Assessing joint action of combination therapy in Synthetic and experimental PDX data using Global-Two-Stage regression and SAEM

Junho Yoon[1] and Min Jin Ha[1*]

[1]Department of Biostatistics and Health Informatics, Graduate School of Public Health, Yonsei University

8 Abstract

9

10 **1. Introduction**

11

12

2. Method

2.1. Growth / Drug effect model

We assume that log tumor growth rate is relevant to natural growth of tumor and its decay by drug treatment. Mathematically, it can be represented as ordinary differential equation:

$$\frac{dL(t)}{dt} = (Tumor\ growth) - (Drug\ effect)$$

where $L(t)$ is the logarithm of tumor volume and t is time. Let $V = exp(L)$ to be the tumor volume, then the changing in tumor size over time dV/dt can be calculated as:

$$\frac{dV(t)}{dt} = \frac{dV(t)}{dL(t)}\frac{dL(t)}{dt} = (Tumor\ growth)V(t) - (Drug\ effect)V(t)$$

This equation aligns with the series of mixed-effect models presented by Ribba et al. (2014) to quantify the effects of anticancer drug treatment. In this model, we assume that the tumor growth follows the Gompertz growth model, and the delay in tumor growth is proportional to the drug concentration at a constant rate. Additionally, we assume that the drugs act independently. Specifically, the tumor growth without drug intervention is modeled using the Gompertz equation, while the drug-induced effects are incorporated through exponential decay terms that represent the reduction in tumor growth rate due to the drugs.

$$\frac{dL(t)}{dt} = \begin{cases} \alpha_1 - \alpha_2(L(t) - L_0) & g = C\ (control) \\ \alpha_1 - \alpha_2(L(t) - L_0) - \beta_g \exp(-k_g t) & g = A\ or\ B \\ \alpha_1 - \alpha_2(L(t) - L_0) - \beta_A \exp(-k_A t) - \beta_B \exp(-k_B t) & g = AB\ (mixture) \end{cases}$$

where α_1 is the growth rate, α_2 is the deceleration rate, and L_0 is the logarithmic volume of the tumor at the initial time. β_A and β_B are the drug effect parameters for drugs A and B respectively,

while k_A and k_B represent the rate constants for the drug effect decay over time.

Integrating the differential equations with respect to t,

$$L(t) = \begin{cases} L_0 + \frac{\alpha_1}{\alpha_2}(1 - \exp(-\alpha_2 t)) + C & g = C \\ L_0 + \frac{\alpha_1}{\alpha_2}(1 - \exp(-\alpha_2 t)) - \frac{\beta_g}{\alpha_2 - k_g}\exp(-k_g t) + C & g = A \text{ or } B, k_g \neq \alpha_2 \\ L_0 + \frac{\alpha_1}{\alpha_2}(1 - \exp(-\alpha_2 t)) - \beta_g t \exp(-\alpha_2 t) + C & g = A \text{ or } B, k_g = \alpha_2 \\ L_0 + \frac{\alpha_1}{\alpha_2}(1 - \exp(-\alpha_2 t)) - \frac{\beta_A}{\alpha_2 - k_A}\exp(-k_A t) - \frac{\beta_B}{\alpha_2 - k_B}\exp(-k_B t) + C & g = AB, k_g \neq \alpha_2 \\ L_0 + \frac{\alpha_1}{\alpha_2}(1 - \exp(-\alpha_2 t)) - \beta_A t \exp(-\alpha_2 t) - \beta_B t \exp(-\alpha_2 t) + C & g = AB, k_g = \alpha_2 \end{cases}$$

2.2. Simulation

To assess the behavior of the growth and drug effect model, synthetic tumor growth data were generated based on the model described in section 2.1. The simulation procedure began by using the model parameters, including the tumor growth rate (α_1), deceleration rate (α_2), drug effect parameters for drugs A (β_A) and B (β_B), the decay rates for the drug effects (k_A and k_B), and the initial tumor volume (L_0). The tumor volume was simulated over a predefined time period, with data points collected at regular intervals as denoted in each simulation, starting from $t = 0$.

