Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Zheng Li Date

Dose Escalation with Over-dose and Under-dose Controls for Phase I/II Clinical Trials

By

Zheng Li

Master of Science in Public Health

Emory University

Rollins School of Public Health

Department of Biostatistics and Bioinformatics

_____

Thesis Advisor

_____

Reader

Dose Escalation with Over-dose and Under-dose Controls for Phase I/II Clinical Trials

By

Zheng Li

Bachelor of Science

Shanghai University of Finance and Economics

2011

Advisor: Zhengjia Chen, Ph.D.

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health

in Department of Biostatistics and Bioinformatics

2014

Abstract

Dose Escalation with Over-dose and Under-dose Controls for Phase I/II Clinical Trials

By: Zheng Li

To save time and resources in new drug development, Phase I/II clinical trials with toxicity response and drug efficacy as dual primary endpoints have become increasingly popular. Escalation with over dose control (EWOC) is a leading Bayesian adaptive Phase I clinical trial design that can accurately estimate maximum tolerated dose (MTD) and control the probability of overdosing patients during dose allocation. To adapt EWOC to Phase I/II clinical trials, we have used it as a framework to incorporate the Gumbel Copula model and additional under-dosing control to guarantee minimum drug efficacy for patients. A utility function is proposed to estimate the composite effect of toxicity and efficacy. The final recommended dose is determined by maximizing this utility. Late onset and missing efficacy data are common in Phase I/II clinical trials, especially during early stages when patients are treated at low doses. Therefore, we further employed the Bayesian data augmentation (DA) algorithm to impute values for late onset or delayed efficacy data. The resulting new Phase I/II design, which can monitor toxicity and efficacy simultaneously, is named Dose Escalation with Over-Dose and Under-Dose Control using Data Augmentation (EWOUC-DA). The underlying theory of EWOUC-DA is elaborated and extensive simulations are conducted to evaluate its performance and operating characteristics. EWOUC-DA has been demonstrated to provide better over-dose control, optimize utility, and reduce the risk of failure in Phase III clinical trials compared to EWOC. Hence, EWOUC-DA can address the new requisites of Phase I/II clinical trials, reduce the cost of clinical trials, and largely shorten the duration of new drug development.

Dose Escalation with Over-dose and Under-dose Controls for Phase I/II Clinical Trials

By

Zheng Li

Bachelor of Science

Shanghai University of Finance and Economics

2011

Advisor: Zhengjia Chen, Ph.D.

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health

in Department of Biostatistics and Bioinformatics

2014

Acknowledgements

I want to thank the faculty, advisors, and staff of the Biostatistics Department at Rollins School of Public Health for the dynamic two years of learning that I have had. This thesis is only a sample of the vast knowledge that was attained and applied through my two years here at Rollins. I would especially like to thank Dr. Zhengjia Chen for all of his advice and support to help me write this thesis. Also a special thanks to Dr. Limin Peng for taking the time to read my thesis.

Lastly I want to give a special thanks to my parents for their loving support and encouragement that motivated me to pursue a degree at the graduate level.

Table of Contents

1 Introduction

The development of new drugs consists of two stages: pre-clinical studies and clinical trials. Clinical trials are further classified into 4 phases: I, II, III, and IV. A Phase I clinical trial is the first trial of an investigational agent in human beings, and its main goal is to find the maximum tolerated dose (MTD) at which the probability of dose limiting toxicity (DLT) is equal or close to the target toxicity level (TTL, i.e., 33%). The primary purpose of subsequent Phase II clinical trials is to screen promising drugs with preliminary evidence of efficacy. To shorten the duration of trial and save resources in new drug discovery, the combination of Phase I and II clinical trials into one Phase I/II clinical trial with toxicity response and drug efficacy as dual primary endpoints has become increasingly popular.

Adaptive Bayesian methods are cutting edge designs for early Phase clinical trials. Some novel Bayesian designs have also been proposed for Phase I/II clinical trials. For example, Thall and Cook (2004) proposed a continual reassessment method (CRM)-based phase I/II clinical trial design, using the trade-off curve drawn by the accepted toxicity and efficacy probability combination. The CRM was first proposed by O'Quigley, Pepe, and Fisher (1990) and has been proven to be more accurate in estimating MTD compared with traditional rule-based designs, such as the 3+3 design.

Babb, Rogatko, and Zacks (1998) proposed the escalation with overdose control (EWOC) Phase I design to control the probability of over-dosing patients and the risk of over-toxic doses. They further proved analytically that treating patients with the optimal Bayesian-feasible dose sequence minimizes the expected risk of over-dosing (doses higher than MTD) and if sample is large enough, the optimal Bayesian-feasible dose converges in probability to the true MTD. Several research advances using EWOC have been achieved in recent years. For example, Babb and Ro-

gatko (2001) extended the EWOC method by incorporating covariates in the model so that it can estimate individual MTDs for different patient subgroups. Chu et al. (2009) proposed a hybrid method, which varies the feasibility bound in EWOC design. Chen et al. (2010) proposed a novel normalized equivalent toxicity score (NETS) system and integrated NETS with EWOC to develop an extended design called EWOC-NETS, which can improve the accuracy of EWOC estimation. Mauguen, Deley and Zohar (2011) proposed another extended version of EWOC called TITE-EWOC to allow staggered patient enrollment and utilize partially completed toxicity data in new dose determination.

While providing considerable advantages, none of these different versions of EWOC design consider efficacy response as a primary outcome, with the potential consequence that too many patients may be treated at dose levels without therapeutic effect. In this thesis, we use EWOC as a framework and implement additionally an under-dose control to optimize the therapeutic effect for patients. In an ideal Phase I/II clinical trial design with toxicity and efficacy as dual endpoints, it is desirable to not only implement over-dose control in order to protect patients from being exposed to an over-toxic dose, but also to conduct under-dose control to optimize the potential therapeutic effect for patients. The new design is named Escalation with Over-dose and Under-dose Controls (EWOUC). The dose efficacy rate increases with elevated dose level, so that MTD is the target dose of Phase I/II trials if we want to maximize the efficacy of a new drug. However, toxicity also increases with elevated dose level and can lead to serious or even fatal side effects. Therefore, we further proposed a utility function to define therapeutic effect, balancing the composite effect of toxicity and efficacy. The final recommended dose level in our design is the one which will optimize the proposed utility under safe administration. EWOUC can address the needs of phase I/II clinical trials, and has the advantage over former CRM-based Phase I/II designs by providing

over-dose control in that the risk of over-dosing patients is decreased while an aggressive algorithm is employed to improve therapeutic effect for patients.

Another barrier to the success of Phase I/II clinical trials is missing efficacy data at the time of new dose assignment. Missing efficacy data can occur due to delayed efficacy or non-efficacy response. Missing efficacy data can be called non-ignorable missing data because the probability of missing-ness of non-efficacy response is higher than the probability of delayed efficacy. Clearly, in contrast to delayed efficacy response, non-efficacy response remains missing until the assessment period end. Therefore, these two types of missing response values should be treated differently. In order to assign doses to next patient under missing efficacy data, we adopted the data augmentation (DA) approach proposed by Liu, Yin, and Yuan (2013) under EWOUC framework, to develop a derived version called EWOUC-DA. Instead of the surrogate response, the DA algorithm is used to solve the problem of missing efficacy data in that 1) the relationship between surrogate response and complete response should be validated; 2) there might be no surrogate response for some diseases. EWOUC-DA can retain all the advantages of EWOUC and meanwhile, improve trial efficiency over EWOUC without sacrificing time to wait for complete data collection.

We have organized the remainder of the thesis as follows. In Section 2, we introduce the study design, model selection, and data augmentation algorithms. In Section 3, extensive simulations are presented to evaluate the performance of EWOUC and EWOUC-DA. Finally, in Section 4, we summarize the operating characteristics of our new designs and discuss some future research topics.

