



Published in final edited form as:

*Stat Med.* 2022 March 30; 41(7): 1205–1224. doi:10.1002/sim.9265.

## BIPSE: A Biomarker-based Phase I/II Design for Immunotherapy Trials with Progression-free Survival Endpoint

Beibei Guo<sup>1</sup>, Yong Zang<sup>2,3</sup>

<sup>1</sup>Department of Experimental Statistics, Louisiana State University, Baton Rouge, LA 70803, U.S.A.

<sup>2</sup>Department of Biostatistics and Health Data Science, Indiana University, Indianapolis, IN 46202, U.S.A.

<sup>3</sup>Center for Computational Biology and Bioinformatics, Indiana University, Indianapolis, IN 46202, U.S.A.

### Abstract

A Bayesian biomarker-based phase I/II design (BIPSE) is presented for immunotherapy trials with a progression-free survival endpoint. The objective is to identify the subgroup-specific optimal dose, defined as the dose with the best risk-benefit tradeoff in each biomarker subgroup. We jointly model the immune response, toxicity outcome, and progression-free survival with information borrowing across subgroups. A plateau model is used to describe the marginal distribution of the immune response. Conditional on the immune response, we model toxicity using probit regression and model progression-free survival using the mixture cure rate model. During the trial, based on the accumulating data, we continuously update model estimates and adaptively randomize patients to doses with high desirability within each subgroup. Simulation studies show that the BIPSE design has desirable operating characteristics in selecting the subgroup-specific optimal doses and allocating patients to those optimal doses, and outperforms conventional designs.

### Keywords

Immunotherapy; subgroups; biomarker; phase I/II trial; dose finding; immune response; progression-free survival; risk-benefit tradeoff; Bayesian adaptive design

## 1. Introduction

Immunotherapy is an innovative treatment that stimulates a patient's immune system to treat cancer. It has provided an alternative and complementary treatment modality to conventional chemotherapy and radiotherapy, and is believed to have the potential to provide higher effectiveness and lower side effects than chemotherapy.<sup>1–3</sup>

A few dose-finding designs have been developed for immunotherapy that can incorporate an immune response.<sup>4-9</sup> However, all these existing designs assume the patient population is homogeneous and use the “one-dose-fits-all” rule for patient allocation and optimal dose identification, which is disconnected from medical practice as population heterogeneity is often expected for clinical trial studies. For example, many studies in various cancer types have demonstrated that the programmed cell death-ligand 1 (PD-L1) expression is a predictive biomarker for checkpoint inhibitor-based immunotherapy, with PD-L1 positive patients showing higher response rates, better progression-free survival and overall survival, compared with PD-L1 negative patients.<sup>10-15</sup> In 2016, the U.S. Food and Drug Administration (FDA) approved pembrolizumab for the treatment of metastatic non-small cell lung cancer (NSCLC) for PD-L1 positive patients only. However, a number of studies have shown that immunotherapy should not be restricted to marker-positive patients as marker-negative patients may also benefit from the therapy.<sup>15-19</sup>

The optimal dose for treating patients with an immunotherapeutic agent may differ according to the patient’s individual characteristics. Goldstein et al.<sup>20</sup> showed that compared with the FDA approved dose, a weight-based personalized dose demonstrated equivalent efficacy in PD-L1 positive NSCLC patients, which would lead to a significant reduction of drug cost. Goldstein et al.<sup>20</sup> called for personalized immunotherapy and emphasized the importance of it. In the conclusion part, Ingles Garces et al.<sup>21</sup> stated that for immunotherapy, “More biomarker data may be useful in selection of dose and regimen in early clinical development, ...”. Therefore, optimizing the treatment benefit for immunotherapy requires the dose-finding method to identify the optimal dose for each subgroup, stratified by patient’s individual characteristics, such as biomarker status. For brevity of description, we use “patient characteristic” and “biomarker” exchangeably hereafter.

Our research is motivated by a phase I/II trial being developed at Indiana University Simon Comprehensive Cancer Center, which uses an anti-PD-L1 immune checkpoint inhibitor to treat patients with recurrent ovarian cancer. Five doses (0.1, 0.3, 0.5, 0.7, 0.9 mg/kg) of the inhibitor will be investigated, and a maximum of 60 patients will be accrued to the trial. The immune response is measured by the count of CD8+ T-cells at week 4 after the treatment. The dose-limiting toxicity is defined according to the National Cancer Institute Common Terminology Criteria for Adverse Events, Version 4.0. The primary efficacy endpoint is the progression-free survival (PFS), which is defined as the time from the initiation of the treatment until disease progression or death, whichever occurs first. The goal of the trial is to identify the optimal dose for each of the two subgroups stratified by the PD-L1 expression. Unlike chemotherapy that works by shrinking the tumor, immunotherapy targets a patient’s immune system so it often delays cancer progression and prolongs survival without achieving rapid tumor shrinkage, and some patients may achieve long-term durable response.<sup>2,22,23</sup> Therefore, some immunotherapy trials prefer to use PFS, rather than the conventional objective response, as the efficacy endpoint.

In this article, we develop a Bayesian biomarker-based phase I/II trial design (BIPSE) to optimize subgroup-specific dose for immunotherapy trials with PFS endpoint. The BIPSE simultaneously considers the continuous immune response, binary toxicity outcome, and PFS. We model the marginal distribution of the immune response using a plateau dose-

response model; and conditional on the immune response, we use a probit regression for toxicity and a mixture cure rate model for PFS to accommodate that some patients may have long-term durable response. Considering the relatively small sample size in a typical phase I/II trial, the models in BIPSE are parsimonious yet reasonably flexible to facilitate information borrowing across different types of outcomes and subgroups. Motivated by the anti-PD-1 therapy, in this article, we focus on the situation where the two biomarker subgroups are ordered in terms of efficacy while discussing how to handle unordered subgroups. We quantify the dose's desirability using a utility function elicited from physicians that accounts for the risk-benefit tradeoff. For each successive cohort of patients, we use adaptive randomization to choose doses based on their estimated utility values, subject to minimum toxicity and efficacy requirements. Patients will not be treated at doses that are unlikely to be safe or efficacious. At the end of the trial, final recommendation is made for each subgroup.

Numerous phase I/II trial designs have been proposed in the literature. Thall and Russell<sup>24</sup>, Braun<sup>25</sup>, Thall and Cook<sup>26</sup>, Zang et al.<sup>27</sup>, and Han et al.<sup>28</sup> considered phase I/II designs that account for both toxicity and efficacy. A few phase I/II designs for subgroup dose finding were also developed. For example, Guo and Yuan<sup>29</sup> developed a phase I/II design that accounts for the patient's biomarker information. Lee, Thall, and Rezvani<sup>30</sup> described a design that aims to find the optimal dose in each prognostic subgroup with five co-primary outcomes. Lee, Thall, and Msaouel<sup>31</sup> described a design that optimizes the dose within subgroups based on toxicity and efficacy. One important difference between these papers and the current paper is that the latter has two co-primary endpoints (i.e., toxicity and PFS) and one ancillary endpoint (i.e., immune response). The toxicity and PFS outcomes are the two co-primary endpoints that are used to determine the optimal dose. The immune response, which is typically quickly-ascertainable, is used as an ancillary endpoint to quickly screen out futile doses and predict the PFS for efficient decision making in the presence of a high percentage of censoring, for example, at the early stage of the trial when the rate of disease progression is relatively slow compared to the rate of patient accrual or when the enrolled patients have been followed for a short period of time. To the best of our knowledge, this article provides the first phase I/II design for immunotherapy trials that aims to identify the subgroup-specific optimal dose by jointly accounting for immune response, toxicity, and PFS.

The remainder of this article is organized as follows. In Section 2, we describe the probability models and present the dose-finding algorithm. In Section 3, we investigate the operating characteristics of the BIPSE design through simulation studies. We provide concluding remarks in Section 4.

## 2. Method

### 2.1. Probability Models

Let  $d_1 < \dots < d_D$  denote  $D$  doses of an immunotherapeutic agent under investigation. Let  $Y$ ,  $X$ , and  $T$  denote the immune response, toxicity outcome, and PFS, respectively. We consider a binary biomarker with subgroup indicator  $Z = 0$  or  $1$  corresponding to the marker-negative or marker-positive subgroup. The objective of the trial is to identify the optimal dose for

each marker subgroup. We model the joint distribution of  $(Y, X, T)$  by first specifying the marginal distribution of  $Y$  and then the distributions of  $X$  and  $T$  conditional on  $Y$  to reflect the fact that in immunotherapy, clinical outcomes rely on the activation of the immune system.

The immune response  $Y$  is taken to be the increase in a log-transformed immune activity (e.g., T-cell resistance) from baseline to post-treatment, which is generally a continuous outcome. We assume the observed immune response  $Y_i$  for patient  $i$  with subgroup indicator  $Z_i$  treated at dose  $d_i$  follows a normal distribution with mean  $\mu_Y(d_i, Z_i)$  and variance  $\sigma^2$ , that is,

$$Y_i | d_i, Z_i \sim N(\mu_Y(d_i, Z_i), \sigma^2) \quad (1)$$

Since the expected value  $\mu_Y(d_i, Z_i)$  represents the mean pre-post difference in the immune activity, it is expected to be 0 when there is no drug, i.e.,  $d_i = 0$ . To satisfy this constraint, we use the following plateau model for  $\mu_Y(d_i, Z_i)$

$$\mu_Y(d_i, Z_i) = \mu_M \exp(\xi Z_i) (1 - \exp(-\alpha d_i)) \quad (2)$$

where  $\mu_M$  and  $\mu_M \exp(\xi)$  are the maximum mean immune responses that the immunotherapeutic agent possibly achieves for marker-negative and positive subgroups, respectively. The specific value  $\xi = 0$  corresponds to the absence of patient heterogeneity, i.e., no subgroup difference exists. We restrict  $\mu_M > 0$  to reflect that immunotherapy typically increases immune activity, so the mean increase is positive. Since marker-positive patients have a higher maximum mean immune response than marker-negative patients in most clinical practices, we assign a prior distribution to  $\xi$  that puts most probability mass to positive values. Note here we allow a small, say 5%, probability on negative values of  $\xi$  to accommodate situations where the observed data contradict the ordering assumption. If the ordering of the subgroups is unknown, no restriction would be placed on the sign of  $\xi$ . We constrain  $\alpha > 0$  such that the immune response increases or first increases and then plateaus when the dose increases. It can be easily seen that  $\mu_Y(d_i, Z_i) = 0$  when  $d_i = 0$  in (2). The mean immune response model (2) provides a parsimonious yet flexible framework for modeling the dose-immune response relationships for the two subgroups. Figure 1 shows the immune response curves generated from this model with different values of  $\mu_M$ ,  $\xi$ , and  $\alpha$  in the dose range considered in our motivating trial. In the left panel, the mean immune response increases for both subgroups; in the middle panel, the mean immune response first increases and then plateaus for both subgroups; in the right panel, the mean immune response plateaus for marker-negative patients, and keeps increasing for marker-positive patients. Compared to the Emax model used in Liu, Guo, and Yuan<sup>4</sup>, model (2) is more parsimonious.

