.596        | 0.004        | 0.318| 0.293|
| Simulation | $\beta_{R1}$ | $\gamma_1$ | $\beta_{R2}$ | $\gamma_2$ | $\beta_T$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\rho_1$ | $\rho_2$ | $\rho_3$ |
|------------|--------------|------------|--------------|------------|----------|----------|----------|----------|--------|-------|-------|
| 4          | 1.6          | -0.2       | 1            | -0.8       | -1.2     | 1        | 1        | 1        | 0.4    | 0.4   | 0.4   |
|            | 1.557        | -0.197     | 0.973        | -0.760     | -1.197   | 0.914    | 0.957    | 0.922    | 0.386  | 0.402 | 0.398 |
|            | 0.043        | 0.003      | 0.027        | 0.040      | 0.003    | 0.086    | 0.043    | 0.078    | 0.014  | 0.002 | 0.002 |
|            | 0.650        | 0.416      | 0.764        | 0.474      | 0.486    | 0.511    | 0.617    | 0.428    | 0.435  | 0.357 | 0.420 |
|            | 0.642        | 0.425      | 0.784        | 0.499      | 0.504    | 0.501    | 0.617    | 0.433    | 0.375  | 0.341 | 0.365 |
| 5          | 1.6          | -0.2       | 1            | -0.8       | -1.2     | 1        | 1        | 1        | 0.2    | 0.2   | 0.2   |
|            | 1.549        | -0.185     | 0.972        | -0.751     | -1.185   | 0.910    | 0.907    | 0.912    | 0.181  | 0.208 | 0.187 |
|            | 0.051        | 0.015      | 0.028        | 0.049      | 0.015    | 0.090    | 0.093    | 0.088    | 0.019  | 0.008 | 0.013 |
|            | 0.655        | 0.416      | 0.764        | 0.471      | 0.486    | 0.512    | 0.601    | 0.425    | 0.498  | 0.381 | 0.440 |
|            | 0.667        | 0.413      | 0.776        | 0.472      | 0.491    | 0.509    | 0.564    | 0.439    | 0.433  | 0.384 | 0.411 |
| 6          | 1.6          | -0.2       | 1            | -0.8       | -1.2     | 1        | 1        | 1        | 0     | 0     | 0     |
|            | 1.550        | -0.178     | 1.021        | -0.737     | -1.236   | 0.923    | 0.907    | 0.904    | 0.181  | 0.003 | 0.000 |
|            | 0.050        | 0.022      | 0.021        | 0.063      | 0.036    | 0.077    | 0.093    | 0.096    | 0.019  | 0.003 | 0.000 |
|            | 0.663        | 0.419      | 0.763        | 0.471      | 0.489    | 0.522    | 0.609    | 0.422    | 0.498  | 0.385 | 0.440 |
|            | 0.664        | 0.410      | 0.761        | 0.465      | 0.501    | 0.524    | 0.578    | 0.438    | 0.433  | 0.386 | 0.419 |
| 7          | 1.6          | -0.2       | 1            | -0.8       | -1.2     | 1        | 1        | 1        | 0     | -0.2  | 0     |
|            | 1.601        | -0.183     | 1.063        | -0.746     | -1.201   | 0.929    | 0.921    | 0.917    | 0.27   | 0.004 | 0.000 |
|            | 0.001        | 0.017      | 0.063        | 0.054      | 0.001    | 0.071    | 0.079    | 0.083    | 0.047  | 0.004 | 0.002 |
|            | 0.669        | 0.421      | 0.766        | 0.472      | 0.489    | 0.528    | 0.611    | 0.428    | 0.574  | 0.381 | 0.377 |
| Simulation 8 |  |  |  |  |  |
|-------------|---|---|---|---|---|
| $\beta_{R1}$   | 1.6 | 1.612 | 0.012 | 0.674 | 0.692 |
| $\gamma_1$     | -0.2 | -0.193 | 0.007 | 0.425 | 0.431 |
| $\beta_{R2}$   | 1   | 1.060 | 0.060 | 0.764 | 0.745 |
| $\gamma_2$     | -0.8 | -0.731 | 0.069 | 0.470 | 0.465 |
| $\beta_T$      | -1.2 | -1.208 | 0.008 | 0.486 | 0.519 |
| $\theta_1$     | 1   | 0.950 | 0.050 | 0.539 | 0.522 |
| $\theta_2$     | 1   | 0.916 | 0.084 | 0.609 | 0.586 |
| $\theta_3$     | 1   | 0.903 | 0.097 | 0.421 | 0.420 |
| $\rho_1$       | -0.4 | -0.298 | 0.102 | 0.425 | 0.372 |
| $\rho_2$       | -0.4 | -0.383 | 0.017 | 0.366 | 0.332 |
| $\rho_3$       | 0.4  | 0.370 | 0.030 | 0.425 | 0.372 |