2 Methods

2.1 Outcome Model and Re-parameterization

Let Y_E and Y_T respectively denote the toxicity and efficacy outcomes with $Y_E = 1$ if the patient experiences efficacy and $Y_T = 1$ if the patient experiences DLT events. We use the Farlie-Gumbel-Morgenstern Copula model (Murtaugh and Fisher, 1990) to model the joint distribution of Y_E and Y_T. The copula model is chosen because under this model, the joint distribution of Y_E and Y_T can be conveniently expressed as a function of their marginal distributions. Specifically, we assume that the marginal distributions of Y_E and Y_T follow logistic regression models as (1) and (2),

$$\pi_T = P(Y_T = 1|X = x) = \frac{1}{1 + exp\{-(\beta_{0,T} + \beta_{1,T}x)\}} \tag{1}$$

$$\pi_E = P(Y_E = 1|X = x) = \frac{1}{1 + exp\{-(\beta_{0,E} + \beta_{1,E}x)\}} \tag{2}$$

where X is the dose level, and $\beta_{0,T}, \beta_{1,T}, \beta_{0,E}, \beta_{1,E}$ are unknown regression parameters. Define $\pi_{ab} = P(Y_T = b, Y_E = a|X = x), a, b \in \{0, 1\}$. Under the Farlie-Gumbel-Morgenstern Copula model (Murtaugh and Fisher, 1990), the joint distribution of Y_E and Y_T conditional on dose X is given by

$$\pi_{11} = \pi_T\pi_E\{1 + \frac{e^{\varphi} - 1}{e^{\varphi} + 1}(1 - \pi_T)(1 - \pi_E)\} \tag{3}$$

and $\pi_{10} = \pi_T - \pi_{11}, \pi_{01} = \pi_E - \pi_{11}, \pi_{00} = 1 - \pi_T - \pi_E - \pi_{11}$.

The marginal logistic models (1) and (2) are familiar to statisticians and practitioners, however parameters $(\beta_{0,T}, \beta_{1,T}, \beta_{0,E}, \beta_{1,E})$ do not possess the interpretations that are directly related to the target of dose finding. Following the approach of the EWOC, we re-parameterize $(\beta_{0,T}, \beta_{1,T}, \beta_{0,E}, \beta_{1,E})$ with new parameters $(\gamma_T, \gamma_E, \rho_T, \rho_E)$, which have more intuitive clinical interpretations. This re-parameterization facilitates the implementation of the over-dose and under-dose control described

later. Specifically, we define γ_T as the MTD, and γ_E as the minimum efficacy dose (MED), which is defined as the lowest dose that satisfies a certain efficacy requirement (i.e., the efficacy lower bound, θ_E). The dose below the MED is deemed clinically futile due to the lack of efficacy. We define ρ_T and ρ_E respectively as the probability of DLT and the probability of efficacy events when the patient is treated at a minimum dose X_{min} under investigation. Mathematically, the dose response curve should satisfy (4) and (5).

$$\begin{cases} logit(\rho_T) = \beta_{0,T} + \beta_{1,T} X_{min} \\ logit(\theta_T) = \beta_{0,T} + \beta_{1,T} \gamma_T \end{cases} \tag{4}$$

$$\begin{cases} logit(\rho_E) = \beta_{0,E} + \beta_{1,E} X_{min} \\ logit(\theta_E) = \beta_{0,E} + \beta_{1,E} \gamma_E \end{cases} \tag{5}$$

The original parameters$(\beta_{0,T}, \beta_{1,T}, \beta_{0,E}, \beta_{1,E})$ are reparameterized in terms of $(\gamma_T, \gamma_E, \rho_T, \rho_E)$ by

$$\beta_{0,j}(\gamma_j, \rho_j) = \frac{1}{\gamma_j - X_{min}} (\gamma_j logit(\rho_j) - X_{min} logit(\theta_j)) \tag{6}$$

$$\beta_{1,j}(\gamma_j, \rho_j) = \frac{1}{\gamma_j - X_{min}} (logit(\theta_j) - logit(\rho_j)), \tag{7}$$

where $j = \{T, E\}$. The reparameterization is a one-to-one transformation which means that each combination of $(\beta_{0,T}, \beta_{1,T}, \beta_{0,E}, \beta_{1,E})$ can only be derived by unique $(\gamma_T, \gamma_E, \rho_T, \rho_E)$ and vice versa.

Let $Y_{ab,i} = 1$ if $Y_{E,i} = a$, and $Y_{T,i} = b$, $a, b \in \{0, 1\}$ for the ith patient. Suppose there are totally N patients in the trial. The likelihood function under the copula model is given by equation

(8),

$$L(\mathbf{V}) = \prod_{i=1}^{N} \prod_{a=0}^{1} \prod_{b=0}^{1} \pi_{ab,i}^{y_{ab,i}}, \tag{8}$$

where $\mathbf{V} = (\rho_T, \rho_E, \gamma_T, \gamma_E, \varphi)$. Let $p(\mathbf{V})$ denote the prior distribution of $\mathbf{V}$, the posterior distribution of $\mathbf{V}$ can be expressed by equation (9), where $\mathbf{D}_N = \{(y_{E,1}, y_{T,1}), ..., (y_{E,N}, y_{T,N})\}$ is the observed outcome.

$$\begin{aligned} p(\mathbf{V}|\mathbf{D}_N) &\propto L(\mathbf{V}) p(\mathbf{V}) \\ &= p(\mathbf{V}) \prod_{i=1}^{N} \prod_{a=0}^{1} \prod_{b=0}^{1} \pi_{ab,i}^{y_{ab,i}} \end{aligned} \tag{9}$$

We assign ρ_T and ρ_E uniform prior distributions in the interval $[0, \theta_T]$ and $[0, \theta_E + \delta]$, where θ_T is the target toxicity level; θ_E is the efficacy lower bound; δ is a small positive value so that patients have higher probability to be treated at a dose above the efficacy lower bound. γ_T and γ_E are assigned uniform prior distributions in the interval $[X_{min}, X_{max}]$, where X_{max} is the maximum dose under investigation or determined by pre-clinical studies. All the priors of parameters in the model are supposed to be mutually independent. We sample the posterior distribution of $\mathbf{V}$ using Gibbs sampler implemented by JAGS.

2.2 Over-dose Control and Under-dose Control

When treating patients, for their benefits, we should avoid exposing patients to the overly toxic dose (i.e., doses above the MTD γ_T) and to a low dose that is deemed therapeutically futile (i.e., doses below the MED γ_E). Ideally, we want to treat patients in doses between (γ_E, γ_T). The relationship between true MTD and true MED should be considered to decide whether a drug can be approved for further study. If the MTD is lower than the MED, we should terminate the trial without selecting any investigational dose for further study because none of the doses are both

safe and efficacious. On the other hand, if MTD is higher than MED, i.e., there is an interval between MED and MTD, we should treat patients with a dose within that interval. This is the basic idea behind the over-dose and under-dose control, which can be formally described as follows. Suppose a total of n patients have been treated in the trial. The observed outcome data $\mathbf{D}_n$ is composed of $\{(y_{E,1}, y_{T,1}), (y_{E,2}, y_{T,2}), ..., (y_{E,n}, y_{T,n})\}$, where $(y_{T,n}, y_{E,n})$ is the nth patient's DLT status and efficacy status. To select a dose x_{n+1} for the incoming (n+1)th patient, we require that x_{n+1} satisfies both the over-dose control condition (10),

$$P(x_{n+1} \geq \gamma_T | \mathbf{D}_n) \leq \alpha_T, \tag{10}$$

and the under-dose control condition (11),

$$P(x_{n+1} \leq \gamma_E | \mathbf{D}_n) \leq \alpha_E \tag{11}$$

where α_T and α_E are respectively called feasibility bounds for toxicity and efficacy. Under these two conditions, the probability of over dosing is less than α_T and the probability of under dosing is less than α_E for the (n+1)th patient based on the observed data. α_T is increased from 0.25 by 0.05 until 0.5 in the trial when we assign dose to the next cohort. On the other hand, α_E is decreased from 0.75 by 0.05 until 0.5 when we assign dose to the next cohort.