Taking the setting of the motivating trial, we model toxicity  $X$  as a binary outcome, with  $X = 1$  indicating toxicity (or severe adverse events), and  $X = 0$  otherwise. Let  $\pi_X(d_i, Z_i, Y_i) \equiv \Pr(X = 1 | d_i, Z_i, Y_i)$  denote the toxicity probability for patient  $i$  treated at dose  $d_i$  with subgroup indicator  $Z_i$  and immune response  $Y_i$ . We model  $\pi_X(d_i, Z_i, Y_i)$  using the probit regression

$$\Phi^{-1}\{\pi_X(d_i, Z_i, Y_i)\} = \beta_{0, Z_i} + \beta_1 d_i + \beta_2 Y_i \tag{3}$$

In equation (3),  $\beta_{0,0}$  and  $\beta_{0,1}$  are the intercepts for marker-negative and marker-positive patients, respectively,  $\beta_1$  is the dose effect, and  $\beta_2$  is the effect of the immune response. In the above toxicity model, we include both dose  $d_i$  and the immune response  $Y_i$  to reflect that the toxicity of an immunotherapeutic agent is closely related to the immune response but not necessarily entirely caused by the immune response. That is, the immune response cannot fully explain the toxicity effect at different doses. In some trials, the toxicity outcome may be coded as a  $K$ -level ordinal variable, for example,  $X = 1, 2,$  and  $3$  (so  $K = 3$ ) denoting toxicity grades 1–2, 3, and 4, respectively. In those cases, we can use the cumulative probit model,

$$\Phi^{-1}[\Pr(X \leq k \mid d_i, Z_i, Y_i)] = \beta_{0, Z_i, k} + \beta_1 d_i + \beta_2 Y_i, \quad k = 1, \dots, K - 1 \tag{4}$$

where  $\beta_{0, Z_1} < \dots < \beta_{0, Z, K-1}$  are intercepts for subgroup  $Z$ .

Let  $S(t \mid d_i, Z_i, Y_i)$  denote the survival function of the PFS (i.e.,  $T$ ) for patient  $i$  in subgroup  $Z_i$  treated at dose  $d_i$  and having immune response  $Y_i$ . To accommodate that a certain percentage of the patients may achieve durable response, we use the following cure rate model to model the time to disease progression  $T$ ,

$$S(t \mid d_i, Z_i, Y_i) = \pi(d_i, Z_i, Y_i) + (1 - \pi(d_i, Z_i, Y_i))S_u(t \mid d_i, Z_i, Y_i) \tag{5}$$

where  $\pi(d_i, Z_i, Y_i)$  is the probability of achieving durable response for patients in subgroup  $Z_i$  treated at dose  $d_i$  with immune response  $Y_i$ , and  $S_u(t \mid d_i, Z_i, Y_i)$  is the survival function for patients in subgroup  $Z_i$  treated at dose  $d_i$  with immune response  $Y_i$  who are susceptible to disease progression. In our model, each patient is either susceptible to PFS or has durable disease. We model the cured proportion  $\pi(d_i, Z_i, Y_i)$  using a probit regression

$$\Phi^{-1}(\pi(d_i, Z_i, Y_i)) = \eta_{c0} + \eta_{c1} Z_i + \eta_{c2} Y_i, \tag{6}$$

and model  $S_u(t \mid d_i, Z_i, Y_i)$  using the Cox proportional hazards model with baseline hazard following the Weibull distribution,

$$S_u(t \mid d_i, Z_i, Y_i) = \exp[-\lambda t^\gamma \exp(\eta_1 Z_i + \eta_2 Y_i)]. \tag{7}$$

where  $\lambda > 0$  and  $\gamma > 0$  are the scale and shape parameters of the baseline Weibull distribution. We restrict  $\eta_{c2} > 0$  and  $\eta_2 < 0$  due to the positive effect of the immune response on efficacy. We assign prior distributions to  $\eta_{c1}$  and  $\eta_1$  that put most probability mass to positive and negative values, respectively, to accommodate that marker-positive patients are expected to have higher efficacy than marker-negative patients, while allowing a small probability on alternative values in case data contradict the ordering assumption. If the ordering between subgroups is unknown, we should use prior distributions for  $\eta_{c1}$  and  $\eta_1$  such that positive and negative values are equally likely. Due to the limited sample size of early phase trials, we take this parametric approach with a Weibull baseline hazard,

rather than the semiparametric approach that does not specify the baseline hazard. Previous studies<sup>6,32</sup> and the sensitivity analysis described later show that our design is not sensitive to this parametric assumption for the purpose of dose finding.

In equations (6) and (7), we do not include dose  $d_j$  as a covariate to reflect the consideration that immunotherapy achieves its treatment effect through activating immune activity. That is, we assume that conditional on the immune response  $Y_j$ ,  $\pi(d_j, Z_j, Y_j)$  and  $S_u(t|d_j, Z_j, Y_j)$  are independent of dose  $d_j$ . For notational brevity, we drop  $d_j$  in  $\pi(d_j, Z_j, Y_j)$  and  $S_u(t|d_j, Z_j, Y_j)$  hereafter. Based on the same consideration and the fact that the correlation between toxicity and efficacy is often weak for targeted agents such as immunotherapy, we assume that conditional on the immune response  $Y$ ,  $T$  and  $X$  are independent. Previous studies (e.g., Cai, Yuan, and Ji<sup>33</sup>) show that ignoring the correlation between toxicity and efficacy has little impact on the performance of phase I-II designs. In cases where these assumptions may be violated, we can simply add  $d_j$  and/or  $X_j$  as covariates in equations (6) and (7).

For the  $i$ th patient, let  $c_i$  denote the censoring time,  $t_i = \min(T_i, c_i)$  denote the actual observed time, and  $\delta_i = I(T_i < c_i)$  be the censoring indicator. The observed outcome from patient  $i$  are  $\mathbf{D}_i = (Y_i, X_i, t_i, \delta_i, Z_i, d_i)$ . Letting  $\Theta = (\alpha, \xi, \sigma^2, \mu_M, \beta_{0,0}, \beta_{0,1}, \beta_1, \beta_2, \eta_{c1}, \eta_{c2}, \eta_1, \eta_2, \lambda, \gamma)$  represent all parameters, the likelihood for the  $i$ th patient is

$$L_i(\mathbf{D}_i; \Theta) = f(Y_i | \Theta)f(X_i | Y_i, \Theta)f(t_i | Y_i, \Theta) \tag{8}$$

where

$$f(Y_i | \Theta) = \phi(Y_i; \mu_M \exp(\xi Z_i)(1 - \exp(-\alpha d_i)), \sigma^2), \tag{9}$$

$$f(X_i | Y_i, \Theta) = \Phi(\beta_0, Z_i + \beta_1 d_i + \beta_2 Y_i)^{X_i} (1 - \Phi(\beta_0, Z_i + \beta_1 d_i + \beta_2 Y_i))^{1 - X_i} \tag{10}$$

$$f(t_i | Y_i, \Theta) = \{(1 - \pi(Z_i, Y_i))f_u(t_i | Z_i, Y_i)\}^{\delta_i} \{(1 - \pi(Z_i, Y_i))S_u(t_i | Z_i, Y_i) + \pi(Z_i, Y_i)\}^{1 - \delta_i} \tag{11}$$

and  $\phi(\cdot; \mu, \sigma^2)$  denotes the probability density function for a normal distribution with mean  $\mu$  and variance  $\sigma^2$ ,  $f_u(t|Z, Y)$  represents the probability density function of PFS for susceptible patients in subgroup  $Z$  with immune response  $Y$ .

Let  $n = 1, \dots, N$  denote an interim sample size when a dose assignment decision is to be made during the trial, and  $\mathcal{D}_n = (\mathbf{D}_1, \dots, \mathbf{D}_n)$  denote the observed data from the first  $n$  patients. The likelihood for the first  $n$  patients in the trial is  $L(\mathcal{D}_n; \Theta) = \prod_{i=1}^n L_i(\mathbf{D}_i; \Theta)$ .

Let  $p(\Theta)$  denote the joint prior distribution of  $\Theta$ . The joint posterior distribution based on the data from the first  $n$  patients is  $p(\Theta | \mathcal{D}_n) \propto L(\mathcal{D}_n; \Theta)p(\Theta)$ . We sample from this posterior distribution using the Markov chain Monte Carlo algorithm with Gibbs sampler.<sup>34</sup>

## 2.2. Prior Distributions

Since  $\mu_M$  has the interpretation of the maximum mean immune response for marker-negative patients, and  $\exp(\xi)$  has the interpretation of the ratio of the maximum immune response of the marker-positive patients relative to the marker-negative patients, we elicit prior estimates of  $\mu_M$  and  $\exp(\xi)$  from clinicians, denoted as  $\hat{\mu}_M$  and  $\hat{r}$ , respectively.  $\mu_M$  is assigned a Gamma prior distributions with mean  $\hat{\mu}_M$  and a relatively large standard deviation (e.g.,  $3\hat{\mu}_M$ ) to obtain a vague prior, and  $\xi$  is assigned a Normal prior distribution with mean  $\log(\hat{r})$  and a variance such that there is a small, say 5%, probability on negative values of  $\xi$  to accommodate situations that the observed data are in contradiction with the ordering assumption. We assign  $\sigma^2$  a vague inverse Gamma prior distribution, e.g.,  $\sigma^2 \sim IG(0.1, 0.1)$  so that data will dominate the posterior distribution.

To specify the prior distribution for  $\alpha$  in equation (2), we use the idea of minimal prior knowledge.<sup>35</sup> By rewriting model (2) as  $\alpha d_i = -\log(1 - \frac{\mu_Y(d_i, Z_i)}{\mu_M \exp(\xi Z_i)})$ , a change of 4.6 moves  $\frac{\mu_Y(d_i, Z_i)}{\mu_M \exp(\xi Z_i)}$  from .01 to 0.99. Considering the range of  $\frac{\mu_Y(d_i, Z_i)}{\mu_M \exp(\xi Z_i)}$  in (0, 1), it is reasonable to assume that the effect of a covariate is unlikely to be more dramatic than that. Therefore, we scale  $d_i$  to have standard deviation 0.5, and assign  $\alpha$  a truncated normal prior distribution  $\alpha \sim \mathcal{N}(0, 2.3^2)I(\alpha > 0)$  so that a change in  $d_i$  from one standard deviation below the mean to one standard deviation above the mean will most likely result in a difference of less than 4.6 for the regression in equation (2).

In toxicity model (3), we assume  $\beta_{0,0}$  and  $\beta_{0,1}$  independently follow Normal prior distributions with mean  $-2$  and standard deviation  $0.5$ , i.e.,  $\mathcal{N}(-2, 0.5^2)$  so that a priori the toxicity probability in each subgroup is centered at  $0.02$  with 95% credible interval  $(0.001, 0.16)$  when there is no drug, i.e., when  $d_i = 0$  and  $Y_i = 0$ . We assign normal prior distributions  $\mathcal{N}(0, 1.25^2)$  to  $\beta_1$  and  $\beta_2$  after scaling  $d_i$  and  $Y_i$  so that each has standard deviation  $0.5$ . This is also based on the idea of minimal prior knowledge on the probit scale with details provided in Guo and Yuan.<sup>29</sup> Here we do not restrict  $\beta_1$  and  $\beta_2$  to be positive to accommodate the fact that toxicity with immunotherapy agents typically increases slowly with the dose, so the true values of  $\beta_1$  and  $\beta_2$  are close to 0. Constraining the values to be positive using, e.g., the Gamma or truncated normal priors, tends to inflate the estimates of  $\beta_1$  and  $\beta_2$  especially at the beginning of the trial when data are sparse, which hinders dose escalation and thus hurts the performance of the design. If toxicity is expected to increase rapidly with the dose, we can use the Gamma or truncated normal priors.