For the drug-treated groups, drugs A and B were administered at specific time points, with the drug concentration decay modeled using exponential functions characterized by the parameters k_A and k_B . The tumor growth for each simulated individual was computed using the Gompertz growth model, incorporating the effects of the drugs as described in the equation from section

2.1. Normal measurement errors were added to the simulated tumor volumes, assumed to be normally distributed $\epsilon_{ij} \sim N(0, \sigma 2)$, which generated synthetic observed data that resemble real-world experimental measurements. The number of simulated individuals, N, was set to 5, 10, and 50, representing different sample sizes to evaluate the model's performance under varying conditions. Each individual was simulated with a varying number of observations, n_i , where 10, 25, and 50 observations were used per individual during the specific time. The detailed parameters used in the synthetic data generation is described in Table 1.

2.3. Estimation of population parameters

2.3.1. Model structure and rationale

To estimate the population-level parameters of the tumor growth model, we employed a modified Global-Two-Stage (GTS) approach. The tumor volume for individual i at time t_{ij} is denoted as y_{ij}, and is described as:

$$y_{ij} = L(\psi_i, t_{ij}) + \epsilon_{ij}$$

where $L(\psi_i, t_{ij})$ is the the predicted log tumor volume from the models described in section 2.1, and $\epsilon_{ij} \sim N(0, \sigma^2)$ is the independent measurement error.

The individual-specific parameter $\psi_i \in \mathbb{R}^6$ which include the tumor growth and drug effect coefficients (α_1, α_2 , β_A, k_A, β_B, k_B) are assumed to follow a multivariate normal distribution:

$$\psi_i = \bar{\psi} + \eta_i$$

$$\eta_i \sim N(0, \Omega)$$

where $\psi_i \in \mathbb{R}^6$ represents the mean vector (fixed effects), and $\Omega \in \mathbb{R}^{6\times 6}$ is the covariance matrix describing inter-individual variability.

The GTS approach proceeds in two stages. In the first stage, we estimate individual parameters $\hat{\psi}_i$ and their associated uncertainty using nonlinear least squares (NLS) fits for each animal. In the second stage, we infer the population-level distribution $(\bar{\psi}, \Omega)$by treating the $\hat{\psi}_i$ as noisy observations and accounting for their estimated variances.

2.3.2. Global Two-Stage (GTS) Regression

For each individual j in the dataset, the vector of observed tumor volumes is defined as:

$$y_j = \begin{pmatrix} y_{j,t_1} \\ y_{j,t_2} \\ \vdots \\ y_{j,t_{n_j}} \end{pmatrix}$$

$$L(t,\theta_j) = \begin{pmatrix} L(t_1,\theta_j) \\ L(t_2,\theta_j) \\ \vdots \\ L(t_{n_j},\theta_j) \end{pmatrix}$$

with additive error:

$$\epsilon_j \sim N_{n_j}(0, \sigma^2 I_{n_j})$$

The individual parameters ψ_j are estimated by minimizing the residual sum of squares:

$$Q(\theta_j) = \frac{1}{2}\sum_{i=1}^{n_j}(y_{j,t_i} - L(t_i,\theta_j))^2$$

Under regularity conditions, the estimator $\hat{\psi}_j$ satisfies the asymptotic distribution:

$$\sqrt{n_j}\left(\hat{\theta}_j - \theta_j\right) \sim N_k\left(0, \hat{\sigma}^2\left[\frac{1}{n_j}\nabla_{\hat{\theta}_j}L\left(t,\hat{\theta}_j\right)\nabla_{\hat{\theta}_j}L\left(t,\hat{\theta}_j\right)^T\right]^{-1}\right)$$

where:

$$\hat{\sigma}^2 = \frac{1}{n_j - k}\sum_{i=1}^{n_j}\left(y_{j,t_i} - L\left(t_i,\hat{\theta}_j\right)\right)^2$$

