CWL: A Conditional Weighted Likelihood Method to Account for the Delayed Joint Toxicity-efficacy Outcomes for Phase I/II Clinical Trials

Yifei Zhang[1] and Yong Zang[1,2*]

1. Department of Biostatistics, Indiana University

2. Center for Computational Biology and Bioinformatics, Indiana University

Corresponding Author: Yong Zang, Department of Biostatistics, Indiana University, 410 W 10th Street, Indianapolis, IN, 46202. Email: zangy@iu.edu

Abstract

The delayed outcome issue is common in early phase dose-finding clinical trials. This problem becomes more intractable in phase I/II clinical trials because both toxicity and efficacy responses are subject to the delayed outcome issue. The existing methods applying for the phase I trials cannot be used directly for the phase I/II trial due to a lack of capability to model the joint toxicity-efficacy distribution. In this paper, we propose a conditional weighted likelihood (CWL) method to circumvent this issue. The key idea of the CWL method is to decompose the joint probability into the product of marginal and conditional probabilities and then weight each probability based on each patient's actual follow-up time. The CWL method makes no parametric model assumption on either the dose-response curve or the toxicity-efficacy correlation and therefore can be applied to any existing phase I/II trial design. Numerical trial applications show that the proposed CWL method yields desirable operating characteristics.

KEY WORDS: Adaptive design; Delayed outcome; Optimal biological dose; Phase I/II clinical trial; Weighted likelihood

1 Introduction

Traditionally, cancer clinical trials are conducted sequentially in different phases. A phase I trial is conducted firstly to identify the maximum tolerated dose (MTD) of the new drug based on the toxicity outcomes. Then, a separate phase II trial is carried out to examine the efficacy of the drug at the identified MTD. The key assumption underlying this sequential paradigm is that both efficacy and toxicity response rates increase monotonically with the dose. Although this assumption is generally true for cytotoxic agents such as chemotherapy and radiotherapy, it has been challenged by the recently developed novel non-cytotoxic agents such as immunotherapy and targeted therapy. Immunotherapy enhances the innate power of the immune system to stop the growth of the cancer cell and to prevent cancer from spreading to other parts of the body. Targeted therapy modulates specific aberrant pathways in cancer cells while sparing normal tissue. As a result, the dose-efficacy curves may not follow monotonic patterns for these non-cytotoxic agents [1, 2, 3, 4, 5].

Because the monotonic pattern does not hold, the MTD may not yield the best therapeutic effect for non-cytotoxic agents. Hence, unlike cytotoxic agents, the primary goal of the dose-finding trial for immunotherapy and targeted therapy is to identify the optimal biological dose (OBD), which is defined as the dose that yields the highest desirability in terms of the toxicity-efficacy trade-off. For this purpose, a lot of phase I/II clinical trials have been proposed for the novel agents which model and monitor the toxicity and efficacy outcomes jointly to find the OBD [6, 7, 8, 9].

Most of the early-phase dose-finding clinical trials require the patients' response outcomes to be quickly ascertainable after receiving the drug. In other words, all the previously treated patients in the trial should have their response information observable during the interim analysis when a new cohort of patients are coming and ready for dose assignment. However, this may not always be the case and delayed response outcomes are common in the practical implementation of dose-finding trials. That means, at least part of the patients' response information is missing at the interim analysis time and may be available long after that time. The simple strategy of using the observable information only will distort the estimate of the response rate due to the not-missing-at-random mechanism [10] and should not be considered in practice. Another practical strategy is to suspend the trial when the delayed response happening and re-open the trial when all the patients have been completely followed. However, this strategy is unethical for the new coming patients waiting for the treatment and frequently suspending the trial will substantially prolong the trial, waste resources and create a tremendous administrative burden.

A lot of statistical methods have been proposed to solve the delayed outcome issue, mostly in the area of phase I clinical trials with toxicity outcomes only. Cheung and Chappell [11] proposed a weighted likelihood method that weighted the likelihood of each patient by his/her actual follow-up time. Lin and Yuan [12] proposed a binomial approximation approach to facilitate the computation of the weighted likelihood. Yuan and Yin [10] developed an EM algorithm to impute the missing outcomes from the observed information and this idea was further extended by Liu et al. [13] to develop a data argumentation method under the Bayesian framework. Recently, Yuan et al. [14] proposed a single imputation approach to address the delayed outcome problem for the BOIN design.

The problem becomes more intractable in phase I/II clinical trials because both toxicity

and efficacy responses are subject to the delayed outcome issue. Unfortunately, the aforementioned methods cannot be used directly for phase I/II clinical trials. That is because these methods can only model the marginal distribution of the toxicity outcome whereas most of phase I/II clinical trials require to modeling the joint toxicity-efficacy outcomes. Hence, a novel method to address the delayed joint efficacy-toxicity outcomes issue for phase I/II clinical trial is in urgent demand.

Our study is motivated by a phase I/II clinical trial being conducted at Indiana University Melvin and Bren Simon Cancer Center. The purpose of this trial is to find the OBD of a KIF18A inhibitor for patients with triple negative breast cancer (TNBC). Sixty patients with TNBC were enrolled and treated with KIF18A inhibitor at 5 dose levels 100 mg, 300 mg, 600 mg, 1000 mg and 1600 mg administered PO once daily (QD). The toxicity outcome is the dose limiting toxicity (DLT) and the efficacy outcome is the objective tumor response (OTR). Every patient in the trial will be followed for a total of three months to evaluate the DLT and OTR outcomes, which may happen any time during the three months follow-up period. Besides, for every two months, a cohort of three patients will be enrolled in the trial for dose allocation before all the toxicity and efficacy outcomes become available for previous patients in the trial. Hence, both toxicity and efficacy outcomes may be delayed at the two months interim analysis time and how to handle these delayed outcomes is the challenge of this trial.

In this paper, we develop a conditional weighted likelihood (CWL) method to account for the delayed joint toxicity-efficacy outcomes for phase I/II clinical trials. The CWL method decomposes the joint probability into the product of marginal and conditional probabilities and weights the probabilities separately based on each patient's actual follow-up time. Then, the complete likelihood function is re-constructed by the weighted probabilities, based on

which the design parameters are derived to conduct the trial. The unique feature of the CWL method is that it does not make any parametric model assumption on either the dose-response curve or toxicity-efficacy correlation so it can be equipped with any existing phase I/II clinical trial design.