At the beginning of the trial, we are not quite confident about the MTD and MED because of the limited information so that the dose with lower probability to be greater than MTD and higher probability to be less than MED is chosen such that more dose levels are included into the accepted dose set. As the trial continues, more data about the MTD and MED are accumulated; hence the posterior median is used to make inference about the true interval and the accepted dose set will be shrunk. Besides, varying feasibility bounds can improve the speed for the posterior estimators of MTD and MED converge to the true MTD and MED (See Chu et al. 2009). When there are

multiple doses in the accepted dose set, we further proposed to choose the dose with the best utility in terms of the toxicity and efficacy tradeoff for treating the (n+1)th patient. The utility function conditional on a dose x is defined as $U(x) = \pi_E(x) - W\pi_T(x)$, where W is a positive number, representing the weight or penalty induced by dose toxicity. W could be 2 for moderate toxic agents so that the trial will not be terminated early. And W could be 3 or 4 for those extremely good or good agents. The higher W is, more favorable the final recommendation will be to the dose level with high efficacy and low toxicity.

The dose-finding algorithm for the proposed design can be summarized as follows:

1. Treat the first cohort at the lowest dose, $d_1 = x_1$. If two patients in the first cohort show DLT status, the clinical trial should stop because it is even too toxic for the minimum dose.

2. Based on the cumulative data $\mathbf{D}_i$, update the posterior distribution $p(\rho_T, \rho_E, \gamma_T, \gamma_E, \varphi|\mathbf{D}_i)$, and assign the next cohort of patients to the dose that satisfies both over-dose and under-dose control conditions (10) and (11) with the highest utility. Utility is estimated by $\hat{U}(x) = E\{U(x)|\mathbf{D}_i\}$.

3. If no dose level satisfy conditions (10), (11) and $P(\gamma_T - \gamma_E > 0|\mathbf{D}_i) < 0.25$, stop the trial. Otherwise, we continue to treat the patient at MTD for under-dose control, assuming that MTD is the most efficacious dose.

4. Repeat steps 1 to 3 until the maximum sample size is reached. Select the dose with the highest utility as the final recommended dose.

2.3 Accommodate Delayed Efficacy

The design proposed above requires that we wait until toxicity and efficacy are observable such that the adaptive dose-finding rule can be used to assign a dose to new patients. However, in many cases, the efficacy outcome is not immediately ascertainable and takes a relatively long time to evaluate, resulting in the so-called delayed efficacy. We assume that toxicity is quickly ascertainable, as often the case in practice for cytotoxic agents. A direct consequence of delayed efficacy is that at the moment of decision making for dose assignment, some patients who have enrolled into the trial might have not finished their time to follow up yet, and thus, their efficacy outcomes are missing. Liu, Yin and Yuan (2013) showed that such missing outcomes are non-ignorable and proposed a Bayesian data augmentation approach to account for these missing data. In this section, by taking the approach of Liu et al. (2013), we extend our methodology to accommodate delayed efficacy. The problem we tackle here is more complicated than the problem proposed by Liu et al. (2013) because we concern the bivariate distribution of toxicity and efficacy, while Liu et al. focus on univariate delayed toxicity only. Although we assume that toxicity is immediately observed and only efficacy is delayed, because of the bivariate structure of the data, the observed toxicity outcome affects the imputation (or data argument) of the missing efficacy outcome, as we show below.

Let T denote follow-up window (or time frame) for assessing efficacy, such that $Y_{E,i} = 1$ if the ith patient experiences efficacy events in (0, T), otherwise $Y_{E,i} = 0$. Let r_i denote the time to efficacy, u_i denote the actual follow-up time at the moment that we need to make the decision of dose assignment for a new patient. Clearly, if $u_i < T$ and $u_i < r_i$ (i.e., the patient has not finished the follow-up assessment and the actual follow-up time is shorter than the time to efficacy), $Y_{E,i}$ is missing. We denote M_i as a missing indicator for $Y_{E,i}$ with $M_i = 1$ indicating

that $Y_{E,i}$ is missing. Following Liu et al (2013), we specify a piecewise exponential model for r_i. We partition the follow-up window $(0, T]$ into K intervals. $[0, h_1), , [h_{K-1}, h_K]$ $(h_K = T)$, and assume a constant hazard λ_k for the kth interval. Define $t_i = min(r_i, u_i)$ as the observed time to event and $\boldsymbol{t} = (t_1, ..., t_n)$ for n treated patients. We assume that the time-to-efficacy distribution is invariant to the dose level. The likelihood function of $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_K)$ conditional on efficacy response based on completed data from n treated patients is given by

$$L(\boldsymbol{\lambda}) = P(\mathbf{t}|\boldsymbol{\lambda}) = \prod_{i=1}^{n} \prod_{k=1}^{K} \lambda_k^{\delta_{ik}} e^{-y_{E,i}\lambda_k s_{ik}} \tag{12}$$

where $s_{ik} = h_k - h_{k-1}$, if $t_i > h_k$, and $t_i - h_{k-1}$, if $h_{k-1} \leq t_i < h_k$, otherwise 0. δ_{ik} is defined as 1 if the ith patient experienced efficacy event in $[h_{k-1}, h_k)$ or 0 otherwise.

Our strategy for handling delayed outcomes is to impute the missing values of y_E, thereby converting the missing problem into a standard complete-data problem. This is completed by using Bayesian data augmentation algorithm, which consists of two steps, namely, imputation (I) step and posterior (P) step. In the I step, we impute the missing values of $y_{E,i}$ conditional on model parameters; and in the P step, we draw samples from the posterior distribution of model parameters conditional on the imputed complete data. Specifically, at the I step, conditioning on $\boldsymbol{\theta} = (\rho_T, \rho_E, \gamma_T, \gamma_E, \varphi, \boldsymbol{\lambda})$ and the observed toxicity outcome $y_{T,i} = b$, where $b \in (0, 1)$, we draw the missing value of $y_{E,i}$ from its posterior predictive distribution (13),

$$f(y_{E,i}|\boldsymbol{\theta}, M_i = 1, Y_{T,i} = b, \boldsymbol{t}) = Bernoulli(\phi) \tag{13}$$

where

$$\begin{aligned}
\phi &= p(y_{E,i} = 1|\boldsymbol{\theta}, M_i = 1, Y_{T,i} = b) \\
&= \frac{p(y_{E,i} = 1, M_i = 1, Y_{T,i} = b|\boldsymbol{\theta})}{p(M_i = 1, Y_{T,i} = b|\boldsymbol{\theta})} \\
&= \frac{p(Y_{E,i} = 1, Y_{T,i} = b|\boldsymbol{\theta})p(M_i = 1|Y_{E,i} = 1, Y_{T,i} = b, \boldsymbol{\theta})}{\sum_{a=0}^{1} p(M_i = 1, Y_{T,i} = b, Y_{E,i} = a|\boldsymbol{\theta})} \\
&= \frac{\pi_{1b,i}p(M_i = 1|Y_{E,i} = 1, Y_{T,i} = b, \boldsymbol{\theta})}{\sum_{a=0}^{1} p(Y_{E,i} = a, Y_{T,i} = b|\theta)p(M_i = 1|Y_{E,i} = a, Y_{T,i} = b, \boldsymbol{\theta})}
\end{aligned} \tag{14}$$

Following Liu et al. (2013), we assume conditional independence $p(M_i = 1|Y_{T,i}, Y_{E,i} = a) = p(M_i = 1|Y_{E,i} = a)$, it follows that

$$\begin{aligned}
\phi &= \frac{\pi_{1b,i}p(M_i = 1|Y_{E,i} = 1, \boldsymbol{\theta})}{\sum_{a=0}^{1} \pi_{ab,i}p(M_i = 1|Y_{E,i} = a, \boldsymbol{\theta})} \\
&= \frac{\pi_{1b,i}p(r_i > u_i|Y_{E,i} = 1, \boldsymbol{\theta})}{\sum_{a=0}^{1} \pi_{ab,i}p(r_i > u_i|Y_{E,i} = a, \boldsymbol{\theta})} \\
&= \frac{\pi_{1b,i}exp(-\sum_{k=1}^{K} \lambda_k s_{ik})}{\pi_{1b,i}exp(-\sum_{k=1}^{K} \lambda_k s_{ik}) + 1 - \pi_{0b,i}}
\end{aligned} \tag{15}$$

Note that the posterior predictive distribution of the missing efficacy outcome $Y_{E,i}$ depends on the observed toxicity outcome $Y_{T,i} = b$ through $\pi_{1b,i}$. After imputing the missing data, at the P step, we essentially handle a standard complete-data problem, and sample $\boldsymbol{\theta}$ from its posterior distribution (conditional on the imputed complete data). A Gibbs sampler of the missing outcomes and

parameters is formed by iteratively sampling between the I and P steps. We iterate the Gibbs sampling algorithm until the MCMC chain converges. The rate of convergence of the Gibbs sampler is controlled by the maximal correlation between the states of two consecutive Gibbs iteration.

