Adaptive Designs for Identifying Optimal Biological Dose for Molecularly Targeted Agents

Yong Zang, J Jack Lee and Ying Yuan*†

Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, TX, USA

∗ Correspondence to : Ying Yuan, Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, TX 77030, U.S.A.

† Email: yyuan@mdanderson.org

Background Traditionally, the purpose of a dose-finding design in cancer is to find the maximum tolerated dose based solely on toxicity. However, for molecularly targeted agents (MTAs), little toxicity may arise within the therapeutic dose range and the dose-efficacy curves may not be monotonic. This challenges the principle that more is better, which is widely accepted for conventional chemotherapy.

Methods We propose three adaptive dose-finding designs for trials evaluating MTAs. The goal of these designs is to find the optimal biological dose (OBD), which is defined as the lowest dose with the highest rate of efficacy while safe. The first proposed design is parametric and assumes a logistic dose-efficacy curve for dose finding; the second design is nonparametric and uses the isotonic regression to identify the optimal biological dose; and the third design has the spirit of a "semiparametric" approach by assuming a logistic model only locally around the current dose.

Results We conducted extensive simulation studies to investigate the operating characteristics of the proposed designs. Simulation studies show that the nonparametric and semiparametirc designs have good operating characteristics for finding the OBD and outperform the existing slope-sign design.

Limitations The proposed designs assume a binary endpoint. Extension of the proposed designs to ordinal and time-to-event endpoints worth further investigation.

Conclusion Among the three proposed designs, the nonparametric and semiparametirc designs yield consistently good operating characteristics and thus are recommended for practical use. The software to implement these two designs is available for free download at *http://odin.mdacc.tmc.edu/∼yyuan/.*

1 Introduction

Traditionally, the primary goal of a phase I cancer clinical trial for a cytotoxic drug is to identify the maximum tolerated dose (MTD) of the new agent, based on the assumption that both efficacy and toxicity increase monotonically with the dose. The recent development of novel molecularly targeted agents (MTAs) has challenged this paradigm as the monotonic assumption of dose-toxicity and dose-efficacy relationships may not hold for MTAs. MTAs are developed to modulate specific aberrant pathways in cancer cells while sparing normal tissue. As a result, the toxicities of the MTAs are often expected to be minimal within the therapeutic dose range and the dose-efficacy curves of MTAs may not follow monotonic patterns [1, 2, 3, 4]. For example, Postel-Vinay et al.[5] investigated the dose-efficacy relationship of monotherapy MTAs based on 135 patients enrolled in phase I trials at the Royal Marsden Hospital from 5 January 2005 to 6 June 2006. The patients were classified into three cohorts A, B and C according to the dose received as a percentage of the MTD (0-33%, 34-65%, >66%). The efficacy endpoint was the non-progression rate, i.e., complete/partial

response or stable disease for at least 3 months. The monotherapy MTAs demonstrated a nonmonotonic dose-efficacy relationship: the NPR for the patients in cohorts A, B and C were 21%, 50% and 31% at 3 months and 11%, 27% and 14% at 6 months after receiving the treatments. That is, the median dose of the MTA (i.e., cohort B), rather than the highest dose (i.e., cohort C), leads to the highest efficacy.

The primary goal of a dose-finding design for MTAs is to find the optimal biological dose (OBD), which can be defined in different ways according to the goal of the clinical trial. In this article, we define the OBD as the lowest dose that has the highest effectiveness efficacy rate while simultaneously safeguarding patients. This definition has been used by some existing trial designs [6, 7]. Due to the nonmonotonic dose-efficacy relationship, the OBD of an MTA is not always the highest dose and may appear in the middle of the investigational dose range. This challenges the principle that more is better, which is widely accepted for conventional chemotherapy [8]. In practice, the dose-efficacy curves for MTAs are often expected to be unimodal (or umbrella-shaped[9]) or to plateau within the therapeutic dose range, that is, the dose-efficacy curves first increase and then remain constant as the dose increases. Although more complicated multimodal dose-efficacy curves are possible, they rarely occur within the therapeutic dose range. Thus, in this paper, we focus on finding the OBD for MTAs with unimodal or plateaued dose-efficacy curves.

In contrast to the rich body of literature on phase I dose-finding trial designs for cytotoxic agents, research on phase I trial designs for MTAs is limited [10, 11]. Under the assumption of minimal toxicity, Hunsberger et al. [6] proposed the slope-sign design to find the biological adequate dose (BAD) for MTAs, which has similar definition as OBD but is defined as the lowest dose in the plateau. The slope-sign design directs the dose finding based on the sign of the estimate of the slope of the dose-efficacy curve using the efficacy data collected from a certain number of adjacent dose levels. If the sign of the estimate of the slope is positive, the dose level is escalated; otherwise, the trial is terminated and the dose level with the

highest efficacy response rate is recommended as the BAD. This slope-sign design is simple, and requires small sample size because it terminates the trial once the estimate of the slope is negative. Due to different trial design goals, some designs adopted other definitions of the OBD for dose finding. Zhang et al. [12] proposed the trinomial continual reassessment method to find the optimal biological dose, defined as the dose with the highest probability of yielding efficacy but no toxicity. Mandrekar et al. [13] extended this method as a way of investigating dual-agent drug combinations. Polley and Cheung [14] proposed a two-stage design with a futility interim analysis for identifying the optimal dose, which was defined as the minimum dose that exhibited adequate drug activity.

In this article, we propose three new dose-finding designs to search for the OBD of MTAs. Be different with Hunsberger et al. [6], we investigate the case in which the MTA under investigation has substantial toxicity; thus both efficacy and toxicity need to considered for dose finding. Our first design models the entire dose-efficacy curve with a Bayesian logistic regression and adaptively assigns patients based on the model estimates. To improve the robustness of the design, the second design fits the dose-efficacy curve with a isotonic regression without making any parametric assumption as to the shape of the curve. The third proposed design has the spirit of a "semiparametric" approach by assuming a logistic model only locally around the current dose.

