Dose Finding Method in Joint Modeling of Efficacy and Safety Endpoints in Phase II Studies

Aiyang Tao[1], Yong Lin[2,3], José Pinheiro[4], & Weichung Joe Shih[2,3]

[1] Novartis Pharmaceuticals, One Health Plaza, East Hanover, New Jersey 07936, U.S.A

[2] Department of Biostatistics, School of Public Health, Rutgers, The State University of New Jersey, 683 Hoes LaneWest, Piscataway, NJ 08854, USA

[3] Biometrics Division, Rutgers Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, 120 Albany Street, New Brunswick, NJ 08901, USA

[4] Johnson & Johnson Pharmaceuticals, 920 Route 202, P.O. Box 300, Raritan, NJ 08869, U.S.A.

Correspondence: Yong Lin, Department of Biostatistics, School of Public Health, Rutgers, The State University of New Jersey, 683 Hoes Lane West, Piscataway, NJ 08854, USA. E-mail: linyo@rutgers.edu

Abstract

Determination of appropriate dose(s) to advance into Phase III trials is one of the most challenging and important tasks during drug development. Selecting a dose too high may result in unacceptable safety problems, while a too low dose may lead to ineffective drugs. Proper estimation of dose-response profiles for relevant safety and efficacy endpoints allows the reliable evaluation of the risk-benefit profile of a drug at the end of Phase II, as well as the selection of appropriate doses to be brought into confirmatory Phase III trials. Thus how to select dose(s) in Phase II trials by combining information about the efficacy and safety in a joint model setting may play a key role in drug development programs and can serve as a gate-keeper for large confirmatory Phase III trials with greater chance of success.

Dose finding methods through joint modeling of both efficacy and safety endpoints are studied in this paper. To be more specific, we extend the popular MCP-Mod dose finding method (Bretz et al., 2005), which considered only the efficacy endpoint, to the method that incorporates both efficacy and safety endpoints through joint modeling. Method of parameter estimation for the extended models, and methods of selection of dose(s) to be brought into confirmatory Phase III trials based on Phase II study data are discussed in the paper. The performances of the proposed methods are evaluated through simulations.

Keywords: dose finding, joint model, efficacy and safety, minimum effective dose, maximum safety dose

1. Introduction

Selection of appropriate dose(s) to carry into confirmatory Phase III trials is one of the most difficult decisions that need to be considered during drug development. It is believed by many that the high attrition rate currently observed in Phase III is largely driven by inadequate dose selection (FDC report, 1991; Bornkamp et al., 2007). Most commonly used dose finding designs and methods today still focus on selection of a target dose out of a fixed, generally small number of dose levels via pairwise hypothesis testing, which is typically inefficient (FDA, 2004). Assessment of dose-response should be an integral component of drug development, with studies designed to assess dose-response as an inherent part of establishing the safety and effectiveness of the drug. Characterization of dose-response relationship can be built into the development process with minimal extra effort and no much loss of time compared to development plans that ignore dose-response estimation. Proper estimation of such dose-response profiles for relevant safety and efficacy endpoints allows reliable evaluation of the risk-benefit profile of a drug at the end of Phase II, as well as the selection of appropriate doses to be brought into confirmatory Phase III trials. Thus how to select dose(s) in Phase II trials by combining information about the efficacy and safety in a joint model setting may play a key role in drug development programs and can serve as a gate-keeper for large confirmatory Phase III trials with greater chance of success.

The primary goals of dose-response studies are to establish the presence of a dose-response signal and to estimate a target dose Bretz et al., 2005; Dragalin et al. 2007, usually the minimum effective dose (*MED*). There are fixed and adaptive approaches for designing and analyzing dose-ranging studies (Bornkamp et al. 2007). Bretz et al. (2005) combined multiple comparison procedure and modeling (MCP-Mod), which has been used extensively for analyzing dose finding trials recently. It includes a PoC (proof-of-concept) assessment and a dose-selection step.

The clear advantage of this approach, compared to traditional multiple comparison dose finding methods, is its added flexibility in selecting an appropriate dose-response model for future drug development. Most approaches in literature so far are targeted only on efficacy for an optimal dose by assuming that a dose range with an acceptable toxicity has been previously determined. However, under a variety of circumstances safety problems may lead to early termination of the drug development in Phase II/III trials. Therefore it is important to address safety and efficacy simultaneously.

There has been some research on how to design dose-finding studies based on efficacy and toxicity responses jointly in early phase (I/II) (Thall and Cook, 2004; Dragalin, 2005; Thall et al., 2008) for oncology trials. These dose finding designs are based on oncology study are different from other dose finding clinical studies in that for oncology early phase studies subjects (exclusively only cancer patients) enter the studies sequentially while for other studies subjects (most of time healthy volunteers) enter the studies simulaneously. There are a few dose finding designs the incorporate the safety and efficacy responses (see for example, Bekele and Shen, 2005; Padmanabhan et al. 2010). But these methods do not take advantage of MCP-Mod approaches of Bretz et al. (2005).

In this paper we will extend the MCP-Mod approach to select the best joint model based on two continuous correlated efficacy and safety outcomes and to get the final optimum dose(s) from the best joint model for the Phase III study. We will address the extension of the MCP-Mod approach for mixed (continuous and discrete) types of correlated efficacy and safety outcomes in a sequel article.

This paper is organized as follows. In Section 2, we propose dose finding methods in joint modeling of efficacy and safety endpoints in Phase II studies. This section includes the methods of finding the maximum safety dose (*MSD*), the method of joint modeling of continuous efficacy and safety responses, the procedures of finding the *MED* and the *MSD*, and the methods of recommending dose(s) to carry into Phase III program development. In Section 3, the proposed methods are evaluated through simulations based on the data from our motivating example. Concluding remarks and recommendations are given in Section 4.

2. Joint Modeling of Efficacy and Safety Endpoints

Bretz et al. (2005) combined multiple comparison procedure and modeling (MCP-Mod), which has been used extensively for analyzing dose finding trials. It includes a PoC (proof-of-concept) assessment and a dose-selection step. The clear advantage of this approach, compared to traditional multiple comparison dose finding methods, is its added flexibility in selecting an appropriate dose-response model for future drug development. Bretz et al. (2005) considered only efficacy to identify the *MED* assuming that all considered doses are within safety tolerance. Assuming the following model for efficacy

$$Y_{ij} \;=\; f(d_i, \boldsymbol{\theta}_Y) \;+\; \epsilon_{ij}^Y, \text{ and } \epsilon_{ij}^Y \sim N\left(0, \sigma_Y^2\right),$$

where $f(d_i, \boldsymbol{\theta}_Y)$ is a non-linear function of dose d_i and vector of the model parameters $\boldsymbol{\theta}_Y$, Y_{ij} is the efficacy response of the jth patient in the ith dose group with dose d_i. Assuming that there are a total of k dose groups with $d_1 < d_2 < \cdots < d_k$, Bretz et al. defined MED as

$$MED = \operatorname*{argmin}_{d \in (d_1, d_k]} \{f(d, \boldsymbol{\theta}_Y) > f(d_1, \boldsymbol{\theta}_Y) + \triangle_e\},$$

where $\triangle_e$ is the clinically meaningful minimun efficacy difference. Let L_d be the lower $1 - 2\gamma$ confidence limit of the predicted mean value $f(d, \widehat{\boldsymbol{\theta}}_Y)$ at dose d. They proposed that *MED* be estimated by

$$\widehat{MED} = \operatorname*{argmax}_{d \in (d_1, d_k]} \left\{f(d, \widehat{\boldsymbol{\theta}}_Y) > f(d_1, \widehat{\boldsymbol{\theta}}_Y) + \Delta_e, L_d > f(d_1, \widehat{\boldsymbol{\theta}}_Y)\right\},$$

where $\widehat{\boldsymbol{\theta}}_Y$ is the estimated $\boldsymbol{\theta}_Y$.

