



R Software for Constructing Optimal Designs for Bivariate Probit Model

Sourish Saha, Richard M. Heiberger and Valerii Fedorov

GSK DDS Technical Report 2011-01

Mar 2011

This paper was reviewed and recommended for publication by

Copyright © 2010 by GlaxoSmithKline Pharmaceuticals

Drug Development Sciences

GlaxoSmithKline Pharmaceuticals

1250 South Collegeville Road, PO Box 5089

Collegeville, PA 19426-0989, USA

R Software for Constructing Optimal Designs for Bivariate Probit Model

Sourish Saha, Richard M. Heiberger and Valerii Fedorov

Research Statistics Unit, Drug Development Sciences
GlaxoSmithKline Pharmaceuticals

ABSTRACT

1. INTRODUCTION

One important goal in the early phase of drug development is to identify the optimal dose(s) that should be carried to the later phase, or to help the investigators understand the dose-response relationship better. The definition of “optimal” dose(s) varies depending on the study goal such as finding the dose which will produce the maximum efficacy response rate without toxicity, or finding the doses that will provide the most accurate information about the model parameters for the dose-response model. In this document, the focus is on the model-based dose finding strategies for which a dose-response relationship is assumed for the drug under investigation. It is also assumed that the dose range such as maximum toxic dose has been determined. These dose-finding strategies are developed in the optimal design framework and for different study goals. As of today, commercial software does not provide options for constructing the optimal designs for non-linear models which are commonly used in the drug development. A set of R programs ([?]) is developed to implement the well-known first-order exchange algorithm for constructing locally optimal designs, L-optimal designs, two-stage design and adaptive optimal designs for bivariate probit model. A user-friendly interface written in RExcel is used to embed the R programs into the Excel environment. RExcel ([?], [?]) is an Excel add-in that allows the full functionality of R to be accessed from Excel. RExcel allows sharing of complex R-based programs with users who have no knowledge of R. The communication between the programs is hidden from the user.

2. BACKGROUND OF OPTIMAL DESIGN THEORY

2.1. Dose-response model. A dose-response model describes the value/probability of a specified response as a function of the dose for a specified population. This function is usually given in the form of a mathematical relationship. For continuous responses, one popular choice of the dose-response model can be linear models. For binary responses, some popular dose-response functions are logistic regression model, Emax model, probit model and so on. In this study, the dose-response relationship does not have

to be monotonic as some dose-finding procedures require. In other words, the response does not need to increase as the dose increases. In clinical trials, the dose-finding goals may be locating the dose that will achieve a pre-specified response/probability (such as ED50, ED90), or locating the doses that will provide the most accurate estimate of the dose-response model. Figure ?? plots the dose-response curve for logistic regression model (blue curve); the green dots denote the ED50 of this model.

2.2. General model. Here let $\eta(x, \theta)$ denote the dose-response model, where x are the covariates of interest such as dose, age, gender etc., and $\theta = (\theta_1, \theta_2, \dots, \theta_m)^T$ denotes the model parameters. Consider the simplest case that the model only contains one covariate dose and there are n dose levels x_1, \dots, x_n and the responses collected at these doses are

$$y_{11}, \dots, y_{1n_1}, y_{21}, \dots, y_{2n_2}, \dots, y_{n1}, \dots, y_{nn_n},$$

where n_1, n_2, \dots, n_n are the the number of observations at dose x_1, x_2, \dots, x_n , respectively, and the total number of subjects is N . In drug development, there are some popular choices of the parametric models such as the linear regression model

$$\eta(x, \theta) = \theta^T f(x),$$

logistic regression model

$$\eta(x, \theta) = \frac{\exp(\theta^T f(x))}{1 + \exp(\theta^T f(x))},$$

where $f(x) = (1, \mathbf{x})$ if only one drug is in the study or $f(x) = (1, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_1 \mathbf{x}_2)$ if drug combinations are considered.

Optimal design theories are based on General Equivalence Theorem ([?],[?],[?]). In the clinical trial setting, the general goal is to decide how many treatments are needed, where in the pre-specified dose region to put those treatments, and how to allocate patients to these treatments (the sample size for each treatment). One can obtain different designs for different study goals such as locate the dose with the maximum efficacy response without toxicity, or one can obtain the unknown parameters as accurately as possible. For each study goal, there is a corresponding design criterion which will be discussed in detail after the introduction of the necessary notations.

Denote the design ξ as

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_n \\ \lambda_1 & \lambda_2 & \dots & \lambda_n \end{array} \right\}, \quad x_1, x_2, \dots, x_n \in \mathcal{X}$$

The ‘‘optimal’’ dose often depends on the unknown parameters and the precision of the estimates is measured by the variance-covariance matrix of the maximum likelihood estimate (MLE) of θ . The log

likelihood function is written as:

$$l(\theta) = \sum_{i=1}^n \sum_{j=1}^{n_i} p(y_{ij}|x_i, \theta),$$

and MLE $\hat{\theta} = \arg \max_{\theta} l(\theta)$.