The rest of this article is organized as follows. In Section 2 we propose three new designs to find the OBD of MTAs. In Section 3 we carry out comprehensive simulation studies to evaluate the performance of the proposed designs. Based on the simulation studies, we provide concluding comments and recommendations in Section 4.

2 Dose-finding methods

2.1 Toxicity monitoring

Compared to conventional cytotoxic agents, toxicity is often of less concern for MTAs and the dose finding of the OBD is mainly driven by efficacy. However, it is still important to monitor toxicity during the dose finding to ensure patients' safety. Let $(d_1, \ldots, d_J)$ denote a set of J pre-specified increasing doses for the MTA under investigation and and q_j denote the toxicity probability of dose level j. Assuming that x_j out n_j patients experienced toxicity at dose level j, we model the toxicity of each dose level independently using a beta-binomial model

$$\begin{aligned} x_j &\sim binom(n_j, q_j) \\ q_j &\sim beta(a, b) \end{aligned}$$

where $binom(\cdot)$ and $beta(\cdot)$ denote binomial and beta distributions, respectively, and a and b are hyperparameters. In most applications, a vague prior should be preferred by setting $a+b$ at a small number, such as 0.5, noting that $a+b$ can be interpreted as a prior sample size.

Let ϕ denote the target toxicity upper bound, the safety of dose level j can be monitored by the posterior probability $\Pr(q_j > \phi | n_j, x_j)$, which has a convenient closed-form $\Pr(q_j > \phi | n_j, x_j) = 1 - Beta(\phi; a + x_j, b + n_j - x_j)$ where $Beta(\cdot)$ is the cumulative density function of a beta distribution evaluated at ϕ with parameters $a + x_j$ and $b + n_j - x_j$. Specifically, we define the admissible set $\mathcal{A}$ as a set of doses satisfying the following safety rule:

$$\mathcal{A} = \{j : \widetilde{\Pr}(q_j > \phi | n_j, x_j) < C_T, j = 1, \ldots, J\} \tag{1}$$

where $\widetilde{\Pr}(q_j > \phi | n_j, x_j)$ is the isotonically transformed posterior probability $\Pr(q_j > \phi | n_j, x_j)$ based on the pool adjacent violators algorithm (PAVA) [16, 17], and C_T is a pre-specified toxicity threshold. The isotonic transformation is used to impose dose-toxicity monotonicity and borrow information across dose levels. As described in the following sections, during the dose finding, we restrict dose assignment and selection within the admissible set $\mathcal{A}$, thereby protecting patients from overly toxic doses. Before treating any patient in the trial, all investigational doses should be in $\mathcal{A}$ and open for testing. This can be done by choosing the values of hyper parameters a and b such that $Beta(\phi; a, b) = 1 - C_T + \delta$, where δ is a small positive number, e.g., $\delta = 0.05$. That is, *a priori*, all doses just satisfy the safety rule given in (1).

2.2 Logistic design

We assume that efficacy is recorded as a binary outcome. Let p_j denote the probability of efficacy at dose level j. We note that like many phase I trial designs (e.g., CRM), the dose values $\{d_j\}$ are not actual clinical dose values, but are rescaled dose values, for example, with some pre-specified mean and standard deviation. The rescaling of the dose improves model estimation stability and also facilitates the prior elicitation, as we will discuss later. Because the dose-efficacy curve for an MTA is often non-monotonic, an intuitive method for fitting this curve is the logistic regression with a quadratic term, which can be written as

$$\log\left(p_j/(1-p_j)\right) = \alpha + \beta d_j + \gamma d_j^2, \quad j = 1, \ldots, J.$$

Assume that at dose level j, y_j out of n_j patients experienced efficacy, the likelihood for the observed data $\mathcal{D} = \{y_j, n_j; j = 1, \ldots, J\}$ is given by

$$L(\mathcal{D}|\alpha, \beta, \gamma) \propto \prod_{j=1}^{J} \left(\frac{e^{\alpha+\beta d_j+\gamma d_j^2}}{1 + e^{\alpha+\beta d_j+\gamma d_j^2}} \right)^{y_j} \left(\frac{1}{1 + e^{\alpha+\beta d_j+\gamma d_j^2}} \right)^{n_j - y_j};$$

and the posterior of unknown regression parameters α, β and γ is

$$f(\alpha, \beta, \gamma|\mathcal{D}) \propto f(\alpha, \beta, \gamma) L(\mathcal{D}|\alpha, \beta, \gamma). \tag{2}$$

where $f(\alpha, \beta, \gamma)$ is the prior distribution of α, β and γ.

The specification of the prior $f(\alpha, \beta, \gamma)$ requires some care. In typical phase I trials, data are sparse, especially at the beginning of the trial when only a few patients are enrolled. As a result, the commonly used noninformative flat prior (or vague normal prior with a large variance) tends to lead to improper posteriors and unstable estimation, and thus should be avoided here. To address the issue of sparse data, we adopt the weakly informative prior proposed by Gelman et al [15]. Specifically, we specify a Cauchy prior distribution with center 0 and scale 10 for α, i.e., $Cauchy(0, 10)$; and two independent Cauchy priors $Cauchy(0, 2.5)$ for β and γ, respectively. These weakly informative priors are regularized such that a dramatic change in efficacy probability (e.g., from 0.01 to 0.5) is unlikely to happen when the dose changes by one level, which substantially improves the estimation stability while also being vague enough to ensure that the data are able to dominate the priors when more data accumulate. In order to use these priors, we need to fix the scale of the covariate d_j, which is done by standardizing the values of $(d_1, \ldots, d_J)$ with a mean of 0 and a standard deviation of 0.5. The posterior distribution (2) can be easily sampled based on the Markov chain Monte Carlo (MCMC) method and used for guiding dose escalation and de-escalation.