In this section, we first show how to estimate the target dose, which is called maximum safety dose (*MSD*), for the safety endpoint. Then we propose the joint modeling of the continuous efficacy and safety responses, This includes

the model formulations, the parameter estimation method, and the procedures of finding the MED and the MSD for combined efficacy and safety data. After that, we propose two strategies for recommending dose(s) to carry into Phase III program development through the joint criteria of success or through utility functions.

2.1 Target Dose Estimation for Safety Endpoint

For a safety endpoint, the MED in the MCP-Mod from Bretz et al. (2005) can be replaced by the MSD. Consider the following model for safety

$$Z_{ij} = g(d_i, \boldsymbol{\theta}_Z) + \epsilon_{ij}^Z, \text{ and } \epsilon_{ij}^Z \sim N\left(0, \sigma_Z^2\right),$$

where $g(d_i, \boldsymbol{\theta}_Z)$ is a non-linear function of d_i and $\boldsymbol{\theta}_Z$, Z_{ij} is the safety response of the jth patient in the ith dose group, $i = 1, ..., k$, and $j = 1, ..., n_i$, d_i is the dose for ith dose group with d_1 for placebo, and $\boldsymbol{\theta}_Z$ refers to the vector of model parameters. The MSD is defined as the maximum safety dose which shows clinically acceptable toxicity.

To be more specifically, given the model $g(., \boldsymbol{\theta}_Z)$, we define

$$MSD = \underset{d \in (d_1, d_k]}{\operatorname{argmax}} \{g(d, \boldsymbol{\theta}_Z) \le g(d_1, \boldsymbol{\theta}_Z) + \triangle_s\},$$

where $\triangle_s$ is the largest clinically acceptable safety response difference from placebo, this implies that, the safety response at the dose is not too worse than that at placebo. Following the above definition, two different rules are proposed to estimate the true MSD. Denote U_d the upper $1 - 2\gamma$ confidence limit of the predicted mean value $g(d, \widehat{\boldsymbol{\theta}}_Z)$ at dose d based on the model $g(., \boldsymbol{\theta}_Z)$ (see Figure 1),
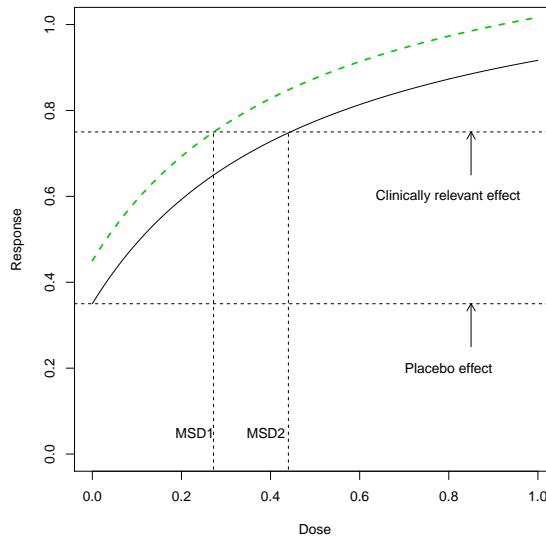
Figure 1. Maximum Safety Dose (MSD)

$$\widehat{MSD}_1 = \underset{d \in (d_1, d_k]}{\operatorname{argmax}} \left\{U_d \le g(d_1, \widehat{\boldsymbol{\theta}}_Z) + \Delta_s\right\},$$

$$\widehat{MSD}_2 = \underset{d \in (d_1, d_k]}{\operatorname{argmax}} \left\{g(d, \widehat{\boldsymbol{\theta}}_Z) \le g(d_1, \widehat{\boldsymbol{\theta}}_Z) + \Delta_s\right\},$$

where $\widehat{\boldsymbol{\theta}}_Z$ is the estimated $\boldsymbol{\theta}_Z$ and can be obtained based on the model $g(., \boldsymbol{\theta}_Z)$ using the nonlinear least squares fitting algorithm *nls* function (Bates and Chambers, 1992) in R (Version 2.11, R Foundation for Statistical Computing, Vienna, Austria, 2010). Based on our simulation in Section 3.2, although $\widehat{MSD}_2$ has a better approximation than $\widehat{MSD}_1$ in terms of bias and coverage rate, $\widehat{MSD}_1$ is still reasonable and is more clinically appropriate to be used to estimate the MSD as a conservative approach since we want to select a dose which has less chance to have toxicity (see Section 3.2 for more details). We will denote $\widehat{MSD}$ for $\widehat{MSD}_1$ in the sequel.

2.2 Joint Nonlinear Modeling of Continuous Efficacy and Safety Outcomes

In this section, a joint nonlinear modeling of continuous efficacy and safety outcomes is introduced, and methods of parameter estimation and strategies of finding *MED* and *MSD* based on the joint model are proposed. The performance of the proposed methods is evaluated through simulation in Section 3.3.

2.2.1 Joint model formulation, log-likelihood function, and estimation and computational methods

Let Y_{ij} and Z_{ij} be the efficacy and safety outcomes respectively for *j*th patient in *i*th dose group with dose d_i, $i = 1, ..., k$, and $j = 1, ..., n_i$, where $d_1 < d_2 < \cdots < d_k$. We assume that the variance-covariance structure between Y_{ij} and Z_{ij} is the same within the same dose group and may be different across dose groups. This formulation results in a simplified version of the extended nonlinear joint regression model as follows:

$$\begin{bmatrix} Y_{ij} \\ Z_{ij} \end{bmatrix} = \begin{bmatrix} f(d_i, \boldsymbol{\theta}_Y) \\ g(d_i, \boldsymbol{\theta}_Z) \end{bmatrix} + \begin{bmatrix} \epsilon_{ij}^Y \\ \epsilon_{ij}^Z \end{bmatrix}, \quad i = 1, ...k, \; j = 1, ...n_i \tag{1}$$

where

$$\boldsymbol{\epsilon}_{ij} = \begin{bmatrix} \epsilon_{ij}^Y \\ \epsilon_{ij}^Z \end{bmatrix} \sim BVN \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Psi_{YY}^i & \Psi_{YZ}^i \\ \Psi_{YZ}^i & \Psi_{ZZ}^i \end{bmatrix} \right),$$

$BVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is for the bivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, and $\begin{bmatrix} \boldsymbol{\theta}^Y \\ \boldsymbol{\theta}^Z \end{bmatrix}$ refers to the vector of model parameters, *i* to the dose group ($i = 1$ corresponds to placebo), and *j* to the patient within dose group *i*.