Likewise, in the cure model (6), we scale  $Y$  to have mean 0 and standard deviation 0.5 and shift the binary  $Z$  to have a mean of 0 and to differ by 1 in their lower and upper conditions, and assign  $\eta_{c2}$  a truncated normal prior distribution  $\eta_{c2} \sim \mathcal{N}(0, 1.25^2)I(\eta_{c2} > 0)$ . We use the same normal prior for the intercept  $\eta_{c0} \sim \mathcal{N}(0, 1.25^2)$ . The prior for  $\eta_{c1}$  is taken to be  $\eta_{c1} \sim \mathcal{N}(1.25, 0.75^2)$  so that there is a 90% prior probability that  $\eta_{c1}$  is between 0 and 2.5 and a 5% prior probability that  $\eta_{c1}$  is negative. With standardized  $Y$ , the elicitation of the prior for  $\eta_2$  in the hazard function (7) is facilitated by the fact that it represents the log of the hazard ratio when the immune response  $Y$  increases by two standard deviations. We elicit from

clinicians a range of hazard ratio that is practically feasible for  $\eta_2$ , say  $(r_{2,1}, r_{2,2})$ , and then assign  $\eta_2$  a uniform prior distribution with range  $(\log(r_{2,1}), \log(r_{2,2}))$ . For example, if the elicited hazard ratio when  $Y$  increases by two standard deviations is believed to be between 0.1 and 1, then the prior for  $\eta_2$  is  $\text{Uniform}(-2.3, 0)$ . Similarly,  $\eta_1$  is the log of the hazard ratio between the two subgroups. We elicit from clinicians a plausible range of the hazard ratio, say  $(r_{1,1}, r_{1,2})$ , that mostly covers values less than 1 with a small portion of values greater than 1. For instance,  $(r_{1,1}, r_{1,2}) = (0.1, 1.2)$  gives  $\eta_1 \sim \text{Uniform}(-2.3, 0.18)$ , so there is a 7% chance that the ordering assumption between the two subgroups is violated.  $\lambda$  and  $\gamma$  are assigned vague Gamma prior distributions  $\text{Gamma}(0.1, 0.1)$ .

Our design requires elicitation of prior information for  $\mu_M, \xi, \eta_1$ , and  $\eta_2$ , which is typically done by consulting the clinicians. Due to the typically small sample sizes of early phase clinical trials, especially at the beginning of the trial, it is common practice for early phase trials to incorporate prior knowledge elicited from clinicians (e.g., O’Quigley, Pepe, and Fisher<sup>36</sup>; Thall and Cook<sup>26</sup>; Lee, Thall, and Rezvani<sup>30</sup>; Lee, Thall, and Msaouel<sup>31</sup>). If clinicians have limited prior information, we recommend to use the vague prior approach, which can tolerate a substantial amount of uncertainty about the elicited estimates. We do not require the prior estimates to be accurately specified. The primary objective of eliciting estimates is to obtain a ballpark estimate of these parameters so that the prior is appropriately centered to avoid extreme (e.g., very large or small) values that may lead to inappropriate action (e.g., terminate the trial too early or escalate the dose too fast) at the beginning of the trial when data are very sparse. As the trial proceeds, the accumulating data will dominate the vague prior and guide dose transition. The simulation described later shows that the BIPSE design is not sensitive to the specifications of these elicited values as long as these priors are very vague. If historical data are available, the approach described in Lee, Thall, and Msaouel<sup>31</sup> could be adopted that uses historical data, possibly along with other elicited values from clinicians, to establish priors. In this paper, we consider the typical setting of a phase I/II trial with no historical data and use vague priors in the simulation studies.

### 2.3. Subgroup-specific Optimal Dose

To define subgroup-specific optimal dose, we first define the admissible dose set to safeguard against treating patients at doses that are futile or overly toxic. Let  $\phi_X$  denote the physician-specified upper limit of the toxicity rate  $\pi_X$ ,  $\phi_Y$  denote the lowest acceptable level of the immune response  $\mu_Y$ , and  $\phi_T$  denote the lowest acceptable progression-free time. A dose is deemed admissible for subgroup  $Z$  if the observed data indicate that there is a reasonable probability that its toxicity probability is less than the upper limit  $\phi_X$  (condition (12) below), its immune response is greater than the lower limit  $\phi_Y$  (condition (13) below), and its PFS is greater than  $\phi_T$  months (condition (14) below). More precisely, based on the data from the first  $n$  patients who have been enrolled into the trial, the admissible dose set  $\mathcal{A}_{Z,n}$  for subgroup  $Z$  is defined as the set of doses that satisfy the following toxicity and efficacy conditions

$$\Pr(\pi_X < \phi_X \mid d, Z, \mathcal{D}_n) > c_X \tag{12}$$

$$\Pr(\mu_Y > \phi_Y \mid d, Z, \mathcal{D}_n) > c_Y \tag{13}$$

$$\Pr(T > \phi_T \mid d, Z, \mathcal{D}_n) > c_T, \tag{14}$$

where  $c_X$ ,  $c_Y$  and  $c_T$  are probability cutoffs that should be tuned through simulation studies to achieve good design operating characteristics. Based on the fact that the immune response is typically quickly observable and an immunotherapeutic agent is deemed futile if it cannot trigger adequate immune response, condition (13) is important because it allows us to quickly screen and eliminate futile doses. In contrast, as the PFS takes a relatively long time to evaluate, condition (14) is more useful in the late stage of the trial to screen out non-optimal doses.

We measure the desirability of a dose using a utility function that accounts for the risk-benefit tradeoff that underlies medical decisions in practice. Let  $t_{max}$  denote the maximum follow-up time for patients. To facilitate the elicitation of the utility, we partition the support of  $T$  into  $R$  intervals,  $(\zeta_1, \zeta_2], \dots, (\zeta_{R-1}, \zeta_R], (\zeta_R, \zeta_{R+1}]$ , where  $\zeta_1 \equiv 0$ ,  $\zeta_R \equiv t_{max}$  and  $\zeta_{R+1} \equiv \infty$ , and let  $\rho(x, r)$  denote the utility for the outcome pair  $(X = x, T \in (\zeta_r, \zeta_{r+1}])$ ,  $x = 0, 1, r = 1, \dots, R$ . For example, if the maximum follow-up time  $t_{max}$  is 12 months, then the support of  $T$  can be partitioned into 5 intervals  $(0, 3], (3, 6], (6, 9], (9, 12], (12, \infty]$ . A convenient way of eliciting the utility  $\rho(x, r)$  is as follows: we fix the desirability of the most desirable outcome pair (i.e.,  $X = 0, T \in (12, \infty]$ ) as  $\rho(0, 5) = 100$  and the desirability of the least desirable outcome pair (i.e.,  $X = 1, T \in (0, 3]$ ) as  $\rho(1, 1) = 0$ . Using these two pairs as reference, we then ask physicians to score the desirability of the other combinations using the scale of  $(0, 100)$ . The elicited utility for our motivating trial is shown in Table 1. If physicians would prefer to increase the PFS by 3 month while sacrificing toxicity, then higher scores can be given to corresponding combinations. For example, if physicians think the combination of  $9 < T < 12$  and  $X = 1$  is more desirable than the combination of  $6 < T < 9$  and  $X = 0$ , then we can give a score of 45 to the combination of  $9 < T < 12$  and  $X = 1$ .

The subgroup-specific optimal dose for biomarker subgroup  $Z$  is defined as the admissible dose with the highest true mean utility

$$U_{true}(d, Z) = \sum_{x=0}^1 \sum_{r=1}^5 \rho(x, r) \Pr(X = x, T \in (\zeta_r, \zeta_{r+1}] \mid d, Z, \Theta).$$

During the trial, given the interim data  $\mathcal{D}_n$  from the first  $n$  patients, we evaluate the desirability of dose  $d$  based on the posterior mean utility,

$$U_n(d, Z) = \sum_{x=0}^1 \sum_{r=1}^5 \rho(x, r) E_{\Theta}[\Pr\{X = x, T \in (\zeta_r, \zeta_{r+1}] \mid d, \Theta\} \mid \mathcal{D}_n].$$

### 2.4. Dose-finding Algorithm

For dose finding involving subgroups, decision-making is difficult at the beginning of the trial due to the small available data. To alleviate this issue, a two-stage dose-finding algorithm is employed. The objective of stage I dose escalation is to quickly explore the dose space and accumulate some data to facilitate stage II model-based dose finding. In this stage, we perform dose escalation based on only the toxicity outcome without considering the immune response or PFS. Stage I dose-escalation continues until reaching the first unsafe dose for either subgroup and then seamlessly switches to stage II dose-finding. Specifically, we treat the first cohort of patients at the lowest dose level and then escalate until we reach a dose that violates the safety requirement  $\Pr(\pi_X < \phi_X | d, Z, \mathcal{D}_n) > c_I$  for either subgroup  $Z$ , where  $c_I$  is a probability cutoff that will be tuned through simulation studies. In stage I of the trial, data are sparse for each subgroup, making the estimates of the toxicity probabilities highly unreliable. As a result, we evaluate the safety requirement based on the simple Beta-Binomial model in stage I. Precisely, for subgroup  $Z$  at dose  $d$ , we assume that the toxicity probability  $\pi_X(Z, d)$  follows a Beta prior distribution  $\text{Beta}(\kappa_1, \kappa_2)$ , and the number of patients who experience toxicity  $m_d^Z$  among  $n_d^Z$  treated patients follows a binomial distribution  $\text{Binom}(n_d^Z, \pi_X(Z, d))$ . We set  $\kappa_1 = 0.1$  and  $\kappa_2 = 0.2$  so that the observed data dominate the posterior distribution. Under the Beta-Binomial model, the posterior distribution of  $\pi_X(Z, d)$  is  $\text{Beta}(\kappa_1 + m_d^Z, \kappa_2 + n_d^Z - m_d^Z)$ . When we reach a dose that violates the safety requirement or reaches the highest dose  $d_D$ , stage I completes, and the trial moves on to stage II. In stage I dose-escalation, dose skipping is not allowed for either subgroup, so if none of the patients in a cohort is in subgroup  $Z$ , then the current dose is repeated for subgroup  $Z$  in the next cohort. Although the stage I dose-escalation strategy does not utilize patient immune response or efficacy data, these data are collected to facilitate the model fitting in stage II.

Stage II dose finding is summarized as follows. Assuming that  $n$  patients have been enrolled in the trial,

1. Based on the currently observed data  $\mathcal{D}_n$ , determine the set of admissible doses  $\mathcal{A}_{Z,n}$  for each subgroup based on conditions (12), (13), and (14).
2. If  $\mathcal{A}_{Z,n}$  is empty for both subgroups, terminate the trial, and no dose is recommended for both subgroups. If  $\mathcal{A}_{Z,n}$  is empty for one of the subgroups, then no dose is recommended for that subgroup, and future patients in that subgroup are treated off protocol.
3. If  $\mathcal{A}_{Z,n}$  is not empty for at least one subgroup, then the patients in the next cohort in subgroup  $Z$  with non-empty  $\mathcal{A}_{Z,n}$  are adaptively randomized to dose  $\{d_j, d_j \in \mathcal{A}_{Z,n}\}$  with randomization probability  $\pi_j$  that is proportional to its posterior mean utility, i.e.,

$$\pi_j = \frac{U_n(d_j, Z)}{\sum_{d_{j'} \in \mathcal{A}_{Z,n}} U_n(d_{j'}, Z)}.$$

4. Repeat steps 1,2 and 3 until reaching the maximum sample size  $N$  or terminating the trial early. Select the admissible dose with the largest posterior mean utility  $U_M(d, Z)$  as the subgroup-specific optimal dose for subgroup  $Z$ .

For safety considerations, any untried dose cannot be skipped when escalating.

### 3. Simulation

#### 3.1. Main Results

We conducted extensive simulation studies to evaluate the performance of the BIPSE design. Taking the setting of the motivating trial, we considered five doses (0.1, 0.3, 0.5, 0.7, 0.9), with a maximum sample size of 60 and a cohort size of 3. The follow-up time for PFS was  $t_{max} = 12$  months, and patient accrual followed a Poisson process with a rate of 3 per month. We assumed the toxicity and immune response were quickly ascertainable. The upper limit for toxicity  $\phi_X = 0.3$ , the lower limit for immune response and PFS were  $\phi_Y = 2$  and  $\phi_T = 3$ , respectively, and the marker-positive prevalence  $\phi = 0.5$ . We set  $\hat{\mu}_M = 20$  as a prior estimate of the maximum immune response for marker-negative patients elicited from clinicians, and assign a vague prior  $\mu_M \sim \text{Gamma}(1/9, 1/180)$ , so that the prior mean was  $\hat{\mu}_M (= 20)$  and the prior standard deviation was 60, which was 3 times the prior mean. The prior for  $\xi$  was taken to be  $\text{Normal}(\log(1.5), 0.25^2)$  to match the physician-elicited ratio of the maximum immune response of the marker-positive patients relative to the marker-negative patients of 1.5 and 5% probability on negative values. We took probability cutoffs  $c_T = 0.3$ ,  $c_X = 0.12$  and  $c_Y = c_T = 0.05$ , which yielded desirable operating characteristics. The utility elicited from physicians is displayed in Table 1. The R program can be obtained by contacting the corresponding author.