3 Simulation

3.1 Simulation Plans and Model Setting

In order to evaluate the performance of EWOUC and its combination with the Bayesian data augmentation algorithm, we use five different scenarios, corresponding to different kinds of agents which may be tested in the real clinical trial. Figure 2 shows a vivid picture for the dose-response curve for these scenarios. Scenario 1 (S1) is corresponding to an extremely good agent, with a steep increasing dose efficacy curve and a flat dose toxicity curve. Scenario 2 (S2), Scenario 3 (S3), Scenario 4 (S4) and Scenario 5 (S5) are respectively corresponding to a good agent, a moderate agent, a bad agent and an extremely bad agent. Five dose levels, (0.2, 0.4, 0.6 ,0.8, 1.0), are pre-specified in our simulation. Along with EWOUC-Comp (EWOUC with competed data), EWOC (with completed data), EWOUC-NW (EWOUC not waiting, treat missing response as non effective), EWOUC-DA design were compared in three main aspects: therapeutic effects, dose recommendation accuracy and trial duration.

Following the notation in section two, the true parameter in the model, $\rho_T$and ρ_E, are respectively 0.03 and 0.08 for all scenarios. The disease status and efficacy status are generated based on these true model parameters. The MTD is corresponding to a target tolerated level, θ_T, 0.33. Meanwhile, the MED is corresponding to the minimum efficacy bound, θ_E, 0.3. The true MTD and MED for each scenario are listed in table 1. The true correlation between toxicity probability and efficacy probability φ is 0 when we compare different designs. The disease status and efficacy status are generated based on these true copula model defined above. Considering the utility function, W is 3 for S1, S2, S4 and S5. W is 2 for S3.

Under Bayesian framework, uniform priors are given for the unknown parameter ρ_T and ρ_E.

We suppose that ρ_T follows a uniform distribution (0, 0.33) and ρ_E follows a uniform distribution (0, 0.5). We also assume that both unknown MTD and MED follow uniform distribution (0.2,1.2). φ is given a standard normal distribution adopted from Thall and Cook (2004). We compared EWOUC-comp, EWOC, EWOUC-NW, EWOUC-DA under a piecewise exponential time to efficacy model. The follow-up window T is supposed to be three months. We spilt it into three intervals with three different hazards $\lambda_1, \lambda_2, \lambda_3$ respectively 0.4, 0.67 and 2.0, derived by $\lambda_k = K/T(K - k + 0.5)(K = 3, T = 3)$. Parameters are chosen so that 70% efficacy response occurs in T/2 to T. The priors for $\lambda_1, \lambda_2, \lambda_3$ are respectively Gamma(0.2,0.5), Gamma(0.33, 0.5) and gamma(1,0.5), adopted from Liu, Yin, and Yuan (2013). We further study different time to efficacy and different correlation models in section 3.5 to evaluate the robustness of the DA algorithm for our design and the sensitivity of our prior selection.

The Metropolis-Hastings algorithm is used to sample from posterior distribution by JAGS. The burn-in iteration is 1000 times. Then, extra 1000 iterations are used as the sample from posterior distribution. We repeat 1000 simulations for each scenario to evaluate the performance of the design. The adaption process of the posterior marginal distribution of MTD and MED (for an extremely good agent) are shown by Figure 1. After the 12th cohort was recruited into the trial, the posterior distributions become a unimodal distribution.

3.2 Dose Recommendation

For all scenarios, EWOUC-Comp (EWOUC with completed data) successfully detected doses with better utility, as shown in table 1 and figure 4. Most recommendations are clearly bounded by the interval composed by MTD and MED. For good agents (S1, S2, S3), EWOUC-Comp will recommend the best utility doses with a quite high accuracy rate. The dose recommendation percentages

for the best utility dose levels are 91.9%, 91% and 72.6% for S1, S2 and S3, respectively. For bad agents (S4, S5), the recommendation accuracy is at least 95.8% (in S4) to claim futility. The recommendations of EWOUC-Comp have a high probability to correctly determine whether a phase III design should be approved.

Compared with the Phase I EWOC design, there is a substantial difference in dose recommendation intention. EWOC tends to recommend doses close to MTD: in S1, 95.7% for dose level 4 and dose level 5 (50.3% for dose level 4 and 45.4% for MTD). By contrast, EWOUC-Comp tends to recommend doses with better utility: 98% for dose level 3 and 4 in S1 (91.9% for best utility dose level 3 and 18% for dose level 4). The MTD detected by EWOC is also the most effective. However, after considering the balance between toxicity and efficacy, the MTD may be not the best choice. Our design EWOUC-Comp balances the toxicity and efficacy by the utility function, choosing a dose maximizing the difference between efficacy and toxicity.

EWOUC-Comp is also compared with EWOUC-NW and EWOUC-DA to evaluate the performance of the DA algorithm. In our study, we find that EWOUC-NW clearly tends to over-estimate the best utility dose. For example, EWOUC-NW recommends 49.60% and 47.30% at dose level 3 and dose level 4 (incorrect doses) in S1 and S2, while EWOUC-DA recommends only 72.30% and 25.90%. In S3, it is clear that 29.20% dose recommendation focuses on dose level 5 compared with 12.7% under EWOUC-DA design. EWOUC-DA is more accurate than EWOUC-NW. For good agents (S1, S2, S3), the performance is very close between EWOUC-DA and EWOUC-Comp.

3.3 Therapeutic Effects

With regarding to the completed data designs, compared with EWOUC-Comp, EWOC again intends to treat patients at higher toxicity doses close to the MTD. For example, 40.46% patients are

treated at dose level 4 in S1 and 37.71% patients are treated at the MTD (also dose level 4) in S2. EWOUC-Comp treats 4.95% and 3.84% at such a dose level in S1 and S2. For S1, S2 and S3, most patients are treated at the best utility doses under EWOUC-Comp design (Figure 2), respectively 73.01% in S1, 70.81% in S2, 50.08% in S3 (table 2). For the bad and the extremely bad agent with high toxicity (S4 and S5), EWOUC-Comp intends to treat more patients close to the MTD or at the MTD, compared with EWOC. For good agents, fewer patients were treated under MED, compared with EWOC. Introduced under-dose control, EWOUC-Comp lowers the probability that patients are treated at futile doses.

EWOUC-DA treats at least 50% at the best utility dose level (respectively 56.13%, 58.23%, and 50.08%), compared with EWOUC-NW (28.01%, 38.76% and 45.86%). EWOUC-NW is more aggressive, treating more patients above the best utility dose. The DA algorithm fixes the problem of over-estimating by imputing the efficacy response, resulting in better therapeutic effect with respect to the utility.