2 Conditional Weighted Likelihood

Let y_{Tdi} be the binary toxicity outcome for the ith patient treated at dose level d ($d = 1, \cdots, D; i = 1, \cdots, n_d$). We define $y_{Tdi} = 1$ if the patient has reported the toxicity event during the entire follow-up period $[0, U_T]$ and $y_{Tdi} = 0$ otherwise. Similarly, we define $y_{Edi} = 1$ if this patient has reported the favorable efficacy event during the entire follow-up period $[0, U_E]$ and $y_{Edi} = 0$ otherwise. Let τ be the inter-arrival time, which means that a new cohort of patients will enter the trial by each time length τ and an interim analysis is needed to determine the dose level for the coming patients. The delayed outcome issue arises if $\tau < \max(U_T, U_E)$, because the existing patients have not been completely followed when the new cohort of patients come in, and the toxicity and efficacy events that may not so far happen are still likely to happen in the remaining follow-up period.

We define v_{Tdi} and v_{Edi} as the actual follow-up time of toxicity and efficacy for the ith patient assigned to dose level d at an interim analysis and we have $v_{Tdi} \leq U_T$ and $v_{Edi} \leq U_E$. The difference between τ and v_{Tdi} and v_{Edi} is that τ is fixed whereas v_{Tdi} and v_{Edi} vary with different interim analyses. For example, assuming $\tau = 1$ month, $U_T = 3$ months and the first patient in the trial is assigned to dose d and will experience DLT at time 2.5 months. Then, when the second, third and fourth cohort of patients come in, we should observe $v_{Td1} = 1$, 2 and 2.5 months respectively.

Let x_{Tdi} be the underlying true toxicity occurring time for the ith patient assigned to dose d and define $p_{Td} = \Pr(x_{Tdi} \leq U_T)$ be the toxicity rate for dose level d. Then, if this patient dose not experience toxicity at the actual follow-up time v_{Tdi}, following Cheung and Chappell [11] we use the weighted likelihood to approximate the probability $\Pr(x_{Tdi} > v_{Tdi})$ as

$$\begin{aligned}
\Pr(x_{Tdi} > v_{Tdi}) &= \Pr(x_{Tdi} > v_{Tdi} | x_{Tdi} > U_T)\Pr(x_{Tdi} > U_T) + \Pr(x_{Tdi} > v_{Tdi} | x_{Tdi} \leq U_T)\Pr(x_{Tdi} \leq U_T) \\
&\approx 1 - p_{Td} + p_{Td}(1 - \frac{v_{Tdi}}{U_T}) \\
&\approx 1 - \frac{v_{Tdi}}{U_T} p_{Td}. \qquad (1)
\end{aligned}$$

This approximation relies on the assumption that the time-to-toxicity outcome is uniformly distributed over the entire follow-up period $[0, U_T]$. According to previous studies [11, 12], this uniform scheme is remarkably robust and yields desirable operating characteristics.

To induce the correlation between the toxicity and efficacy outcomes, we model the efficacy outcome conditional on the toxicity output y_{Tdi}. Let x_{Edi} be the underlying true efficacy occurring time for the ith patient assigned to dose level d and define $p_{E1d} = \Pr(x_{Edi} \leq U_E | y_{Tdi} = 1)$ and $p_{E0d} = \Pr(x_{Edi} \leq U_E | y_{Ti} = 0)$ be the efficacy rate for this patient conditional on the toxicity happening or not. Similar to the toxicity outcome, we approximate $\Pr(x_{Edi} > v_{Edi} | y_{Tdi} = 1)$ and $\Pr(x_{Edi} > v_{Edi} | y_{Tdi} = 0)$ as

$$\Pr(x_{Edi} > v_{Edi} | y_{Tdi} = 1) \approx 1 - \frac{v_{Ei}}{U_E} p_{E1d} \quad ; \quad \Pr(x_{Edi} > v_{Edi} | y_{Tdi} = 0) \approx 1 - \frac{v_{Ei}}{U_E} p_{E0d} \qquad (2)$$

Let us define $\Phi_{1di} = \Pr(x_{Tdi} \leq v_{Tdi}, x_{Edi} \leq v_{Edi})$, $\Phi_{2di} = \Pr(x_{Tdi} \leq v_{Tdi}, x_{Edi} > v_{Edi})$, $\Phi_{3di} = \Pr(x_{Tdi} > v_{Tdi}, x_{Edi} > v_{Edi})$ and $\Phi_{4di} = \Pr(x_{Tdi} > v_{Tdi}, x_{Edi} \leq v_{Edi})$ as the joint toxicity-efficacy probabilities for the ith patient assigned to dose level d when he/she has been followed v_{Tdi} for toxicity and v_{Edi} for efficacy. With formulas (1) and (2) at hand, after some algebra based on the Bayes' theorem, we get their expression as

$$\begin{aligned}
\Phi_{1di} &\approx \frac{v_{Tdi}v_{Edi}}{U_T U_E} p_{Td} p_{E1d} \\
\Phi_{2di} &\approx \frac{v_{Tdi}}{U_T} p_{Td}(1 - \frac{v_{Edi}}{U_E} p_{E1d}) \\
\Phi_{3di} &\approx \Big(1 - \frac{v_{Tdi}}{U_T} p_{Td}\Big)\Big\{1 - \frac{v_{Edi}}{U_E}\big(p_{E1d}\frac{(1 - \frac{v_{Tdi}}{U_T})p_{Td}}{(1 - \frac{v_{Tdi}}{U_T})p_{Td} + 1 - p_{Td}} + p_{E0d}\frac{1 - p_{Td}}{(1 - \frac{v_{Tdi}}{U_T})p_{Td} + 1 - p_{Td}}\big)\Big\} \\
\Phi_{4di} &\approx \Big(1 - \frac{v_{Tdi}}{U_T} p_{Td}\Big)\Big\{\frac{v_{Edi}}{U_E}\big(p_{E1d}\frac{(1 - \frac{v_{Tdi}}{U_T})p_{Td}}{(1 - \frac{v_{Tdi}}{U_T})p_{Td} + 1 - p_{Td}} + p_{E0d}\frac{1 - p_{Td}}{(1 - \frac{v_{Tdi}}{U_T})p_{Td} + 1 - p_{Td}}\big)\Big\} \quad (3)
\end{aligned}$$

Finally, let $\mathrm{I}(\cdot)$ be the indicator function and define $\mathrm{I}_{1di} = \mathrm{I}(x_{Tdi} \leq v_{Tdi}, x_{Edi} \leq v_{Edi})$, $\mathrm{I}_{2di} = \mathrm{I}(x_{Tdi} \leq v_{Tdi}, x_{Edi} > v_{Edi})$, $\mathrm{I}_{3di} = \mathrm{I}(x_{Tdi} > v_{Tdi}, x_{Edi} > v_{Edi})$ and $\mathrm{I}_{4di} = \mathrm{I}(x_{Tdi} > v_{Tdi}, x_{Edi} \leq v_{Edi})$, the conditional weighted likelihood function for all the patients in the trial can be written as

$$\mathrm{CWL} = \prod_{d=1}^{D} \prod_{i=1}^{n_d} \prod_{j=1}^{4} \Phi_{jdi}^{\mathrm{I}_{jdi}}. \quad (4)$$

Hence, during the interim analysis, we can derive the estimates of p_{Td}, p_{E1d} and p_{E0d} from the CWL, based on which we can identify the OBD and assign it to the next coming patients. More detail will be given in the next section.