The proposed Bayesian logistic model-based dose-finding method can be summarized as

follows:

1. The first cohort of patients is treated at the lowest dose d_1, or at the physician-specified dose.

2. At the current dose level j, based on the toxicity outcomes, applying safety rule (1) to find the admissible set $\mathcal{A}$.

3. Identify the dose level j^* which has the highest posterior estimate of efficacy probability within $\mathcal{A}$. If $j^* > j$, the dose level is escalated to $j + 1$; if $j^* < j$, the dose level is de-escalated to $j - 1$; otherwise, the dose level remains at j.

4. Once the maximum sample size is reached, the dose that has the highest estimate of efficacy probability within $\mathcal{A}$ is selected as the OBD.

This parametric design is straightforward and easy to implement, but may be sensitive to model misspecifications. In the subsection that follows, we propose a nonparametric approach based on the isotonic regression.

2.3 Isotonic design

Within conventional phase I trial designs for cytotoxic agents, various isotonic designs have been proposed to find the MTD without making any parametric assumption beyond the monotonicity on the dose-toxicity curve. These designs have been based on isotonic regression, which is typically implemented using the PAVA. For example, Leung and Wang [18] proposed an isotonic design that fits an isotonic regression to accumulated data to determine dose escalation based on the PAVA. Conaway et al. [19] suggested isotonic designs for drug combination trials. Yuan and Chappell [20] proposed isotonic designs for trials with ordered groups. The existing isotonic designs cannot be directly used to find the OBD for MTAs because in order to conduct isotonic regression, these designs require

that the dose-response order constraint is known (e.g., monotonicity), which may not be satisfied for MTAs.

Specifically, for the MTAs with unimodal or plateaued dose-efficacy curves, our goal is to find the OBD, the dose level j^* such that

$$p_1 \leq \cdots \leq p_{j^*} \geq \cdots \geq p_J. \tag{3}$$

In other words, before we identify the OBD, the order constraint (3) is actually unknown. Therefore, the standard isotonic regression cannot be directly applied to find the OBD. To overcome this difficulty, we take the approach of double-sided isotonic regression [21]. In this approach, we first enumerate all J possible locations of j^*, $j^* = 1, \ldots, J$. Given each of the locations, say $j^* = k$, the isotonic estimates $\{\hat{p}_j^{(k)}; j = 1, \ldots, J\}$ can be obtained by fitting two separate standard isotonic regressions: one for $p_1, \ldots, p_{j^*}$ with the constraint $p_1 \leq \cdots \leq p_{j^*}$ and the other one for $p_{j^*+1}, \ldots p_J$ with the constraint $p_{j^*+1} \geq \cdots \geq p_J$. Each of these two isotonic regressions can be done using the PAVA algorithm[16, 17], which yields isotonic estimates by replacing any adjacent observations violated the monotonicity assumption with their (weighted) average. For example, assuming that the observed efficacy rates at three doses are (1/10, 3/10, 1/5), under the monotonically increasing order assumption, the efficacy rates of dose levels 2 and 3 (i.e., 3/10 and 1/5) violate the monotonicity assumption. The PAVA algorithm replaces 3/10 and 1/5 with their average 4/15, yielding the isotonic estimates (1/10, 4/15, 4/15).

After applying this procedure to each of J possible locations of j^*, we obtain J sets of possible isotonic estimates $\{\hat{p}_j^{(k)}\}, k = 1, \ldots, J$. We select as the final isotonic estimates $\{\hat{p}_j\} = \{\hat{p}_j^{(k^*)}\}$, the set of isotonic estimates with the smallest sum of the square error, i.e.,

$$k^* = \text{argmin}_{k \in (1,\ldots,J)} \sum_{j=1}^{J} (\hat{p}_j^{(k)} - \frac{y_k}{n_k})^2$$

Our isotonic design can be described as follows:

1. Treat the first cohort of patients at the lowest dose d_1, or at the physician-specified dose.

2. At the current dose level j, based on the toxicity outcomes, applying safety rule (1) to find the admissible set $\mathcal{A}$.

3. Identify the dose level j^* that has the highest isotonic estimate of efficacy probability among the tried doses within $\mathcal{A}$. Where there are ties, we select j^* as the lowest dose level among the ties. Let j^t denote the highest dose level tried thus far. If $j^* > j$, we escalate the dose level to $j+1$; if $j^* < j$, we de-escalate the dose level to $j-1$; if $j^* = j = j^t$, we escalate the dose level to $j+1$ given that $j+1$ is in $\mathcal{A}$, otherwise we retain the dose level j.

4. Once the maximum sample size is reached, we select as the OBD the lowest dose that has the highest estimate of efficacy probability within $\mathcal{A}$.

One limitation of the isotonic regression is that it cannot estimate the efficacy probabilities for the untried doses at which no patients have been treated. That is, the isotonic regression cannot extrapolate the dose-efficacy curve beyond the range of the observed data. Therefore, during the trial conduct, when the dose with the highest estimate of efficacy is the highest tried dose (i.e., $j^* = j = j^t$), there is no information to determine whether we have achieved the maximum point of the dose-efficacy curve or not. To overcome this limitation, in the above dose-finding algorithm, when $j^* = j = j^t$ we automatically escalate the dose level to further explore the dose-efficacy curve and search the maximum point, given that the next higher dose level is safe (i.e., within the admissible set $\mathcal{A}$).

2.4 Local logistic design

We have proposed designs based on parametric logistic regression and nonparametric isotonic regression. We now propose the third approach, which we call the L-logistic design to denote the local logistic regression design. The L-logistic design can be regarded as a "semi-parametric" approach in the sense that, rather than imposing a parametric dose-efficacy curve that spans the whole dose range, it assumes a parametric form around only the neighbors of the current dose. It is worthy noting that for our purpose, we are not interested in estimating the whole dose-response curve but only the local trend around the current dose in order to direct the dose escalation or de-escalation. Our approach can also be viewed as a Bayesian extension of the slope-sign design proposed by Hunsberger et al. [6].