Efficacy and DLT rate were calculated in table 2 to compare these designs. It is noting that although there is under-dose control, the efficacy probability of EWOUC-Comp decreased, compared with EWOC (decreased from 0.80 to 0.58 in S1). The main reason for the decrease is that EWOC treats more patients at higher and potentially risky doses leading to a more efficacious result. It can be shown that the DLT rate of EWOC is much higher than corresponding DLT rate in EWOUC designs (decreased from 0.17 to 0.10 in S1). EWOUC aims to keep a balance between toxicity and efficacy by treating patients at higher utility. Consequently, the expected utility is increased from 0.21 to 0.28 (increased about 34%) in S1 under EWOUC-Comp design, as shown in table 2. More clearly, the expected utility (see figure 3) increased sharply (from 0.02 to 0.14) in S2 under EWOUC-Comp design. EWOC loses utility because it treats many patients near the MTD

and the utility is negative at such dose levels. For bad and extremely bad agents, EWOUC-Comp terminates the trial quite early. The average sample size is 13.79 for EWOUC-Comp, 20.364 for EWOUC-DA. The sample size is smaller for EWOUC-NW because EWOUC-NW overestimates the MED. Considering the overall performance, EWOUC-DA is more reliable for real clinical trial because we don't know which kind of agent is being tested.

3.4 Expected Trial Duration

EWOUC-DA and EWOUC-NW are thought to be similar when calculating trial duration. These designs do not require clinicians to collect all data before they treat next cohort, which means that there is no waiting time. In order to compare the expected time for a trial among EWOUC-DA, EWOUC-NW and EWOUC-Comp, we performed a simulation study based on different mean cohort inter-arrival times (Table 4) for S1 (extremely good agent). Suppose that the inter-cohort arrival time is distributed exponentially and the follow up window is three months. The means of the inter-arrival time are respectively 0.25, 0.5, 1, 2, and 3 months so that the corresponding accrual rates (assessment time period / recruitment time) for the mean inter-arrival times would be 12, 6, 3, 1.5, and 1 in table 4. When the accrual rate is high, EWOUC-DA is considerably more efficient. It will take an expected 8.98 months to complete the trial using EWOUC-DA. By contrast, EWOUC-Comp will take 28.27 months to complete the trial. When the accrual rate is nearly one, the durations of different trial designs are nearly the same. This is because all responses have been confirmed when the next cohort arrives; thus, there is nearly no delay in treating the next cohort.

3.5 Sensitivity Study

Since we have tested our model by the different locations of best utility dose, in this section, different efficacy lower bounds θ_E and correlation φ are used to test the EWOUC-DA model in detecting the best utility dose. The φ is set to be 1,2 or 3 in the copula model to generate the patients' responses. θ_E is set to be 0.3 or 0.4. There are six combinations of θ_E and φ. We simulated S1, S2 and S3 based on these combinations so that we try to detect a best utitlity dose level corresponding to different efficacy lower bounds and true correlation between efficacy events and toxicity events (See table 5) . Each combination of (φ, θ_E) is simulated with 1000 replicates. Under all scenarios, EWOUC-DA design can detect the correct best utility dose at least 68% and treats most patients at that dose level (dose level 3).

We used log logistic and Weibull distribution as the true time to efficacy model. As shown in table 6, we found that both dose recommendation and patient distribution are very similar among different time to efficacy distributions under S1 and S2. The assumed piece-wise exponential model (K=3) can be fitted well when the true time to efficacy distributions are the log logistic and Weibull distribution. Compared with completed data, the behavior of the DA algorithm is very similar with EWOUC.

4 Conclusion and Discussion

Through extensive simulation, we have demonstrated that EWOUC performs well in detecting a dose between MED and MTD with best utility. For bad agents, it can be used to claim futility for high toxicity and low efficacy agents. The EWOUC design can reduce the risk of failing phase III clinical trial and save more expenses if the agent is futile.

Another advantage of EWOUC is to provide both over-dose and under-dose protection in clinical trials. Patients tend to be treated at higher utility doses. The over-toxic protection with EWOUC designs is even better than with EWOC design after considering utility. In EWOUC designs, patients will not be treated at a dose level quite close to the MTD but at a dose level with low toxicity and high efficacy. Thus, the DLT rate in the trial will drop down and the over-toxic protection is better. Unlike the EWOC design, patients can attain higher utility without higher risk of DLT with the EWOUC design, a more ethically favorable situation.

Implementation of the DA algorithm renders the EWOUC design more practical by largely shortening the duration of a clinical trial when the accrual rate is high. When EWOUC ignores incomplete responses, it will lead to large bias, overestimating the dose with best utility and treating most patients at a lower utility dose. EWOUC-DA can be used for late onset efficacy data, and can maintain accuracy without sacrificing time and cost.

The accuracy of EWOUC-DA and EWOUC can be further improved by adjusting the threshold of feasibility bounds conditioned on the same sample size. If we fix both feasibility bounds at 0.5 from the beginning of the trial, the performance of EWOUC-DA and EWOUC can be further improved for good and extremely good agents in prediction accuracy. However, such a method will lead to higher DLT probability for both extremely good and extremely bad agents. The authors have chosen both 0.25 as threshold so that the probability of DLT is lower when we cannot confirm

which type of agent is to be tested. If clinicians have relatively confirmative information for the agent, the feasibility bound threshold can be increased.

The joint probability of toxicity events and efficacy events is modeled by the Farlie-Gumbel-Morgenstern Copula model (Murtaugh and Fisher, 1990). There are many other copula models (See Nelson, 2006). The model selection will be studied in the future. Our method can only estimate a marginal best utility dose level for all the patients in the trial. The model can be extended to include patients' covariates so that our design can be applied to find personalized best utility dose level.

Overall, we propose a novel phase I/II clinical trial design based on the traditional EWOC design by incorporating utility consideration and under-dose control. Using the method proposed by Tanner and Wong (1986), we have implemented a Bayesian data augmentation algorithm on EWOUC design to solve the problem of late-onset efficacy data, maintaining accuracy for dose recommendation without the requirement for completed data.

5 References

Babb, J., Rogatko, A., Zacks, S. (1998). Cancer phase I clinical trials: efficient dose escalation with overdose control. Statistics in Medicine, 17(10), 1103-1120.

Babb, J. S., Rogatko, A. (2001). Patient specific dosing in a cancer phase I clinical trial. Statistics in Medicine, 20(14), 2079-2090. doi: 10.1002/sim.848

Berry, S. M. (2011). Bayesian adaptive methods for clinical trials. Boca Raton: CRC Press.

Chen, Z., Krailo, M. D., Azen, S. P., Tighiouart, M. (2010). A novel toxicity scoring system treating toxicity response as a quasi-continuous variable in Phase I clinical trials. Contempary Clinical Trials, 31(5), 473-482.

Chu, P.-L., Lin, Y., Shih, W. J. (2009). Unifying CRM and EWOC designs for phase I cancer clinical trials. Journal of Statistical Planning and Inference, 139(3), 1146-1163.

Liu, S., Yin, G., and Yuan, Y. (2013). Bayesian data augmentation dose finding with continual reassessment method and delayed toxicity. The Annals of Applied Statistics, 7(4), 2138-2156.

Mauguen, A., Le Deley, M. C., Zohar, S. (2011). Dose-finding approach for dose escalation with overdose control considering incomplete observations. Stat Med, 30(13), 1584-1594.

Murtaugh, P. A., and Fisher, L. D. (1990). Bivariate binary models of efficacy and toxicity in dose-ranging trials. Communications in Statistics-Theory and Methods, 19(6), 2003-2020.

Nelsen, R. B. (2006). An introduction to copulas (2nd ed.). New York: Springer.

O'Quigley, J., Pepe, M., Fisher, L. (1990). Continual reassessment method: a practical design for phase 1 clinical trials in cancer. Biometrics, 46(1), 33-48.

Tanner, M. A., Wong, W. H. (1987). The Calculation of Posterior Distributions by Data Augmentation. Journal of the American Statistical Association, 82(398), 528-540.

Thall, P. F., Cook, J. D. (2004). Dose-Finding Based on Efficacy-Toxicity Trade-Offs. Biometrics, 60(3), 684-693.

Yin, G., Yuan, Y. (2009). Bayesian dose finding in oncology for drug combinations by copula regression. Journal of the Royal Statistical Society: Series C (Applied Statistics), 58(2), 211-224.