| Simulation 9 |  |  |  |  |  |
|-------------|---|---|---|---|---|
| $\beta_{R1}$   | 1.6 | 1.584 | 0.016 | 0.678 | 0.681 |
| $\gamma_1$     | -0.2 | -0.168 | 0.032 | 0.427 | 0.421 |
| $\beta_{R2}$   | 1   | 1.001 | 0.001 | 0.766 | 0.792 |
| $\gamma_2$     | -0.8 | -0.691 | 0.109 | 0.473 | 0.492 |
| $\beta_T$      | -1.2 | -1.227 | 0.027 | 0.484 | 0.504 |
| $\theta_1$     | 1   | 0.954 | 0.046 | 0.545 | 0.548 |
| $\theta_2$     | 1   | 0.948 | 0.052 | 0.623 | 0.584 |
| $\theta_3$     | 1   | 0.939 | 0.061 | 0.433 | 0.441 |
| $\rho_1$       | -0.6 | -0.468 | 0.132 | 0.591 | 0.366 |
| $\rho_2$       | -0.6 | -0.580 | 0.020 | 0.338 | 0.289 |
| $\rho_3$       | 0.6  | 0.553 | 0.047 | 0.387 | 0.318 |

| Simulation 10 |  |  |  |  |  |
|---------------|---|---|---|---|---|
| $\beta_{R1}$   | 1.6 | 1.601 | 0.001 | 0.685 | 0.673 |
| $\gamma_1$     | -0.2 | -0.164 | 0.036 | 0.429 | 0.427 |
| $\beta_{R2}$   | 1   | 1.055 | 0.055 | 0.771 | 0.788 |
| $\gamma_2$     | -0.8 | -0.711 | 0.089 | 0.476 | 0.500 |
| $\beta_T$      | -1.2 | -1.222 | 0.022 | 0.483 | 0.502 |
| $\theta_1$     | 1   | 0.955 | 0.045 | 0.549 | 0.559 |
| $\theta_2$     | 1   | 0.980 | 0.020 | 0.637 | 0.626 |
| $\theta_3$     | 1   | 0.950 | 0.050 | 0.436 | 0.452 |
| $\rho_1$       | -0.8 | -0.627 | 0.173 | 0.589 | 0.302 |
| $\rho_2$       | -0.8 | -0.758 | 0.042 | 0.309 | 0.215 |
| $\rho_3$       | 0.8  | 0.727 | 0.073 | 0.343 | 0.246 |

| Simulation 11 |  |  |  |  |  |
|---------------|---|---|---|---|---|
| $\beta_{R1}$   | 1.6 | 1.634 | 0.034 | 0.689 | 0.693 |
| $\gamma_1$     | -0.2 | -0.169 | 0.031 | 0.432 | 0.421 |
| $\beta_{R2}$   | 1   | 1.061 | 0.061 | 0.775 | 0.781 |
| $\gamma_2$     | -0.8 | -0.728 | 0.072 | 0.481 | 0.492 |
| $\beta_T$      | -1.2 | -1.194 | 0.006 | 0.482 | 0.496 |
| $\theta_1$     | 1   | 0.985 | 0.015 | 0.562 | 0.533 |
| $\theta_2$     | 1   | 1.039 | 0.039 | 0.662 | 0.641 |
| $\theta_3$     | 1   | 0.980 | 0.020 | 0.447 | 0.458 |
| $\rho_1$       | -0.9 | -0.730 | 0.170 | 0.567 | 0.266 |
| $\rho_2$       | -0.9 | -0.830 | 0.070 | 0.288 | 0.177 |
2.2 Simulation study considering TMB measurement errors