3 Extend Existing Designs

In this section, we illustrate how to extend existing phase I/II designs to handle the delayed joint toxicity-efficacy outcomes by using the proposed CWL method. We first extend the U-BOIN design [15]. The U-BOIN design is a Bayesian two-stage phase I/II design identifying the OBD. The BOIN design [16] is used in the first stage to explore the dose space and collect preliminary toxicity and efficacy data. In the second stage, a multinomial-Dirichlet model is used to model the toxicity-efficacy outcomes jointly and a utility function is employed to measure dose risk-benefit trade-off. The posterior estimate of the utility for each dose is to keep updating in the second stage by using all the accumulated information, which is then used to guide the dose assignment and OBD selection. The U-BOIN design is developed for binary response outcomes. Although the authors briefly discussed how to handle the delayed efficacy outcome, no work has been done for the delayed joint outcomes.

We first adapt the first stage of the U-BOIN design. Because only the toxicity outcome is monitored in the first stage, it is not necessary to use the CWL method. Instead, we propose to use the binomial approximation method [12] to accommodate the delayed toxicity outcomes, which can remarkably alleviate the computational burden. Specifically, based on the toxicity information for all the n patients at the interim analysis time, p_{Td} can be estimated as

$$\widehat{p_{Td}} \approx \frac{\sum_{i=1}^{n_d} \mathrm{I}(x_{Tdi} \leq v_{Tdi})}{\sum_{i=1}^{n_d} \left(\mathrm{I}(x_{Tdi} \leq v_{Tdi}) + \mathrm{I}(x_{Tdi} > v_{Tdi}) \frac{v_{Tdi}}{U_T} \right)}.$$

Then, in the current dose d, we determine the next dose d^* for the coming patients by comparing $\widehat{p_{Td}}$ with a pair of boundary (λ_e, λ_d). λ_e and λ_d are pre-determined before the trial conduction which can minimize the probability of incorrect dose-assignment decisions. In particular, we escalate the dose $d^* = d + 1$ if $\widehat{p_{Td}} \leq \lambda_e$, de-escalate the dose $d^* = d - 1$ if

$\widehat{p_{Td}} \geq \lambda_d$ and retain the dose $d^* = d$ otherwise.

In addition to the dose-assignment rule, a toxicity control rule is imposed as follows. Let $\mathcal{D}$ be all the accumulated outcome information and $\bar{\pi}_T$ be the maximum tolerable toxicity rate. From a Bayesian perspective, if $\Pr(p_{Td} > \bar{\pi}_T|\mathcal{D}) > \nu_T$ and at least three patients have been treated at dose d, dose level d and higher are eliminated from the trial; the trial is terminated if the lowest dose level is eliminated. By assigning p_{Td} a non-informative prior beta(α, β), the posterior distribution of p_{Td} can be approximated as

$$p_{Td}|\mathcal{D} \sim \text{beta}\Big(\sum_{i=1}^{n_d} \text{I}(x_{Tdi} \leq v_{Tdi}) + \alpha, \sum_{i=1}^{n_d} \text{I}(x_{Tdi} > v_{Tdi})\frac{v_{Tdi}}{U_T} + \beta\Big),$$

based on which we can calculate the posterior probability $\Pr(p_{Td} > \bar{\pi}_T|\mathcal{D}_T)$.

We now adapt to the second stage of the U-BOIN design. Once the number of patients treated on one of the doses reaches a certain value (12 as recommended by the authors), the design moves to the second stage. The purpose of the second stage is to evaluate the joint toxicity-efficacy distribution at each dose level and determine the OBD. For this purpose, a utility function is introduced to measure the dose risk-benefit trade-off. Let $\pi_{lkd} = \Pr(y_{Tdi} = l, y_{Edi} = k)$ $(l, k = 0, 1)$ be the joint toxicity-efficacy probability at dose d, the utility function is defined as

$$\text{UT}_d = \sum_{l,k=0}^{1} \omega_{lk}\pi_{lkd}.$$

Here ω_{lk} is ascribed to outcome $(y_{Tdi} = l, y_{Edi} = k)$ to reflect the risk-benefit trade-off. Typically, we define $\omega_{10} = 0$ to reflect the worst outcome and $\omega_{01} = 100$ to reflect the best outcome. Then, the values of ω_{00} and ω_{11} lay between 0 and 100 and should be elicited by the physicians to reflect the specific biological mechanism of the drug to be used in the trial.

The toxicity control rule used in the first stage is kept in the second stage to exclude overly toxic doses. In addition, as the efficacy outcome is introduced, the U-BOIN design also imposes a futility rule to exclude less efficacious doses. Let $\bar{\pi}_E$ be the lowest acceptable efficacy rate and $p_{Ed} = \Pr(y_{Edi} = 1)$ be the efficacy rate at dose d, the futility rule exclude any dose d with a posterior probability $\Pr(p_{Ed} < \bar{\pi}_E|\mathcal{D}) > \nu_E$. Furthermore, An admissible set is constructed as

$$\mathcal{A} = \{d : \Pr(p_{Td} > \bar{\pi}_T|\mathcal{D}) \leq \nu_T \cap \Pr(p_{Ed} < \bar{\pi}_E|\mathcal{D}) \leq \nu_E\}.$$

At each interim analysis, the dose that yields the highest estimate of UT_d within the admissible set $\mathcal{A}$ is identified as the OBD.

Under the Bayesian paradigm, following the original design, we speculate $\pi_d = (\pi_{00d}, \pi_{01d}, \pi_{10d}, \pi_{11d})$ to follow a non-informative Dirichlet distribution as

$$\pi_d \sim \mathrm{Dir}(\delta_{00}, \delta_{01}, \delta_{10}, \delta_{11}),$$

To accommodate the delayed outcome, we just need to re-parametrize p_{Td}, p_{Ed}, p_{E1d} and p_{E0d} in the form of π as

$$\begin{aligned} p_{Td} &= \pi_{10d} + \pi_{11d} \\ p_{E1d} &= \frac{\pi_{11d}}{\pi_{10d} + \pi_{11d}} \\ p_{E0d} &= \frac{\pi_{01d}}{\pi_{00d} + \pi_{01d}}. \end{aligned} \tag{5}$$

Then, during an interim analysis, after plugging in the CWL function (4) with the formula (5), we can obtain the posterior distributions of π_d through the MCMC algorithm, which

can be used to construct the admissible set $\mathcal{A}$, update the posterior estimate of UT_d, and identify the OBD. The next cohort of patients should be assigned to the identified OBD. When the trial ends and all the patients have been completely followed, the lastly identified OBD based on the complete data is the OBD recommended by the trial. We denote this extension of the U-BOIN design as the CWL-U-BOIN design.