2.3. Information matrix. For each individual in the trial we accumulate some information which can be described by the information matrix $\mu(x, \theta)$ as

$$\mu(x, \theta) = E \left[\frac{\partial p(y|x, \theta)}{\partial \theta} \frac{\partial p(y|x, \theta)}{\partial \theta^T} \right] = -E \left[\frac{\partial^2 p(y|x, \theta)}{\partial \theta^2} \right].$$

The total information matrix

$$M_N(\xi, \theta) = \sum_{i=1}^n n_i \mu(x_i, \theta)$$

can be defined as the summation of the individual information matrices. The precision of the unknown parameter estimates is measured by the variance-covariance matrix. When the sample size is large enough, one can observe the asymptotic property that the variance-covariance matrix of the unknown parameters estimate $\hat{\theta}$ is approximately equal to the inverse of $M_N(\xi, \theta)$, i.e.

$$\text{Var}(\hat{\theta}) \approx M_N^{-1}(\xi, \theta).$$

A practical design should be exact, but here it is assumed that the design ξ is continuous. That is to say, the sample size for a given dose doesn't have to be a integer, just for simplicity of calculation. In optimal design framework, it is generally assumed the sample size is given with a defined normalized information matrix.

$$M(\xi, \theta) = 1/N \sum_{i=1}^n n_i \mu(x_i, \theta) = \sum_{i=1}^n \lambda_i \mu(x_i, \theta)$$

2.4. Design criterion and sensitivity function. To reach the ultimate goal of getting the unknown parameter estimates as precisely as possible, one needs to minimize the design criterion Ψ which is a function of the information matrix $M(\xi, \theta)$ where Ψ is a convex function and differentiable. The ultimate goal is to locate the design ξ^* which satisfies the following:

$$\xi^* = \arg \min_{\xi} \Psi[M(\xi, \theta)], \tag{1}$$

where Ψ is the convex function (design criterion) for the corresponding study goal. Table ?? lists some popular design criteria ([?],[?]),

- D-optimality is the most popular criterion which minimizes the general variance of all the unknown parameters ($M^{-1}(\xi, \theta)$) and is the most robust one.

TABLE 1. Some popular design criteria and corresponding

Criterion	Ψ	Sensitivity Function	Boundary under ξ^*
D	$\ln M^{-1}(\xi, \theta) $	$\text{tr}[\mu(x, \theta)M^{-1}(\xi, \theta)]$	m
D_s	$\ln A^T M^{-1}(\xi, \theta)A $	$\text{tr}[\mu(x, \theta)M^{-1}(\xi, \theta)] - \text{tr}[\mu_2(x, \theta)M_2^{-1}(\xi^*, \theta)]^\#$	
A	$\text{tr}(AM^{-1}(\xi, \theta))$	$\text{tr}[\mu(x, \theta)M^{-1}(\xi, \theta)AM^{-1}(\xi, \theta)]$	$\text{tr}[AM^{-1}(\xi^*, \theta)]$
c	$c^T M^{-1}(\xi, \theta)c$	$\text{tr}[\mu(x, \theta)M^{-1}(\xi, \theta)c^T cM^{-1}(\xi, \theta)]$	$\text{tr}[c^T cM^{-1}(\xi^*, \theta)]$
$L(\theta)$	$L(\theta)^T M^{-1}(\xi, \theta)L(\theta)$	$\text{tr}[\mu(x, \theta)M^{-1}(\xi, \theta)L(\theta)^T L(\theta)M^{-1}(\xi, \theta)]^*$	$\text{tr}[L(\theta)^T L(\theta)M^{-1}(\xi^*, \theta)]$

$^\# \mu_2(x, \theta)$ and $M_2^{-1}(\xi^*, \theta)$ denote the sub-block of the information matrix that corresponds to the parameters that are not of interest.

* $L(\theta)^T = \partial \eta^{-1}(x, \theta) / \partial \theta$, where $\eta^{-1}(x, \theta)$ is the target dose (e.g. for a target response rate) derived from the dose-response model.

- D_s -optimality criterion, the criterion to minimize the determinant of the variance-covariance matrix of a subset of the unknown parameters. For instance, the precision of the regression coefficients for the first order term in a quadratic model. Where $A^T \theta$ is the subset of the parameters that are of interest.
- In A-optimality criterion, A is a $m \times m$ non-negative definite matrix, recall here m is the number of unknown parameters in the model. If A equals the identity matrix I , this criterion minimizes the average variance of the unknown parameters.
- c-optimality minimizes the variances of a linear combination of the unknown parameters, i.e. $c^T \theta$, where c is an m -dimension vector.
- $L(\theta)$ -optimality minimizes the variance of the estimate of a quantity, for instance, a target response (rate) of the dose-response model.
- D -optimal designs minimize the volume of the asymptotic confidence region for θ . This property is easy to explain to practitioners in various fields.
- D -optimal designs are invariant with respect to non-degenerate transformations of parameters (e.g., changes of the parameter scale).
- D -optimal designs are attractive in practice because they often perform well according to other optimality criteria; see Atkinson and Donev (1992, Chapter 11) or Fedorov and Hackl (1997).

In general, it would be a mathematical challenge to locate the optimal design ξ^* by minimizing function Ψ directly. However, the Kiefer-Wolfowitz (cite) equivalence theorem provides a necessary and sufficient condition for ξ^* that satisfies Equation (??). In another words, this theorem provides the condition—the so called sensitivity function $\psi(x, \xi, \theta)$ —that is the derivative of Equation (??) which equals zero at a minimum of the function. Table ?? lists the sensitivity function for the corresponding optimal criteria. For example, the sensitivity function for D-optimal criterion is

$$\psi(x, \xi, \theta) = \text{tr}[\mu(x, \theta)M^{-1}(\xi, \theta)] \leq m, \quad (2)$$

where m is the number of unknown parameters. The inequality holds if and only if ξ is the optimal design and equal sign holds at the optimal design points. The property sets a good guideline for checking whether the design is optimal or not. For other optimal criteria, there are also corresponding upper boundaries for identifying optimal designs. These boundaries are also listed in Table ?. The following text will focus on two criteria: D-optimality and $L(\theta)$ -optimality.