and:

$$\nabla_{\hat{\theta}_j}L\left(t,\hat{\theta}_j\right) = \begin{pmatrix} \frac{\partial L\left(t_1,\hat{\theta}_j\right)}{\partial\theta_j^{(1)}} & \cdots & \frac{\partial L\left(t_1,\hat{\theta}_j\right)}{\partial\theta_j^{(k)}} \\ \vdots & \ddots & \vdots \\ \frac{\partial L\left(t_{n_j},\hat{\theta}_j\right)}{\partial\theta_j^{(1)}} & \cdots & \frac{\partial L\left(t_{n_j},\hat{\theta}_j\right)}{\partial\theta_j^{(k)}} \end{pmatrix}$$

The Hessian matrix is approximated by:

$$\nabla^2_{\hat{\theta}_j}Q\left(\hat{\theta}_j\right) \approx \nabla_{\hat{\theta}_j}L\left(t,\hat{\theta}_j\right)\nabla_{\hat{\theta}_j}L\left(t,\hat{\theta}_j\right)^T$$

We defined $\hat{S}_j^{-1}$ as the individual Fisher information matrix. To align parameter dimensionality across groups, the estimated parameters $\hat{\psi}_j$ are mapped into a 6-dimensional vector:

$$\theta_j = \begin{pmatrix} \alpha_{1,j} \\ \alpha_{2,j} \\ \beta_{A,j} \\ k_{A,j} \\ \beta_{B,j} \\ k_{B,j} \end{pmatrix}$$

For the groups where certain parameters are not estimable (e.g. no drug A in group B), zeros are inserted in the corresponding positions, and zero rows/columns are added to $\hat{S}_j^{-1}$ accordingly.

For example, in the control group (group C):

$$\hat{\theta}_j = \begin{pmatrix} \alpha_{1,j} \\ \alpha_{2,j} \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\hat{S}_j^{-1} = \begin{pmatrix} s_{11} & s_{12} & 0 & 0 & 0 & 0 \\ s_{21} & s_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

In group B (only drug B is treated):

$$\hat{\theta}_j = \begin{pmatrix} \alpha_{1,j} \\ \alpha_{2,j} \\ 0 \\ 0 \\ \beta_{B,j} \\ k_{B,j} \end{pmatrix}$$

$$\hat{S}_j^{-1} = \begin{pmatrix} s_{11} & s_{12} & 0 & 0 & s_{15} & s_{16} \\ s_{21} & s_{22} & 0 & 0 & s_{25} & s_{26} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ s_{51} & s_{52} & 0 & 0 & s_{55} & s_{56} \\ s_{61} & s_{62} & 0 & 0 & s_{65} & s_{66} \end{pmatrix}$$

Then, the individual level estimation was used to estimate population level parameters. Assume that

$$\psi_j \sim N_6(\psi, C_0)$$

105 we estimate the population mean θ_0 and covariance matrix C_0 by minimizing the following
106 objective function:

107 $$O(\xi) = \sum_{j=1}^{M}\left[(\hat{\psi}_j - \psi_0)^T(\hat{S}_j + C_0)^{-1}(\hat{\psi}_j - \psi_0) + \log\left(\det\left(\hat{S}_j + C_0\right)\right)\right]$$

108 This is solved iteratively using the following rules. At iteration n + 1, the empirical Bayes
109 estimate for each individual is updated as:

110 $$\hat{\psi}_j^{(n+1)} = (\hat{S}_j^{-1} + \left(C_0^{(n)}\right)^{-1})^{-1}\,(\hat{S}_j^{-1}\hat{\theta}_j + \,(C_0^{(n)})^{-1}\psi_0^{(n)})$$

111 The population mean is then updated by:

112 $$\psi_0^{(n+1)} = \frac{1}{M}\sum_{j=1}^{M}\hat{\psi}_j^{(n+1)}$$

113 and the covariance matrix is updated by:

114 $$C_0^{(n+1)} == \frac{1}{M}\sum_{j=1}^{M}(\hat{\psi}_j^{(n+1)} - \psi_0^{(n+1)})(\hat{\psi}_j^{(n+1)} - \psi_0^{(n+1)})^T + \frac{1}{M}\sum_{j=1}^{M}\left(\hat{S}_j^{-1} + \,(C_0^{(n)})^{-1}\right)^{-1}$$

115 These steps are repeated until ψ_0 and C_0 converges, yielding estimates of the fixed effects and
116 their inter-individual variability.

