ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Applied Statistics
Series C

Appl. Statist. (2018)
67, *Part* 3, *pp.* 723–740

BLAST: Bayesian latent subgroup design for basket trials accounting for patient heterogeneity

Yiyi Chu

University of Texas School of Public Health, Houston, USA

and Ying Yuan

University of Texas MD Anderson Cancer Center Houston, USA

[Received April 2017. Revised October 2017]

Summary. The basket trial refers to a new type of phase II cancer trial that evaluates the therapeutic effect of a targeted agent simultaneously in patients with different types of cancer that involve the same genetic or molecular aberration. Although patients who are enrolled in the basket trial have the same molecular aberration, it is common for the targeted agent to be effective for patients with some types of cancer, but not others. We propose a Bayesian latent subgroup trial (BLAST) design to accommodate such treatment heterogeneity across cancer types. We assume that a cancer type may belong to the sensitive subgroup, which is responsive to the treatment, or the insensitive subgroup, which is not responsive to the treatment. Conditionally on the latent subgroup membership of the cancer type, we jointly model the binary treatment response and the longitudinal biomarker measurement that represents the biological activity of the targeted agent. The BLAST design makes the interim go–no-go treatment decision in a group sequential fashion for each cancer type on the basis of accumulating data. The simulation study shows that the BLAST design outperforms existing trial designs. It yields high power to detect the treatment effect for sensitive cancer types that are responsive to the treatment and maintains a reasonable type I error rate for insensitive cancer types that are not responsive to the treatment.

Keywords: Basket trials; Bayesian adaptive design; Longitudinal biomarker; Precision medicine; Subgroups; Targeted therapy

1. Introduction

With tremendous advances in biomarker development and genomic medicine, the forefront of oncology therapeutic research has shifted from conventional chemotherapy to targeted therapy that treats cancer by targeting a specific genetic or molecular aberration. The basket trial is a novel type of clinical trial that accommodates this paradigm shift (Redig and Jänne, 2015). In contrast with traditional phase II trials that focus on evaluating a treatment in patients with a certain type of cancer, basket trials seek to evaluate simultaneously the effects of a particular targeted therapy on patients with the same genetic or molecular aberration, regardless of their cancer types. In other words, a basket trial enrols patients with different types of cancer into the trial, as long as they bear the same genetic aberration. This provides great potential for patients with rare cancers to be eligible to participate in clinical trials. It also enables the incorporation of precision medicine in clinical trials, even for mutations that are difficult to study solely within

Address for correspondence: Ying Yuan, Department of Biostatistics, MD Anderson Cancer Center, Unit 1411, 1400 Pressler Street, Houston, TX 77030, USA.
E-mail: yyuan@mdanderson.org

 0035–9254/18/67723

a disease-specific context. In addition, unlike traditional clinical trials, for which a large number of patients is needed to establish therapeutic efficacy, the basket trial often requires a smaller number of patients and a shorter duration to identify a favourable response to the targeted therapy (Redig and Jänne, 2015; Berry, 2015). This ensures that basket trials optimize the translation of findings in a timely and safe way (Simon and Roychowdhury, 2013).

Despite increasing awareness and importance of basket trials in clinical research, designing efficient basket trials remains challenging. A straightforward approach is the pooled analysis, in which we treat patients with different types of cancer as a homogeneous population and simply pool the data across cancer types for designing the trial and performing interim analysis. This approach, however, leads to large biases if the treatment effect actually is heterogeneous across different cancer types. Another straightforward approach is the independent analysis, where we regard each cancer type as an independent treatment arm and conduct the analysis in each cancer type arm independently. This approach acknowledges the potential heterogeneity across cancer types but is less efficient and often lacks sufficient power to detect the treatment effect due to the limited sample size in each cancer type. Thall *et al.* (2003) proposed and Berry *et al.* (2013) advocated the use of a Bayesian hierarchical model (BHM) to borrow information across different cancer types in basket trials, assuming that the treatment effects of the cancer types are exchangeable and centre near a common mean. The exchangeability assumption underlying the BHM approach, however, is often violated in practice. In basket trials, it is common that some cancer types are sensitive to a targeted agent, whereas others are not sensitive, although they carry the same genetic aberration. In this paper, we use 'sensitive' to refer to patients who are responsive to and benefiting from the treatment and 'insensitive' to refer to patients who are unresponsive to and not benefiting from the treatment. For example, BRAF mutant gene melanoma and hairy cell leukaemia are sensitive to the BRAF inhibitor PLX4032 (vemurafenib), whereas BRAF mutant gene colon cancer is not (Flaherty *et al.*, 2010; Tiacci *et al.*, 2011; Prahallad *et al.*, 2012). Trastuzumab is effective for treating human epidermal growth factor receptor 2 (HER2) positive breast cancer but shows little clinical benefit for HER2 positive recurrent endometrial cancer (Fleming *et al.*, 2010) or HER2 positive non-small-cell lung cancer (Gatzemeier *et al.*, 2004). In other words, different cancer types in a basket trial may not be exchangeable regarding the treatment effect; instead, they often consist of different subgroups that respond differently to the treatment. Another issue is that, because the BHM assumes that the treatment effects for different cancer types centre near a common mean, it tends to overshrink the treatment effect of cancer types towards the common mean, resulting in substantially inflated type I error rates for the cancer types that are insensitive to the treatment (Freidlin and Korn, 2013; Chu and Yuan, 2017). Motivated by that, Chu and Yuan (2017) proposed a calibrated BHM for basket trials to obtain better type I error control than the BHM. Cunanan *et al.* (2017) proposed a strategy to allow a binary interim decision of either pooling data across all cancer types (i.e. fully borrowing) or treating each of them independently (i.e. no borrowing).