2.5. Numerical procedure. The well known numerical procedure—first order exchange algorithm—is developed based on Kiefer–Wolfowitz’s Equivalence Theorem and it usually performs pretty well ([?]). It is an interactive procedure which shuffles the points among the current design and the candidate design points set to achieve the optimal condition. All our R programs are based on this numerical algorithm. The main idea is that at each step of the algorithm, the sensitivity function $\psi(x, \xi, \theta)$ is maximized over the design region to determine the “most informative” (i.e. the point which has the largest variation) new support point (forward step). It is then minimized over the support points of the current design to remove the “least informative” (i.e. the point which has the smallest variation) point in the current design (backward step). See [?] or [?] for more details.

Any numerical procedure for constructing optimal designs requires two key elements:

- The information matrix $\mu(x, \theta)$ or, equivalently, sensitivity function $\psi(x, \xi, \theta)$.
- The design region \mathcal{X} (the set of candidate design points).

In all examples discussed in this report, the design region is defined as a compact set, but nonetheless we search for optimal points on a pre-defined discrete grid. This grid can be rather fine in order to guarantee that the resulting design is close to the optimal one.

To define a design algorithm, let $\xi_s = \{X_{is}, \lambda_{is}\}$, $i = 1, \dots, n_s$, be the design at Step s . Here $\{X_{is}\}$ is the vector of support points in the current design and w_{is} is the vector of weights assigned to X_{is} . The

iterative algorithm is of the following form:

$$\xi_{s+1} = (1 - \alpha_s)\xi_s + \alpha_s\xi(X_s),$$

where $\xi(X)$ is a one-point design supported on point X .

- **Forward step.** At Step s , a point x_s^+ that maximizes $\psi(x, \xi, \theta)$ over $x \in \mathcal{X}$ is added to the design ξ_s with weight $\alpha_s = \gamma_s$, where $\gamma_s = 1/(n_0 + s)$ and n_0 is the number of points in the initial design.
- **Backward step.** After that a point x_s^- that minimizes $\psi(x, \xi, \theta)$ over all support points in the current design is deleted from the design with weight

$$\alpha_s = \begin{cases} -\gamma_s, & \lambda_s \geq \gamma_s, \\ -\lambda_s/(1 - \lambda_s), & \lambda_s < \gamma_s. \end{cases}$$

In general, the user can change γ_s to $c_1/(n_0 + c_2s)$, where c_1 and c_2 are two constants. The default values of the constants are $c_1 = c_2 = 1$.

2.6. Penalty function. In practice, it would be unethical to assign patients to potentially non-efficacious doses (close to 0) or toxic doses (close to the high end of the design region). To incorporate such constraints in the design, one can introduce penalty function $\varphi(x, \theta) \mathcal{F}_p$. Note that the efficacy and toxicity constraints are just one example of the penalty function. In real problem, depending on the study goal, any function of the constraints can be defined as the penalty function([?], [?]). Some examples of the penalty function are:

- Restricted design region which is a special case of penalty function that all observations should be taken inside the design region, that the probability of toxicity is lower than a pre-specified value and the probability of efficacy is larger than a pre-specified value. Note that most of the time the design region depends on the unknown model parameters.
- Penalty function is based on toxicity and efficacy probabilities and controlled by tuning parameters.

The general optimization criterion Equation (??) becomes:

$$\xi^*(\theta) = \arg \min_{\xi} \Psi[M(\xi, \theta)/\mathcal{F}_{pn}(\xi, \theta)], \quad (3)$$

where \mathcal{F}_{pn} is the total normalized penalty

$$\mathcal{F}_{pn}(\xi, \theta) = \int_{\mathcal{X}} \varphi(x, \theta)\xi(dx). \quad (4)$$

For example, a design ξ^* is locally D-optimal if and only if its sensitivity function $\psi(x, \xi^*, \theta)$ satisfies the inequality:

$$\psi(x, \xi^*, \theta) / \varphi(x, \theta) = \text{tr}\{\mu(x, \theta)M^{-1}(\xi^*, \theta)\} / \varphi(x, \theta) \leq m / \mathcal{F}_{pn}(\xi^*, \theta), \quad (5)$$

3. LOCALLY OPTIMAL DESIGN AND ADAPTIVE DESIGN

Unless the model is linear with constant variance, the information matrix $M(\xi, \theta)$ depends on the unknown parameters θ . Therefore the optimal design ξ^* which satisfies (??) also depends on the unknown parameters θ . For a given value of θ , the corresponding optimal design is defined as the locally optimal design. Theoretically, one can construct a locally optimal design. But in practice, the practitioner is faced with the dilemma that the goal of the study is to estimate the quantities which are either a function of the unknown parameters or the unknown parameters themselves. Simultaneously, one has to know those unknown values in advance to obtain a reasonable design. To solve this problem, two-stage composite designs or adaptive designs will be reasonable choices, and are recommended for use. Nevertheless, locally optimal designs may be constructed as the benchmark for the adaptive designs. Usually the practitioner will have some idea what the unknown parameters might be, or a range of values that may contain the true parameters, and this information can be used to construct the locally optimal design as the initial design in the adaptive procedure.

3.1. Two-stage composite design. Two-stage composite design is a practical alternative for locally optimal designs because the two stages have different intentions. The goal of the initial design is to obtain preliminary parameter estimates. The second stage locally optimal design construction is based on these. The unknown parameters will be estimated twice in this type of design, once after the first stage for getting the preliminary parameter estimate, and once after all the observations are taken. As a result, there is only one data analysis needed in the middle of the trial which is easier to implement than fully adaptive design in practice. If needed, however, this design idea can be generalized to three or four stage designs ([?]).

The procedure of two-stage composite design is very similar to the fully adaptive design:

- (1) Identify the dose-response model and design region
- (2) Select a penalty function if necessary
- (3) Select the parameter(s) to be estimated (e.g. θ) and choose an optimization criterion
- (4) Build a locally optimal design (or a collection of them for different “true” parameter values to pick up the “robust” one), and assign a part of available subjects (N_1) to this design. If the

investigator has little evidence what the “true” parameters are, any “reasonable” design (for instance, uniform design) can be used.