In addition to the U-BOIN design, we also briefly discussed how to extend another phase I/II design based on toxicity-efficacy odds ratios [17] (denoted as the OR design hereafter) to handle the delayed joint outcomes. The OR design is a one-stage design, the same toxicity and futility rules are used to construct the admissible set. But this design relies on a three-dimensional volume ratio to find the OBD. The volume ratio is defined as

$$\mathrm{R}_d = \frac{p_{Td}(1-p_{Ed})(1-p_{E0d})}{(1-p_{Td})p_{Ed}p_{E0d}}, \quad d = 1, \cdots, D.$$

The dose that yields the lowest value of R_d within the admissible set is the OBD.

In different with the U-BOIN design which makes no dose-response assumption and models each dose level independently, the OR design borrows strength across different dose levels through a Bayesian logistic regression. In particular, the OR design formulates the marginal toxicity and efficacy probabilities p_{Td} and p_{Ed} as

$$\begin{aligned} p_{T1} &= \frac{e^{\phi_1}}{1+e^{\phi_1}}, \quad p_{Td} = \frac{e^{\phi_1}+\cdots+e^{\phi_d}}{1+e^{\phi_1}+\cdots+e^{\phi_d}}, \quad d = 2, \cdots, D; \\ p_{E1} &= \frac{e^{\psi_1}}{1+e^{\psi_1}}, \quad p_{Ed} = \frac{e^{\psi_1+\cdots+\psi_d}}{1+e^{\psi_1+\cdots+\psi_d}}, \quad d = 2, \cdots, D. \end{aligned} \tag{6}$$

Formula (6) indeed specifies a monotonic dose-toxicity pattern. Then, let $\phi = (\phi_1, \cdots, \phi_D)'$ and $\psi = (\psi_1, \cdots, \psi_D)'$, the vector (ϕ, ψ) is assumed to follow a 2D-dimensional multivariate

normal distribution as

$$(\phi, \psi) \sim N(0, \Sigma).$$

The global cross-ratio model is used in the OR design to induce the correlation between the toxicity and efficacy outcomes. The cross ratio is defined as

$$\theta_d = \frac{\pi_{00d}\pi_{11d}}{\pi_{01d}\pi_{10d}}, \quad d = 1, \cdots, D,$$

and each θ_d is assigned to a non-informative gamma prior $\theta_d \sim \text{Gamma}(\zeta, \eta)$.

Similar to the U-BOIN design, to derive the posterior distribution of (ϕ, ψ) and $\theta = (\theta_1, \cdots, \theta_D)$ in the presence of jointly delayed outcomes, we just need to re-parametrize p_{E1d} and p_{E0d} as follows

$$\begin{aligned} a_d &= 1 + (p_{Ed} + p_{Td})(\theta_d - 1) \\ b_d &= -4\theta_d(\theta_d - 1)p_{Ed}p_{Td} \\ p_{E1d} &= \frac{a_d - \sqrt{a_d^2 + b_d}}{\{2(\theta_d - 1)\}p_{Td}}\text{I}(\theta_d \neq 1) + p_{Ed}\text{I}(\theta_d = 1) \\ p_{E0d} &= \frac{p_{Ed} - p_{E1d}}{1 - p_{Td}}. \end{aligned} \tag{7}$$

Then, after plugging in the CWL function (4) with formulas (6) and (7), we should be able to derive the posterior distributions for (ϕ, ψ) and θ and further identify the OBD. We denote this extension of the OR design as the CWL-OR design.

4 Trial Application

4.1 Operating characteristics

In this section, we apply the proposed CWL-U-BOIN design to the motivating TNBC trial and evaluate its operating characteristics. Reporting operating characteristics is often required in trial protocols when a new design is involved. As we mentioned in the introduction, sixty patients will be enrolled in this trial and assigned to five dose levels. Every patient will be followed for a total of three months time period to determine the DLT and OTR outcomes. A new cohort of patients will enter the trial every two months with a cohort size of three and an interim analysis is required at that time to determine the dose level assigned to the new patients.

We use the CWL-U-BOIN design to generate the operating characteristics. The parameter settings used to generate the results in Table 3 of the original U-BOIN paper [15] is also used here to specify p_{Td}, p_{E1d} and p_{E0d}. Figure S1 in the online supplementary materials shown the setting for all the 8 scenarios. The highest acceptable DLT rate is set at $\bar{\pi}_T = 0.3$ and the lowest OTR rate is set at $\bar{\pi}_E = 0.2$. Also, for the toxicity and futility control rules, we specify $\nu_T = 0.95$ and $\nu_E = 0.9$ as recommended by the authors. We use the Weibull distribution to generate the time-to-event outcomes at each dose level. The Weibull distributions are specified to match the response rate under the condition that about two-third of the toxicity and efficacy outcomes will be observable during the first two months follow-up. We compare the proposed CWL-U-BOIN design with two practical designs, the first one uses the observable data only during the interim analysis and we name it the PARTIAL-U-BOIN design. The second one suspends the trial during the interim analysis in the presence of delayed outcome and reopens the trial until all the patients have been completely followed

and we name it the FULL-U-BOIN design.

Table 1 summarized the operating characteristics of the TNBC trial using different designs with 5,000 replicates, including the dose selection probability, the average percentage of patients treated at each dose level, the percentages of patients that experienced DLT and OTR, the average sample size (N) and the average duration of the trial. We specified $\omega_{00} = 30$, $\omega_{01} = 100$, $\omega_{10} = 0$ and $\omega_{11} = 50$ for utility calculation. A variety of dose-efficacy curves have been considered. Scenarios 1 and 2 represent the situation that the dose-efficacy curve increases first and then plateau. Scenarios 3 and 4 represent the umbrella shape situation where the efficacy rate increases first and then keep decreasing. The efficacy rate levels off at the first dose level in scenario 5. The efficacy rate increases monotonically in scenarios 6, 7 and 8. There are two OBDs in scenario 7 with the same utility value of 51.0. Scenario 8 represents the null situation where there is no dose to be recommended. The FULL-U-BOIN yields the best performance in terms of the OBD selection and patient allocation, which is as expected because this design utilizes all the patients' response information from the trial. In general, the difference between the FULL-U-BOIN design and the proposed CWL-U-BOIN is slight, For example, the differences of the OBD selection percentage are 0.5%, 1% and 0.9% for scenarios 1, 2 and 3. On the other hand, the CWL-U-BOIN substantially shorten the trial duration by one third. The CWL-U-BOIN design outperforms the PARTIAL-U-BOIN designs in terms of the OBD selection rate. For example, the former one yields 2.9%, 3.7% and 4.5% higher OBD selection percentages than the latter one in scenarios 1, 2 and 3 and the difference can be as large as 8.5% (scenario 6). The CWL-U-BOIN and PARTIAL-U-BOIN designs report similar trial duration time. When there is no promising dose exiting, the FULL-U-BOIN has almost 90% chance to terminate the trial and the values for CWL-U-BOIN and PARTIAL-U-BOIN designs are about 86% and 84%. In summary,