We considered 10 scenarios in our simulation studies. For each scenario, we first specified the mean immune response  $\mu_Y(d, Z)$  at each dose  $d$  for each subgroup  $Z$ . Given these mean immune responses, the immune response  $Y_i$  for patient  $i$  treated at dose  $d_i$  in subgroup  $Z_i$  was generated from the normal model (1). Conditional on  $Y_i$ , the toxicity and PFS outcomes  $X_i$  and  $T_i$  were simulated from models (3) and (5), respectively. Table 2 shows the true mean immune response  $\mu_Y$ , toxicity probability  $\pi_X$ , survival probabilities  $\Pr(T > 3)$  and  $\Pr(T > 12)$ , and utility at each dose for the two subgroups for the 10 scenarios; and Figure 2 displays the true dose-response curves for the mean immune response, toxicity probability, and survival curves of PFS for the two subgroups under these scenarios. We can see that the 10 scenarios varied in the number of target doses, location of the target doses, and the patterns of the dose-response curves. We simulated 1,000 trials under each scenario. Data sharing is not applicable to this paper as all data in this article is computer simulated.

We compared the BIPSE design with two conventional phase I/II designs. The first conventional design (conventional I) considered only efficacy and toxicity, as in most existing phase I/II designs. To make a fair comparison, we used the same toxicity and efficacy models as in the proposed BIPSE design, but with the immune response term dropped such that

$$\Phi^{-1}\{\pi_X(d_i, Z_i, Y_i)\} = \beta_0, Z_i + \beta_1 d_i \tag{15}$$

$$S(t | d_i, Z_i, Y_i) = \pi(d_i, Z_i, Y_i) + (1 - \pi(d_i, Z_i, Y_i))S_u(t | d_i, Z_i, Y_i) \tag{16}$$

$$\Phi^{-1}(\pi(d_i, Z_i, Y_i)) = \eta_{c0} + \eta_{c1}Z_i + \eta_{c2}d_i, \tag{17}$$

$$S_u(t | d_i, Z_i, Y_i) = \exp[-\lambda t^\gamma \exp(\eta_1 Z_i + \eta_2 d_i)]. \tag{18}$$

To demonstrate the importance of accounting for the subgroup effect in immunotherapy dose finding, the second conventional design (conventional II) did not include the subgroup effects. Precisely, the immune response, toxicity, and efficacy models were

$$\mu_Y(d_i, Z_i) = \mu_M(1 - \exp(-\alpha d_i)) \tag{19}$$

$$\Phi^{-1}\{\pi_X(d_i, Z_i, Y_i)\} = \beta_0 + \beta_1 d_i + \beta_2 Y_i \tag{20}$$

$$S(t | d_i, Z_i, Y_i) = \pi(d_i, Z_i, Y_i) + (1 - \pi(d_i, Z_i, Y_i))S_u(t | d_i, Z_i, Y_i) \tag{21}$$

$$\Phi^{-1}(\pi(d_i, Z_i, Y_i)) = \eta_{c0} + \eta_{c2}Y_i, \tag{22}$$

$$S_u(t | d_i, Z_i, Y_i) = \exp[-\lambda t^\gamma \exp(\eta_2 Y_i)]. \tag{23}$$

Table 3 summarizes the operating characteristics of the BIPSE design and the two conventional designs. In scenarios 1–5, there is one target dose in each subgroup. In the first three scenarios, the target dose is the same for the two subgroups. In scenario 1, the mean immune response first increases and then plateaus for marker-negative subgroup and keeps increasing for marker-positive subgroup; toxicity increases rapidly while the two survival probabilities  $Pr(T > 3)$  and  $Pr(T > 12)$  increase slowly as dose increases. In scenario 2, the mean immune response and the two survival probabilities all plateau as dose increases. In scenario 3, toxicity remains low in the dose range under investigation; the immune response and survival probabilities increase with dose. For these three scenarios, the BIPSE design outperformed the conventional I design and yielded comparable performance with the conventional II design. For example, for scenario 1, the optimal dose is dose level 2 for both subgroups. The percentage of correct optimal dose selection (PCS) of the target dose was 61.5% and 75.2% for marker-negative and marker-positive subgroups, respectively, under the BIPSE design, 33.8% and 53.6% under the conventional I design that ignored the immune response and 66.4% and 66.4% under the conventional II design that ignored the subgroups. The BIPSE design assigned 8.3 and 10.8 patients to the two subgroups, and the

numbers were 8.7 and 9.9 under conventional I design and 10 and 10.2 under conventional II design.

In scenarios 4–5, each subgroup has only one target dose, but the target doses are different for the two subgroups. For these two scenarios, the proposed design resulted in higher PCS of the target doses and at least a comparable number of patients treated at the target doses than both conventional designs. For example, in scenario 5, the target dose is dose level 5 for marker-negative patients and dose level 3 for marker-positive patients. The proposed design yielded much higher PCS of the target doses than the two conventional designs (75% and 65.6% under the proposed design, 60.8% and 44.4%, and 42% and 12.4% under the two conventional designs), and assigned slightly more patients to the target doses (4.9 and 7.2 under the proposed design, and 4.6 and 7.2, and 4.7 and 7.1 under the two conventional designs).

In scenarios 6–8, one subgroup has one target dose and the other subgroup has two target doses. In scenario 9, both subgroups have two target doses. In these scenarios, the BIPSE design resulted in overall better performance than the two conventional designs, and the improvement can be substantial. For example, in scenario 7, the target doses are dose level 3 and 4 for marker-negative patients and dose level 3 for marker-positive patients. The PCS under BIPSE were 45% and 40.6% for marker-negative patients and 63.1% for marker-positive patients. The counterpart PCS were 1% and 7.8%, and 12% for the conventional I design and 38.4% and 49.4%, and 38.4% for the conventional II design.

In scenario 10, none of the doses is admissible for both subgroups because all doses are too toxic and futile. The trial was terminated 100% of the time under BIPSE and conventional II designs and nearly 100% for conventional I design.

### 3.2. Sensitivity Analyses

We carried out sensitivity analyses to assess the robustness of our proposed design's performance by using 1) a smaller sample size, 2) an alternative data generating process, 3) different marker-positive prevalences, 4) an alternative partition of the support of  $T$  and corresponding utility function, 5) alternative utility functions, and 6) alternative prior distributions.

When the maximum sample size dropped from 60 to 48, our design's performance was slightly worse, as summarized in Table 4, but the BIPSE design still selected the target dose(s) with a high probability.

We conducted additional simulation studies to investigate the performance of the proposed BIPSE design when data were generated from alternative models. Specifically, we matched the mean and variance of the immune response at each dose to those shown in Table 2. For toxicity, we used the logistic model

$$\text{logit}\{\pi_X(d_i, Z_i, Y_i)\} = \beta_0 + \beta_1 d_i + \beta_2 Y_i \quad (24)$$

For the PFS model, we used the same mixture cure model (5), (6) and (7) but with the baseline hazard in (7) replaced by the log logistic distribution with a shape parameter 4 so that the hazard function is hump-shaped. The parameters in the toxicity and efficacy models were selected to match the toxicity probabilities,  $P_T(T > 3)$ , and the utilities shown in Table 2. Both the admissible dose set and the optimal dose are the same as in the original setting. The operating characteristics of the proposed model are shown in Table 5. As can be seen, the results are very similar to the original results in Table 3.

We also examined the sensitivity of the BIPSE design to the marker prevalence ratio. We evaluated two additional values of marker-positive prevalence, 0.3 and 0.7. The operating characteristics of the BIPSE design are provided in Tables S1 and S2 in the Supplementary Materials. The selection percentages of the doses are very similar to the original ones shown in Table 3 when the prevalence is 0.5. These results indicate that the BIPSE design is not sensitive to marker prevalence.

We evaluated the performance of the BIPSE design when the support of  $T$  is partitioned into four-month intervals:  $(0, 4]$ ,  $(4, 8]$ ,  $(8, 12]$ ,  $(12, \infty)$  with corresponding utility function shown in Table S3 in the Supplementary Materials. Under this setting, the target doses under each scenario are the same as in the original setting except for the marker-positive subgroup in scenario 9. With this alternative partition and utility function, the utility values of dose levels 2 and 3 are 84.7 and 85.3 for  $Z = 1$  under scenario 9, which are very close. Table S4 in the Supplementary Materials shows the operating characteristics of the BIPSE design. The results are very similar to the original ones. For the marker-positive subgroup in scenario 9, the selection percentages of dose levels 2 and 3 are 0.316 and 0.524, respectively. So when the utility function is appropriately elicited that reflects the clinicians' opinion and judgment regarding the tradeoff between toxicity and efficacy, the results from the BIPSE design are in general robust to the partition of the support of  $T$ .

We also evaluated the sensitivity of the BIPSE design to the specification of the utility function. We considered two alternative utility functions displayed in Table S5 in the Supplementary Materials. Compared with the original utility function, the first (second) new utility in Table S5 gave higher (lower) desirability scores to DLT outcomes. As shown in Tables S6 and S7 in the Supplementary Materials, the results under the two alternative utility functions are generally similar to those reported in Table 3, and the BIPSE design resulted in comparable or better performance than the two conventional designs. This suggests that the proposed BIPSE design is not sensitive to the elicited utility.

Additional simulation analyses were performed to evaluate the robustness of the BIPSE design to different prior distributions. In our design, most parameters are assigned the standard non-informative priors or default priors. Four parameters,  $\mu_M$ ,  $\xi$ ,  $\eta_1$ , and  $\eta_2$ , require elicitation of prior information from clinicians. For these four parameters, we set  $\hat{\mu}_M = 20$ ,  $\hat{r} = 1.5$ ,  $r_{2,1} = 0.1$ ,  $r_{1,1} = 0.1$  in our original simulation setting. In the new sensitivity analyses, we evaluated two alternative sets of prior estimates. For the first set,  $\hat{\mu}_M = 15$ ,  $\hat{r} = 1.3$ ,  $r_{2,1} = 0.2$ ,  $r_{1,1} = 0.2$ ; and for the second set,  $\hat{\mu}_M = 25$ ,  $\hat{r} = 1.7$ ,  $r_{2,1} = 0.3$ ,  $r_{1,1} = 0.3$ . The corresponding alternative prior distributions are  $\mu_M \sim \text{Gamma}(1/9, 1/135)$ ,  $\xi \sim \text{Normal}(0.26, 0.16^2)$ ,  $\eta_1 \sim \text{Unif}(-1.6, 0.12)$ ,  $\eta_2 \sim \text{Unif}(-1.6, 0)$  for the first set; and  $\mu_M \sim$

$\text{Gamma}(1/9, 1/225)$ ,  $\xi \sim \text{Normal}(0.53, 0.32^2)$ ,  $\eta_1 \sim \text{Unif}(-1.2, 0.09)$ ,  $\eta_2 \sim \text{Unif}(-1.2, 0)$  for the second set. The operating characteristics of the BIPSE design with these alternative prior distributions are given in Tables S8 and S9 in the Supplementary Materials. As can be seen from these tables, these new results are similar to the original ones given by Table 3, suggesting that our design is not sensitive to prior estimates elicited by clinicians.

### 3.3. Trial illustration

We illustrate the BIPSE design under the setting of the motivating trial. The trial investigated five dose levels with a sample size of 60 and a cohort size of 3. Other settings, e.g., patient accrual rate, lower and upper limits for efficacy and toxicity, probability cutoffs, follow-up time for PFS, and marker prevalence ratio, were all taken to be the same as in the main simulation setting. Patient's outcomes were generated from scenario 5 in the main simulation setting (Table 2), in which the target dose is dose level 5 for marker-negative patients and dose level 3 for marker-positive patients.