117

118 **2.3.3. Stochastic Approximation Expectation Maximization (SAEM)**

119 The individual- and population- level parameters can also be estimated simultaneously by
120 utilizing SAEM algorithm. Let $\theta = \left(\hat{\psi}, \Omega, \sigma^2\right)$. Maximum likelihood estimation of θ consists of

maximizing with respect to θ:

$$L(\theta, y) = p(y; \theta) = \int p(y, \psi; \theta) d\psi = \prod_{i=1}^{N} \int p(y_i|\psi_i; \theta) p(\psi_i; \theta) d\psi$$

SAEM was implemented to maximize the likelihood function, using saemix package in R.

2.4. Equivalence to Bliss Independence Model

Bliss independence principle is based on the probabilistic interpretation of the drug effect, where the presence of one drug does not affect the probability of another drug's effect on tumor growth decay. We note that the additive model is in concordance with Bliss independence model using the Treatment-to-control ratio. Let E(A) and E(B) represent the probability of cell growth inhibited by drug A and B respectively. Under Bliss independence, $E_{AB} = E_A + E_B - E_A E_B$

where the effect $E_g = 1 - \frac{V_g(t)}{V_C(t)}$, indicating the proportion of tumor being inhibited, and $V_g(t)\left(L_g(t)\right)$ is the log tumor volume in g, where $g \in A, B, C, AB$. This equation is equivalent to

$$\frac{V_{AB}(t)}{V_C(t)} = \frac{V_A(t)}{V_C(t)} \frac{V_B(t)}{V_C(t)}$$

$$\leftrightarrow \log\left(V_{AB}(t)\right) = \log\left(V_A(t)\right) + \log\left(V_B(t)\right) - \log\left(V_C(t)\right)$$

$$\leftrightarrow L_{AB}(t) = L_A(t) + L_B(t) - L_C(t)$$

$$\leftrightarrow L_{AB}(t) = \begin{cases} L_0 + \alpha_1 t + D_A(t) + D_B(t), & \alpha_2 = 0 \\ L_0 + \frac{\alpha_1}{\alpha_2}(1 - \exp(-\alpha_2 t)) + D_A(t) + D_B(t), & \alpha_2 \neq 0 \end{cases}$$

2.5. Hypothesis testing and model evaluation

After estimation of population parameters through the Global-Two-Stage (GTS) regression, we performed hypothesis testing and model evaluation to determine the statistical significance and adequacy of the proposed tumor growth model under combination treatments.

2.5.1. Normalized Prediction Distribution Errors (NPDE)

We first evaluated the adequacy of the fitted model to capture tumor growth curves in the combination therapy group, assuming the two drugs act independently under the Bliss independence hypothesis (null hypothesis). We used Normalized Prediction Distribution Errors (NPDE) to assess model fitting. NPDE transforms observations to values expected to follow a standard normal distribution, thus allowing model diagnostics through established statistical methods.

The predictive distribution of the observations can be computed as:

$$p(y_i;\ \hat{\theta}) = \int p(y_i|\psi_i;\hat{\theta})p(\psi_i;\hat{\theta})d\psi_i$$

where $\hat{\theta}$ denotes the maximum likelihood estimator obtained from the population parameter estimation.