For computational convenience, the response variables Y_{ij} or Z_{ij} can be combined into a single variable V_{ijt} with the indicator variable *t*, i.e.,

$$V_{ijt} = \begin{cases} Y_{ij} \text{ if } t = 0 \\ Z_{ij} \text{ if } t = 1, \end{cases}$$

then the model can be reformulated as

$$V_{ijt} = f(d_i, \boldsymbol{\theta}_Y)^{1-t} \, g(d_i, \boldsymbol{\theta}_Z)^t + (1 - t, \, t) \, \boldsymbol{\epsilon}_{ij}, \tag{2}$$

where $\boldsymbol{\epsilon}_{ij} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{\Lambda_i})$ and

$$\sigma^2 \boldsymbol{\Lambda}_i = \begin{bmatrix} \Psi_{YY}^i & \Psi_{YZ}^i \\ \Psi_{YZ}^i & \Psi_{ZZ}^i \end{bmatrix},$$

$\boldsymbol{\Lambda}_i$ is a positive-definite matrix. The matrices $\boldsymbol{\Lambda}_i$ are determined by fixed, generally a small set of parameters $\boldsymbol{\lambda}$. Let $\boldsymbol{\Lambda}_i = \left(\boldsymbol{\Lambda}_i^{1/2} \right)^T \boldsymbol{\Lambda}_i^{1/2}$, $\boldsymbol{\Lambda}_i^{-1} = \boldsymbol{\Lambda}_i^{-1/2} \left(\boldsymbol{\Lambda}_i^{-1/2} \right)^T$, $\boldsymbol{V}_{ij} = \begin{bmatrix} Y_{ij} \\ Z_{ij} \end{bmatrix}$, $\boldsymbol{V}_{ij}^* = \left(\boldsymbol{\Lambda}_i^{-1/2} \right)^T \boldsymbol{V}_{ij}$,

$$\begin{pmatrix} f^*(d_i, \boldsymbol{\theta}_Y) \\ g^*(d_i, \boldsymbol{\theta}_Z) \end{pmatrix} = \left(\boldsymbol{\Lambda}_i^{1/2} \right)^T \begin{pmatrix} f(d_i, \boldsymbol{\theta}_Y) \\ g(d_i, \boldsymbol{\theta}_Z) \end{pmatrix}$$

and $\boldsymbol{\epsilon}_{ij}^* = \left(\boldsymbol{\Lambda}_i^{-1/2} \right)^T \boldsymbol{\epsilon}_{ij}$.

Estimation and inference under a single response nonlinear model have been studied in Pinheiro and Bates (2000). When the $\boldsymbol{\Lambda}_i$ matrices are known, this is referred to the generalized nonlinear least-squares (GNLS) model. The model in (1) can be re-expressed as a "classic" nonlinear model:

$$V_{ijt}^* = f^*(d_i, \boldsymbol{\theta}_Y)^{1-t} \, g^*(d_i, \boldsymbol{\theta}_Z)^t + (1 - t, \, t) \, \boldsymbol{\epsilon}_{ij}^*,$$

where $\boldsymbol{\epsilon}_{ij}^* \sim N(0, \sigma^2 \boldsymbol{I}_2)$, $\boldsymbol{I}_2$ is a 2×2 identity matrix.

The log-likelihood function for the GNLS model in (2) can be written as:

$$\begin{aligned} l(\boldsymbol{\theta}_Y, \boldsymbol{\theta}_Z, \boldsymbol{\lambda}, \sigma^2 | \boldsymbol{V}) = & -\frac{1}{2} N \log(2\pi) \\ & -\frac{1}{2} \left\{ \sum_{i=1}^k \sum_{j=1}^{n_i} \left[\sum_{t=0}^1 \frac{\|V_{ijt}^* - f^*(d_i, \boldsymbol{\theta}_Y)^{1-t} \, g^*(d_i, \boldsymbol{\theta}_Z)^t\|^2}{\sigma^2} + \log |\boldsymbol{\Lambda}_i| \right] \right\}, \end{aligned}$$

$$\tag{3}$$

where $N = \sum_{i=1}^{k} n_i$, N represents the total number of observations.

The estimation and inference methods discussed in Chapter 7.5 of Pinheiro and Bates (2000) can be applied to find the MLE estimates based on the log-likelihood function in (3). The *gnls* function (Pinheiro and Bates (2000) in R can be used to fit the extended nonlinear regression model using maximum likelihood. It can fit the reformulated model in (2) that is equivalent to the model in (1). Efficacy or safety data for the same patient are correlated and variances are usually unequal across dose groups. These two components can be reflected in the within-patient error variance structure by using the weight and correlation arguments in *gnls* .

2.2.2 Procedures of finding *MED* and *MSD* for combined efficacy and safety data

In PoC clinical trials, efficacy is usually tested first. If a drug does not have the PoC for efficacy, then there is no continued assessment for the safety endpoint. The significance level α for safety dose-response testing can be relaxed. The following procedures provide the testing sequence when both efficacy and safety endpoints need to be considered for the drug development.

Denote the discrete dose set $D = \{d_1, ..., d_k\}$ under investigation and the entire dose range $(d_1, d_k]$. Let $d_1 < d_2 < ... < d_k$. The first step is to confirm whether PoC exists for the efficacy, say, with Type I error $\alpha = 0.05$. If there is no PoC established for efficacy, which indicates no dose-response relationship exists for efficacy or $\widehat{MED} > d_k$, then there will be no dose-finding continued for this drug, i.e., the drug will not be carried to Phase III development. When the PoC of efficacy is established and also $\widehat{MED} \leq d_k$ is satisfied, the second step is to test PoC of the safety response, say, with Type I error $\alpha = 0.2$. When there is no established PoC for the safety response, the efficacy response alone is studied further to identify the *MED* for the Phase III program. In other word, we only need to focus on the dose finding for efficacy when the dose-response curve for safety is flat and no dose-response relationship is present for safety. When there is established PoC for the safety response, if $\widehat{MSD} < \widehat{MED}$, there will be no dose-finding continued for this drug. On the other hand, when both PoC for efficacy and safety responses exist and $\widehat{MSD} \geq \widehat{MED}$, joint modeling for efficacy and safety responses are performed.

The methods of finding *MED* and *MSD* under separate model fitting and joint model fitting are as follows:

A. With separate model fitting:

The next paragraph describe how to estimate the *MED* and *MSD* from separate model fittings when ignoring the correlation between the efficacy and safety responses:

1. The following set of candidate *efficacy* models are chosen for fitting the data by MCP-Mod (Bretz et al. 2005): Linlog, Emax, exponential, and quadratic, while the Linglog, linear, Emax and exponential models are selected for fitting *safety* data.