In the L-logistic design, we fit a local Bayesian linear logistic model based on the local data $\mathcal{D}_j = \{y_k, n_k; k = j - l + 1, \ldots, j\}$ collected from the current dose level j up to the previous $l - 1$ dose levels as,

$$\begin{aligned} \log\left(p_k/(1-p_k)\right) &= \alpha + \beta d_k, \quad k = j - l + 1, \ldots, j. \\ \alpha \sim Cauchy(0, 10), &\qquad \beta \sim Cauchy(0, 2.5), \end{aligned}$$

where $2 \leq\ l \leq J$. We take the weakly informative Cauchy prior distribution for α and β, as described in Section 2.1. Again, dose d_k is standardized with a mean of 0 and a standard deviation of 0.5. We require $l \geq 2$ to ensure that there are adequate data points to identify the two regression parameters α and β. When $l = J$, the model uses all data for estimation and becomes a fully parametric approach. In phase I trials, the number of investigational dose levels is typically small (e.g., ≤ 10), therefore a value of l between 2 and 4 is often a reasonable choice. We recommend $l = 2$ because we find that increasing the value of l yields similar or even worse operating characteristics in our simulation studies.

One may question how reliable the estimation of the parameters for the local logistic

model is based on data observed from 2 doses. We note that for the purpose of dose finding, our goal here is not to obtain precise estimates of the model parameters, but to capture the local trend (e.g., increasing or decreasing) of the dose-response curve for directing dose escalation/deescalation. As long as the estimates correctly identify the trend of the curve (i.e., the sign of the slope), they lead to appropriate dose escalation and deescalation. Simulation studies in Section 3 show that the L-logistic design based on two local doses yields good operating characteristics, suggesting that the estimation using local data is adequately stable to identify the local trend of the dose-response curve.

To direct the dose escalation/de-escalation in a trial, we calculate the posterior probability $\Pr(\beta > 0|\mathcal{D}_j)$ based on the local logistic model. Let C_{E1} and C_{E2} be pre-specified efficacy cutoffs and $C_{E1} > C_{E2}$. If $\Pr(\beta > 0|\mathcal{D}_j) > C_{E1}$ (i.e., the current trend of the dose-efficacy curve is increasing), we escalate the dose because the next higher dose level is expected to have higher efficacy. In contrast, if $\Pr(\beta > 0|\mathcal{D}_j) < C_{E2}$, which indicates a decreasing dose-efficacy curve, we de-escalate the dose because the lower dose level is expected to have higher efficacy. Otherwise, i.e., $C_{E2} \leq Pr(\beta > 0|\mathcal{D}_j) \leq C_{E1}$, we stay at the current dose to accumulate more data. The values of C_{E1} and C_{E2} can be calibrated by simulation to obtain good operating characteristics. Typically, C_{E1} should be larger than C_{E2} by a certain reasonable margin, such as 10% to 20%. By using the posterior probability $Pr(\beta > 0|\mathcal{D}_j)$ as the criterion of dose escalation, we automatically account for the uncertainty associated with parameter estimation. That is, we escalate or de-escalate the dose only when we achieve a certain degree of confidence that $\beta > 0$ or $\beta < 0$.

The dose-finding algorithm for the proposed L-logistic design is described as follows:

1. Starting from the lowest l dose levels, treat one cohort of patients at each dose levels.

2. At the current dose level j, based on the toxicity outcomes, applying safety rule (1) to find the admissible set $\mathcal{A}$.

3. Based on the efficacy outcomes from the current and previous $l-1$ dose levels, calculate the posterior probability $Pr(\beta > 0|\mathcal{D}_j)$. If $Pr(\beta > 0|\mathcal{D}_j) > C_{E1}$, the dose level is escalated to $j+1$ when $j+1$ is in $\mathcal{A}$ and otherwise retain at j; if $Pr(\beta > 0|\mathcal{D}_j) < C_{E2}$, the dose level is de-escalated to $j-1$; otherwise, i.e., $C_{E2} \leq Pr(\beta > 0|\mathcal{D}_j) \leq C_{E1}$, the dose level j is retained.

4. Once the maximum sample size is reached, based on all the observed data, we carry out a double-sided isotonic regression and select the lowest dose that has the highest estimate of the efficacy probability as the OBD.

One potential problem of the above dose-finding algorithm is that the dose movement may bounce back and forth between dose levels j and $j+1$ when the dose level j is the maximum point of the dose-efficacy curve. This is because in this case, it may happen that $Pr(\beta > 0|\mathcal{D}_j) > C_{E1}$ (i.e., the dose-efficacy curve is increasing from dose level $j-1$ to j) and $Pr(\beta > 0|\mathcal{D}_{j+1}) < C_{E2}$ (i.e., the dose-efficacy curve is decreasing from dose level j to $j+1$). Therefore, when the current dose level is j, we will escalate the dose level to $j+1$ because $Pr(\beta > 0|\mathcal{D}_j) > C_{E1}$; and once we move to the dose level $j+1$, at the next dose assignment, we will de-escalate back to the dose level j because $Pr(\beta > 0|\mathcal{D}_{j+1}) < C_{E2}$; and so on. To avoid this problem, before conducting any dose escalation, we will determine whether $Pr(\beta > 0|\mathcal{D}_{j+1}) < C_{E2}$ whenever the dose level $j+1$ has been used to treat any patients. If $Pr(\beta > 0|\mathcal{D}_{j+1}) < C_{E2}$, indicating that the dose level j is the maximum point of the curve, we will retain the current dose level. The values of C_{E1} and C_{E2} are calibrated according to the desirable operating characteristics.