- (5) Estimate the parameters $\hat{\theta}$ based on the initial design and construct the locally optimal design based on $\hat{\theta}$.
- (6) Assign the remaining N_2 subjects according to the locally optimal design.
- (7) Analysis data using both stage data.

For two-stage composite design, the original optimization problem should be modified to take into account all of the information obtained in the initial design, i.e. information Σ_0^{-1} and (??) becomes:

$$\xi^* = \arg \min_{\xi} \Psi[\Sigma_0^{-1} + N_2 M(\xi, \theta)]. \quad (6)$$

If penalty function is considered, (??) becomes:

$$\xi^* = \arg \min_{\xi} \Psi[\Sigma_0^{-1} + N_2 M(\xi, \theta)] / \mathcal{F}_{pn}(\xi, \theta). \quad (7)$$

and the inequality for the sensitivity function becomes

$$\text{Trace} \left\{ \frac{\mu(x, \hat{\theta})}{\varphi(x, \hat{\theta})} [\Sigma_0^{-1} + N_2 M(\xi, \hat{\theta})]^{-1} \right\} \leq \text{Trace} \left\{ \frac{M(\xi^*, \hat{\theta})}{\mathcal{F}_{pn}(\xi^*, \hat{\theta})} [\Sigma_0^{-1} + N_2 M(\xi, \hat{\theta})]^{-1} \right\}, \quad (8)$$

or if $\Sigma_0^{-1} = N_1 M(\xi_0)$, then

$$\text{Trace} \left\{ \frac{\mu(x, \hat{\theta})}{\varphi(x, \hat{\theta})} [wM(\xi_0, \hat{\theta}) + (1-w)M(\xi, \hat{\theta})]^{-1} \right\} \leq \text{Trace} \left\{ \frac{\mu(x, \hat{\theta})}{\varphi(x, \hat{\theta})} [wM(\xi_0, \hat{\theta}) + (1-w)M(\xi, \hat{\theta})]^{-1} \right\}, \quad (9)$$

where $w = N_1 / (N_1 + N_2)$.

3.2. Adaptive design. The idea of adaptive design is to constantly update the values of unknown parameters as the new observations are made, until the procedure converges or all the resources are consumed. For model-based adaptive design, the major steps can be summarized as follows:

- (1) Identified the dose-response model and design region
- (2) Select a penalty function if necessary
- (3) Select the parameter(s) to be estimated (e.g. θ) and choose an optimization criterion
- (4) Build a locally optimal design (or a collection of them for different “true” parameter values to pick up the “robust” one), and assign a part of available subjects (N_1) to this design. If the investigator has little evidence what the “true” parameters are, any “reasonable” design (for instance, uniform design) can be used.
- (5) Estimate the parameters based on the initial design and assign the next cohort of patient(s) to the most informative dose.

- (6) Update the parameter estimates and assign the next cohort of patients to the most informative dose given the current parameter estimation.
- (7) Repeat step 5 and 6 until all the resources are consumed.

In this procedure, the most informative dose is the dose which maximizes the sensitivity function, i.e.

$$x_s = \arg \max_{x \in \mathcal{X}} \psi(x, \xi_s, \hat{\theta}_s), \quad (10)$$

where ξ_s denotes the current design and $\hat{\theta}_s$ denotes the current parameter estimates. If penalty function need to be included, (??) should be modified as:

$$x_s = \arg \max_{x \in \mathcal{X}} [\psi(x, \xi_s, \hat{\theta}_s) / \varphi(x, \hat{\theta}_s)] \quad (11)$$

Actually, this procedure is the modification of the exchange algorithm that only contains the forward step.

4. IMPLEMENTATION IN R

There are several commercial software packages available for constructing optimal designs such as PROC OPTEX of the SAS/QC module, JMP, and Statistica. However, these are only available for D or C optimal criterion for simple linear regression model and factorial designs. In pharmacokinetic applications, the ADAPT II program developed at the University of Southern California implements *c*- and *D*-optimal and partially optimal design generation for individual pharmacokinetic models; however it does not support more complex population pharmacokinetic models (D'Argenio and Schumitzky, 1997). Ogungbenro et al. (2005) implemented in Matlab the classical exchange algorithm for population PK experiments using D-optimality (note that the exchange algorithm improves the initial design with respect to selected optimality criterion, but, in general, does not converge to the optimal design).

We developed a set of R programs that implement the optimal design algorithm which can be applied to different design criteria and various linear and non-linear models. Similarly, [?] developed a set of SAS programs and various examples for constructing locally optimal designs for different models.

4.1. Bivariate probit model. In clinical trial analysis, the toxicity and efficacy responses usually occur together and it may be useful to assess them together. However, in practice the assessment of toxicity and efficacy is sometimes separated. For example, determining the maximum tolerated dose (MTD) is based on toxicity alone and then efficacy is evaluated in Phase II trials over the predetermined dose range. Obviously, the two responses from the same patient are correlated which will introduce complexity into the analysis. But if one studies these outcomes simultaneously, more information will be gained for

future trials and treatment effects will be understood more thoroughly. In drug-response relationship, the correlation between efficacy and toxicity can be negative or positive depending on the therapeutical area. Two commonly used models, the Gumbel model (Kotz et al, 2000) and the bivariate binary Cox model (Cox, 1970), have been introduced to incorporate the two dependent outcomes, toxicity and efficacy, when both of them are dichotomous. In those two models, one needs to model the probabilities of different outcome combinations separately. When both outcomes are binary, the total number of combinations is four, but when outcomes contain more than two categories, the number of unknown parameters may increase dramatically. Here we propose a bivariate probit model which incorporates the correlated responses naturally via the correlation structure of the underlying bivariate normal distribution; see Lesaffre and Molenberghs (1991). When the number of responses is more than two, the multivariate probit model may be used in a similar fashion.