We propose a Bayesian latent subgroup trial (BLAST) design for basket trials. Taking a latent class modelling approach, we aggregate different cancer types into subgroups. Within each subgroup, the treatment effect is similar and approximately exchangeable such that information borrowing can be carried out by using a BHM. Another innovation of the BLAST design is that we leverage longitudinal biomarker measurements that are routinely taken in clinical trials to improve the efficiency of the basket trial. In this paper, 'biomarker' refers to the type of biomarkers that measure the biological activity of the targeted agent, i.e. measure how well the targeted agent hits its molecular target and triggers the downstream biological activities (e.g. expression of a certain gene, proliferation of certain cells or an increase in enzyme activity). For example, in immunotherapy, the number of CD8+ T-cells, CD4+ T-cells or the concentration

of cytokines (e.g. IFN-α, IL-6 or IL-8) are routinely measured to assess the biological activity (i.e. immunogenicity) of immune checkpoint inhibitors. As the biological activity of the targeted agent is closely related to its clinical efficacy (Galon *et al.*, 2006; Zuiverloon *et al.*, 2012; Pardoll, 2012), the longitudinal biomarker measurements provide rich information to aggregate disease types into subgroups. We use a semiparametric model to model jointly the longitudinal biomarker measurements with the binary clinical outcome to facilitate adaptive decision making in the trial. To the best of our knowledge, the BLAST design is the first basket trial design that jointly utilizes longitudinal biomarker measurements and clinical outcome. In this paper, we do not consider subgroups within a cancer type because of the small sample size that is available for each cancer type and thus very limited power to identify the within-tumour-type subgroups reliably. This assumption of homogeneity within each cancer type is often reasonable by using appropriate patient eligibility criteria.

Our research is motivated by a phase II basket trial that is currently in design at the MD Anderson Cancer Center for patients with CDKN2A deficient advanced solid tumours. The trial studied a novel aurora kinase inhibitor that targets CDKN2A in the CDKN2A signalling pathway, which plays a major role in cell cycle checkpoint regulation. CDKN2A aberration has been found in several types of advanced cancer and contributes to tumorigenesis by driving uncontrolled cell cycles (Okada and Mak, 2004). CDKN2A deletion appears in bladder cancer (43.8%), brain tumours (29.1%), melanoma (27.6%), pancreatic cancer (24.1%), oesophageal cancer (21.3%) and gastric cancer (13.80%). The goal of the trial is to evaluate the efficacy of the aurora kinase inhibitor in patients with cancers that harbour CDKN2A deletion or mutation. The trial includes patients with the six cancer types listed above and will enrol up to 20 patients for each cancer type. The treatment efficacy will be scored by using the 'Response evaluation criteria in solid tumors', version 1.1, and coded as 'response' if the patient achieves complete remission or partial remission or otherwise 'no response'. The targeted agent will be regarded as unpromising if the response rate is lower than 20%, and promising if the response rate is higher than 30%. The immunofluorescent intensity of phospho-aurora-A (Thr288) will be measured repeatedly over time to evaluate the biological activity of the aurora kinase inhibitor, i.e. the effectiveness of the inhibitor to hit the target and to inhibit the CDKN2A pathway.