3 Simulation studies

3.1 Dose-finding studies

We investigated the operating characteristics of the proposed designs through simulation studies under eight efficacy and toxicity scenarios, as listed in Table 1. We considered five dose levels and started the trials at the lowest dose level. We assumed a target toxicity upper bound of $\phi = 0.3$, a toxicity threshold of $C_T = 0.8$ and a maximum sample size of 30 in cohorts of size 3. Under each scenario, we simulated 5,000 trials. For the L-logistic design, we specified efficacy cutoffs $C_{E1} = 0.4$ and $C_{E2} = 0.3$ according to the calibration study. We used two adjacent doses (i.e., $l = 2$) to fit the local logistic model in the proposed L-logistic design. We examined other choices for the number of adjacent doses and found very similar performance levels (results not shown). We compared our designs with the slope-sign design and a traditional design. In the slope-sign design, three adjacent dose levels were used to estimate the slope, and safety monitoring was carried out using the method described in Section 2.1. Under the traditional approach, we first conducted the dose escalation using the conventional "3+3" design, and once the MTD is identified, we then randomized the remaining patients between the MTD and the dose one level lower than the MTD.

Table 1 shows the simulation results, including the dose selection probability, the average percentage of patients treated at each dose level, the average percentages of patients experienced efficacy and toxicity, and the averaged sample size under the slope-sign, logistic, isotonic and L-logistic designs. In practice, besides the OBD, the other doses with high efficacy and low toxicity are often interested to investigators as well. Therefore, we also reported the selection percentage of the OBD region, which is defined as the two doses that have the highest response rates within the admissible set $\mathcal{A}$.

In scenario 1, the dose-efficacy curve is unimidal and the fourth dose level is the OBD

with the highest efficacy rate and acceptable toxicity rate. The traditional design yield the lowest OBD selection percentage of 19.8%, The logistic and Isotonic designs outperformed the slope-sign design and the traditional design with a selection percentage of 31.1% and 45.6% respectively. The L-logistic design performed best and selected the OBD 54.4% percent of the time, respectively. The three proposed designs also yielded the highest selection percentages of the OBD region, ranging from 69% to 79%. In addition, the proposed designs assigned higher percentages of patients to the OBD than the other two designs. Specifically, the percentages of patients allocated at the OBD under the logistic, isotonic and L-logistic designs were 6.6%, 12.2% and 13.9% higher than that under the slope-sign design and 6.2%, 11.8% and 13.5% higher than that under the traditional design. However, the slope-sign design used the smallest sample size. In scenario 2, the OBD was located at dose level 2 whereas the MTD was located at dose level 5. As the traditional design focuses on evaluating the doses around the MTD, it selected dose 4 as the OBD with a percentage of 42.1% whereas only 22.8% of selecting dose 2 as the OBD. All the other designs performed significantly better than the traditional design. Specifically, the isotonic and L-logistic designs outperformed other designs and selected dose level 2 with the percentages of 78.6% and 77.0%, respectively.

Scenarios 3 to 4 also considered unimodal patterns, with different toxicity profiles. Across these two scenarios, the logistic design demonstrated a large variation:it yielded the highest OBD selection (51.8%) in scenario 4 but the lowest selection percentage (36.4%) in scenario 3, suggesting the sensitivity of this parametric approach. In contrast, the proposed isotonic and L-logistic designs performed consistently and outperformed the slope-sign design and the traditional design across these two scenarios.

Scenarios 5 examined the performance of the designs when the dose-efficacy curve was monotonic. We can see that the proposed designs consistently outperformed the other two designs and the L-logistic design performed best. Indeed, the OBD selection percentages of

the L-logistic design was 19.6% higher than that of the slop-sign design; and the percentages of patients treated at the OBD using the L-logistic design was also 13.6% higher than that of the slope-sign design. Scenarios 6 simulated the cases in which efficacy initially increased with dose and then plateaued. In these cases, the target OBD was the lowest dose yielding the highest efficacy rate. The slope-sign design performed worst and other designs yielded similar percentage of OBD selection.

In summary, the proposed designs outperformed the slope-sign design and the traditional design in terms of selecting the target doses and allocating patients to the efficacious doses. The slope-sign design has the advantage of using smaller sample sizes. Among the three proposed designs, the isotonic and L-logistic designs yielded consistently good operating characteristics, and are recommended for practical use. The software to implement these two designs is available for free download at *http:// odin.mdacc.tmc.edu/∼yyuan/.* We do not recommend the parametric logistic design because of its sensitivity for the dose-efficacy and dose-toxicity curves. In addition, we note that finding OBD without the monotonicity assumption is substantially more difficult than finding the MTD for cytotoxic agents, which assume monotonicity. This is because without the monotonicity assumption, the uncertainty on the shape of the dose-response curve is substantially higher and at each decision making of dose assignment, we face two possible directions (i.e., the dose higher or lower than the current dose) for dose transition. In our simulation, with a sample size of 30, the selection percentage of the OBD is generally lower than 50%. If we aim a higher selection percentage, a larger sample size should be used.

3.2 Sensitivity analysis

In the proposed logistic and L-logistic designs, we adopted the weakly informative $Cauchy(0, 2.5)$ prior recommended by Gelman et al [15] for regression parameters β and γ. To assess the sensitivity to this prior, we examined the operating characteristics of the proposed methods

under a tighter prior $\beta, \gamma \sim Cauchy(0, 1.25)$ and a more vague prior $\beta, \gamma \sim Cauchy(0, 5)$. Table 2 showed the results where the true efficacy rate was (0.4,0.6,0.8,0.7,0.55) and we assumed minimal toxicity with the dose level range. We can see that the results are rather stable across different priors, suggesting that our designs are not sensitive to the prior distribution.

We conducted another sensitivity analysis to assess the robustness of the proposed designs with respect to the number of dose levels under investigation. Starting from 4 dose levels, we gradually increased the number of dose levels to 8 by inserting additional doses under the minimal toxicity assumption. As shown in Figure 1. the parametric logistic design was sensitive to the number of dose levels and its performance deteriorated when the number of dose levels increased. This is because when the number of dose levels was large, the dose-efficacy curve tended to have more local features and consequently the parametric logistic model was more likely subject to the model misspecification. Comparatively, the isotonic and L-logistic designs was less sensitive to the number of dose levels as they utilized more robust nonparametric and semiparametric approaches to modeling the dose-efficacy curve. When the number of dose levels increased, the performance the isotonic and L-logistic designs downgraded somewhat, but that was mainly because under a larger number dose levels, the designs had a smaller average sample size (per dose) to search the OBD, not because of the model misspecification. Hence, we recommend the isotonic and L-logistic designs to be used in practice when the number of dose levels is large.