the FULL-U-BOIN design is slightly better than the CWL-U-BOIN in OBD selection and patient allocation but substantially prolongs the trial duration. The CWL-U-BOIN design also yields better operating characteristics than the PARTIAL-U-BOIN. Hence, based on the results of Table 1 we recommend the proposed CWL-U-BOIN design to be used in practice.

In addition to the extensions of the U-BOIN design, we also report the operating characteristics of the TNBC trial using different extensions of the OR design. Table 2 summarized the results for the CWL-OR, PARTIAL-OR and FULL-OR designs. These designs use the volume ratio as the criteria to select the OBD and a smaller volume ratio indicates better desirability of the dose level. When there is no promising dose (scenarios 7 and 8), all the designs have about 95% chance to terminate the trial. Otherwise, the FULL-OR design in general yields a little higher OBD selection percentage than the CWL-OR design at the price of the substantially prolonged trial duration. When the OBD is located at dose 1 or 2, the CWL-OR design and PARTIAL-OR design have comparable performances. Besides, The CWL-OR design outperforms the PARTIAL-OR design. Indeed, for scenarios 3 to 6, the CWL-OR design yields at least 6% more OBD selection percentage and assigns at least 3 more patients at the OBD. Therefore, based on the results from Table 2 the proposed CWL-OR design is the overall best extension of the OR design. To simplify the presentation, Figures S2 and S3 in the online supplementary materials summarized the OBD selection percentages of Tables 1 and 2 using different designs.

4.2 Sensitivity analysis

We conduct sensitivity analysis to investigate the robustness of the proposed CWL method. Due to the page limit, we focus on the CWL-U-BOIN design and report its OBD selection percentage under a variety of settings. All the results were summarized in the online sup-

plementary materials. First of all, in Figure S4, we depict the OBD selection percentages of the CWL-U-BOIN design with different data generating distributions such as the Weibull, Gamma, and Lognormal. We select the same scenarios as used in Table 1. Then, in Figure S5, we report the OBD selection percentages with different inter-arrival time $\tau = 2$, 1.5 and 1. After that, in Figure S6 we report the OBD selection percentages with different entire follow-up time $(U_T = 3, U_E = 3)$, $(U_T = 3, U_E = 4)$ and $(U_T = 2, U_E = 3)$. Noticing that when $U_T = 2$ and $\tau = 2$, the toxicity outcome can be fully observed and only the efficacy outcomes may be delayed. The results show that the CWL-U-BOIN performs robustly under different data generating distribution, inter-arrival time and total follow-up time with comparable OBD selection percentages.

In the trial application study we generated the time-to-event outcomes in such a way that about two-third of the toxicity and efficacy outcomes were observable during the first two months follow-up (e.g., a missing rate of 33%) when the complete follow-up time is three months, which implied the uniform distribution assumption. In Figure S7 we report the OBD selection percentages with different missing rates violating the uniform distribution assumption. Based on Figure S7 we conclude that the proposed CWL method is a robust method against the uniform distribution assumption.

5 Conclusion

We develop a simple yet efficient conditional weighted likelihood (CWL) method to account for the delayed joint toxicity-efficacy outcomes for phase I/II clinical trials. The proposed CWL method incorporates the toxicity-efficacy correlation into consideration through a multinominal distribution and makes no parametric model assumption on either the dose-

response curve or the toxicity-efficacy correlation. Therefore, the CWL method is highly flexible and can be equipped to any phase I/II clinical trials monitoring the toxicity and efficacy outcomes simultaneously. Also, although the CWL method is originally proposed for the delayed joint outcomes, it can be applied without modification to the setting where only toxicity or efficacy outcome is delayed and the joint distribution is needed to identify the optimal dose. The R code to implement the CWL method is freely available at https://github.com/yongzang2020/CWL.

The CWL method is extendable in several directions. In clinical practice, immnuotherapy and targeted therapy are commonly combined with other drugs. Then, an optimal dose pair is needed to maximize the therapeutic effect due to the synergistic effect and the trial identifying the optimal dose pair is typically referred to as the drug-drug combination trial. A lot of drug-drug combination trial designs have been proposed in the literature [19, 20, 21, 22, 23] yet little attention has been given to the jointly delayed outcome, which can be the first extension of the CWL method. Moreover, in this paper each patient is treated identically for any given dose level. However, an increased understanding of population heterogeneity of cancer has already brought us to the era of personalized medicine, providing the clinicians unbeatable opportunity to select individually tailored treatment taking into account each subject's variability. A lot of personalized clinical trial designs have been proposed with the primary interest in the binary toxicity and efficacy outcomes [24, 25, 26, 27]. Therefore, it is of intense interest to extend the proposed method to integrate personalized information into the trial. We used non-informative priors through this paper under the assumption that the trials under investigation are new and no similar studies had been conducted before. However, as a Bayesian method, the proposed designs retain the flexibility to borrow historical information through prior specification. Specifically, if there are related historical data or real world

evidence available, then the informative priors based on the ideal of prior effective sample size [28] can be constructed to incorporate these information, which are expected to improve the efficiency of the proposed designs.

Acknowledgment

The research of Yong Zang is partially supported by NIH P30 grant CA082709. The authors thank two referees for helpful comments.

Table 1: Operating characteristics of the TNBC trials using the CWL-U-BOIN, PARTIAL-U-BOIN and FULL-U-BOIN designs. The probability pairs in parentheses are the probabilities of occurring DLT and OTR at each dose level (p_{Td}, p_{Ed}). The percentage of stopping the trial with no dose selected is denoted by "None". The percentages of patients experiencing DLT and OTR is denoted by DLT/OTR(%).