2. Fit the *efficacy* data separately to choose the best model with Type I error $\alpha = 0.05$ and the lowest Akaike information criterion (AIC). If PoC of *efficacy* is established, fit *safety* data separately and choose best model with Type I error $\alpha = 0.2$ and the lowest AIC.

3. Get the $\widehat{MED}$ and $\widehat{MSD}$ based on the models from Step 2.

B. With joint model fitting:

After PoC for both efficacy and safety are established, we first get all the significant efficacy and safety models, of which t-statistics are greater than the critical values (q^Y, q^Z) while controlling the family-wise error rate (FWER) from separate fittings. Type I error $\alpha = 0.05$ is pre-specified for PoC of efficacy and Type I error $\alpha = 0.2$ is pre-specified for PoC of safety, then the following two strategies can be used to obtain the *MED* and *MSD* from the joint model fitting:

Strategy I: Keep the most significant model obtained from separate efficacy and safety model fitting

1. Select the most significant efficacy and safety models based on the lowest AIC criteria from the set of significant efficacy and safety models obtained from separate fittings, keep the parameter estimates to use as start values for joint model fittings.

2. Fit Joint model by generalized nonlinear least squares model.

3. Obtain $\widehat{MED}$ and $\widehat{MSD}$ based on updated efficacy and safety model from joint fitting.

Strategy II: Keep all the significant models obtained from separate efficacy and safety models

1. The parameter estimates for all the significant efficacy and safety models obtained from separate fittings are kept and used as initial values for the joint model fitting.

2. Joint model fitting for all the combinations of efficacy and safety models selected from separate model fitting in Step 1, choose the best combination based on the lowest AIC from all the joint models fitting.

3. Obtain $\widehat{MED}$ and $\widehat{MSD}$ based on the updated efficacy and safety models from the best joint model.

We evaluate these two strategies in Section 3.3.4.

2.3 Suggested Dose(s) for the Phase III Program Development based on Joint Continuous Safety and Efficacy Responses

After $\widehat{MED}$ and $\widehat{MSD}$ are obtained with $\widehat{MSD} \geq \widehat{MED}$, we propose two different methods to select an optimal dose or a dose range for Phase III. The first method will focus on the joint success criteria for the efficacy and safety in Phase III; the second method will use the utility function to identify the final dose(s) to carry into Phase III program.

2.3.1 Method I: Identify the dose(s) through the joint criteria of continuous efficacy and safety responses for Phase III program

The recommended dose(s) will be determined by

$$\underset{d \in [MED,\ MSD]}{\arg\max} \quad P(Y > a, Z < b \mid d) \geq\ c,$$

or

$$\text{a dose range in } [MED, MSD] \text{ such that } P(Y > a, Z < b \mid d) \geq\ c,$$

where Y is the efficacy variable and Z is the safety variable, a and b are the criteria for Phase III success and c is the success probability. The probabilities are calculated based on the joint bivariate distribution estimated from the joint fitted model in Section 2.2.1.

2.3.2 Method II: Utility function based on trade off of continuous efficacy and safety responses for Phase III program

Another method to determine the recommended dose is based on the following utility function.

$$F(d) = \text{eff}(d) - k \times \text{saf}(d).$$

The dose is determined by maximizing the utility function $F(d)$, i.e., $\underset{d \in [MED,\ MSD]}{\arg\max} \quad F(d)$. Here, $k > 0$ is some weight to discount efficacy for safety. When k increases, we put more weight to safety, thus expect to lower the dose. The functions eff(d)and saf(d) can represent either

1) $P(Y > a \mid d)$ and $P(Z < b \mid d)$ respectively, with estimated marginal density function derived from either the separate model fitting or the joint model fitting, a and b are criteria for Phase III success probabilities, or

2) Standardized mean responses: $\text{eff}(d) = \widehat{Y}(d) = \frac{\widehat{E(Y(d))}}{\sqrt{\widehat{Var(Y(d))}}} = \frac{f(d,\widehat{\theta_Y})}{\widehat{\sigma}_Y}$ and $\text{saf}(d) = \widehat{Z}(d) = \frac{\widehat{E(Z(d))}}{\sqrt{\widehat{Var(Z(d))}}} = \frac{g(d,\widehat{\theta_Z})}{\widehat{\sigma}_Z}$, where the parameter estimates are based on the joint or separate model fitting.

3. Evaluation of Proposed Methods through Simulations

In this section, simulation studies are performed to evaluate the proposed methods in Section 2. The parameters used in the simulations are based on the data from our motivating example, a real clinical trial in which both safety and efficacy outcomes need to be considered simultaneously.

3.1 Motivating Example

Angiotensin-converting enzyme (ACE) inhibitors are used to treat hypertension and congestive heart failure. They have been clinically shown superior to other classes of drugs in the reduction of morbidity and mortality for cardiovascular disorders and hypertension (Thomas, 2000; Rossi, 2004). However, renal impairment is a significant adverse effect of all ACE inhibitors. The reason for this is still unknown. Some suggest that it is associated with their effect on angiotensin II-mediated homeostatic functions such as renal blood flow. Renal blood flow may be affected by angiotensin II because it vasoconstricts the efferent arterioles of the glomeruli of the kidney, thereby increases glomerular filtration rate (GFR). Hence, by reducing angiotensin II levels, ACE inhibitors may reduce GFR, a marker of renal function.

In one clinical trial an ACE inhibitor (Drug A) is used to treat hypertension. The efficacy endpoint is the change of sitting diastolic blood pressure from baseline. Decreasing GFR is the undesirable effect and the main safety measure is the change of GFR from baseline. Both response variables are assumed to be normally distributed. When the bivariate continuous outcomes are considered to find a suitable dose, joint model fitting that accounts for the correlation between efficacy and safety outcomes should be superior to the separate modeling that ignores the correlation.

3.2 Evaluation of Target Dose Estimation for Safety Endpoint

We evaluate the two proposed estimation methods discussed in Section 2.1 through simulation. We assume that the mean dose-response model $g(d_i, \theta_Z)$ for the safety responses Z_{ij} (change in GFR from baseline) is described by an exponential model with mean $0.163 + 0.037 \exp(0.5912 \times dose)$. The standard deviation for the change in GFR from baseline is 8 ml/min/1.73m^2, and the clinical relevance for MSD is the change in GFR from baseline less than 5 ml/min/1.73m^2. The placebo effect for mean change of GFR from baseline is assumed to be 0.2 ml/min/1.73m^2. The true MSD from the safety mean model is 0.829.

Let $\sigma_Z^2 = 64$, $n = 100$ per dose group and dose = 0, 0.05, 0.2, 0.4, 0.6, 0.8 and 1. Data are simulated 1000 times from the above safety model. The performance of the two proposed estimates of MSD are evaluated.