4 Discussion

The introduction of molecularly targeted agents has changed the drug development of the traditional phase I clinical trials. Compared with the dose-finding designs for conventional chemotherapy which escalates or de-escalates the dose based on the toxicity endpoint, the optimal biological dose is the primary goal of a dose-finding design for MTAs which is

based mainly on the efficacy endpoint. However, the dose-efficacy relationship for molecularly targeted agents may not be monotonic, which restricts the adoption of the traditional dose-finding designs (e.g. 3+3 or CRM) under this circumstance. Hence, novel dose-finding designs for identifying optimal biological dose for molecularly targeted agents without making any monotonic assumption need to be developed.

In this article, we have proposed three dose-finding trial designs, namely, the logistic, isotonic and L-logistic designs to determine the optimal biologic dose for molecularly targeted agents. The logistic design is purely parametrical; the isotonic design is based on the nonparametric isotonic regression; and the L-logistic assumes a logistic dose-response model only locally around the current dose. During the trial, the three designs continuously update the estimate of the dose-response curve, which in turn is used to determine the dose assignment for new patients. Simulation studies show that the isotonic and L-logistic designs have good operating characteristics for finding the OBD and outperform the slope-sign design, which is one of the trial designs currently available in practice. However, the logistic design is sensitive to the parametric model assumption and is not recommended for practical use. In addition, all the proposed design use more patients than the slope-sign design.

The proposed designs are appropriate for trials in which the efficacy outcome is binary and observable shortly after the initiation of the treatment. They cannot be applied directly to cases in which the efficacy outcome requires a long follow-up time to be assessed. To address this delayed outcome issue, a simple method is to use the observed data to fit the model and make the decision of dose assignment for newly enrolled patients. This method is simple and works reasonably well in the case that the accrual rate is slow [22]. However, when the accrual rate is fast, it may subject to large estimation bias. A better strategy is to treat the delayed efficacy outcome as a missing data problem and using missing data methodology to handle it. For example, Yuan and Yin [23] proposed an EM algorithm to

estimate the toxicity rate based on delayed outcomes for dose finding, and Liu, Yin and Yuan [22] developed the Bayesian data augmentation method to deal with delayed toxicity for finding the MTD of cytotoxic agents. The similar approach can be adopted here for handle the delayed outcome for finding OBD, and future research in this direction is warranted.

Acknowledgment

The research was supported in part by grant from the National Cancer Institute CA016672 (Y.Y., J.J.L.), CA154591 (Y.Y.), CA097007 (J.J.L.) and 5P50CA098258-09 (Y.Y., Y.Z.). The authors thank two referees and the associated editor for their valuable comments and LeeAnn Chastain for her editorial assistance.

References

[1] 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.

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

[3] 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.

[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] Postel-Vinay S, Arkenau HT, Olmos D, Ang J, Barriuso J, Ashley S, Banerji U, De-Bono J, Judson I, Kaye S. Clinical benefit in Phase-I trials of novel molecularly targeted agents: does dose matter? *British Journal of Cancer* 2009; **100**: 1373-1378.

[6] Hunsberger S, Rubnstein LV, Dancey J, Korn EL. Dose escalation trial designs based on a molecularly targeted endpoint. *Statistics in Medicine* 2005; **24**: 2171-2181.

[7] Hoering A, Lebilanc M, Crowley J. Seamless phase I-II trial design for assessing toxicity and efficacy for targeted agents. *Clinical Cancer Research* 2011; **17**: 640-646.

[8] Jain RK, Lee JJ, Hong D, Markman M, Gong J, Naing A, Wheler J, Kurzrock R. Phase I oncology studies: Evidence that in the era of targeted therapies patients on lower doses do not fare worse. *Clinical Cancer Research* 2010; **16**: 1289-1297.

[9] Reynolds AR. Potential relevance of bell-shaped and U-shaped dose-responses for the therapeutic targeting of angiogenesis in cancer. *Dose Response* 2010; **8**: 253-284.

[10] Brunetto AT, Kristeleit RS, De Bono JS. Early oncology clinical trial design in the era of molecular-targeted agents. *Future Oncology* 2010; **6**: 1339-1352.

[11] Mandrekar SJ, Qin R, Sargent DJ. Model-based phase I designs incorporating toxicity and efficacy for single and dual agent drug combinations: Methods and challenges. *Statistics in Medicine* 2010; **29**: 1077-1083.

[12] Zhang W, Sargent DJ, Mandrekar SJ. An adaptive dose-finding design incorporating both toxicity and efficacy. *Statistics in Medicine* 2006; **25**: 2365-2383.

[13] Mandrekar SJ, Cui Y, Sargent DJ. An adaptive phase I design for identifying a Biologically optimal dose for dual agent drug combinations. *Statistics in Medicine* 2007; **26**: 2317-2330.

[14] Polley MY, Cheung YK. Two-stage designs for dose-finding trials with a biologic endpoint using stepwise tests. *Biometrics* 2008; **64**: 232-241.

[15] Gelman A, Jakulin A, Pittau MG, Su, Y. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics* 2008; **2**: 1360-1383.

[16] Barlow RE, Bartholomew DJ, Bremner JM, Brunk HD. Statistical Inference under Order Restrictions. Wiley, London 1972.

[17] Robertson T, Wright FT, Dykstra, RL. Order Restricted Statistical Inference. Wiley, New York 1988.

[18] Leung DH, Wang YG. Isotonic designs for phase I trials. *Controlled Clinical Trials* 2001; **22**: 126-138.