Let $Y = \{0 \text{ or } 1\}$ denote the efficacy response and $U = \{0 \text{ or } 1\}$ denote the toxicity response in a clinical trial. Here 0 indicates no response and 1 indicates a response. Let d denote the dose of a drug and

$$p_{yu}(x) = P(Y = y, U = u | d = x), \quad y, u = 0 \text{ or } 1.$$

Assume that Z_1 and Z_2 follow bivariate normal distribution with zero mean and variance-covariance matrix

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

where ρ may be interpreted as the correlation measure between toxicity and efficacy for the same patient. This set-up may be viewed as a standardization of the observed responses Y and U since under the natural scale the mean and variance vary from study to study. After simple transformation, the correlated responses follow the standard bivariate normal distribution,

$$p_{11} = F(\theta_1^T f_1, \theta_2^T f_2) = \int_{-\infty}^{\theta_1^T f_1} \int_{-\infty}^{\theta_2^T f_2} \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}Z^T \Sigma^{-1}Z\right) dz_1 dz_2 \quad (12)$$

$$p_{1.} = \Phi(\theta_1^T f_1) \quad (13)$$

$$p_{.1} = \Phi(\theta_2^T f_2), \quad (14)$$

where θ_1, θ_2 are unknown parameters and $f_1(x)$ and $f_2(x)$ contain the covariates of interest. Here we assume a simple linear model defined as follows

$$\theta_1^T f_1 = \theta_{11} + \theta_{12}x, \quad \theta_2^T f_2 = \theta_{21} + \theta_{22}x.$$

Since $f_1 = f_2 = (1, x)^T$, in the information matrix, f is used to denote f_1 and f_2 . $p_{1.}$ and $p_{.1}$ denote the efficacy and toxicity response rates, respectively, which can be expressed as the marginals of the bivariate normal distribution, where $\Phi(z)$ is the cumulative distribution function of the standard normal

distribution. Note that p_{11} , $p_{1\cdot}$ and $p_{\cdot 1}$ uniquely define the joint distribution of Y and U , i.e., $p_{10} = p_{1\cdot} - p_{11}$, $p_{01} = p_{\cdot 1} - p_{11}$ and $p_{00} = 1 - p_{1\cdot} - p_{\cdot 1} + p_{11}$.

Assume that the $\{y_i, u_i\}$'s are independent for different i 's. Then the likelihood function for $\{Y, U\}$ is given by

$$L(Y, U|\theta) = \prod_{i=1}^N y_i u_i \log p_{11} + y_i(1 - u_i) \log p_{10} + u_i(1 - y_i) \log p_{01} + (1 - y_i)(1 - u_i) \log p_{00}.$$

The information matrix of a single observation is

$$\begin{bmatrix} [C_1 C_2] \otimes f \\ \phi_2 & -\phi_2 & -\phi_2 \end{bmatrix} (P - pp^T)^{-1} \begin{bmatrix} [C_1 C_2] \otimes f \\ \phi_2 & -\phi_2 & -\phi_2 \end{bmatrix}^T,$$

where

$$C_1 = \begin{pmatrix} \phi(\theta_1^T f) & 0 \\ 0 & \phi(\theta_2^T f) \end{pmatrix}, C_2 = \begin{pmatrix} F(u_1) & 1 - F(u_1) & -F(u_1) \\ F(u_2) & -F(u_2) & 1 - F(u_2) \end{pmatrix},$$

$$u_1 = \frac{\theta_2^T f_2 - \rho \theta_1^T f_1}{\sqrt{1 - \rho^2}}, u_2 = \frac{\theta_1^T f_1 - \rho \theta_2^T f_2}{\sqrt{1 - \rho^2}},$$

$$P = \begin{pmatrix} p_{11} & 0 & 0 \\ 0 & p_{10} & 0 \\ 0 & 0 & p_{01} \end{pmatrix}, \quad p = (p_{11} \ p_{10} \ p_{01})^T,$$

$\phi(v)$ denotes the probability density function of the standard normal distribution, $\phi_2 = \phi(\theta_1^T f, \theta_2^T f, \rho)$ denotes the probability density function of bivariate normal distribution with mean $\theta_1^T f_1$ and $\theta_2^T f_2$, variance 1 and correlation coefficient ρ .

It will be unethical to assign patients to the non-eficacious dose or potentially toxic dose. To incorporate this concern into account, Dragalin and Fedorov (2005) proposed a rather flexible penalty function $\varphi(x, \theta)$:

$$\varphi(x, \theta; C_E, C_T) = \{p_{10}(x, \theta)\}^{-C_E} \{1 - p_{\cdot 1}(x, \theta)\}^{-C_T}. \quad (15)$$

This function of the dose and the model parameters penalizes more when the probability of success ($Y = 1, U = 0$) gets lower/or toxicity gets higher. The magnitude of penalty is controlled by the parameters C_E and C_T . When $C_T = 0$, the penalty is added only for the observations taken at the doses with low probability of success; when $C_E = 0$, the penalty is added only for the observations taken at the doses with high probability of toxicity.

Although model ?? will be used throughout the following sections and examples, the R programs can be easily modified to address other non-linear models.

This example is motivated by a dose-finding trial for prevention of cardioembolism in patients with atrial fibrillation or prosthetic heart valves. The efficacy endpoint for the new anticoagulant compound is NO venous thromboembolism (VTE) events and one can expect the toxicity endpoint is the major bleeding events. It is necessary to consider the two endpoints simultaneously and since both endpoints are binary, the bivariate probit model is the proper model for model the dose-response relationship. Here we only consider the simple model that the first order linear function of dose is included in the model.