Figure S1 in the Supplementary Materials shows the dose allocation and marker subgroups for patients in the 20 cohorts during the trial. No patients in the first three cohorts experienced toxicity, so we kept escalating the dose and treated the fourth cohort at dose level 4. Two marker-positive patients experienced toxicities in cohort 4 at dose level 4, which violated the safety requirement for the marker-positive subgroup. Therefore, Stage I dose escalation completed and we switched to Stage II dose finding. To provide a sense of the within-trial behavior, we report the posterior mean utility of each dose in each subgroup during the trial as a summary value of the multiple endpoints and parameters. Table S10 of the Supplementary Materials displays these posterior mean utility values when each cohort was enrolled in stage II, as well as the final posterior mean utilities at the end of the trial when all patients had finished follow-up. As Table S10 shows, at the early and middle stages of the trial, the true target dose may not always have the highest posterior mean utility. As the trial proceeds, at the late stage of the trial, the target dose has the highest posterior mean utility and the BIPSE design successfully recommended the target dose for both subgroups at the end of the trial.

## 4. Discussion

We have proposed the BIPSE design, a Bayesian biomarker-based phase I/II design to determine the subgroup-specific optimal dose for immunotherapy trials with progression-free survival endpoint. We develop a parsimonious yet flexible model to jointly model the immune response, toxicity, and PFS that allows strength borrowing across biomarker subgroups and utilize a utility function to quantify the desirability of the doses within each subgroup. We propose a two-stage dose-finding algorithm to assign patients to desirable doses based on their biomarker status and select the subgroup-specific optimal doses for future patients. Simulation studies show that the BIPSE design has desirable operating characteristics. It reliably identifies the subgroup-specific optimal dose with a large probability and assigns most patients to the optimal doses. Optimizing subgroup-specific dose has the potential of being more ethical and lowering the drug cost.

As we mention in the Introduction, the proposed design is motivated by an ovarian cancer trial being developed at Indiana University Cancer Center. Originally, this trial planned to use a conventional two-stage design. In the first stage of dose finding, the 3+3 design will be used to identify the MTD solely based on the DLTs and the biomarker information will be ignored. In the second stage of cohort expansion, more patients will be randomized to the MTD and the lower dose. The PFS information will be collected for all patients in the trial. At the end of the trial, subgroup-specific optimal dose will be selected by comparing the PFS between the MTD and the lower dose within each biomarker subgroup. Considering the desirable operating characteristics of the BIPSE design as shown in the simulation study, we are now collaborating with the PIs to re-design the trial using the proposed BIPSE design. The main simulation results in Table 3 will be included in the statistical plan section of the protocol to demonstrate the practical superiority of the BIPSE design.

In this article, we model binary toxicity to mimic the motivating trial and discuss how to handle ordinal toxicity. In some immunotherapy trials, it is more appropriate to model toxicity with various types and grades. In those cases, we can use the total toxicity burden for toxicity (Bekele and Thall<sup>37</sup>) We assume in this article that the toxicity outcome is quickly ascertainable. For some immunotherapy trials, the toxicity may be late-onset and requires a longer time to score. This will cause accrual suspension as the toxicity may not be observed soon enough to apply decision rules to choose treatments for new patients. One approach to address this issue is to apply the methodology proposed by Yuan and Yin<sup>38</sup> and Jin et al.<sup>39</sup>, which accommodates the delayed outcome using Bayesian data augmentation or the expectation-maximization algorithm. We assume there are two subgroups with ordered efficacy. As discussed in Section 2, the BIPSE design can be straightforwardly extended to multiple-subgroup situations with possibly unknown orders among subgroups. However, one limitation of the BIPSE design is that we speculate pre-determined subgroup status based on each patient's biomarker information. It is of vital interest to extend the BIPSE design to handle unknown subgroup situations. One possible solution is to use the latent class model proposed by Chapple and Thall<sup>40</sup> and Lin et al.<sup>41</sup> that utilizes a “spike-and-slab” prior to adaptively combine and/or collapse subgroups based on a data-driven latent subgroup membership variable.

In this article, we assume the two intercepts  $\beta_{00}$  and  $\beta_{01}$  are two separate parameters in the toxicity model (3) to accommodate the situation of heterogeneous toxicity between the two subgroups. In some cancer trials, the toxicity may be homogeneous across subgroups. In that case, we can do Bayesian model averaging (Hoeting et. al<sup>42</sup>) to account for model uncertainty and better borrow strength across subgroups. Specifically, let model  $M_1$  be the toxicity model (3) used in Section 2.1 so the two subgroups could have heterogeneous toxicity

$$\Phi^{-1}\{\pi_X(d_i, Z_i, Y_i)\} = \beta_0 + \beta_1 d_i + \beta_2 Y_i;$$

and model  $M_2$  be the homogeneous model

$$\Phi^{-1}\{\pi_X(d_i, Z_i, Y_i)\} = \beta_0 + \beta_1 d_i + \beta_2 Y_i,$$

i.e.,  $\beta_{00} = \beta_{01} \equiv \beta_0$ , so the two subgroups have homogeneous toxicity. Then the posterior distribution of any quantity of interest can be obtained as a weighted average of the posterior distributions under each model, weighted by the corresponding posterior model probabilities (Hoeting et al.<sup>42</sup>). If data suggest the two subgroups have heterogeneous (homogeneous) toxicity, then the posterior model probability of model  $M_1$  ( $M_2$ ) will be higher than that of model  $M_2$  ( $M_1$ ), and as a result, the posterior probability of a quantity will be dominated by that under model  $M_1$  ( $M_2$ ). Bayesian model averaging models both parameter uncertainty and model uncertainty.

The proposed BIPSE design employs the cure rate model for PFS, as it's often the case for immunotherapy that some patients may achieve long-term durable response. We recommend to use the cure model when there is strong biological evidence of such a cure fraction. As Ying et al.<sup>43</sup> showed, the cure model provided better prediction than a non-cure model especially at the late stage of the trial when evidence for a cure became clear. With little evidence of a cure fraction, the standard Cox proportional hazards model can be used instead.

One concern may be that the cure rate model has a relatively large number of parameters compared with the typically limited sample size of a phase I/II trial, which would lead to highly variable parameter estimates, especially at the beginning of the trial with sparse data. There are several reasons that our design can afford such a complex model. First, as noted by Guo, Garret, and Liu<sup>8</sup>, the primary objective of a phase I/II trial is to identify the optimal dose among a set of pre-specified doses, not to obtain accurate parameter estimates. This renders the design high tolerance to the variability of parameter estimates. As long as the method obtains the rank of estimated utility correctly, it will accurately select the target dose. Second, since our dose-finding algorithm does not allow skipping untried doses when escalating, highly variable parameter estimates do not cause any issue. At the beginning of the trial when the parameter estimates are highly variable, the dose-finding algorithm acts somewhat “semi-randomly” by trying the doses sequentially from low to high. In some anti-intuitive sense, such “randomness” is helpful because it provides the design freedom to explore the dose space, and to avoid getting stuck at a local optimum dose. When the trial proceeds and data accumulate, we obtain more reliable estimates and the dose assignment becomes more stable and converges to the target dose. Therefore, as long as we can make reasonable estimates at the middle or late stage of the trial, we are likely to make correct selection of the target dose at the end of the trial. Indeed, as the simulation results show, the BIPSE design is able to correctly identify the subgroup-specific optimal dose at the end of the trial with a large probability.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgment

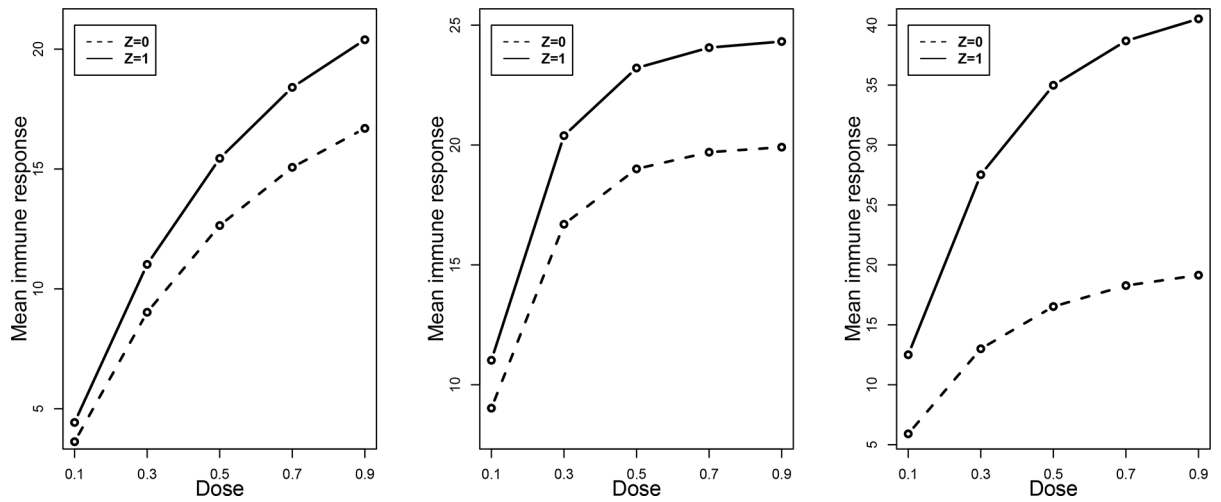
The authors thank the Associate Editor and two Referees for their valuable comments which substantially improved the presentation of this paper. The research of Beibei Guo is supported by the R & D Research Competitiveness Subprogram of Louisiana Board of Regents, Contract number LEQSF(2020–21)-RD-A-04. Yong Zang's research was partially supported by NIH/NCI grants P30 CA082709; R21 CA264257 and the Ralph W. and Grace M. Showalter Research Trust award.