Zacks, S., Rogatko, A., Babb, J. (1998). Optimal Bayesian-feasible dose escalation for cancer phase I trials. Statistics Probability Letters, 38(3), 215-220.

Zhong, W., Carlin, B. P., Koopmeiners, J. S. (2013). Flexible link continual reassessment methods for trivariate binary outcome phase I/II trials. Journal of Statistical Theory and Practice(just-accepted).

6 Appendix A

Figure 1: Posterior distributions under scenario 1 after 12th cohort

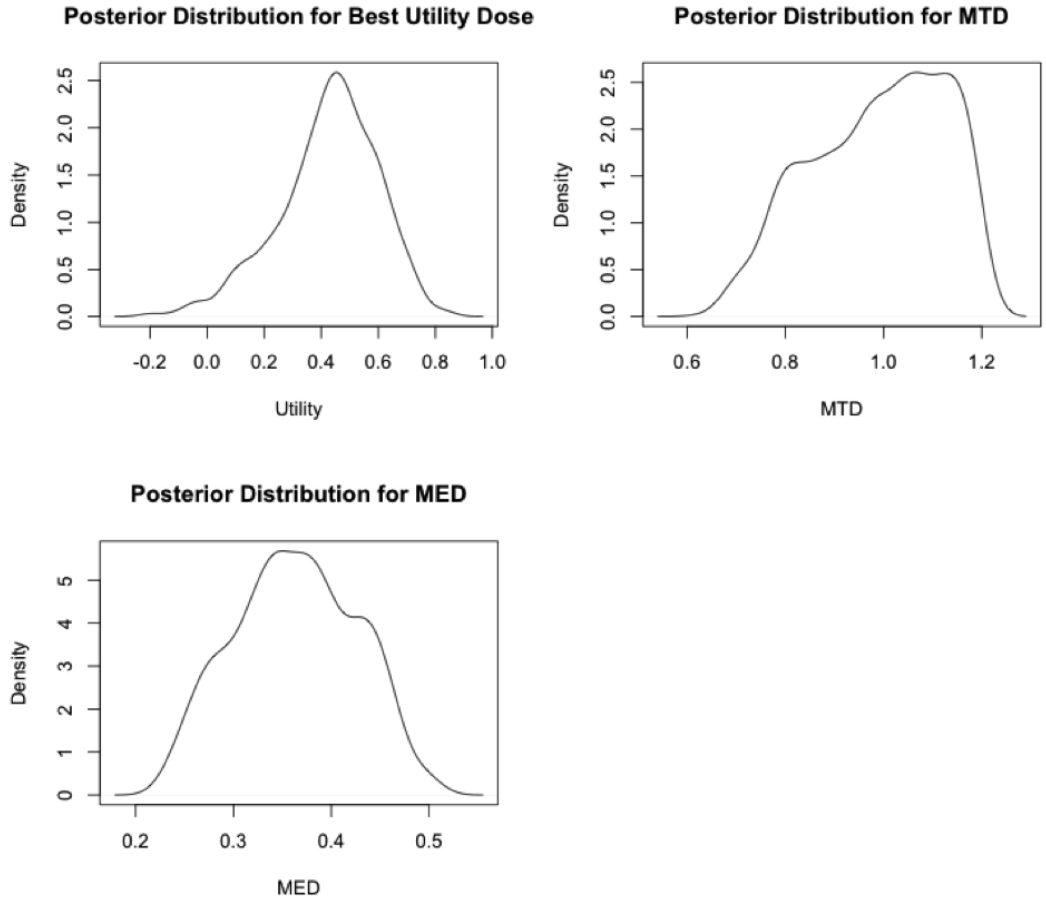
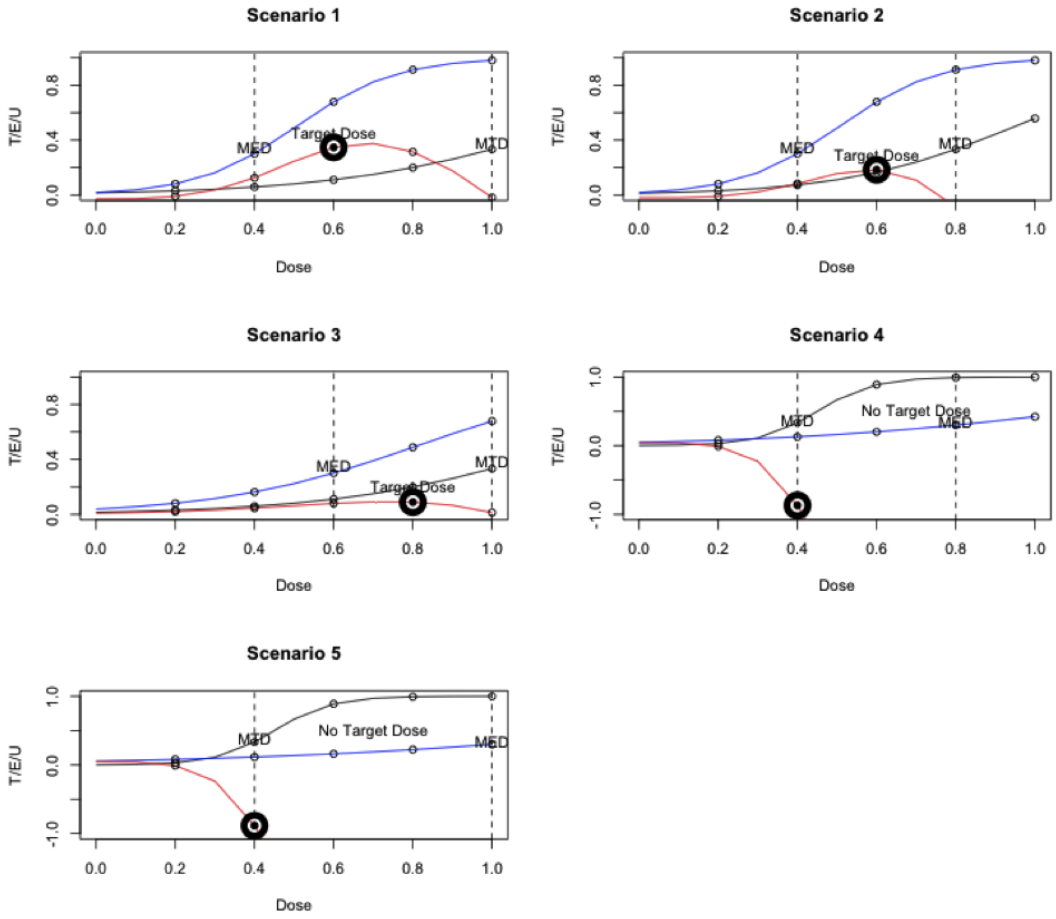


Figure 2: Dose response curves for five scenarios



Black curve is toxic dose response curve. Blue curve is efficacy dose response curve. Red curve is the utility curve. True MTD and MED are respectively noted on two curves. Solid dot is the dose which most patients are treated at.