		Dose Level								
Design		1	2	3	4	5	None	DLT/OTR (%)	N	Duration (month)
Scenario 1	(p_{Td}, p_{Ed})	(0.02,0.2)	(0.15,0.65)	(0.3,0.65)	(0.45,0.65)	(0.6,0.65)				
	Utility	43.0	69.0	63.0	56.0	50.0				
CWL-U-BOIN	Selection %	2.7	**72.9**	21.6	1.8	0.3	0.7			
	No. of patients	6.1	**31.5**	15.3	5.0	1.8		21.5/60.3	59.7	40.8
PARTIAL-U-BOIN	Selection %	4.9	**70.0**	20.1	2.2	0.5	2.3			
	No. of patients	8.3	**30.9**	14.0	4.2	1.9		20.1/58.5	59.3	41.5
FULL-U-BOIN	Selection %	1.6	**73.4**	22.5	1.9	0.3	0.3			
	No. of patients	6.2	**32.3**	15.2	4.6	1.6		20.9/60.4	60.0	60.0
Scenario 2	(p_{Td}, p_{Ed})	(0.03,0.1)	(0.08,0.22)	(0.15,0.6)	(0.28,0.6)	(0.4,0.6)				
	Utility	36.0	43.0	66.0	60.0	55.0				
CWL-U-BOIN	Selection % d	1.1	3.0	**67.3**	22.6	4.8	1.2			
	No. of patients	4.6	6.9	**27.2**	14.5	6.5		19.2/51.6	59.6	40.7
PARTIAL-U-BOIN	Selection %	0.5	5.1	**63.6**	24.8	4.8	1.2			
	No. of patients	5.5	8.6	**25.7**	14.0	5.8		18.1/50.0	59.7	41.8
FULL-U-BOIN	Selection %	0.2	2.1	**68.3**	24.6	4.6	0.2			
	No. of patients	4.3	7.1	**27.8**	14.6	6.1		18.7/52.0	59.9	59.9
Scenario 3	(p_{Td}, p_{Ed})	(0.05,0.08)	(0.15,0.46)	(0.3,0.25)	(0.45,0.2)	(0.6,0.1)				
	Utility	34.0	56.0	37.0	29.0	18.0				
CWL-U-BOIN	Selection %	1.9	**91.4**	3.4	0.8	0.1	2.4			
	No. of patients	6.7	**36.1**	9.9	4.5	2.0		20.0/34.9	59.3	40.5
PARTIAL-U-BOIN	Selection %	1.6	**86.9**	4.9	0.6	0.2	5.8			
	No. of patients	7.6	**35.5**	9.1	4.0	1.9		19.4/34.9	58.1	40.7
FULL-U-BOIN	Selection %	1.2	**92.3**	4.5	0.6	0.1	1.3			
	No. of patients	6.6	**36.9**	9.8	4.5	1.8		19.8/35.5	59.5	59.5
Scenario 4	(p_{Td}, p_{Ed})	(0.15,0.15)	(0.25,0.45)	(0.4,0.3)	(0.45,0.25)	(0.5,0.2)				
	Utility	36.0	52.0	36.0	32.0	27.0				
CWL-U-BOIN	Selection %	13.0	**69.7**	4.6	2.1	1.8	8.8			
	No. of patients	13.4	**29.7**	7.3	3.8	2.4		26.9/33.6	56.6	38.7
PARTIAL-U-BOIN	Selection %	13.3	**64.5**	6.7	3.0	1.3	11.2			
	No. of patients	15.6	**26.5**	6.9	4.2	3.0		27.1/32.0	56.3	39.4
FULL-U-BOIN	Selection %	11.8	**73.2**	5.4	3.5	2.0	4.1			
	No. of patients	13.3	**30.7**	7.6	4.5	2.8		27.3/33.5	59.0	59.0
Scenario 5	(p_{Td}, p_{Ed})	(0.1,0.45)	(0.3,0.45)	(0.5,0.45)	(0.55,0.45)	(0.65,0.45)				
	Utility	58.0	50.0	42.0	40.0	36.0				
CWL-U-BOIN	Selection %	**78.6**	18.0	1.1	0.5	0.6	1.2			
	No. of patients	**36.0**	15.4	4.8	2.1	1.0		21.3/44.7	59.4	40.6
PARTIAL-U-BOIN	Selection %	**78.4**	17.8	1.2	0.6	0.4	1.6			
	No. of patients	**36.5**	14.6	4.4	2.2	1.5		20.9/44.8	59.2	41.4
FULL-U-BOIN	Selection %	**79.5**	18.3	0.7	0.7	0.7	0.1			
	No. of patients	**36.1**	16.8	4.6	1.6	0.8		20.7/45.2	59.9	59.9
Scenario 6	(p_{Td}, p_{Ed})	(0.05,0.35)	(0.07,0.45)	(0.1,0.5)	(0.12,0.55)	(0.16,0.75)				
	Utility	53.0	59.0	61.0	64.0	75.0				
CWL-U-BOIN	Selection %	3.0	7.7	9.7	13.9	**65.4**	0.3			
	No. of patients	6.1	7.9	8.3	10.4	**27.2**		12.3/60.0	59.8	40.9
PARTIAL-U-BOIN	Selection %	4.5	11.2	13.0	14.2	**56.9**	0.2			
	No. of patients	7.2	9.3	10.0	10.4	**23.0**		11.8/58.0	59.9	41.9
FULL-U-BOIN	Selection %	2.4	6.5	9.7	11.2	**70.2**	0.0			
	No. of patients	5.6	7.7	8.5	9.5	**28.7**		12.1/60.8	60.0	60.0

Table 1 continues.

Design		Dose Level 1	2	3	4	5	None	OTR/DLT (%)	N	Duration (month)
Scenario 7	(p_{Td}, p_{Ed})	(0.03,0.15)	(0.16,0.38)	(0.27,0.45)	(0.45,0.6)	(0.55,0.7)				
	Utility	40.0	51.0	51.0	53.0	55.0				
CWL-U-BOIN	Selection %	6.9	**46.6**	**34.3**	7.4	2.9	1.9			
	No. of patients	8.0	**21.3**	**17.1**	8.8	4.2		24.4/42.5	59.4	40.6
PARTIAL-U-BOIN	Selection %	7.3	**46.5**	**29.5**	8.4	3.9	4.4			
	No. of patients	9.8	**22.6**	**15.7**	6.8	3.9		22.7/40.7	58.7	41.1
FULL-U-BOIN	Selection %	7.3	**46.9**	**33.1**	8.7	3.1	0.9			
	No. of patients	8.5	**22.4**	**17.0**	8.0	3.8		24.0/41.8	59.8	59.8
Scenario 8	(p_{Td}, p_{Ed})	(0.22,0.03)	(0.45,0.1)	(0.55,0.2)	(0.65,0.35)	(0.7,0.4)				
	Utility	25.0	23.0	25.0	30.0	31.0				
CWL-U-BOIN	Selection %	2.4	6.8	3.9	0.7	0.6	85.6			
	No. of patients	12.9	11.2	6.8	4.6	3.6		43.9/15.5	39.1	26.3
PARTIAL-U-BOIN	Selection %	2.7	7.4	3.5	1.8	1.0	83.6			
	No. of patients	12.7	10.0	6.4	4.7	4.3		44.4/15.7	38.1	26.6
FULL-U-BOIN	Selection %	1.7	4.7	2.3	1.1	0.7	89.5			
	No. of patients	12.8	10.6	6.1	4.9	4.0		44.1/15.6	38.5	38.5

Table 2: Operating characteristics of the TNBC trials using the CWL-OR, PARTIAL-OR and FULL-OR designs. The probability pairs in parentheses are the probabilities of occurring DLT and OTR at each dose level (p_{Td}, p_{Ed}). The percentage of stopping the trial with no dose selected is denoted by "None". The percentages of patients experiencing DLT and OTR is denoted by DLT/OTR(%).