## References

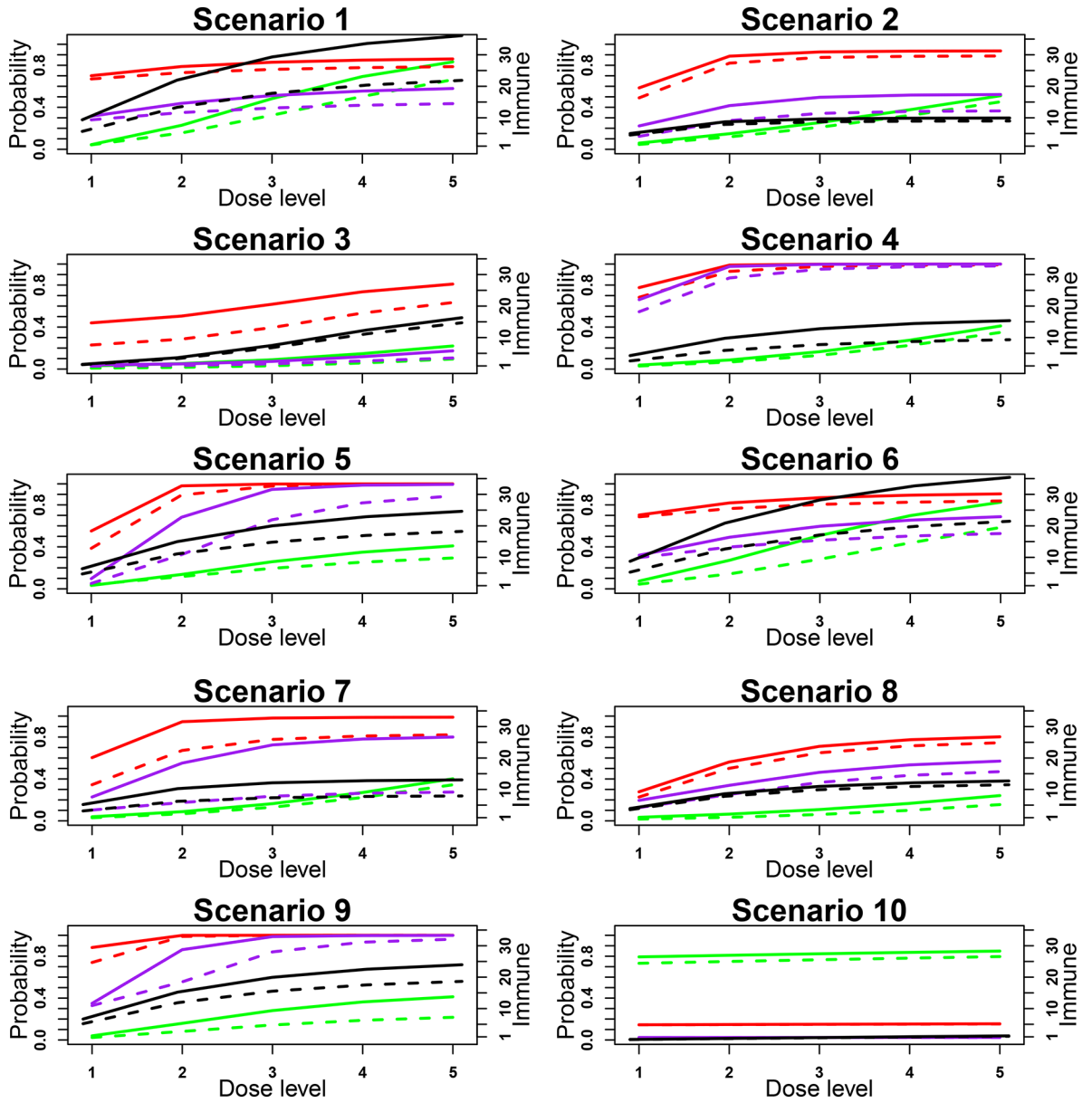
- [1]. Couzin-Frankel J (2013), Cancer immunotherapy. *Science*, 324, 1432–1433.
- [2]. Topalian SL, Weiner GJ and Pardoll DM (2011), Cancer immunotherapy comes of age. *Journal of Clinical Oncology*, 23, 4828–4836.
- [3]. Makkouk A, and Weiner GJ (2015), Cancer immunotherapy and breaking immune tolerance: new approaches to an old challenge. *Cancer Research*, 75, 5–10. [PubMed: 25524899]
- [4]. Liu S, Guo B, Yuan Y (2017) A Bayesian Phase I/II Design for Immunotherapy Trials. *Journal of the American Statistical Association*, 113: 1016–1027.
- [5]. Guo B, Li D, Yuan Y (2018) SPIRIT: A Seamless Phase I/II Randomized Design for Immunotherapy Trials. *Pharmaceutical Statistics*, 17: 527–540. [PubMed: 29882388]
- [6]. Guo B, Park Y, Liu S (2019) Utility-based Bayesian phase I/II design for immunotherapy trials with progression-free survival endpoint. *Journal of the Royal Statistical Society: Series C*, 68: 411–425.
- [7]. Wang C, Rosner G, Roden R (2019) A Bayesian design for phase I cancer therapeutic vaccine trials. *Statistics in Medicine*, 38: 1170–1189. [PubMed: 30368868]
- [8]. Guo B, Garrett E, Liu S (2021) A Bayesian phase I/II design for cancer clinical trials combining an immunotherapeutic agent with a chemotherapeutic agent. *Journal of the Royal Statistical Society: Series C*, In Press.
- [9]. Shi H, Cao J, Yuan Y, Lin R (2021) uTPI: A utility-based toxicity probability interval design for phase I/II dose-finding trials. *Statistics in Medicine*, DOI: 10.1002/sim.8922
- [10]. Topalian S, Hodi F, Brahmer J, et al. (2012), Safety, Activity, and Immune Correlates of Anti-PD-1 Antibody in Cancer. *The New England Journal of Medicine*, 366: 2443–2454. [PubMed: 22658127]
- [11]. Manhoney KM, Atkins MB (2014), Prognostic and predictive markers for the new immunotherapies. *Oncology*, 28: 39–48. [PubMed: 25384886]
- [12]. Borghaei H, Paz-Ares L, Horn L, et al. (2015), Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *The New England Journal of Medicine*, 373: 1627–39. [PubMed: 26412456]
- [13]. Garon EB, Rizvi NA, Hui R, et al. (2015), Pembrolizumab for the treatment of non-small-cell lung cancer. *The New England Journal of Medicine*, 372: 2018–28. [PubMed: 25891174]
- [14]. Larkin J, Chiarion-Sileni V, Gonzales R, et al. (2015), Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *The New England Journal of Medicine*, 373: 23–34. [PubMed: 26027431]
- [15]. Aguiar P, Santoro S, Tadokoro H, et al. (2016), The role of PD-L1 expression as a predictive biomarker in advanced non-small-cell lung cancer: a network meta-analysis. *Immunotherapy*, 8, 479–488. [PubMed: 26973128]
- [16]. Gibney G, Weiner L, Atkins M (2016) Predictive biomarkers for checkpoint inhibitor-based immunotherapy. *Lancet Oncology*, 17: 542–551.
- [17]. Abdel-Rahman O (2016), Correlation between PD-L1 expression and outcome of NSCLC patients treated with anti-PD-1/PD-L1 agents: a meta-analysis. *Critical Reviews in Oncology/Hematology*, 101, 75–85. [PubMed: 26969107]
- [18]. Wu K, Yi M, Qin S, Chu Q, Zheng X, Wu K (2019), The efficacy and safety of combination of PD-1 and CTLA-4 inhibitors: a meta-analysis. *Experimental Hematology & Oncology*, 8:26. [PubMed: 31673481]

- [19]. Bai R, Lv Z, Xu D, Cui J (2020), Predictive biomarkers for cancer immunotherapy with immune checkpoint inhibitors. *Biomarker Research*, 8–34. [PubMed: 32128213]
- [20]. Goldstein D, Gordon N, Davidescu M et al. (2017) A Pharmacoeconomic Analysis of Personalized Dosing vs Fixed Dosing of Pembrolizumab in Firstline PD-L1-Positive Non-Small Cell Lung Cancer. *Journal of the National Cancer Institute*, 109(11): djx063.
- [21]. Ingles Garcés A, Au L, Mason R, Thomas J, Larkin J (2019), Building on the anti-PD1/PD-L1 backbone: combination immunotherapy for cancer. *Expert Opinion on Investigational Drugs*, 28: 695–708. [PubMed: 31359805]
- [22]. Hodi FS, O’Day SJ, McDermott DF, et al. (2010), Improved survival with ipilimumab in patients with metastatic melanoma. *The New England Journal of Medicine*, 363: 711–723. [PubMed: 20525992]
- [23]. Robert C, Schachter J, Georgina VL, and et al. (2015) Pembrolizumab versus Ipilimumab in Advanced Melanoma. *The New England Journal of Medicine*, 372: 2521–2532. [PubMed: 25891173]
- [24]. Thall P, Russell K (1998) A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. *Statistics in Medicine*, 27, 4895–4913.
- [25]. Braun TM. The bivariate continual reassessment method: extending the CRM to phase I trials of two competing outcomes. *Controlled clinical trials* 2002; 23: 240–256. [PubMed: 12057877]
- [26]. Thall P, Cook J (2004) Dose-finding based on efficacy-toxicity trade-offs. *Biometrics*, 60, 684–693. [PubMed: 15339291]
- [27]. Zang Y, Lee JJ, Yuan Y. (2014) Adaptive designs for identifying optimal biological dose for molecularly targeted agents. *Clin Trials* 11: 319–327. [PubMed: 24844841]
- [28]. Han Y, Liu H, Cao S, Zhang C, Zang Y. (2021), TSNP: A two-stage nonparametric phase I/II clinical trial design for immunotherapy. *Pharmaceutical Statistics*, 20: 282–296. [PubMed: 33025762]
- [29]. Guo B, Yuan Y (2017) Bayesian phase I/II biomarker-based dose finding for precision medicine with molecularly targeted agents. *Journal of the American Statistical Association*, 112: 508–520. [PubMed: 32863478]
- [30]. Lee J, Thall P, Rezvani K (2019), Optimizing natural killer cell doses for heterogenous cancer patients on the basis of multiple event times. *Journal of the Royal Statistical Society, Series C*, 68: 461–474.
- [31]. Lee J, Thall P, Msaouel P (2021), Precision Bayesian phase I-II dose-finding based on utilities tailored to prognostic subgroups. *Statistics in Medicine*, DOI: 10.1002/sim.9120
- [32]. Yuan Y, Yin G (2009) Bayesian dose finding by jointly modeling toxicity and efficacy as time-to-event outcomes. *Journal of the Royal Statistical Society, Series C*, 58, 719–736.
- [33]. Cai C, Yuan Y, Ji Y (2014) A Bayesian dose finding design for oncology clinical trials of combinational biological agents. *Applied Statistics*, 63:159–173. [PubMed: 24511160]
- [34]. Robert C, Casella G (2004). *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 2nd edition.
- [35]. Gelman A, Jakulin A, Pittau MG, Su YS (2008) A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2: 1360–1383.
- [36]. O’Quigley J, Pepe M, Fisher L, (1990) Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics* 46: 33–48. [PubMed: 2350571]
- [37]. Bekele B, Thall P (2004), Dose-finding based on multiple toxicities in a soft tissue sarcoma trial. *Journal of the American Statistical Association*, 99:26–35.
- [38]. Yuan Y, Yin G (2011) Robust EM Continual Reassessment Method in Oncology Dose Finding. *Journal of the American Statistical Association*, 106: 818–831. [PubMed: 22375092]
- [39]. Jin I, Liu S, Thall P, Yuan Y (2014) Using Data Augmentation to Facilitate Conduct of Phase I/II Clinical Trials with Delayed Outcomes. *Journal of the American Statistical Association*, 109: 525–536. [PubMed: 25382884]
- [40]. Chapple A, Thall P (2018) Subgroup-specific dose finding in phase I clinical trials based on time to toxicity allowing adaptive subgroup combination. *Pharmaceutical Statistics*, 17: 734–749. [PubMed: 30112806]

- [41]. Lin R, Thall P, Yuan Y (2021) BAGS: A Bayesian Adaptive Group Sequential Trial Design With Subgroup-Specific Survival Comparisons. *Journal of the American Statistical Association*, 116: 322–334. [PubMed: 35582047]
- [42]. Hoeting J, Madigan D, Raftery A, Vilinsky C (1999) Bayesian model averaging: a tutorial. *Statistical Science*, 14: 382–417.
- [43]. Ying G, Zhang Q, Lan Y, Li Y, Heitjan D (2017) Cure modeling in real-time prediction: how much does it help?. *Contemporary Clinical Trials*, 59: 30–37. [PubMed: 28545934]



**Figure 1:** The mean immune response curves for the two subgroups generated from model (2). The parameters are  $\mu_M = 20, \xi = 0.2, \alpha = 2$ ;  $\mu_M = 20, \xi = 0.2, \alpha = 6$ ; and  $\mu_M = 20, \xi = 0.75, \alpha = 3.5$  for the three panels from left to right. Dashed and solid lines represent marker-negative ( $Z=0$ ) and marker-positive ( $Z=1$ ) subgroups, respectively.



**Figure 2:** Dose-response curves for the 10 scenarios in the simulation study. The dashed and solid lines are for subgroups  $Z = 0$  and  $Z = 1$ , respectively; and green, red, purple, and black curves are the toxicity probability  $\pi_X$ , survival probabilities  $Pr(T > 3)$  and  $Pr(T > 12)$ , and mean immune response  $\mu_Y$  curves, respectively. All probabilities are plotted against the left y-axis, and the immune response is plotted against the right y-axis.

**Table 1:**

Utility for the motivating trial.

	0	$T < 3$	3	$T < 6$	6	$T < 9$	9	$T < 12$	$T$	12
$X=0$	15		25		40		60		100	
$X=1$	0		5		15		25		50	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2:**

True mean immune response  $\mu_Y$ , toxicity probability  $\pi_X$ , survival probabilities  $Pr(T > 3)$  and  $Pr(T > 12)$ , and utility at each dose for the two subgroups. The boldface numbers are the subgroup-specific optimal dose.