4.2. Fully adaptive design and simulations. In this particular design scheme it is assumed that at every step a patient is added, and the values correlated from each response is regularly updated. This process is continual and cannot be reversed and therefore the ideal scenario is to implement the exchange algorithm, see (11). This program can best understood by using the same bivariate probit model.

5. REXCEL INTERFACE

This section is an illustration of how to obtain optimal designs for bivariate probit model through the use of the RExcel interface. We show the complete interaction between the user and the interface program.

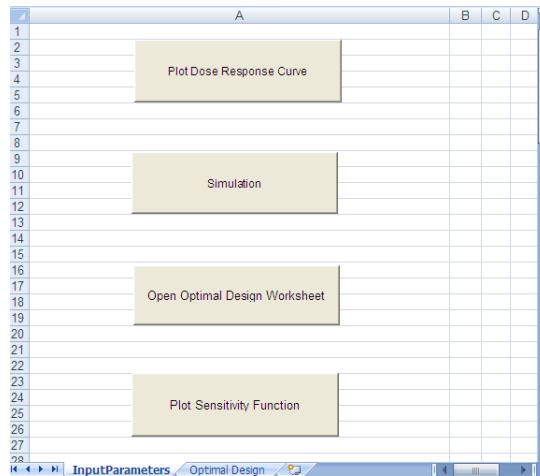


FIGURE 1. Screenshot:

Dose Response Curve

Show Defaults

Enter parameters for the bivariate probit model

t1

t2

t3

t4

t1

t2

t3

t4

Penalty Function

Ce

Ct

Plot Dose Response Curve

Close form

FIGURE 2. Screenshot:

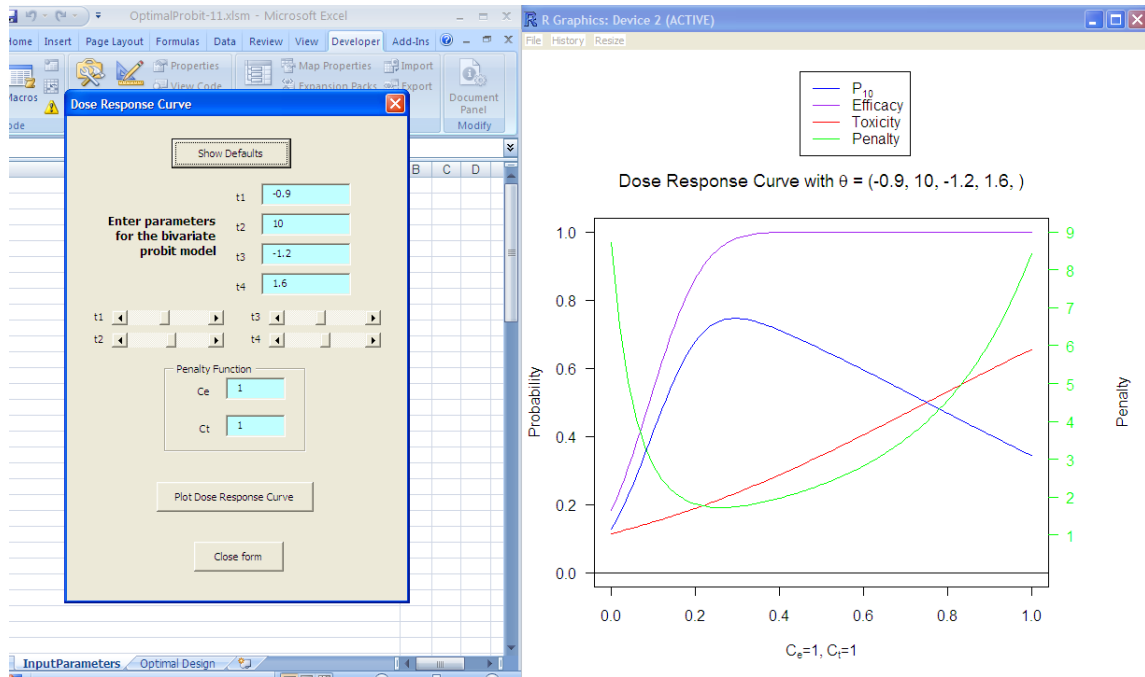


FIGURE 3. Screenshot:

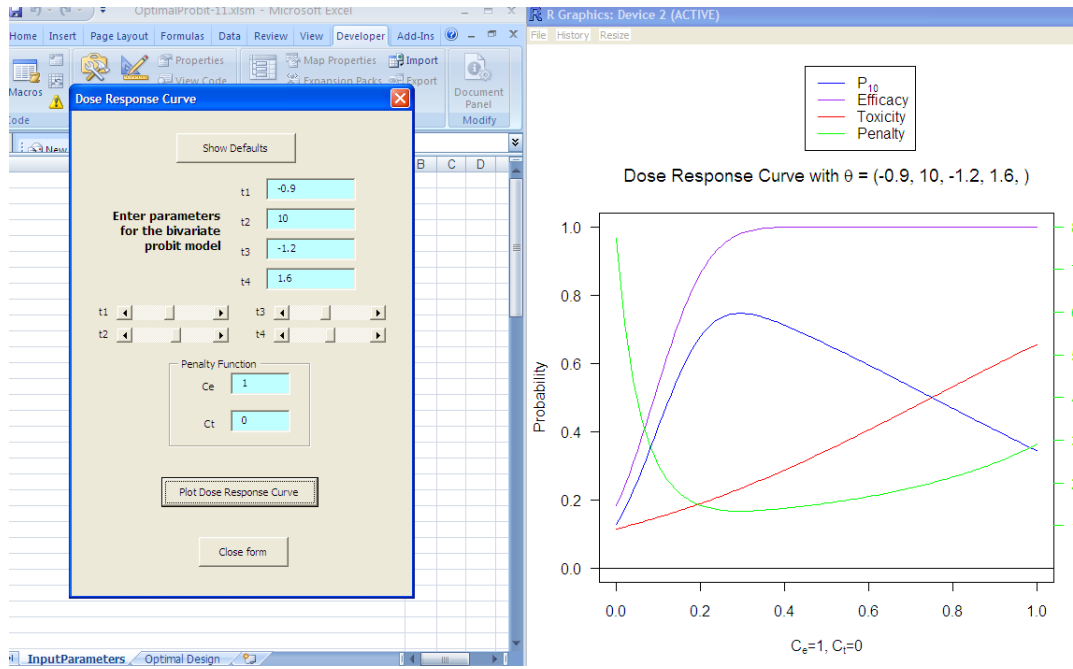


FIGURE 4. Screenshot:

Simulation

Show Defaults

Enter parameters for the bivariate probit model

t1: -0.9

t2: 10

t3: -1.2

t4: 1.6

Enter number of Monte Carlo simulations: 10

Enter number of observations: 100

Run Simulation & Plot Histogram

Close form

FIGURE 5. Screenshot:

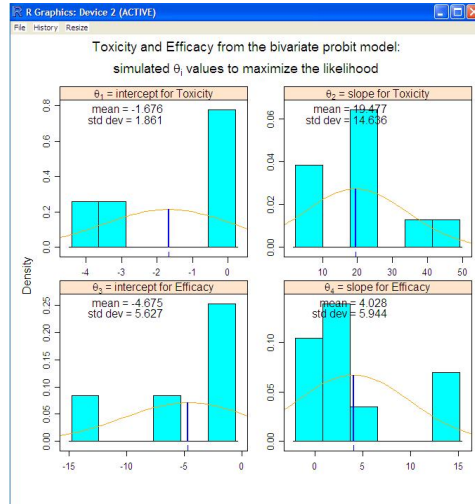
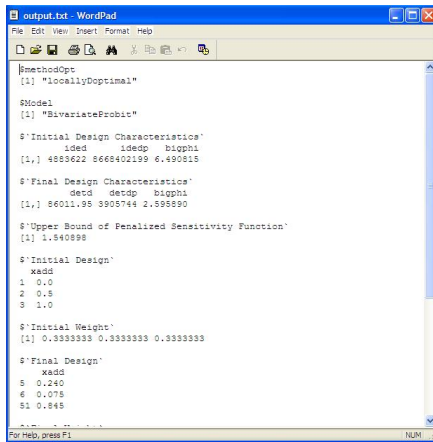


FIGURE 6. Screenshot:

	A	B	C	D	E	F	G
1	INPUT	PARAMETERS					
2							
3	inputk	1	Copy Default Parameter Values from Initial Values				
4	t11	-0.9					
5	t12	1.6					
6	t21	-1.2					
7	t22	1.6					
8							
9	initial	2					
10	method	1					
11	sample	3					
12	shape	1					
13	a1	0					
14	a2	1					
15	candidn	201					
16	lin	2					
17	convic	1E-10					
18	maximit	1000					
19	const1	1					
20	const2	4					
21	cmarg	4					
22	pmax	0.7					
23	rho	0.5					
24	Ce	1					
25	Ct	1					
26							
27	initialDose	0	0.2	1			
28	initialWeight	0.4	0.4	0.2			
29	nobs	100					
30	k	100					
31							
32	methodOpt	local,Optimal					
33							
34							

FIGURE 7. Screenshot:



```
output.txt - WordPad
File Edit View Insert Format Help
[MethodOpt]
[1] "locallyOptimal"
$Model
[1] "BivariateProbit"
$Initial Design Characteristics
      ided      idedp      bigph1
[1,] 4883622 8668402199 6.490815
$Final Design Characteristics
      deds      detsp      bigph1
[1,] 86211.98 3993744 2.598890
$Upper Bound of Penalized Sensitivity Function
[1] 1.540898
$Initial Design
      xads
1      0.0
2      0.5
3      1.0
$Initial Weights
[1] 0.3333333 0.3333333 0.3333333
$Final Design
      xads
5      0.240
6      0.075
51     0.845
For help, press F1
```

FIGURE 9. Screenshot:

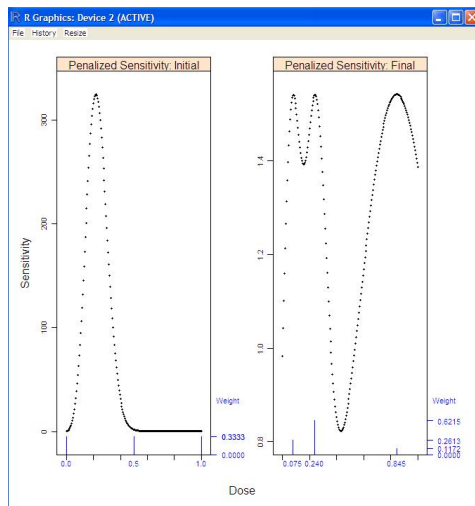


FIGURE 10. Screenshot:

6. CONCLUSION

In this article, we have considered construction of optimal designs, namely, locally optimal designs, $L(\theta)$ -optimal, two stage composite designs and the fully adaptive designs for the bivariate probit model. One can easily change or modify the R programs in order to construct optimal designs pertaining to other non-linear response models.