Figure 2 (Boxplots) shows that, as expected, $\widehat{MSD}_2$ is less biased and less dispersed than $\widehat{MSD}_1$. However, for safety endpoints, it is more desirable and more clinically appropriate to use more conservative approaches so that the dose selected has less chance to be toxic in practice. From the Figure, we see that $\widehat{MSD}_1$ is still reasonably well to be an estimate of MSD. Hence $\widehat{MSD}_1$ is used in our simulation and is denoted as $\widehat{MSD}$.
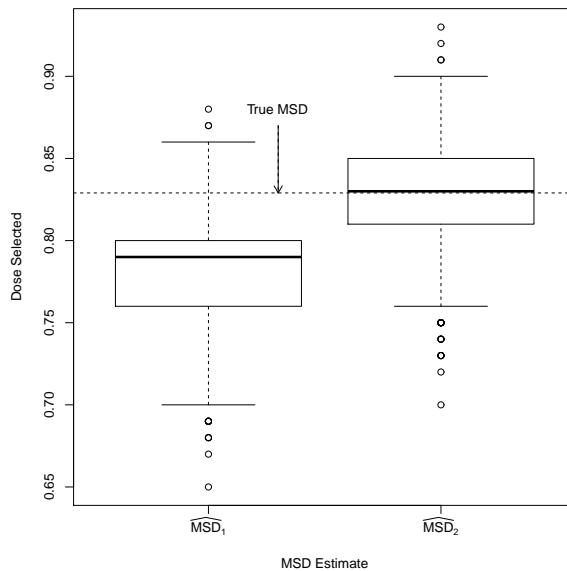

Figure 2. Box Plot of Maximum Safety Dose (MSD)

3.3 Evaluation of Joint Nonlinear Modeling of Safety and Efficacy Outcomes

We now evaluate the methods discussed in Section 2.2 through simulations.

3.3.1 Simulation Setups

We assume the joint model for the efficacy and safety responses in (1) with

$$
\begin{bmatrix} \Psi_{YY}^i & \Psi_{YZ}^i \\ \Psi_{YZ}^i & \Psi_{ZZ}^i \end{bmatrix} = \begin{bmatrix} \Psi_{YY} & \rho \Psi_Y \Psi_Z \\ \rho \Psi_Z \Psi_Y & \Psi_{ZZ} \end{bmatrix}
$$

for all i, where $\Psi_Y = \sqrt{\Psi_{YY}}$ and $\Psi_Z = \sqrt{\Psi_{ZZ}}$. The joint model fitting is conducted as described in Section 2.2.1.

The parameters used in the simulations are based on the related values from the motivating example. For the simulation purpose, we assumed that the change in diastolic blood pressure (DBP) from baseline follows the Emax model $f(d_i, \theta_Y) = e_0 + e_{max} \times d_i/(e_{d50} + d_i) = 2.5 + 14.5 \times d_i/(0.2 + d_i)$, and that the change in GFR from baseline follows the exponential model $g(d_i, \theta_Z) = e_{s0} + e_1 \exp(d_i/\delta) = 0.163 + 0.037 \times \exp(3.3 \times \log(6) \times d_i)$. The standard deviations for the changes from baselines in diastolic blood pressure and GFR are 7 mmHg and 8 $ml/min/1.73m^2$ respectively. The clinical relevance for MED is more than 3 mmHg of the difference in the change in DBP from baseline between treatment and placebo group, and for MSD is less than 5 $ml/min/1.73m^2$ of the difference in the change in GFR from baseline between treatment and placebo group, respectively. The placebo effect for the mean change in DBP from baseline is usually around 2.5 mmHg and the mean change in GFR from baseline is 0.2 ml/min/1.73 m^2. Based on the above, the true MED from the efficacy mean model is 0.056 and the true MSD from the safety mean model is 0.829.

Let $\Psi_{YY} = 49$, $\Psi_{ZZ} = 64$, and dose = 0, 0.05, 0.2, 0.4, 0.6, 0.8 and 1. Data are simulated 1000 times with $n = 50$ per dose group for evaluation of parameter (θ_Y and θ_Z) estimates and 500 times with $n = 50$ per dose group for evaluation of MED and MSD estimates for each $\rho = 0$, 0.4, 0.8, from the above joint efficacy and safety models.

After data are simulated, models are fitted separately and jointly with Emax model for efficacy response and Exponential model for safety response for evaluation of model parameter estimations. The mean, standard deviation, median, bias, percent of bias, MSE and relative efficiency for parameter estimates are computed. For evaluations of the estimates of MED and MSD, the strategies described in Section 2.2.2 are used with candidate models for efficacy and safety listed therein. Joint model fitting is performed using *gnls* function (Pinheiro and Bates, 2000) in R.

3.3.2 Evaluation Methods

The median, mean, percent bias, standard deviation, and mean squared error of the parameter (θ_Y and θ_Z) estimates, and MED and MSD estimates from the separate models and the joint model are computed from the 1000 and the 500 simulations, respectively. The estimates of the correlation between bivariate responses and the variances of the efficacy and safety responses from the *gnls* output in R are used to verify the correctness of simulation and goodness of the covariance matrix estimation of the model.

The parameter (θ_Y and θ_Z) estimates for the joint model and the separate models are compared by the percent of bias and the relative efficiency. In addition, the $\widehat{MED}$ and $\widehat{MSD}$ from the joint or the separate fitting are compared with the true mean from the selected dose-response models. The mean/median bias, percent bias, bootstrap 95% confidence interval coverage for MED and MSD are computed. Also the percentage of simulated trials for which:

- $\widehat{MED} \leqslant \widehat{MSD}$,

- true $MED \leqslant \widehat{MED} \leqslant true\ MSD$,

- $\widehat{MED} > true\ MSD$, and

- $0 < \widehat{MED} \leqslant true\ MED$

are calculated.

3.3.3 Simulation Results for Parameter (θ_Y and θ_Z) Estimates

Parameter estimates for Emax and exponential models either from the separate or the joint fitting are shown in Table 1. When the efficacy and safety responses are independent, the parameter estimates of Emax and exponential models are similar between separate and joint model fitting. This can be seen from the mean bias, percent of bias, mean squared error and relative efficiency of parameter estimates. When data are not correlated ($\rho = 0$), the mean

bias, percent of bias, mean squared error are a little bit bigger for joint fitting than for separate fitting because the responses for the underlying model are independent and joint fitting has an additional redudant parameter to estimate. But the differences between the two types of model fittings are minimal. When data are correlated ($\rho = 0.4$, or 0.8), the mean bias, percent of bias, mean squared error are smaller for joint fitting than for separate fitting. In addition, the efficiencies for joint modelling increase when the correlation between the efficacy and safety responses increases. As seen in Table 1, efficiencies increase from $0.98 - 1.0$ to $1.2 - 3.8$ for all parameter estimates when the correlation increases from 0 to 0.8.