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

[20] Yuan Z, Chappell R. Isotonic designs for phase I cancer clinical trials with multiple risk groups. *Clinical Trials* 2004; **1**: 499-508.

[21] Turner TR, Wollan PC. Locating a maximum using isotonic regression. *Computational Statistics and Data Analysis* 1997; **25**: 305-320.

[22] Liu S, Yin GS, and Yuan Y. Bayesian Data Augmentation Dose Finding with Continual Reassessment Method and Delayed Toxicities. *Annal of Applied Statistics*, 2013, to appear.

[23] Yuan Y and Yin GS. Robust EM continual reassessment method in oncology dose finding. *Journal of the American Statistical Association* 2011; **106**: 818-831.

Table 1: The dose selection percentage, average percentage of patients treated at each dose level, average percentage of efficacy and toxicity, average percentage of promising dose selection and average sample size under the slope-sign, logistic, isotonic and L-logistic designs.

Design		Dose level 1	2	3	4	5	% of efficacy	% of toxicity	% of promising dose	# of sample
Scenario 1										
	True efficacy	0.2	0.4	0.6	0.8	0.55				
	True toxicity	0.08	0.12	0.2	0.3	0.4				
Slope-sign	Selection (%)	3.4	23.6	36.5	**30.3**	6.2				
	Patient(%)	26.3	26.3	26.3	14.1	7.0	47.4	17.8	66.8	12.0
Traditional	Selection (%)	22.8	27.4	27.0	**21.6**	1.1				
	Patient(%)	32.4	26.1	21.8	13.8	5.9	44.4	16.5	48.6	30.0
Logistic	Selection (%)	12.8	11.4	39.9	**31.1**	4.7				
	Patients (%)	22.0	18.3	32.0	20.7	7.3	51.7	19.7	71.0	30.0
Isotonic	Selection (%)	12.5	15.8	23.4	**45.6**	2.7				
	Patients (%)	21.0	22.0	21.0	26.3	9.7	52.0	20.3	69.0	30.0
L-logistic	Selection (%)	6.8	10.0	24.9	**54.4**	3.9				
	Patients (%)	16.7	19.3	20.3	28.0	15.7	54.3	22.3	79.3	30.0
Scenario 2										
	True efficacy	0.6	0.8	0.5	0.4	0.2				
	True toxicity	0.01	0.05	0.10	0.15	0.3				
Slope-sign	Selection (%)	13.3	**59.1**	23.9	3.5	0.2				
	Patient(%)	31.2	31.2	31.2	6.0	0.4	62.5	6.4	72.4	9.8
Traditional	Selection (%)	4.3	**24.4**	25.9	40.0	5.4				
	Patient(%)	14.9	18.2	22.4	27.0	17.5	49.0	12.6	28.7	30.0
Logistic	Selection (%)	25.3	**64.4**	9.4	0.4	0.5				
	Patients (%)	32.9	45.1	19.3	2.6	0.2	66.4	5.0	89.7	30.0
Isotonic	Selection (%)	14.0	**78.6**	5.4	2.0	0.0				
	Patients (%)	21.8	58.4	15.0	4.1	0.7	68.6	5.6	92.6	30.0
L-logistic	Selection (%)	11.4	**77.0**	9.1	2.5	0.0				
	Patients (%)	14.9	49.3	22.4	9.8	3.8	64.4	6.2	88.4	30.0
Scenario 3										
	True efficacy	0.2	0.4	0.6	0.8	0.55				
	True toxicity	0.06	0.08	0.14	0.2	0.3				
Slope-sign	Selection (%)	5.7	16.5	29.4	**38.7**	9.7				
	Patient(%)	25.4	25.4	25.4	14.9	8.9	47.4	13.3	68.1	12.4
Traditional	Selection (%)	12.6	18.1	22.7	**43.6**	2.9				
	Patient(%)	23.2	21.4	21.4	21.0	12.9	50.0	14.2	66.3	30.0
Logistic	Selection (%)	12.3	9.2	35.4	**36.4**	6.7				
	Patients (%)	21.3	17.0	30.3	22.7	8.3	52.3	14.0	71.8	30.0
Isotonic	Selection (%)	12.0	13.7	16.8	**52.7**	4.7				
	Patients (%)	20.3	20.7	18.7	29.0	11.0	53.0	14.7	69.5	30.0
L-logistic	Selection (%)	4.1	7.7	18.6	**62.8**	6.8				
	Patients (%)	14.3	18.0	18.3	30.0	19.3	55.7	16.7	81.4	30.0

Table 1 continues.