In order to assess the performance of the proposed joint model with corrected-score function, we also conducted a series of simulation studies under TMB measurement errors. The simulation setting is similar with the above section 2.1. $\beta_{Rz} = -1.8$, $\beta_{Rm} = 0.3$, $\beta_{Tz} = 2.2$, $\beta_{Tm} = -0.4$. The variance for the error term is set to be 0.5, 0.75, 1.0, respectively, in order to evaluate the performance of the estimators with different measurement error levels. With the specified parameters, for each dataset, based on joint model, the true-data estimator, the naive estimator and the correct-score estimator with Monte Carlo approximation $J=10$, were computed 1000 replications. As a comparison, we also based the standard regression models, the true-data estimator and the naive estimator were computed. Results of the simulations are presented in Table S2. We report the fitted value, average bias, SD and SE for each parameter, where SD and SE are defined as the standard error of the estimates over 1000 simulations and the average of the standard error of the estimates, respectively.

According to Table S2, the regression parameter estimates for the two function components perform reasonably well for a variety of measurement error conditions. In the absence of measurement errors, the joint model outperforms ordinary regression models in calculating regression coefficients because it more precisely reflects the potential connections between several endpoints. When considering different levels of measurement errors, the performance of the estimator based on corrected-score was significantly superior to that of the naive estimator and only marginally poorer than that of the true-data estimator. Clearly, the performance of the naive estimator deteriorates with increasing error magnitude, which further suggests that measurement error introduces a more significant bias effect on the parameter estimates. Overall, the results of the simulation experiments support the proposed joint multi-endpoint model and the iterative numerical estimation procedure, as well as the applicability of the associated random effects. Additionally, comparing SE and SD, the precision of the stated standard errors is generally satisfactory. The biases of the joint assessments compared to the standard separate regressions emphasizes that the dependence among clinical endpoints could be an important and non-negligible con-founder in analyzing the factors determining the treatment effect.

### Table S2: Bias and standard error of estimators for joint model considering varying measurement errors.

| Parameter | True value | Fitted value | Average Bias | SE   | SD   |
|-----------|------------|--------------|--------------|------|------|
| Correct-score $\sigma_e = 0.5$ | | | | | |
| $\beta_{Rz}$ | -1.8 | -1.854 | 0.054 | 0.503 | 0.495 |
| $\beta_{Rm}$ | 0.3 | 0.278 | 0.022 | 0.092 | 0.091 |
| $\beta_{Tz}$ | 2.2 | 2.117 | 0.083 | 0.296 | 0.373 |
| $\beta_{Tm}$ | -0.4 | -0.425 | 0.025 | 0.06 | 0.081 |
| Correct-score $\sigma_e = 0.75$ | | | | | |
| $\beta_{Rz}$ | -1.8 | -1.798 | 0.002 | 0.502 | 0.477 |
| $\beta_{Rm}$ | 0.3 | 0.271 | 0.029 | 0.091 | 0.089 |
| $\beta_{Tz}$ | 2.2 | 2.108 | 0.092 | 0.293 | 0.37 |
| $\beta_{Tm}$ | -0.4 | -0.418 | 0.018 | 0.059 | 0.076 |
| Correct-score $\sigma_e = 1.0$ | | | | | |
| $\beta_{Rz}$ | -1.8 | -1.79 | 0.01 | 0.498 | 0.444 |
| $\beta_{Rm}$ | 0.3 | 0.274 | 0.026 | 0.09 | 0.081 |
| $\beta_{Tz}$ | 2.2 | 2.027 | 0.173 | 0.288 | 0.362 |
| $\beta_{Tm}$ | -0.4 | -0.39 | 0.01 | 0.058 | 0.074 |