Table 1. Parameter estimates based on separate and joint modeling -Emax for efficacy and Exponential for safety

Para meter	True value	Separate Fitting				Joint Fitting				RE.* (Joint vs. Sep.)
		Median	Mean (Std)	Bias (%)	MSE†	Median	Mean (Std)	Bias (%)	MSE†	
Model Correlation $\rho = 0$ (simulated value is -0.001):										
e_0	2.5	2.420	2.409 (0.9234)	-0.09 (-3.7)	1.720	2.340	2.388 (0.9208)	-0.11 (-4.49)	1.396	1.006
e_{max}	14.5	14.850	15.081 (2.1319)	0.58 (4.0)	9.756	14.586	15.126 (2.424)	0.63 (4.3)	8.081	0.990
e_{d50}	0.2	0.202	0.231 (0.1415)	0.03 (15.3)	0.042	0.203	0.231 (0.1464)	0.03 (15.6)	0.036	0.934
e_1	0.037	0.072	0.180 (0.3303)	0.14 (386.1)	0.212	0.072	0.182 (0.3358)	0.15 (391.6)	0.217	0.967
e_{s0}	0.163	-0.049	-0.128 (0.8488)	-0.29 (-178.5)	1.321	-0.052	-0.128 (0.8601)	-0.29 (-178.6)	1.337	0.974
δ	0.169	0.191	0.199 (0.0741)	0.03 (17.4)	0.010	0.199	0.199 (.0746)	0.03 (17.6)	0.010	0.987
Model Correlation $\rho = 0.4$ (simulated value is -0.397):										
e_0	2.5	2.450	2.411 (0.9176)	-0.09 (-3.5)	1.698	2.320	2.340 (0.8641)	-0.15 (-6.4)	1.260	1.127
e_{max}	14.5	14.934	15.064 (2.1493)	0.56 (3.9)	9.866	14.984	15.132 (2.000)	0.63 (4.4)	7.182	1.155
e_{d50}	0.2	0.203	0.230 (0.1387)	0.03 (14.9)	0.040	0.197	0.223 (0.1245)	0.03 (11.4)	0.026	1.241
e_1	0.037	0.073	0.183 (0.3267)	0.15 (393.4)	0.212	0.066	0.165 (0.2999)	0.13 (346.2)	0.174	1.187
e_{s0}	0.163	-0.054	-0.135 (0.8448)	-0.30 (-182.8)	1.327	-0.043	-0.113 (0.8339)	-0.27 (-169.4)	1.260	1.026
δ	0.169	0.190	0.199 (0.0746)	0.03 (17.6)	0.011	0.187	0.195 (.0699)	0.03 (15.5)	0.009	1.141
Model Correlation $\rho = 0.8$ (simulated value is -0.797):										
e_0	2.5	2.477	2.451 (0.9381)	-0.05 (-2.0)	1.759	2.424	2.405 (0.7238)	-0.09 (-3.8)	0.940	1.680
e_{max}	14.5	14.748	14.959 (2.0296)	0.46 (3.2)	8.635	14.757	14.936 (1.6763)	0.44 (3.0)	5.292	1.466
e_{d50}	0.2	0.200	0.231 (0.1332)	0.03 (15.3)	0.037	0.204	0.215 (0.0702)	0.01 (7.3)	0.009	3.604
e_1	0.037	0.066	0.170 (0.2742)	0.15 (358.8)	0.156	0.055	0.105 (0.1406)	0.07 (182.6)	0.043	3.801
e_{s0}	0.163	-0.039	-0.119 (0.7943)	-0.28 (-172.8)	1.195	-0.044	-0.013 (0.7239)	-0.17 (-107.8)	0.979	1.204
δ	0.169	0.188	0.196 (0.0725)	0.03 (15.8)	0.010	0.181	0.181 (0.0536)	0.01 (6.9)	0.005	1.833

Note: outputs are from 1000 simulations and $n = 50$ per dose group.

*: RE = relative efficiency. †: MSE = mean squared error.

3.3.4 Simulation Results for $\widehat{MED}$ and $\widehat{MSD}$ from Separate and Joint Model Fittings

Table 2 shows the results of $\widehat{MED}$ and $\widehat{MSD}$ from joint and separate fitting when correlations ρ of efficacy and safety responses are 0, 0.4, and 0.8. When $\rho = 0$ (independent), percent of bias and mean squared error for $\widehat{MED}$ are similar between separate model fitting and joint model fitting methods described in Section 2.2.2; mean squared error and percent bias for $\widehat{MSD}$ are slightly better for the joint model fitting (using the approach in Strategy I) than separate fitting. mean squared error decreases from 0.00323 to 0.00183.

When the correlation ρ between efficacy and safety increases, percent of bias and mean squared error decrease more for both approaches of joint fitting than separate fitting in $\widehat{MED}$ and $\widehat{MSD}$ estimation. As shown in Table 2, both separate and joint model approaches overesetimate the MED and underestimate the MSD within a reasonable range. This is desired in clinical studies as we prefer a conservative appproach to get more clinically appropriate estimates of MED and MSD. With joint model approaches, both $\widehat{MED}$ and $\widehat{MSD}$ are better estimated when the data are correlated. In general, joint fitting with the Strategy II approach is better than joint fitting with the Strategy I approach in terms of percent of bias and mean squared error.

Furthermore, the consistency of correlation and variance parameter estimates for both efficacy and safety responses are also shown in the simulations. With the true values of $\Psi_{YY} = 49$ and $\Psi_{ZZ} = 64$, the average estimated values are $\widehat{\Psi}_{YY} = 49.13$, 49.13 and 49.14, $\widehat{\Psi}_{ZZ} = 64.13$, 63.96 and 63.84, and $\widehat{\rho} = -0.001$, 0.399 and 0.799 for $\rho = 0$, 0.4, and 0.8 respectively.

Table 2. Estimates $\widehat{MED}$ and $\widehat{MSD}$ from separate fitting and joint fitting

		True	Fitting	Mean	SD	Median	bias	% bias	MSE
				$\rho = 0$					
Efficacy	MED	0.052	Separate	0.062	0.019	0.060	0.011	20.8	0.00047
Model:			Joint (1)	0.062	0.019	0.060	0.010	19.7	0.00045
Emax			Joint (2)	0.063	0.020	0.060	0.011	21.2	0.00050
Safety	MSD	0.829	Separate	0.783	0.033	0.790	-0.040	-5.6	0.00323
Model:			Joint (1)	0.803	0.035	0.810	-0.026	-3.2	0.00183
Exp.			Joint (2)	0.798	0.055	0.810	-0.037	-3.7	0.00401
				$\rho = 0.4$					
Efficacy	MED	0.052	Separate	0.063	0.020	0.060	0.011	21.1	0.00054
Model:			Joint (1)	0.063	0.019	0.060	0.010	19.9	0.00047
Emax			Joint (2)	0.063	0.019	0.060	0.010	20.2	0.00049
Safety	MSD	0.829	Separate	0.782	0.034	0.790	-0.047	-5.6	0.0033
Model:			Joint (1)	0.801	0.035	0.810	-0.028	-3.4	0.0019
Exp.			Joint (2)	0.801	0.035	0.800	-0.028	-3.4	0.0020
				$\rho = 0.8$					
Efficacy	MED	0.052	Separate	0.063	0.020	0.060	0.011	20.8	0.00052
Model:			Joint (1)	0.061	0.016	0.060	0.009	17.8	0.00032
Emax			Joint (2)	0.060	0.014	0.060	0.008	16.1	0.00028
Safety	MSD	0.829	Separate	0.782	0.034	0.780	-0.047	-5.7	0.0034
Model:			Joint (1)	0.799	0.033	0.800	-0.029	-3.6	0.0019
Exp.			Joint (2)	0.800	0.031	0.800	-0.029	-3.4	0.0019