The rest of this paper is organized as follows. In Section 2, we propose the probability model and decision rule for the BLAST design. In Section 3, we present simulation studies to examine the performance of the BLAST design. We conclude with a brief discussion in Section 4.

The software for implementing the BLAST design is available from `http://odin.mdacc.tmc.edu/~yyuan/index_code.html`.

2. Methods

2.1. Probability model

Consider a phase II basket trial that is designed to evaluate the efficacy of a targeted agent in I different cancer types that carry the same genetic or molecular aberration. Let p_i denote the treatment response rate for the ith cancer type, $i = 1, \ldots, I$. The objective of the basket trial is to test whether the targeted agent is effective in each of the cancer types:

$$H_0 : p_i \leqslant q_0 \quad \textit{versus} \quad H_a : p_i \geqslant q_1 \qquad \text{for } i = 1, \ldots, I,$$

where q_0 is the response rate cut-off under which the agent is deemed futile, and q_1 is the target response rate under which the agent is regarded as promising.

Assume that, at an interim go–no-go treatment decision time, n_i patients with the ith cancer

type have been enrolled. Let Y_{ij} denote a binary variable for the treatment response of the jth patient in the ith cancer type, with $Y_{ij}=1$ denoting the favourable treatment response (e.g. complete remission or partial remission). Let Z_{ijl} denote the biomarker measurement for the jth patient in the ith cancer type at the time t_l, for $l=1,\ldots,L$. For notational brevity, we assume that the biomarker is measured according to the same time schedule across patients, but our method allows different patients to have different numbers of measurements taken on different time schedules.

We assume that I cancer types can be classified into K subgroups, $1 \leqslant K \leqslant I$, such that, within each subgroup, patients respond similarly to the treatment. A simple but practically important case is $K=2$, with a sensitive subgroup consisting of cancer types that are sensitive to the targeted agent, and an insensitive subgroup consisting of cancer types that are not sensitive to the targeted agent. Our methodology is not limited to two subgroups (sensitive or insensitive) but allows for multiple subgroups with varying levels of sensitivity, e.g. $K=3$ subgroups representing insensitive, somewhat sensitive and highly sensitive. We temporarily assume that the number of subgroups K is known and discuss how to determine the value of K later. Let C_i denote the latent subgroup membership indicator, with $C_i=k$ denoting that the ith cancer type belongs to the kth subgroup, $k=1,\ldots,K$. We assume that C_i follows a multinomial distribution,

$$C_i \sim \text{multinomial}(\pi_1,\ldots,\pi_K), \tag{1}$$

where $\pi_k=\Pr(C_i=k)$, $k=1,\ldots,K$, is the probability that the ith cancer type belongs to the kth subgroup, with $\Sigma_{k=1}^{K}\pi_k=1$. As C_i is latent, its value is never observed and is estimated jointly with other model parameters by using the Markov chain Monte Carlo method as described later.

Conditionally on C_i, we model the joint distribution of (Y,Z) by first specifying a model for Y and then a model for Z conditional on Y. Specifically, we assume that treatment response Y_{ij} follows a latent subgroup hierarchical model

$$\left.\begin{aligned} Y_{ij}|p_i &\sim \text{Ber}(p_i),\\ \theta_i &= \log\left(\frac{p_i}{1-p_i}\right),\\ \theta_i|C_i=k &\sim N(\theta_{(k)},\tau^2_{(k)}), \end{aligned}\right\} \tag{2}$$

where $\text{Ber}(\cdot)$ denotes a Bernoulli distribution, $N(\cdot)$ denotes a normal distribution, θ_i is the logit transformation of response rate p_i and $\theta_{(k)}$ is the mean of θ_i in the kth subgroup. We assume that θ_i is random and follows a normal distribution to accommodate that, although the response rates of the cancer types in the kth subgroup are generally similar, they may deviate from the subgroup mean $\theta_{(k)}$. A more parsimonious but slightly more restrictive model is to treat θ_i as a fixed effect by setting $\theta_i=\theta_{(k)}$, or equivalently $\tau^2_{(k)}=0$, which indicates that all cancer types in the kth subgroup have the same response rate $\theta_{(k)}$. Here, we focus on the case in which Y_{ij} is a binary response variable. Our approach can be easily extended to accommodate a continuous outcome Y_{ij} as follows:

$$\begin{aligned} Y_{ij} &\sim N(\theta_i,\sigma^2_y),\\ \theta_i|C_i=k &\sim N(\theta_{(k)},\tau^2_{(k)}). \end{aligned}$$