		Dose Level								
Design		1	2	3	4	5	None	DLT/OTR (%)	N	Duration (month)
Scenario 1	(p_{Td}, p_{Ed})	(0.05,0.4)	(0.25,0.45)	(0.28,0.45)	(0.35,0.4)	(0.5,0.3)				
	Volume Ratio	0.12	0.95	1.43	2.42	7.00				
CWL-OR	Selection %	**67.1**	14.5	8.9	4.0	0.9	4.6			
	No. of patients	**32.3**	10.5	6.9	5.0	3.0		16.2/40.7	57.8	40.5
PARTIAL-OR	Selection %	**65.4**	15.4	6.8	3.4	1.2	7.8			
	No. of patients	**31.9**	10.7	5.9	4.6	3.1		15.8/40.6	56.1	39.3
FULL-OR	Selection %	**68.4**	15.2	9.1	3.5	0.9	2.9			
	No. of patients	**33.2**	10.6	7.2	5.0	2.9		16.3/40.8	58.9	58.9
Scenario 2	(p_{Td}, p_{Ed})	(0.15,0.3)	(0.2,0.5)	(0.22,0.4)	(0.25,0.35)	(0.4,0.3)				
	Volume Ratio	1.24	0.20	0.79	1.44	3.63				
CWL-OR	Selection %	9.3	**63.9**	12.2	5.6	1.7	7.3			
	No. of patients	9.8	**27.3**	8.6	5.7	4.7		21.4/41.6	56.1	39.3
PARTIAL-OR	Selection %	9.0	**64.5**	9.5	3.4	2.3	11.3			
	No. of patients	9.5	**27.3**	7.7	5.0	4.4		21.4/42.1	53.9	37.8
FULL-OR	Selection %	10.0	**64.5**	13.7	3.2	1.5	7.1			
	No. of patients	9.8	**28.2**	8.6	5.4	4.3		21.2/41.8	56.3	56.3
Scenario 3	(p_{Td}, p_{Ed})	(0.1,0.15)	(0.15,0.3)	(0.2,0.6)	(0.25,0.45)	(0.28,0.35)				
	Volume Ratio	5.67	1.65	0.07	0.41	0.88				
CWL-OR	Selection %	0.6	2.1	**74.8**	11.0	4.7	6.8			
	No. of patients	4.5	4.9	**31.8**	8.9	6.4		20.4/48.7	56.5	39.6
PARTIAL-OR	Selection %	1.2	3.1	**66.4**	10.7	5.8	12.8			
	No. of patients	5.2	5.6	**28.0**	8.5	6.2		20.2/47.2	53.4	37.5
FULL-OR	Selection %	0.6	2.9	**75.0**	12.5	3.8	5.2			
	No. of patients	4.1	5.3	**32.8**	9.6	5.5		20.7/48.9	57.3	57.3
Scenario 4	(p_{Td}, p_{Ed})	(0.1,0.2)	(0.15,0.3)	(0.18,0.4)	(0.2,0.7)	(0.25,0.6)				
	Volume Ratio	2.52	1.24	0.77	0.03	0.18				
CWL-OR	Selection %	1.2	2.2	1.6	**76.8**	12.1	6.1			
	No. of patients	5.4	5.3	4.6	**32.0**	9.5		19.2/57.5	56.8	39.8
PARTIAL-OR	Selection %	1.2	1.1	2.1	**70.0**	11.5	14.1			
	No. of patients	5.4	5.5	4.8	**28.6**	8.2		19.3/56.3	52.5	36.9
FULL-OR	Selection %	1.4	2.4	2.8	**77.8**	12.0	3.6			
	No. of patients	4.8	5.3	4.8	**33.9**	9.3		19.4/58.2	58.0	58.0
Scenario 5	(p_{Td}, p_{Ed})	(0.12,0.32)	(0.13,0.35)	(0.15,0.5)	(0.18,0.55)	(0.2,0.7)				
	Volume Ratio	0.87	0.65	0.22	0.18	0.03				
CWL-OR	Selection %	3.2	3.4	9.6	9.3	**70.3**	4.2			
	No. of patients	6.7	5.7	8.2	8.4	**28.6**		17.5/57.1	57.7	40.4
PARTIAL-OR	Selection %	4.0	2.6	8.6	10.7	**63.6**	10.5			
	No. of patients	7.5	6.1	8.0	8.2	**24.5**		17.0/55.9	54.3	38.1
FULL-OR	Selection %	2.3	3.1	8.9	11.1	**71.3**	3.3			
	No. of patients	6.0	5.5	8.2	8.4	**30.1**		17.7/57.7	58.1	58.1
Scenario 6	(p_{Td}, p_{Ed})	(0.1,0.27)	(0.15,0.3)	(0.2,0.4)	(0.25,0.6)	(0.4,0.6)				
	Volume Ratio	1.20	1.24	0.70	0.07	0.44				
CWL-OR	Selection %	7.7	4.8	7.6	**67.1**	8.3	4.5			
	No. of patients	8.5	6.3	7.2	**27.9**	7.8		23.0/49.5	57.7	40.4
PARTIAL-OR	Selection %	7.0	5.1	7.9	**60.9**	8.5	10.6			
	No. of patients	8.1	6.8	7.1	**25.0**	7.5		22.9/48.6	54.6	38.3
FULL-OR	Selection %	7.0	5.1	7.3	**68.2**	8.1	4.3			
	No. of patients	7.6	6.6	7.1	**29.0**	7.4		23.0/49.7	57.7	57.7

Table 2 continues.