Notes: a. Outputs are from 500 simulations and n=100 per dose group.

b. Joint (1) for $\widehat{MED}$ and $\widehat{MSD}$ are based on the joint model of most significant model from separate fitting of efficacy and safety data by lowest AIC criteria. The results consist of 460 emax–exponential, and 29 quadratic-exponential final joint models when $\rho = 0$; 453 emax–exponential, and 37 quadratic-exponential and 6 of linlog-exponential final joint models when $\rho = 0.4$; 448 emax-exponential, and 35 quadratic-exponential final joint models when $\rho = 0.8$.

c. Joint (2) for $\widehat{MED}$ and $\widehat{MSD}$ are based on all joint models by combinations of significant models from separate fitting of efficacy and safety data, the final model is selected from all joint models by lowest AIC criteria. The results consist of 460 emax-exponential, 34 quadratic-exponential, 5 emax-linear and 1 quadratic-linear final joint models when $\rho = 0$; 464 emax-exponential, 24 quadratic-exponential and 12 linlog-exponential final joint models when $\rho = 0.4$; 476 emax-exponential, 24 quadratic-exponential final joint models when $\rho = 0.8$.

3.3.5 Evaluation of Bias and Precision of the Target Dose Estimate

Table 3 displays the outputs of the relationship among $\widehat{MED}$, $\widehat{MSD}$, true values of efficacy and safety means. In general it shows the consistency of the estimated $\widehat{MED}$ and $\widehat{MSD}$ for both the joint model and separate model fitting. All the estimated $\widehat{MED}$ are lower than the estimated $\widehat{MSD}$ value for this simulated example, 60% of $\widehat{MED}$ are lower than the true safety mean and higher than the true efficacy mean.

Table 3. Comparisons of $\widehat{MED}$, $\widehat{MSD}$ to true mean based on separate and joint fittings

Fitting	Efficacy (Model=Emax)		Safety (Model=Exp.)		% of $\widehat{MED} \leq \widehat{MSD}$	% of $\widehat{MED} \geq$ True MSD	% of Eff. Mean $\leq \widehat{MED} \leq$ True MSD	% of $0 < \widehat{MED} \leq$ True MED
	True MED	$\widehat{MED}$	True MSD	$\widehat{MSD}$				
Model Correlation $\rho = 0$ (estimated $\widehat{\rho} = -0.060$)								
Separate	0.052	0.062	0.829	0.783	100	0	59.8	40.2
Joint	0.052	0.063	0.829	0.798	100	0	60.0	40.0
Model Correlation $\rho = 0.4$ (estimated $\widehat{\rho} = 0.399$)								
Separate	0.052	0.063	0.829	0.782	100	0	59.8	40.2
Joint	0.052	0.063	0.829	0.801	100	0	60.0	40.0
Model Correlation $\rho = 0.8$ (estimated $\widehat{\rho} = 0.799$)								
Separate	0.052	0.063	0.829	0.782	100	0	59.8	40.2
Joint	0.052	0.060	0.829	0.800	100	0	60.0	40.0
Note: outputs are from 500 simulations and n=100 per dose group by separate and joint model fitting.								

3.4 Evaluations of Suggested Dose(s) for the Phase III program Development based on Joint Safety and Efficacy Response

3.4.1 Simulation Procedure

The following is the simulation setup:

1. σ_y, σ_z, ρ are from the previous joint model fitting (these values will be retained from the previously described Strategy II, which select the most significant joint model from all combinations). μ_y and μ_z will be derived from the mean models for efficacy and safety through the simulation results based on the joint model fitting. The true correlation ρ between efficacy and safety response is 0.8.

2. Criteria values a and b are given such that the DBP change from baseline is more than certain value ($a = 3$ mmHg) and GFR change from baseline is less than certain value ($b = 6$ ml/min/1.73 m^2) in order to claim the success of the Phase III program. d_e and d_s are the doses which fit the mean efficacy and safety model to satisfy criteria values a and b respectively.

3. d_e and d_s are 0.056 and 0.824 respectively, which are derived from the updated mean efficacy and safety model from the joint model fitting to satisfy the mean efficacy change of a and mean safety change of b.