	<b>Z=0</b>					<b>Z=1</b>				
	Scenario 1									
$\mu_Y$	5.7	<b>13.3</b>	17.8	20.4	21.9	9.4	<b>22.0</b>	29.3	33.6	36.1
$\pi_X$	0.042	<b>0.156</b>	0.324	0.503	0.663	0.044	<b>0.229</b>	0.482	0.693	0.832
$Pr(T > 3)$	0.670	<b>0.730</b>	0.761	0.778	0.787	0.701	<b>0.788</b>	0.829	0.849	0.860
$Pr(T > 12)$	0.281	<b>0.350</b>	0.393	0.419	0.435	0.315	<b>0.438</b>	0.512	0.555	0.579
Utility	45.8	<b>48.3</b>	46.2	42.1	37.6	48.8	<b>52.5</b>	48.5	43.2	39.2
	Scenario 2									
$\mu_Y$	4.5	7.9	<b>8.7</b>	8.9	9.0	5.0	8.7	<b>9.6</b>	9.9	9.9
$\pi_X$	0.046	0.118	<b>0.210</b>	0.323	0.452	0.059	0.148	<b>0.254</b>	0.377	0.510
$Pr(T > 3)$	0.490	0.821	<b>0.875</b>	0.886	0.889	0.585	0.887	<b>0.927</b>	0.935	0.937
$Pr(T > 12)$	0.123	0.274	<b>0.342</b>	0.361	0.366	0.223	0.416	<b>0.495</b>	0.516	0.522
Utility	30.2	46.1	<b>49.0</b>	46.6	42.6	38.3	55.6	<b>57.8</b>	54.6	49.8
	Scenario 3									
$\mu_Y$	1.4	3.3	6.8	11.2	<b>14.7</b>	1.5	3.6	7.5	12.4	<b>16.3</b>
$\pi_X$	0.009	0.017	0.031	0.058	<b>0.096</b>	0.031	0.052	0.089	0.147	<b>0.220</b>
$Pr(T > 3)$	0.229	0.285	0.395	0.533	<b>0.634</b>	0.440	0.505	0.617	0.736	<b>0.809</b>
$Pr(T > 12)$	0.028	0.035	0.050	0.077	<b>0.106</b>	0.039	0.049	0.073	0.118	<b>0.174</b>
Utility	19.3	20.3	22.5	26.0	<b>29.2</b>	22.3	23.7	26.8	31.8	<b>36.1</b>
	Scenario 4									
$\mu_Y$	2.6	5.9	<b>7.8</b>	8.8	9.3	4.3	<b>9.8</b>	12.8	14.5	15.4
$\pi_X$	0.029	0.067	<b>0.131</b>	0.226	0.350	0.038	<b>0.087</b>	0.166	0.276	0.411
$Pr(T > 3)$	0.684	0.930	<b>0.978</b>	0.989	0.993	0.776	<b>0.990</b>	0.999	1.000	1.000
$Pr(T > 12)$	0.546	0.868	<b>0.951</b>	0.974	0.982	0.661	<b>0.977</b>	0.998	0.999	1.000
Utility	63.4	87.1	<b>90.1</b>	87.0	81.5	72.2	<b>94.1</b>	91.6	86.2	79.5
	Scenario 5									
$\mu_Y$	4.7	11.1	14.8	17.0	<b>18.2</b>	6.4	15.0	<b>20.0</b>	22.9	24.6
$\pi_X$	0.040	0.117	0.196	0.254	<b>0.294</b>	0.032	0.137	<b>0.258</b>	0.349	0.409
$Pr(T > 3)$	0.386	0.894	0.979	0.993	<b>0.996</b>	0.551	0.981	<b>0.999</b>	1.000	1.000
$Pr(T > 12)$	0.052	0.326	0.658	0.819	<b>0.885</b>	0.097	0.682	<b>0.947</b>	0.988	0.995
Utility	23.1	51.2	71.2	77.8	<b>79.5</b>	29.3	75.2	<b>84.4</b>	81.9	79.3
	Scenario 6									
$\mu_Y$	5.3	<b>12.7</b>	<b>17.1</b>	19.8	21.5	8.8	<b>20.9</b>	28.2	32.7	35.4
$\pi_X$	0.045	<b>0.142</b>	<b>0.283</b>	0.438	0.585	0.074	<b>0.272</b>	0.508	0.697	0.825
$Pr(T > 3)$	0.686	<b>0.765</b>	<b>0.804</b>	0.826	0.837	0.703	<b>0.819</b>	0.868	0.892	0.905
$Pr(T > 12)$	0.298	<b>0.398</b>	<b>0.463</b>	0.503	0.527	0.320	<b>0.492</b>	0.596	0.654	0.687
Utility	47.3	<b>52.5</b>	<b>52.5</b>	49.6	45.7	48.3	<b>54.8</b>	52.6	48.2	44.6
	Scenario 7									

	<b>Z=0</b>				<b>Z=1</b>					
$\mu_Y$	3.1	6.2	<b>7.3</b>	<b>7.8</b>	7.9	5.2	10.2	<b>12.1</b>	12.8	13.0
$\pi_X$	0.030	0.067	<b>0.130</b>	<b>0.223</b>	0.344	0.039	0.087	<b>0.164</b>	0.270	0.401
$Pr(T > 3)$	0.345	0.672	<b>0.777</b>	<b>0.810</b>	0.822	0.604	0.947	<b>0.982</b>	0.988	0.990
$Pr(T > 12)$	0.103	0.175	<b>0.235</b>	<b>0.263</b>	0.275	0.227	0.552	<b>0.725</b>	0.781	0.800
Utility	26.6	37.2	<b>42.2</b>	<b>42.2</b>	39.6	39.6	68.4	<b>76.7</b>	75.2	70.3
Scenario 8										
$\mu_Y$	3.5	7.8	9.9	11.0	<b>11.5</b>	3.9	8.6	11.0	<b>12.1</b>	<b>12.7</b>
$\pi_X$	0.016	0.034	0.061	0.100	<b>0.155</b>	0.034	0.064	0.108	<b>0.166</b>	<b>0.241</b>
$Pr(T > 3)$	0.227	0.501	0.650	0.716	<b>0.746</b>	0.277	0.562	0.711	<b>0.774</b>	<b>0.802</b>
$Pr(T > 12)$	0.126	0.252	0.367	0.435	<b>0.470</b>	0.195	0.339	0.463	<b>0.534</b>	<b>0.570</b>
Utility	27.0	40.4	50.2	54.7	<b>55.7</b>	32.1	46.3	56.0	<b>59.6</b>	<b>59.6</b>
Scenario 9										
$\mu_Y$	5.2	11.9	15.5	<b>17.6</b>	<b>18.7</b>	6.7	<b>15.2</b>	<b>20.0</b>	22.5	24.0
$\mu_X$	0.024	0.082	0.143	<b>0.188</b>	<b>0.216</b>	0.039	<b>0.159</b>	<b>0.281</b>	0.363	0.413
$Pr(T > 3)$	0.740	0.989	0.999	<b>1.000</b>	<b>1.000</b>	0.883	<b>0.999</b>	<b>1.000</b>	1.000	1.000
$Pr(T > 12)$	0.329	0.555	0.841	<b>0.934</b>	<b>0.962</b>	0.349	<b>0.863</b>	<b>0.987</b>	0.997	0.999
Utility	47.7	71.1	85.1	<b>87.5</b>	<b>87.5</b>	52.5	<b>85.4</b>	<b>85.4</b>	81.7	79.3
Scenario 10										
$\mu_Y$	0.1	0.4	0.7	0.9	1.2	0.2	0.5	0.8	1.0	1.3
$\pi_X$	0.733	0.750	0.766	0.782	0.797	0.794	0.809	0.823	0.837	0.849
$Pr(T > 3)$	0.145	0.147	0.149	0.151	0.153	0.145	0.148	0.150	0.152	0.155
$Pr(T > 12)$	0.023	0.023	0.023	0.024	0.024	0.024	0.024	0.024	0.024	0.024
Utility	6.3	6.1	5.8	5.6	5.4	5.4	5.1	4.9	4.7	4.6

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3:**

Selection percentage and the average number of patients treated at each dose for the two subgroups under the BIPSE design and the two conventional designs which ignore the immune response  $Y$  or the subgroup  $Z$ . The numbers in boldface are target doses.

	<b>Z=0</b>					<b>Z=1</b>				
Scenario 1										
Sel % (BIPSE)	0.075	<b>0.615</b>	0.254	0.038	0.019	0.092	<b>0.752</b>	0.156	0.000	0.000
# of patients	7.9	<b>8.3</b>	7.4	4.4	2.1	10.2	<b>10.8</b>	6.3	2.0	0.6
Sel % (conventional I)	0.036	<b>0.338</b>	0.462	0.150	0.014	0.100	<b>0.536</b>	0.338	0.024	0.002
# of patients	8.3	<b>8.7</b>	7.8	3.9	1.1	9.6	<b>9.9</b>	7.6	2.6	0.6
Sel % (conventional II)	0.042	<b>0.664</b>	0.286	0.002	0.006	0.042	<b>0.664</b>	0.286	0.002	0.006
# of patients	9.2	<b>10.0</b>	7.1	2.5	1.0	9.2	<b>10.2</b>	7.1	2.5	1.0
Scenario 2										
Sel % (BIPSE)	0.000	0.046	<b>0.738</b>	0.181	0.035	0.000	0.110	<b>0.765</b>	0.098	0.027
# of patients	6.0	7.2	<b>7.3</b>	6.0	3.9	6.2	7.6	<b>7.1</b>	5.5	3.2
Sel % (conventional I)	0.000	0.018	<b>0.104</b>	0.308	0.570	0.000	0.032	<b>0.212</b>	0.472	0.284
# of patients	6.1	6.6	<b>7.0</b>	6.2	3.9	6.8	7.4	<b>7.1</b>	5.9	3.0
Sel % (conventional II)	0.000	0.060	<b>0.792</b>	0.110	0.038	0.000	0.060	<b>0.792</b>	0.110	0.038
# of patients	5.9	7.6	<b>7.5</b>	5.6	3.6	6.0	7.7	<b>7.1</b>	5.8	3.3
Scenario 3										
Sel % (BIPSE)	0.000	0.002	0.004	0.021	<b>0.967</b>	0.000	0.000	0.000	0.019	<b>0.979</b>
# of patients	3.5	6.7	6.9	6.7	<b>6.0</b>	5.0	6.0	6.6	6.5	<b>5.8</b>
Sel % (conventional I)	0.000	0.004	0.002	0.004	<b>0.990</b>	0.000	0.000	0.000	0.008	<b>0.992</b>
# of patients	4.8	5.6	6.4	6.6	<b>6.2</b>	5.2	5.7	6.4	6.6	<b>6.2</b>
Sel % (conventional II)	0.000	0.000	0.002	0.006	<b>0.988</b>	0.000	0.000	0.002	0.006	<b>0.988</b>
# of patients	4.3	6.2	6.6	6.6	<b>6.2</b>	4.4	6.0	6.7	6.7	<b>6.1</b>
Scenario 4										
Sel % (BIPSE)	0.004	0.129	<b>0.633</b>	0.185	0.046	0.062	<b>0.785</b>	0.142	0.002	0.006
# of patients	6.2	6.9	<b>6.7</b>	5.9	4.2	6.7	<b>7.0</b>	6.7	5.6	4.0
Sel % (conventional I)	0.004	0.092	<b>0.472</b>	0.380	0.052	0.118	<b>0.454</b>	0.384	0.040	0.004
# of patients	6.3	6.6	<b>6.8</b>	5.9	4.1	6.8	<b>6.8</b>	6.9	5.9	3.9
Sel % (conventional II)	0.004	0.338	<b>0.590</b>	0.048	0.018	0.004	<b>0.338</b>	0.590	0.048	0.018
# of patients	6.4	6.9	<b>6.7</b>	5.5	4.2	6.5	<b>6.8</b>	6.6	5.9	4.4
Scenario 5										
Sel % (BIPSE)	0.000	0.002	0.033	0.210	<b>0.750</b>	0.000	0.042	<b>0.656</b>	0.169	0.129
# of patients	5.1	6.5	7.0	6.2	<b>4.9</b>	5.8	7.5	<b>7.2</b>	5.5	4.0
Sel % (conventional I)	0.000	0.000	0.048	0.344	<b>0.608</b>	0.000	0.006	<b>0.444</b>	0.528	0.022
# of patients	5.3	6.4	7.1	6.6	<b>4.6</b>	6.2	7.4	<b>7.2</b>	5.7	3.6
Sel % (conventional II)	0.000	0.002	0.124	0.454	<b>0.420</b>	0.000	0.002	<b>0.124</b>	0.454	0.420
# of patients	5.2	6.9	7.1	6.0	<b>4.7</b>	5.4	6.8	<b>7.1</b>	6.0	4.8
Scenario 6										
Sel % (BIPSE)	0.040	<b>0.415</b>	<b>0.369</b>	0.117	0.058	0.112	<b>0.727</b>	0.156	0.000	0.002