APPENDIX A. INSTALLATION OF SOFTWARE

The OptimalDesigns toolbox consists of a workbook written for Microsoft Excel on Windows that communicates through the RExcel interface with a package written in R. The *Download and Installation* section gives detailed instructions on

- (1) downloading R with RExcel and related packages included. The instructions assume that you have Excel on your Windows computer, but do not yet have R on your computer.
- (2) downloading and installing the OptimalDesigns toolbox.

A.1. Download and Installation.

R and RExcel. Although R by itself can be installed on any computer, RExcel requires administrator privileges for installation. Go to the RExcel website <http://rcom.univie.ac.at>, click the Download tab, and download the RAndFriendsSetup* installer file. This is a Windows executable .exe file that installs R and several related packages. Run the installer from an account with administrator privileges. Accept

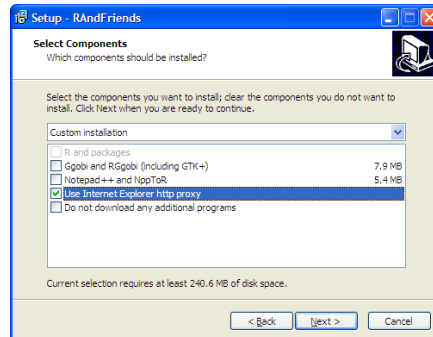


FIGURE 11. You *must* check the “Use Internet Explorer http proxy” checkbox when installing RAndFriends.

almost all defaults. The exception is that you *must* check the checkbox

Use Internet Explorer http proxy

on the installation dialog box as in Figure ???. If you forget, then you will not be able to get through the GSK firewall, you will get messages like "Unable to execute file c:\...\RExcelInst.latest.exe", and you will need to cancel the installation and start over.

OptimalDesigns Package and Workbook. The *OptimalDesigns Package and Workbook* are provided as a single ZIP file containing three files. Download the ZIP file to a directory on your computer. Unzip it and it will create an *OptimalDesigns* subdirectory with three files.

`OptimalDesigns_version.zip`

`OptimalDesigns-version.xlsm`

`OptimalDesigns_version.tar.gz`

Install the *OptimalDesigns* by starting R from the R icon. In the R Console window, click

Package > Install package(s) from local Zip files...

Navigate the Select Files window to the *OptimalDesigns* directory and double-click on

`OptimalDesigns_version.zip`

Close R with the File > Exit > No menu item.

The software is now installed and ready to use.

A.2. Using the *OptimalDesigns Toolbox*.

Open Windows Explorer to the *OptimalDesigns* directory and double-click

`OptimalDesigns-version.xlsm` for Excel 2007

The file will open in Excel and start R.

Questions. If you need additional help, please contact Sourish Saha sourish.c.saha@gsk.com and include two necessary pieces of information.

- (1) The information that you get from clicking About RExcel.

In Excel 2007, click on the main Excel menu

Add-Ins > RExcel > About RExcel > Copy to Clipboard

Paste that information into the email.

- (2) When RExcel is running, the R Console is visible on the Windows Taskbar.

Click the R Console icon and enter the line

```
packageDescription("OptimalDesigns")
```

into the R Console. Copy and paste the lines it displays into the email.

R. R is a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the R project homepage: <http://www.r-project.org> for further information.

The current version of R (R-2.12.0) was released in November, 2010. The OptimalDesigns software may not work with earlier releases of R.

RExcel. RExcel is an Excel add-in using `statconn` (D)COM or `rcom` to allow Excel to call R from within Excel. The RExcel website is <http://rcom.univie.ac.at>. Much detail is available at the Wiki there.

REFERENCES

- [1] Atkinson, A.C. and Donev, A. *Optimum Experimental Designs*, 1992, Oxford: Clarendon Press.
- [2] Ashford, J.R. and Sowden, R. R. (1970). Multi-variate Probit Analysis, *Biometrics*, **26**: 535-546.
- [3] Dragalin, V. and Fedorov, V. (2006). Adaptive designs for dose-finding based on efficacy-toxicity response. *Journal of Statistical Planning and Inference*, **136**: 1800-1823.
- [4] Dragalin, V., Fedorov, V., and Wu, Y. (2005). Optimal Designs for Bivariate Probit Model. *GSK Technical Report 2005-07*. 62 pages.
- [5] Dragalin, V., Fedorov, V., and Wu, Y. (2006). Adaptive designs for selecting drug combinations based on efficacy-toxicity response. Accepted by *Journal of Statistical Planning and Inference*.
- [6] Fedorov V, Gagnon R, Leonov S, Wu Y. (2007) Optimal Design of Experiments in Pharmaceutical Applications. In: Dmitrienko A, Chuang-Stein C, D'Agostino R, (Eds), *Pharmaceutical Statistics Using SAS: A Practical Guide*, SAS Books by Users series, SAS Press, 151195.
- [7] Fedorov, V.V. *Theory of Optimal Experiments*, 1972, New York: Academic Press.
- [8] Fedorov, V.V. and Hackl, P. *Model-Oriented Design of Experiments*; Lecture Notes in Statistics 125; Springer-Verlag, New York, 1997.

- [9] Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics* **12**: 363-366.
- [10] Lesaffre, E. and Molenberghs, G. (1991). Multivariate Probit Analysis: A Neglected Procedure in Medical Statistics. *Statistics in Medicine*, **10**: 1391-1403.
- [11] Baier, T. and Neuwirth, E. (2007) Excel :: Com :: R. *Computational Statistics*, 22 (1): 91–108.
- [12] Microsoft (2002–2007). Microsoft Office Excel.
- [13] Neuwirth, E., with contributions by Heiberger, R., Ritter, C., Pieterse, J., and Volkering, J. (2009) *RExcelInstaller: Integration of R and Excel, (use R in Excel, read/write XLS files)*. R package version 3.0-12.
- [14] R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.