Design		Dose Level 1	2	3	4	5	None	DLT/OTR (%)	N	Duration (month)
Scenario 7	(p_{Td}, p_{Ed})	(0.15,0.03)	(0.2,0.05)	(0.23,0.08)	(0.28,0.1)	(0.45,0.1)				
	Volume Ratio	184.5	90.3	65.3	40.4	66.3				
CWL-OR	Selection %	0.0	0.2	0.9	2.8	0.6	95.5			
	No. of patients	3.8	4.2	5.0	5.5	4.2		26.5/7.4	22.7	16.2
PARTIAL-OR	Selection %	0.2	0.6	1.3	3.6	0.5	93.8			
	No. of patients	4.0	5.3	5.2	6.2	5.5		26.8/7.7	26.2	18.5
FULL-OR	Selection %	0.2	0.1	1.4	1.6	0.7	96.0			
	No. of patients	3.9	4.2	4.8	5.4	4.1		26.6/7.4	22.3	22.3
Scenario 8	(p_{Td}, p_{Ed})	(0.45,0.28)	(0.5,0.3)	(0.6,0.35)	(0.7,0.4)	(0.8,0.5)				
	Volume Ratio	8.4	8.3	8.4	8.2	9.3				
CWL-OR	Selection %	3.5	0.8	0.2	0.0	0.0	95.5			
	No. of patients	9.4	4.5	1.8	0.7	0.1		48.9/30.5	16.5	12.1
PARTIAL-OR	Selection %	3.4	0.8	0.2	0.0	0.1	95.5			
	No. of patients	10.2	5.2	1.9	0.6	0.1		48.6/30.1	18.0	13.1
FULL-OR	Selection %	2.9	0.6	0.2	0.0	0.0	96.3			
	No. of patients	9.4	4.7	1.8	0.6	0.1		48.9/30.3	16.5	16.5

References

[1] Korn EL. Nontoxicity endpoints in phase I trial designs for targeted, non-cytotoxic agents. *Journal of the National Cancer Institution* 2004; **96**: 977-978.

[2] Parulekar WR, Eisenhauer EA. Phase I trial design for solid tumor studies of targeted, non-cytotoxic agents: Theory and practice. *Journal of the National Cancer Institution* 2004; **96**: 990-997.

[3] Tourneau CL, Lee, JJ, Siu LL. Dose escalation methods in phase I cancer clinical trials. *Journal of the National Cancer Institution* 2009; **101**: 708-720.

[4] LoRusso PM, Boerner SA, Seymour, L. An overview of the optimal planning, design, and conduct of phase I studies of new therapeutics. *Clinical Cancer Research* 2010; **16**: 1710-1718.

[5] Reynolds A. Potential relevance of bell-shaped and u-shaped dose-responses for the therapeutic targeting of angiogenesis in cancer.*Dose-response* 2010; **8(3)**.

[6] deLima M, Champlin RE, Thall PF et al. Phase I/II study of gemtuzumab ozogamicin added to fludarabine, melphalan and allogeneic hematopoietic stem cell transplantation for high-risk CD33 positive myeloid leukemias and myelodysplastic syndrome. *Leukemia* 2008; **22**: 258-264.

[7] Whelan HT, Cook JD, Amlie-Lefond CM et al. Practical model-based dose-finding in early phase clinical trials: Optimizing tissue plasminogen activator dose for treatment of ischemic stroke in children. *Stroke* 2008; **39(9)**: 2627-2636

[8] Konopleva M, Thall PF, Yi CA et al. Phase I/II Study of PR104, hypoxiaactivated prodrug in refractory/relapsed acute myeloid leukemia and acute lymphoblastic leukemia. *Haematologica* 2015; **100(7)**: 375-385.

[9] Wang P, Burikhanov R, Jayswal R, et al. Neoadjuvant administration of hydroxychloroquine in a phase 1 clinical trial induced plasma Par-4 levels and apoptosis in diverse tumors. *Genes Cancer* 2018; **9**: 190-197.

[10] Yuan Y, Yin G. Robust EM Continual Reassessment Method in Oncology Dose Finding. *Journal of the American Statistical Association* 2011; **106(495)**: 818-831.

[11] Cheung YK, Chappell R..Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* 2000; **56**: 1177-1182.

[12] Lin R, Yuan Y. Time-to-event model-assisted designs for dose-finding trials with delayed toxicity. *Biostatistics* 2019; in press.

[13] Liu S, Yin G, Yuan Y. Bayesian data augmentation dose finding with continual reassessment method and delayed toxicity. *Annals of Applied Statistics* 2013; **4**: 2138-2156.

[14] Yuan Y, Lin R, Li D, Nie L, Warren K. Time-to-event Bayesian optimal interval design to accelerate phase I trials. *Clinical Cancer Research* 2018; **24**: 4921-4930.

[15] Zhou Y, Lee JJ, Yuan Y. A utility-based Bayesian optimal interval (U-BOIN) phase I/II design to identify the optimal biological dose for targeted and immune therapies. *Statistics in Medicine* 2019; in press.

[16] Liu S, Yuan Y. Bayesian optimal interval designs for phase I clinical trials. *Journal of the Royal Statistical Society: Series C* 2015; **64(3)**: 507-523.

[17] Yin G, Li Y, Ji Y. Bayesian dose-finding in phase I/II clinical trials using toxicity and efficacy odds ratios. *Biometrics* 2006; **62**: 777-787.

[18] Dale JR. Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics* 1986; **42**: 909-917.

[19] Conaway MR, Dunbar S, Peddada SD. Designs for single- or multiple-agent phase I trials. *Biometrics* 2004; **60(3)**: 661-669.

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

[21] Braun TM, Wang SF. A hierarchical bayesian design for phase I trials of novel combinations of cancer therapeutic agents. *Biometrics* 2010; **66(3)**: 805-812.

[22] Wages NA, Conaway MR, OQuigley J. Continual reassessment method for partial ordering. *Biometrics* 2011; **67(4)**: 1555-1563.

[23] Riviere MK, Yuan Y, Dubois F, Zohar S. A bayesian dose-finding design for clinical trials combining a cytotoxic agent with a molecularly targeted agent. *Journal of the Royal Statistical Society: Series C* 2015; **64**: 215-229.

[24] Piantadosi S, Liu G. Improved designs for dose-escalation studies using pharmacokinetic measurements. *Statistics in Medicine* 1996; **15**: 1605-1618.

[25] Babb JS, Rogatko A. Patient specific dosing in a phase I cancer trial. *Statistics in Medicine* 2001; **20**: 2079-2090.

[26] Thall P, Nguyen H, Estey E. Patient-specific dose finding based on bivariate outcomes and covariates. *Biometrics* 2008; **64**: 1126-1136.

[27] Guo B, Yuan Y. Bayesian phase I/II biomarker-based dose finding for precision medicine with molecularly targeted agents. *Journal of the American Statistical Association* 2017; **112**: 508-520.

[28] Morita S, Thall P, Muller P. Determining the effective sample size of a parametric prior. *Biometrics* 2008; **64**: 595-602.