	<b>Z=0</b>					<b>Z=1</b>				
# of patients	7.5	<b>7.9</b>	<b>7.2</b>	4.8	2.7	10.6	<b>10.8</b>	5.8	2.0	0.5
Sel % (conventional I)	0.012	<b>0.210</b>	<b>0.454</b>	0.272	0.052	0.112	<b>0.518</b>	0.350	0.020	0.000
# of patients	7.9	<b>8.4</b>	<b>7.6</b>	4.5	1.5	9.9	<b>10.2</b>	6.9	2.6	0.5
Sel % (conventional II)	0.032	<b>0.602</b>	<b>0.330</b>	0.018	0.014	0.032	<b>0.602</b>	0.330	0.018	0.014
# of patients	9.5	<b>10.0</b>	<b>6.4</b>	2.7	1.3	9.6	<b>10.1</b>	6.6	2.7	1.1
Scenario 7										
Sel % (BIPSE)	0.000	0.010	<b>0.450</b>	<b>0.406</b>	0.131	0.000	0.031	<b>0.631</b>	0.273	0.062
# of patients	5.4	6.9	<b>6.9</b>	<b>6.1</b>	4.5	5.7	7.0	<b>6.9</b>	6.1	4.4
Sel % (conventional I)	0.000	0.000	<b>0.010</b>	<b>0.078</b>	0.912	0.000	0.010	<b>0.120</b>	0.456	0.414
# of patients	5.2	6.1	<b>6.6</b>	<b>6.4</b>	5.7	6.0	6.7	<b>6.7</b>	6.0	4.6
Sel % (conventional II)	0.000	0.000	<b>0.384</b>	<b>0.494</b>	0.122	0.000	0.000	<b>0.384</b>	0.494	0.122
# of patients	5.2	6.9	<b>6.7</b>	<b>6.3</b>	4.6	5.5	6.6	<b>7.0</b>	6.4	4.9
Scenario 8										
Sel % (BIPSE)	0.000	0.000	0.006	0.188	<b>0.806</b>	0.000	0.000	0.029	<b>0.350</b>	<b>0.621</b>
# of patients	4.7	6.4	6.8	6.7	<b>5.5</b>	5.0	6.5	6.9	<b>6.2</b>	<b>5.3</b>
Sel % (conventional I)	0.000	0.000	0.000	0.008	<b>0.992</b>	0.000	0.000	0.000	<b>0.020</b>	<b>0.980</b>
# of patients	4.6	5.7	6.4	6.7	<b>6.4</b>	5.3	6.1	6.7	<b>6.5</b>	<b>5.6</b>
Sel % (conventional II)	0.000	0.000	0.006	0.256	<b>0.732</b>	0.000	0.000	0.006	<b>0.256</b>	<b>0.732</b>
# of patients	4.8	6.2	6.6	6.4	<b>5.6</b>	4.9	6.2	6.9	<b>6.3</b>	<b>5.7</b>
Scenario 9										
Sel % (BIPSE)	0.000	0.002	0.144	<b>0.333</b>	<b>0.521</b>	0.008	<b>0.498</b>	<b>0.383</b>	0.013	0.098
# of patients	5.6	6.8	7.0	<b>6.1</b>	<b>4.8</b>	6.4	<b>7.3</b>	<b>6.6</b>	5.4	3.9
Sel % (conventional I)	0.000	0.006	0.084	<b>0.598</b>	<b>0.312</b>	0.008	<b>0.176</b>	<b>0.668</b>	0.144	0.004
# of patients	6.2	6.7	6.6	<b>6.0</b>	<b>4.6</b>	6.9	<b>7.4</b>	<b>6.7</b>	5.5	3.5
Sel % (conventional II)	0.000	0.036	0.426	<b>0.208</b>	<b>0.330</b>	0.000	<b>0.036</b>	<b>0.426</b>	0.208	0.330
# of patients	5.9	6.8	6.8	<b>5.9</b>	<b>4.7</b>	6.0	<b>7.3</b>	<b>6.5</b>	5.7	4.3
Scenario 10										
Sel % (BIPSE)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
# of patients	1.8	1.0	0.3	0.0	0.0	1.9	1.5	0.3	0.0	0.0
Sel % (conventional I)	0.020	0.006	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000
# of patients	5.8	3.2	0.9	0.1	0.0	10.7	2.8	0.5	0.1	0.0
Sel % (conventional II)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
# of patients	1.9	1.1	0.3	0.0	0.0	1.7	1.1	0.3	0.0	0.0

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 4:**

Selection percentage and the average number of patients treated at each dose for the two subgroups under the BIPSE design when the maximum sample size is 48.

	<b>Z=0</b>				<b>Z=1</b>						
	Scenario 1										
Sel %	0.092	<b>0.524</b>	0.302	0.062	0.018	0.102	<b>0.736</b>	0.158	0.002	0.000	
# of patients	6.3	<b>6.6</b>	5.7	3.4	1.6	8.1	<b>8.2</b>	5.2	2.1	0.6	
	Scenario 2										
Sel %	0.000	0.042	<b>0.634</b>	0.258	0.066	0.000	0.094	<b>0.710</b>	0.168	0.028	
# of patients	5.0	6.0	<b>5.6</b>	4.5	2.8	5.3	6.1	<b>5.7</b>	4.5	2.5	
	Scenario 3										
Sel %	0.002	0.004	0.006	0.032	<b>0.946</b>	0.000	0.000	0.008	0.052	<b>0.934</b>	
# of patients	3.3	5.4	5.6	5.3	<b>4.4</b>	4.4	5.1	5.3	5.0	<b>4.0</b>	
	Scenario 4										
Sel %	0.018	0.166	<b>0.536</b>	0.208	0.072	0.126	<b>0.686</b>	0.168	0.012	0.008	
# of patients	5.2	5.6	<b>5.3</b>	4.6	3.1	5.8	<b>5.8</b>	5.4	4.4	2.7	
	Scenario 5										
Sel %	0.000	0.000	0.050	0.270	<b>0.676</b>	0.000	0.086	<b>0.630</b>	0.158	0.122	
# of patients	4.4	5.4	5.5	5.0	<b>3.8</b>	4.8	6.0	<b>5.5</b>	4.5	2.9	
	Scenario 6										
Sel %	0.040	<b>0.386</b>	<b>0.358</b>	0.126	0.090	0.128	<b>0.710</b>	0.156	0.004	0.002	
# of patients	6.1	<b>6.6</b>	<b>5.6</b>	3.7	2.0	8.6	<b>8.1</b>	4.9	1.9	0.6	
	Scenario 7										
Sel %	0.000	0.022	<b>0.400</b>	<b>0.422</b>	0.156	0.002	0.046	<b>0.600</b>	0.284	0.068	
# of patients	4.7	5.6	<b>5.5</b>	<b>4.8</b>	3.2	4.9	5.7	<b>5.8</b>	4.7	3.1	
	Scenario 8										
Sel %	0.000	0.000	0.026	0.188	<b>0.786</b>	0.000	0.000	0.046	<b>0.326</b>	<b>0.628</b>	
# of patients	3.9	5.3	5.6	5.2	<b>4.1</b>	4.3	5.2	5.5	<b>5.0</b>	<b>3.7</b>	
	Scenario 9										
Sel %	0.002	0.010	0.182	<b>0.358</b>	<b>0.448</b>	0.020	<b>0.480</b>	<b>0.386</b>	0.030	0.084	
# of patients	4.8	5.6	5.6	<b>4.5</b>	<b>3.3</b>	5.7	<b>6.1</b>	<b>5.4</b>	4.2	2.8	
	Scenario 10										
Sel %	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
# of patients	1.8	1.1	0.2	0.0	0.0	2.1	1.4	0.3	0.0	0.0	

**Table 5:**

Selection percentage and the average number of patients treated at each dose for the two subgroups under the BIPSE design with an alternative data-generating process.

	<b>Z=0</b>				<b>Z=1</b>						
	Scenario 1										
Sel %	0.036	<b>0.680</b>	0.172	0.002	0.004	0.362	<b>0.627</b>	0.006	0.000	0.000	
# of patients	7.5	<b>7.4</b>	6.4	3.7	1.8	10.0	<b>9.0</b>	5.3	1.9	0.5	
	Scenario 2										
Sel %	0.000	0.108	<b>0.692</b>	0.138	0.032	0.000	0.132	<b>0.716</b>	0.126	0.022	
# of patients	4.5	7.3	<b>7.5</b>	6.0	3.7	6.1	7.8	<b>7.2</b>	5.5	3.2	
	Scenario 3										
Sel %	0.002	0.002	0.004	0.012	<b>0.978</b>	0.000	0.000	0.000	0.018	<b>0.980</b>	
# of patients	3.5	6.4	6.8	6.8	<b>6.2</b>	4.9	6.0	6.6	6.6	<b>5.8</b>	
	Scenario 4										
Sel %	0.000	0.066	<b>0.584</b>	0.282	0.066	0.028	<b>0.668</b>	0.292	0.010	0.000	
# of patients	6.0	6.8	<b>6.8</b>	5.8	4.2	6.7	<b>7.1</b>	6.8	5.7	4.1	
	Scenario 5										
Sel %	0.000	0.002	0.038	0.248	<b>0.712</b>	0.002	0.038	<b>0.600</b>	0.204	0.156	
# of patients	4.8	6.3	7.0	6.7	<b>5.1</b>	5.8	7.4	<b>7.1</b>	5.7	4.1	
	Scenario 6										
Sel %	0.066	<b>0.458</b>	<b>0.314</b>	0.078	0.068	0.132	<b>0.648</b>	0.196	0.020	0.000	
# of patients	7.0	<b>8.0</b>	<b>7.1</b>	4.6	2.7	10.6	<b>11.2</b>	5.4	2.1	0.6	
	Scenario 7										
Sel %	0.000	0.006	<b>0.514</b>	<b>0.386</b>	0.094	0.000	0.122	<b>0.726</b>	0.124	0.028	
# of patients	5.5	6.8	<b>6.9</b>	<b>6.1</b>	4.4	6.0	6.9	<b>6.9</b>	6.1	4.3	
	Scenario 8										
Sel %	0.000	0.002	0.014	0.238	<b>0.744</b>	0.000	0.000	0.048	<b>0.304</b>	<b>0.646</b>	
# of patients	4.8	6.2	6.8	6.7	<b>5.5</b>	4.9	6.4	7.0	<b>6.4</b>	<b>5.1</b>	
	Scenario 9										
Sel %	0.000	0.000	0.196	<b>0.380</b>	<b>0.422</b>	0.008	<b>0.456</b>	<b>0.412</b>	0.032	0.090	
# of patients	5.2	6.9	6.9	<b>6.2</b>	<b>4.8</b>	6.6	<b>7.4</b>	<b>6.8</b>	5.2	3.9	
	Scenario 10										
Sel %	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
# of patients	1.7	0.8	0.1	0.0	0.0	2.1	1.5	0.3	0.0	0.0	