3.4.2 Simulation Results

Simulation results are shown in Figure 3. The simulated parameter estimates for efficacy and safety response models are based on the Emax and exponential model simulation results respectively. There are 4 lines in each plot. The upper and lower lines represent upper 97.5% and lower 2.5% of 500 values of the probabilities or utility indices based on the simulated parameter estimates. The middle 2 lines are the true probability or utility indices based on the true parameter value and the mean of simulated probability or utility indeces, respectively.
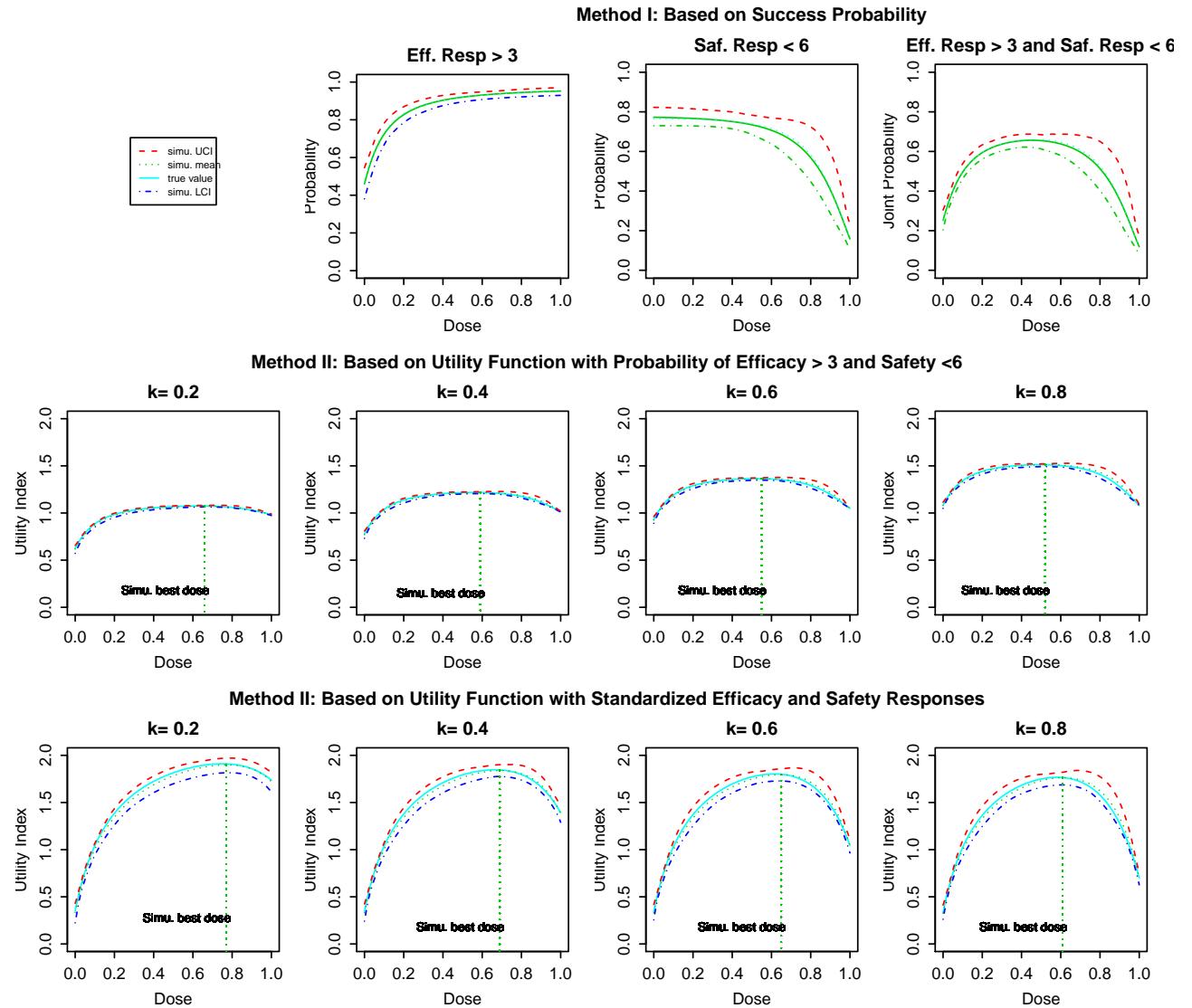
Figure 3. Suggested doses based on success probabilities (Method I) or utility functions (Method II)

In the top row of Figure 3, the first two plots show the probability of efficacy response (change from baseline) > 3 and the probability of safety response (change from baseline) < 6 from separate fittings, respectively. The right most plot shows the joint probability for efficacy response (change from baseline) > 3 and safety response (change from baseline) < 6. The simulated probability is very close to the true probability, which results in the similar final best dose satisfying the joint efficacy and safety criteria. The best dose that satisfies the joint criteria based on the true parameter value is 0.47 with 66.03% of success probability while the best simulated dose is 0.45 with 65.65% of success probability. The set of doses with probability > 0.6 of satisfying the joint criteria are from 0.21 to 0.71 based on true parameter, and from 0.21 to 0.69 based on the simulated parameter estimates.

The middle row of Figure 3 shows the utility index based on the probability of the efficacy response (change from baseline) > 3, the safety response (change from baseline) < 6 and different weights k for the safety response. As seen in Figure 3, the simulation shows that when k increases from 0.2 to 0.8, which means the efficacy is discounted more for safety, the maximum utility index increases from 1.074 to 1.508 and the best dose to satisfy this criteria decreases from 0.63 to 0.48.

Similarly, the simulated best dose for standardized efficacy and safety response is very similar to the true dose based on the true parameter value (third row of the plots in Figure 3). Furthermore, the simulation shows that when

k increases from 0.2 to 0.8, which means the efficacy is discounted more for safety, the maximum utility index decreases from 1.919 to 1.757 and the best dose to satisfy this criteria decreases from 0.76 to 0.58.

4. Conclusions

Determining optimal dose(s) for the Phase III program is an important task for the success of a drug development. In the past, this precess is carried out by considering efficacy and safety endpoints separately, even though it is well understood that these endpoints are correlated and depend on dose. In this paper, we develop the methodology to identify the final dose for the Phase III program based on joint nonlinear models for continuous safety and efficacy endpoints. The procedure is easily implemented by using available functions in R. Simulation results show that the proposed approaches provide more efficient parameter estimates than the separate model fitting. The identification of the final dose(s) for the Phase III program naturally relies on the individual profiles of the drug. We have explicitly considered the joint criteria for efficacy and safety, and a utility function expressing the trade-off between efficacy and safety. These criteria and function involve parameters that facilitate the inputs from clinicians.

Acknowledgements

Reference

Bates, D, Chambers, J. M. (1992). Nonlinear models. *Statistical Models* in S., J.M. Chambers and T.J. Hastie(eds), 421-454.

Bekele, B. N., Shen, Y. (2005). A Bayesian Approach to Jointly Modeling Toxicity and Biomarker Expression in a Phase I/II Dose Finding Trial. *Biometrics* , *61*(2), 343-354. http://dx.doi.org/10.1111/j.1541-0420.2005.00314.x

Bretz, F., Pinheiro, J., & Bransosn, M.(2005). Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics*, *61*, 738-748. http://dx.doi.org/10.1111/j.1541-0420.2005.00344.x

Bornkamp, B., Bretz, F., Dmitrienko, A., Enas, G., Gaydos, B., Hsu, C. H., ... Shen, F. (2007). Innovative approaches for designing and analyzing adaptive dose-ranging trials. *Journal of Biopharmaceutical Statistics*, *17*(6), 965-995. http://dx.doi.org/10.1080/10543400701643848

Dragalin, V., Hsuan, F., Padamanabhan, S. K. (2007). Adaptive designs for dose-finding studies based on Sigmoid Emax model. *Journal of Biopharmaceutical Statistics*, *17*, 1051-1070. http://dx.doi.org/10.1080/10543400701643954

FDA. (2004). Innovation/Stagnation: Challenge and opportunity on the critical path to new medical products. *FDA report from March*.

FDC Report.(1991). *The Pink Sheet*, May 6, 53(18), 14-15.

Pinheiro, J., & Bates, D. (2000). Mixed-effects models in S and S-PLUS. New York: Springer. http://dx.doi.org/10.1007/978-1-4419-0318-1

Padmanabhan, S. K., Hsuan, F., & Dragalin, V. (2010). Adaptive Penalized D-Optimal Designs for Dose Finding Based on Continuous Efficacy and Toxicity. *Statistics in Biopharmaceutical Research*, *2*(2), 182-198. http://dx.doi.org/10.1198/sbr.2009.0053

R Development Core Team. (2008). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org.*.

Rossi, S. (Ed.). (2004). Australian Medicines Handbook. ISBN 0-9578521-4-2.

Thomas, M. C. (2000). Diuretics, ACE inhibitors and NSAIDs - the triple whammy. *Med J*, *172*(4), 184-185.

Copyrights