Naive estimator $\sigma_e = 0.5$
3 COHORT INFORMATION

3.1 Patient characteristics

Radiographic tumor assessments were taken at baseline and approximately every 6 weeks. ORR was the proportion of patients with CR or PR. Treatment continued until confirmed PD, intolerable toxicities, death or withdrawal of consent. Censored data documented the last radiographic assessment before cut-off, loss of follow-up, or change of treatment. Progression-free survival (PFS) was defined as the time from the first dose to PD or prior death. Overall survival (OS) was duration from the first dose to death, and patients who remained alive were censored at the date of their last follow-up. This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethical Review Committee of Sun Yat-sen University Cancer Center. Written informed consents were provided by all patients. Tumor samples from all NPC patients were analyzed using both WES and Panel. FFPE samples were consistently used in the analysis due to the paucity of biopsy samples and given the limited intra-tumor heterogeneity they represent. The clinical characteristics of this cohort are summarized in Table 1 with details in Supplementary Table S1. For lung cancer, 60% of the patients had adenocarcinoma, followed by squamous carcinoma (31%). Almost all patients (99%) were stage IV at diagnosis; the median ages of patients with NSCLC and NPC at the treatment initiation were 54 and 46 years, respectively. 49% of the NSCLC patients and 25% of the NPC patients had a smoking history, and there were more males in both cohorts (68% vs. 32% for NSCLC, 80% vs. 20% for NPC). ORR of the study cohorts was 18% and 12%. The median progression-free survival (mPFS) for lung cancer and nasopharyngeal carcinoma was 60 and 67.5 days, respectively.

3.2 Library preparation and sequencing

For whole-exome sequencing (WES), genomic DNAs were extracted from Formalin-fixed, paraffin-embedded (FFPE) or biopsy tumor and blood samples by QIAamp DNA FFPE Tissue Kit and DNeasy Blood & Tissue Kit (Qiagen), respectively, and quantified by Qubit 3.0 using the dsDNA HS Assay Kit (ThermoFisher Scientific). Library preparations were performed with KAPA Hyper Prep Kit (KAPA Biosystems). Target enrichment was performed using the xGen Exome Research Panel and Hybridization and Wash Reagents Kit (Integrated DNA Technology). Sequencing was performed on Illumina HiSeq4000 platform using PE150 sequencing chemistry (Illumina). For targeted-panel, customized xGen lockdown probes (Integrated DNA Technologies) targeting 422 cancer-relevant genes were used for hybridization enrichment. The capture reaction was performed with Dynabeads M-270 (Life Technologies) and xGen Lockdown Hybridization and Wash Kit (Integrated DNA Technologies). Captured libraries were on-beads PCR amplified with Illumina p5 and p7 primers in KAPA HiFi HotStart ReadyMix (KAPA Biosystems), followed by purification using Agencourt AMPure XP beads. Libraries were quantified by qPCR using KAPA Library Quantification Kit (KAPA Biosystems). Library fragment size was determined by Bioanalyzer 2100 (Agilent Technologies). The target-enriched library was then sequenced on HiSeq4000 or HiSeq4000 NGS platforms.
(Illumina). The average coverage depth was 140X and 1341X for tumors (64X and 143X for normal controls) using WES and Panel, respectively.

### 3.3 Bioinformatics procedure for sequencing data analysis

FASTQ files were processed using Trimmomatic for quality control. Leading/trailing low quality (below 20) or N bases were removed. The high quality paired-end reads were aligned to the human hg19 reference genome using the Burrows-Wheeler Aligner. The resulting alignment files were cleaned by Picard and standardized GATK3, including deduplication, BQSR and Indel realignment. Cross-sample contamination was estimated using ContEst (Broad Institute, contamination rate <0.02). Somatic single nucleotide variant (SNV) and insertion/deletions (indels) calling was performed by Mutect and Scalpel, respectively. Variants called were further filtered using the following criteria: (i) 4 reads and 2% variant allele frequency (VAF) supporting the variant, (ii) filtered if present in >1% population frequency in the 1,000 g or ExAC database, (iii) filtered through an internally collected list of recurrent sequencing errors (3 variant reads and 20% VAF in at least 30 out of 2,000 normal samples) on the same sequencing platform. Final list of mutations was annotated using vcf2maf. Tumor mutation burden (TMB) for WES was defined as the total number of missense mutations. In addition, we also used a targeted next-generation sequencing panel (Geneseq) to profile TMB of these samples and compare it to WES results. Panel TMB was calculated by adding up all base substitutions and indels in the coding region of targeted genes. Synonymous variants were counted to reduce sampling noise, and known driver mutations were excluded due to over-representation, as described in previous study.