Figure 3: Expected utility comparison for different designs

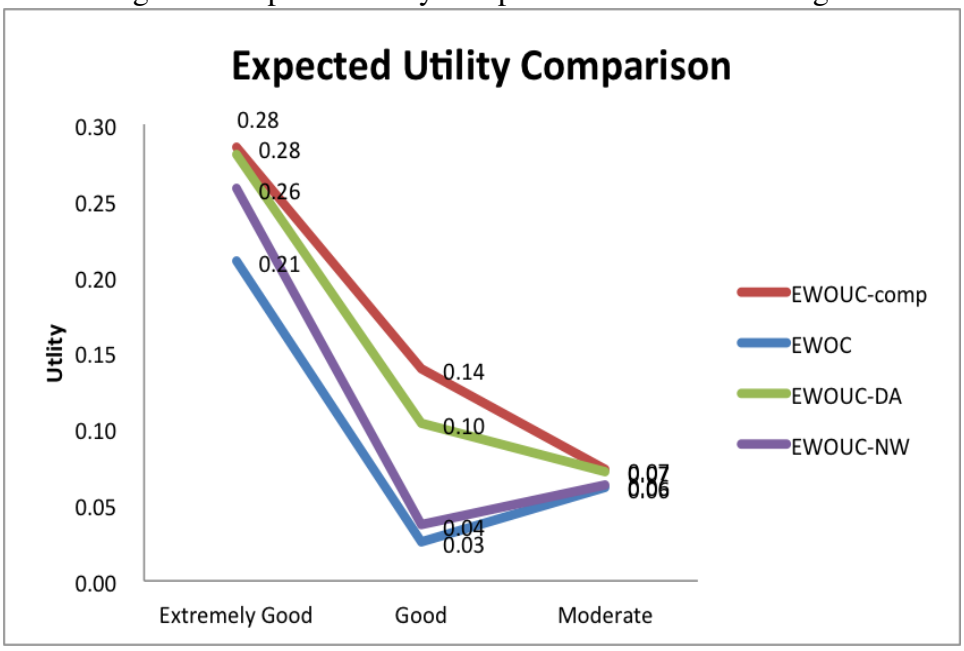
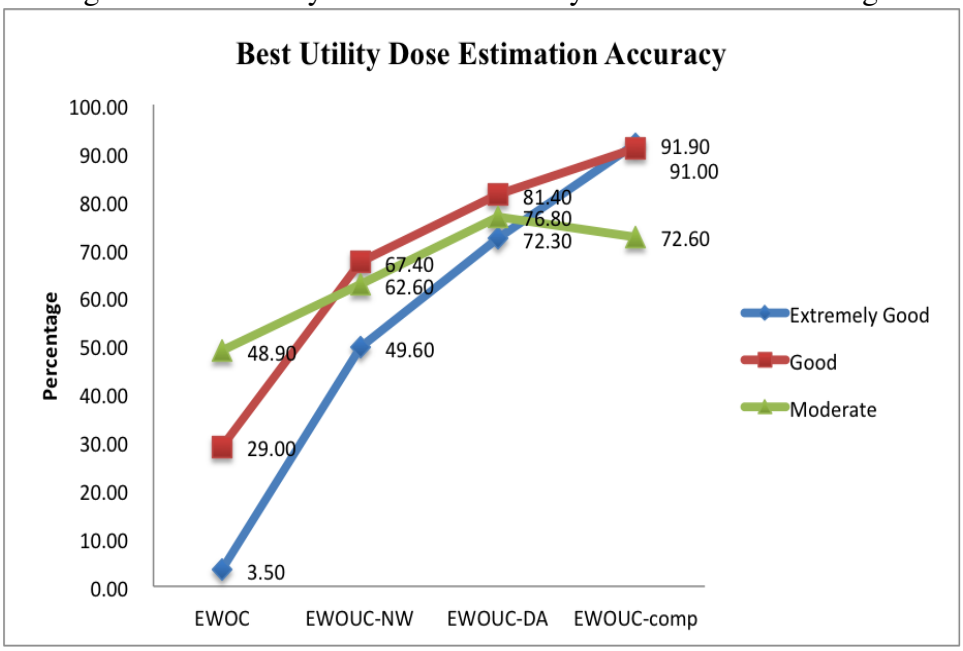


Figure 4: Best utility estimation accuracy rate for different designs



7 Appendix B

Table 1: Dose recommendation percentage(%) under each dose level

Extremely Good						
			Dose Level			
	1	**2**	**3**	**4**	**5**	None
Probability of DLT	0.03	0.06	0.11	0.20	0.33	0.00
Probability of Efficacy	0.08	0.30	0.68	0.91	0.98	0.00
Utility	-0.01	0.12	0.35	0.31	-0.02	0.00
EWOC	0.20	0.60	3.50	50.30	45.40	0.00
EWOUC-NW	0.00	0.90	49.60	47.30	0.80	1.40
EWOUC-DA	0.10	0.90	72.30	25.90	0.10	0.70
EWOUC-comp	0.00	1.40	91.90	6.10	0.10	0.50
Good						
			Dose Level			
	1	**2**	**3**	**4**	**5**	None
Probability of DLT	0.03	0.07	0.16	0.33	0.56	0.00
Probability of Efficacy	0.08	0.30	0.68	0.91	0.98	0.00
Utility	-0.01	0.08	0.18	-0.09	-0.69	0.00
EWOC	0.40	1.80	29.00	64.80	4.00	0.00
EWOUC-NW	0.00	2.40	67.40	25.90	0.00	4.30
EWOUC-DA	0.00	4.00	81.40	13.80	0.10	0.70
EWOUC-comp	0.20	3.60	91.00	4.40	0.00	0.80
Moderate						
			Dose Level			
	1	**2**	**3**	**4**	**5**	None
Probability of DLT	0.03	0.06	0.11	0.20	0.33	0.00
Probability of Efficacy	0.08	0.16	0.30	0.49	0.68	0.00
Utility	0.02	0.05	0.08	0.09	0.01	0.00
EWOC	0.20	0.50	3.40	48.90	47.00	0.00
EWOUC-NW	0.00	0.00	3.90	62.60	29.20	4.30
EWOUC-DA	0.00	0.10	9.20	76.80	12.70	1.20
EWOUC-comp	0.00	0.20	21.00	72.60	5.40	0.80
Bad						
			Dose Level			
	1	2	**3**	**4**	**5**	None
Probability of DLT	0.03	0.33	0.89	0.99	1.00	NA
Probability of Efficacy	0.08	0.13	0.20	0.30	0.42	NA
Utility	-0.01	-0.87	-2.47	-2.68	-2.58	NA
EWOC	53.00	43.30	3.70	0.00	0.00	0.00
EWOUC-NW	0.30	0.60	0.00	0.00	0.00	99.10
EWOUC-DA	0.30	4.00	0.20	0.00	0.00	95.50
EWOUC-comp	0.40	3.70	0.10	0.00	0.00	95.80
Extremely Bad						
			Dose Level			
	1	**2**	**3**	**4**	**5**	None
Probability of DLT	0.03	0.33	0.89	0.99	1.00	NA
Probability of Efficacy	0.08	0.11	0.16	0.22	0.30	NA
Utility	-0.01	-0.88	-2.51	-2.75	-2.70	NA
EWOC	53.40	42.70	3.90	0.00	0.00	0.00
EWOUC-NW	0.10	0.40	0.10	0.00	0.00	99.40
EWOUC-DA	0.10	3.00	0.50	0.00	0.00	96.40
EWOUC-comp	0.10	1.60	0.20	0.00	0.00	98.10

Table 2: Toxicity and efficacy probability under each design

Scenario	**Toxicity**			
	EWOUC-comp	EWOC	EWOUC-DA	EWOUC-NW
Extremely good	0.58	0.80	0.63	0.70
Good	0.57	0.73	0.60	0.67
Moderate	0.37	0.45	0.37	0.41
Bad	0.11	0.11	0.12	0.11
Extremely Bad	0.09	0.10	0.10	0.09
	Efficacy			
	EWOUC-comp	EWOC	EWOUC-DA	EWOUC-NW
Extremely good	0.10	0.17	0.11	0.15
Good	0.14	0.21	0.17	0.21
Moderate	0.14	0.17	0.15	0.17
Bad	0.32	0.22	0.31	0.32
Extremely Bad	0.32	0.21	0.31	0.32