Design		Dose level 1	2	3	4	5	% of efficacy	% of toxicity	% of promising dose	# of sample
Scenario 4										
	True efficacy	0.2	0.4	0.6	0.8	0.55				
	True toxicity	0.05	0.1	0.25	0.5	0.6				
Slope-sign	Selection (%)	4.1	22.9	**48.1**	23.0	1.9				
	Patient(%)	26.8	26.8	26.8	15.0	4.7	46.5	22.1	71.0	11.7
Traditional	Selection (%)	18.8	41.7	**30.4**	9.0	0.1				
	Patient(%)	30.5	33.9	25.3	8.9	1.4	42.6	20.5	72.1	30.0
Logistic	Selection (%)	12.5	21.1	**51.8**	12.2	2.3				
	Patients (%)	21.7	21.7	36.0	15.3	5.3	50.0	23.3	72.9	30.0
Isotonic	Selection (%)	12.4	25.4	**49.5**	12.1	0.6				
	Patients (%)	20.7	25.0	27.7	19.3	7.7	50.3	24.7	74.9	30.0
L-logistic	Selection (%)	5.8	21.4	**50.0**	21.7	1.1				
	Patients (%)	15.3	23.3	29.7	21	10.7	52.7	27.3	71.4	30.0
Scenario 5										
	True efficacy	0.05	0.25	0.45	0.65	0.8				
	True toxicity	0.05	0.1	0.15	0.2	0.5				
Slope-sign	Selection (%)	2.1	20.0	19.7	**35.0**	23.2				
	Patient (%)	25.0	25.0	25.0	15.1	9.8	35.6	16.0	54.7	12.6
Traditional	Selection (%)	13.7	20.2	27.4	**31.7**	7.0				
	Patient (%)	24.8	22.3	23.8	20.5	8.7	37.6	15.5	59.0	30.0
Logistic	Selection (%)	6.1	12.3	30.6	**44.5**	6.4				
	Patients (%)	15.3	18.7	28.7	24.7	12.7	44.3	18.3	75.1	30.0
Isotonic	Selection (%)	6.1	14.3	19.9	**53.3**	6.4				
	Patients (%)	15.0	21.3	20.7	26.0	17.0	45.7	20.0	73.2	30.0
L-logistic	Selection (%)	3.5	5.6	16.1	**64.6**	10.2				
	Patients (%)	13.3	16.3	19.3	28.7	22.3	50.0	22.2	80.7	30.0
Scenario 6										
	True efficacy	0.1	0.3	0.5	0.5	0.5				
	True toxicity	0.1	0.2	0.4	0.5	0.6				
Slope-sign	Selection (%)	7.8	**36.7**	38.8	12.2	4.5				
	Patient (%)	28.3	28.3	28.3	11.9	3.0	33.3	29.1	44.5	11.0
Traditional	Selection (%)	44.5	**41.9**	12.3	1.2	0.2				
	Patient (%)	52.2	30.1	13.6	3.5	0.6	23.1	23.5	86.4	30.0
Logistic	Selection (%)	19.0	**42.2**	31.1	6.4	1.3				
	Patient (%)	23.7	30.6	32.0	11.3	2.5	34.2	28.7	61.2	30.0
Isotonic	Selection (%)	18.3	**42.3**	31.0	7.9	0.5				
	Patients (%)	23.0	32.1	27.5	12.8	4.7	34.2	29.3	60.6	30.0
L-Logistic	Selection (%)	18.1	**39.3**	27.7	12.4	2.5				
	Patient (%)	23.4	28.4	24.9	15.9	7.4	35.3	30.2	57.4	30.0

Table 2: Prior sensitivity analysis for the logistic and L-logistic designs.

Design		Dose level					% of efficacy
		1	2	3	4	5	
	True efficacy	0.4	0.6	**0.8**	0.7	0.55	
		$\beta, \gamma \sim Cauchy(0, 1.25)$					
Logistic	Selection (%)	17.5	5.8	**61.4**	11.1	4.2	
	Patients (%)	27.0	15.3	41.3	13.4	3.0	63.7
L-logistic	Selection (%)	4.1	13.5	**48.1**	27.9	6.4	
	Patients (%)	15.3	18.3	25.3	24.3	16.8	63.3
		$\beta, \gamma \sim Cauchy(0, 2.5)$					
Logistic	Selection (%)	12.9	10.4	**60.7**	11.5	4.5	
	Patients (%)	22.7	17.3	42.7	14.0	3.3	65.1
L-logistic	Selection (%)	4.6	11.7	**51.1**	26.8	5.9	
	Patients (%)	14.0	17.3	27.3	24.3	16.7	64.3
		$\beta, \gamma \sim Cauchy(0, 5)$					
Logistic	Selection (%)	13.7	11.7	**59.4**	10.7	4.5	
	Patients (%)	25	18.3	40.7	12.7	3.3	64.1
L-logistic	Selection (%)	5.6	11.0	**51.5**	26.9	5.0	
	Patients (%)	16.0	19.3	30.3	22.3	12.1	64.3

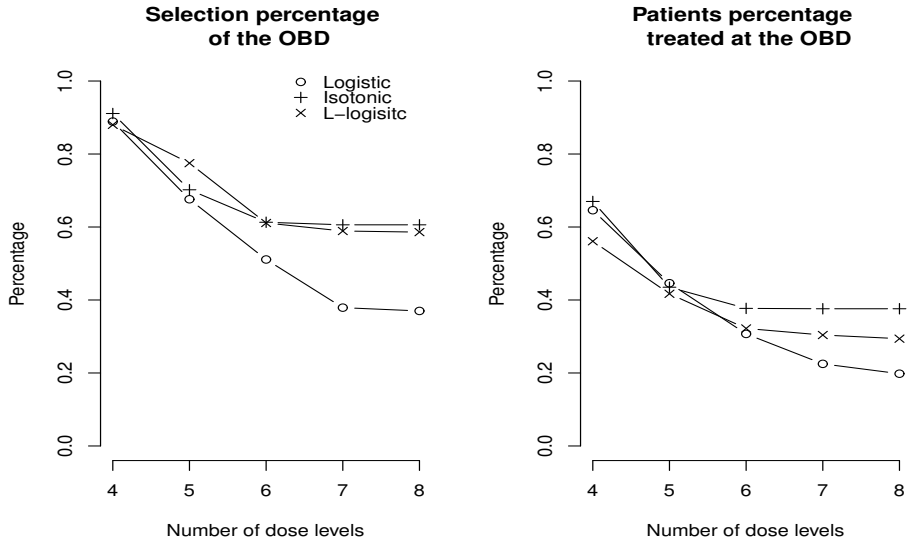

Figure 1: The OBD selection percentages (left panel) and the percentages of patients treated at the OBD (right panel) under different numbers of dose levels. The efficacy rates are (0.4, 0.8, 0.45, 0.2) for 4 dose levels; (0.4, 0.55, 0.8, 0.45, 0.2) for 5 dose levels; (0.4, 0.55, 0.8, 0.6, 0.45, 0.2) for 6 dose levels; (0.4, 0.55, 0.8, 0.6, 0.45, 0.4, 0.2) for 7 dose levels, and (0.4, 0.55, 0.8, 0.6, 0.45, 0.4, 0.2, 0.1) for 8 dose levels.