For Monte Carlo approximation, this predictive distribution was computed as:

$$pd_{ij} = F_{ij}(y_{ij}) \approx \frac{1}{K}\sum_{k=1}^{K} I[y_{ij}^{sim(k)} < y_{ij}]$$

with $y_{ij}^{sim(k)}$ simulated from the predictive distribution. The pd_{ij} values thus represents the percentile of each observation y_{ij} within its predictive distribution.

We then transformed these percentiles into NPDE as follows:

$$npde_{ij} = \Phi^{-1}(pd_{ij})$$

Due to repeated measurements within individuals, correlation between observations was expected. To address this, decorrelation was performed using the empirical mean and variance-covariance matrix of simulated data:

$$a = E[Y_i] = \frac{1}{K}\sum_{k=1}^{K} Y_i^{sim(k)}$$

$$var[Y_i] = \frac{1}{K-1}\sum_{k=1}^{K}(Y_i^{sim(k)} - a)(Y_i^{sim(k)} - a)^T$$

The decorrelated data were then obtained by:

$$Y_i^{sim(k)*} = var[Y_i]^{-\frac{1}{2}}(Y_i^{sim(k)} - a)$$

We conducted the following statistical tests on the NPDE distribution; rank test to evaluate if the mean equals zero, Fisher variance test to determine if the variance equals one, and Shaprio-Wilk test to verify normality. A global test combining the above tests with a Bonferroni correction was applied. Alternatively, a Kolmogorov-Smirnov test was used to directly assess normality (mean = 0, variance = 1). The Type-I error rate was computed as the proportion of simulations incorrectly rejected at a significance level $\alpha = 0.05$.

2.5.2. Distance-based statistics

We then employed a distance-based measure to quantify discrepancies between observed tumor volume data and the values predicted by the fitted model. For each individual mouse *j*, we define the distance statistics D_j as:

$$D_j = \frac{1}{n_j}[y_j - L(t, \theta_j)]^T \, [Cov(y_j)]^{-1} \, [y_j - L(t, \hat{\theta}_0)]$$

This distance statistics follows an asymptotic chi-square distribution:

$$n_j D_j \sim \chi^2_{nj}$$

To assess the goodness-of-fit of the nonlinear mixed-effects model across control and single-drug treatment groups, we calculated the aggregated distance:

$$\sum_{j \in A,B,C} n_j D_j \sim \chi^2_{\sum j \in A,B,C} {}^{n_j}$$

For the combination group (AB), the distance statistic specifically tested whether the additive model was correctly specified under the null hypothesis of additive drug interaction:

$$\sum_{j \in A,B,C} n_j D_j \sim \chi^2_{\sum j \in AB} {}^{n_j}$$

These aggregated statistics were compared to their respective chi-square distributions to formally evaluate the model adequacy. Type-I error rate was calculated as the proportion of simulations with distance statistics exceeding the critical chi-square value at $\alpha = 0.05$.

2.5.3. Prediction interval

193 Finally, we assessed the model's predictive performance using a prediction interval approach.
194 We derived the 90 % prediction interval from the predictive distribution generated through
195 Monte Carlo simulation. Specifically, for each time point and individual, we simulated 1,000
196 tumor volume observations based on the fitted model parameters. We tested whether the
197 proportion of observed data points falling outside a pre-defined 90% prediction interval
198 significantly differed from the expected proportion (10%) using a binomial test. The type-I error
199 rate was determined by the proportion of simulations showing significant binomial test results (P
200 < 0.05).

201

202 **2.6. Implementation with Real PDX data**

203 To demonstrate the practical applicability of our approach, we applied the proposed
204 methodology to real patient-derived xenograft (PDX) data from an in vivo preclinical experiment
205 involving tumor-bearing mice randomly assigned to control, single-drug (A or B), and
206 combination treatment (AB) groups. Individual and population-level parameters were estimated
207 using the previously described Global-Two-Stage (GTS) method. Model adequacy and the
208 validity of the additive assumption (Bliss independence) were evaluated through NPDE analysis,
209 distance-based tests, and prediction interval diagnostics as outlined above.