Table 3: Patient distribution(%) treated at each dose level

Extremely Good	**Dose Level**				
	1	**2**	**3**	**4**	**5**
Probability of DLT	0.03	0.06	**0.11**	0.20	0.33
Probability of Efficacy	0.08	0.30	**0.68**	0.91	0.98
Utility	-0.01	0.12	**0.35**	0.31	-0.02
EWOC	8.80	11.96	**20.83**	40.46	17.95
EWOUC-NW	8.59	10.86	**28.94**	46.36	5.26
EWOUC-DA	8.81	12.77	**56.13**	22.25	0.04
EWOUC-comp	8.70	13.32	**73.01**	4.95	0.03
Good	**Dose Level**				
	1	**2**	**3**	4	**5**
Probability of DLT	0.03	0.07	**0.16**	0.33	0.56
Probability of Efficacy	0.08	0.30	**0.68**	0.91	0.98
Utility	-0.01	0.08	**0.18**	-0.09	-0.69
EWOC	9.15	14.82	**35.64**	37.71	2.68
EWOUC-NW	8.75	12.21	**38.76**	38.88	1.41
EWOUC-DA	9.05	15.39	**58.23**	17.32	0.02
EWOUC-comp	9.06	16.29	**70.81**	3.84	0.00
Moderate	**Dose Level**				
	1	**2**	**3**	**4**	**5**
Probability of DLT	0.03	0.06	0.11	**0.20**	0.33
Probability of Efficacy	0.08	0.16	0.30	**0.49**	0.68
Utility	0.02	0.05	0.08	**0.09**	0.01
EWOC	8.80	11.84	20.29	**40.53**	18.54
EWOUC-NW	8.78	11.02	17.22	**45.86**	17.12
EWOUC-DA	8.66	11.17	26.75	**48.64**	4.78
EWOUC-comp	8.57	11.02	27.28	**50.08**	3.05
Bad	**Dose Level**				
	1	**2**	**3**	**4**	**5**
Probability of DLT	0.03	0.33	0.89	0.99	1.00
Probability of Efficacy	0.08	0.13	0.20	0.30	0.42
Utility	-0.01	-0.87	-2.47	-2.68	-2.58
EWOC	47.42	47.75	4.83	0.00	0.00
EWOUC-NW	30.07	54.47	15.44	0.02	0.00
EWOUC-DA	28.18	59.90	11.92	0.00	0.00
EWOUC-comp	29.20	56.68	14.13	0.00	0.00
Extremely Bad	**Dose Level**				
	1	**2**	**3**	**4**	**5**
Probability of DLT	0.03	0.33	0.89	0.99	1.00
Probability of Efficacy	0.08	0.11	0.16	0.22	0.30
Utility	-0.01	-0.88	-2.51	-2.75	-2.70
EWOC	48.18	47.02	4.80	0.00	0.00
EWOUC-NW	30.82	53.78	15.40	0.00	0.00
EWOUC-DA	28.59	59.74	11.67	0.00	0.00
EWOUC-comp	30.30	55.61	14.09	0.00	0.00

Table 4: Expected time for a trial (months) under different observation window and mean inter cohort arrival times for 12 cohorts under scenario 1

Mean Arrival	**Obs Window**	**EWOC-Comp**	**EWOUC-DA**	**Change**
0.25	3	28.05	5.99	78.65%
0.5	3	28.27	8.98	68.23%
1	3	28.36	14.96	47.25%
2	3	34.65	26.92	22.31%
3	3	42.37	38.88	8.24%

Table 5: Patient distribution and dose recommendation under different dose levels for extremely good, good and moderate agents under EWOUC-DA. Different φ and θ_E values are used for the model under the same drug. The best utility dose is underlined.

Extremely Good	**Dose level**				
	1	**2**	**3**	**4**	**5**
$\varphi = 1, \theta_E = 0.3$					
Patient Distribution	8.66	12.40	**52.95**	25.91	0.08
Dose Recommendation	0.00	0.70	**68.20**	30.50	0.20
$\varphi = 2, \theta_E = 0.3$					
Patient Distribution	8.68	12.68	**53.24**	25.25	0.14
Dose Recommendation	0.10	0.60	**71.60**	27.00	0.10
$\varphi = 3, \theta_E = 0.3$					
Patient Distribution	8.93	11.82	**53.11**	26.06	0.08
Dose Recommendation	0.10	0.90	**67.50**	31.10	0.10
$\varphi = 1, \theta_E = 0.4$					
Patient Distribution	8.54	14.53	**61.53**	15.38	0.03
Dose Recommendation	0.00	1.60	**94.00**	4.10	0.00
$\varphi = 2, \theta_E = 0.4$					
Patient Distribution	9.05	14.55	**61.19**	15.19	0.03
Dose Recommendation	0.30	2.10	**93.60**	3.10	0.10
$\varphi = 3, \theta_E = 0.4$					
Patient Distribution	8.65	15.21	**61.40**	14.70	0.04
Dose Recommendation	0.00	1.90	**94.80**	2.60	0.10
Good	**1**	**2**	**3**	**4**	**5**
$\varphi = 1, \theta_E = 0.3$					
Patient Distribution	8.95	14.73	**56.19**	20.12	0.01
Dose Recommendation	0.10	3.60	**77.70**	17.70	0.10
$\varphi = 3, \theta_E = 0.3$					
Patient Distribution	8.84	15.21	**56.12**	19.81	0.03
Dose Recommendation	0.10	3.70	**78.50**	17.00	0.00
$\varphi = 3, \theta_E = 0.3$					
Patient Distribution	8.73	14.54	**56.49**	20.20	0.03
Dose Recommendation	0.00	2.90	**80.00**	16.40	0.00
$\varphi = 1, \theta_E = 0.4$					
Patient Distribution	8.77	18.62	**60.30**	12.31	0.00
Dose Recommendation	0.00	6.70	**90.90**	1.90	0.00
$\varphi = 2, \theta_E = 0.4$					
Patient Distribution	8.85	17.42	**60.83**	12.89	0.02
Dose Recommendation	0.30	4.40	**91.90**	3.30	0.00
$\varphi = 3, \theta_E = 0.4$					
Patient Distribution	9.08	18.00	**60.95**	11.96	0.01
Dose Recommendation	0.20	5.10	**91.50**	2.70	0.00
Moderate	**1**	**2**	**3**	**4**	**5**
$\varphi = 1, \theta_E = 0.3$					
Patient Distribution	8.75	10.94	25.32	**49.82**	5.18
Dose Recommendation	0.00	0.20	8.90	**76.30**	13.80
$\varphi = 2, \theta_E = 0.3$					
Patient Distribution	8.60	10.64	26.28	**49.21**	5.27
Dose Recommendation	0.10	0.00	11.00	**73.00**	15.90
$\varphi = 3, \theta_E = 0.3$					
Patient Distribution	8.76	10.76	25.71	**49.07**	5.69
Dose Recommendation	0.10	0.10	8.60	**74.10**	16.20
$\varphi = 1, \theta_E = 0.4$					
Patient Distribution	8.63	11.17	29.71	**48.63**	1.85
Dose Recommendation	0.00	0.50	14.80	**81.00**	3.20
$\varphi = 2, \theta_E = 0.4$					
Patient Distribution	8.52	11.07	29.92	**48.49**	2.00
Dose Recommendation	0.00	0.30	15.70	**81.40**	2.30
$\varphi = 3, \theta_E = 0.4$					
Patient Distribution	8.58	10.60	29.35	**49.52**	1.95
Dose Recommendation	0.00	0.10	14.90	**81.30**	3.60

Table 6: Patient distribution and dose recommendation under each dose level for S1 and S2 in EWOUC-DA design. True Time to efficacy model is set to be log logistic or Weibull distribution for each scenario. The best utility dose is dose level 3

Scenario	Dose level				
	1	**2**	**3**	**4**	**5**
Extremely Good					
Log logistic					
Patient Distribution	8.54	11.51	**43.43**	36.17	0.35
Dose Recommendation	0.00	0.40	**68.10**	30.80	0.60
Weibull					
Patient Distribution	8.72	11.49	**42.96**	36.34	0.50
Dose Recommendation	0.00	0.50	**66.30**	32.10	0.50
Good					
Log logistic					
Patient Distribution	8.64	13.33	**49.21**	28.65	0.17
Dose Recommendation	0.10	4.00	**78.90**	16.30	0.20
Weibull					
Patient Distribution	8.84	13.47	**49.63**	27.93	0.13
Dose Recommendation	0.10	3.20	**80.90**	14.80	0.10