Conditionally on C_i and Y_{ij}, we model the longitudinal biomarker measures Z_{ijl} by using a semiparametric mixed model as follows:

$$\left.\begin{aligned} Z_{ijl}|(Y_{ij}, C_i = k) &= \mu_{(k)}(t_l) + v_i + w_{ij} + \beta Y_{ij} + \epsilon_{ijl}, \\ v_i &\sim N(0, \sigma_v^2), \\ w_{ij} &\sim N(0, \sigma_w^2), \end{aligned}\right\} \tag{3}$$

which reflects the unique data structure of the basket trial; patients are nested in a cancer type, and cancer types are nested in a subgroup. Specifically, $\mu_{(k)}(t_l)$ is a non-parametric function of time t_l that specifies the mean trajectory of the biomarker for the kth subgroup, v_i is the cancer-type-specific random effect accounting for the fact that the mean biomarker trajectory for a cancer type may deviate from the mean trajectory for its subgroup and w_{ij} is a subject-specific random effect to allow the biomarker trajectory for an individual patient to deviate from the mean trajectory for his or her cancer type. Regression parameter β captures the relationship between the biomarker Z and treatment response Y. We assume residuals $\epsilon_{ijl} \sim N(0, \sigma_\epsilon^2)$. A semiparametric mixed model similar to model (3) has been used to model longitudinal data (Fitzmaurice *et al.*, 2008; Li *et al.*, 2010). When appropriate, a more complicated model can be entertained, for example, letting the cancer type level random variation v_i and patient level random variation w_{ij} be time dependent, i.e. $v_i(t)$ and $w_{ij}(t)$.

We model the non-parametric function $\mu_{(k)}(t_l)$ by using penalized splines (Eilers and Marx, 1996; Ruppert *et al.*, 2003) because of its flexibility and close ties to the BHM. Other smoothing methods, such as smoothing splines (Green and Silverman, 1993) and local polynomials (Fan and Gijbels, 1996) can also be used. Let $\kappa_1 < \kappa_2 < \ldots < \kappa_S$ denote S prespecified knots that partition the time interval $[t_1, t_L]$ into $S+1$ subintervals, and define the truncated power function as

$$(t_l - \kappa_s)_+^d = \begin{cases} (t_l - \kappa_s)^d & \text{if } t_l > \kappa_s, \\ 0 & \text{otherwise.} \end{cases}$$

The penalized spline with the dth-degree truncated power basis function for $\mu_{(k)}(t_l)$ can be expressed as

$$\mu_{(k)}(t_l) = \gamma_{0(k)} + \gamma_{1(k)} t_l + \gamma_{2(k)} t_l^2 + \ldots + \gamma_{d(k)} t_l^d + \sum_{s=1}^{S} a_{s(k)} (t_l - \kappa_s)_+^d, \qquad a_{s(k)} \sim N(0, \sigma_{a(k)}^2), \tag{4}$$

where $\gamma_{0(k)}, \ldots, \gamma_{q(k)}$ are unknown parameters, and $a_{1(k)}, \ldots, a_{S(k)}$ are random effects that follow a normal distribution with mean 0 and variance $\sigma_{a(k)}^2$. The smoothness of $\mu_{(k)}(t_l)$ is controlled by the smoothing parameter $\sigma_{a(k)}^2$. One advantage of the Bayesian approach is that, by treating $\sigma_{a(k)}^2$ as a variance parameter, it can be estimated simultaneously with other model parameters. Ruppert *et al.* (2003) showed that a penalized spline is generally robust to the choice of knots and basis functions. In practice, different basis functions with reasonably spaced knots often provide similar results.

We now discuss how to choose the number of latent subgroups. We choose the value of K such that the corresponding model has the best goodness of fit according to a certain model selection statistic, such as the deviance information criterion DIC (Spiegelhalter *et al.*, 2002). In principle, the selection of K can be done by fitting the model with $K = 1, \ldots, I$, and then selecting the value of K that yields the smallest value of DIC as the number of latent subgroups. However, because the number of cancer types that are included in a basket trial is typically small (e.g. 4–15) and all enrolled patients carry the same genetic or molecular aberration, in practice, it is often adequate to restrict the search space of K to $\{1, 2, 3\}$. This also facilitates the interpretation of the results. For example, $K = 1$ means that all cancer types are sensitive or insensitive to the treatment; $K = 2$ accommodates the most common case in which some

cancer types are sensitive to the treatment (i.e. the sensitive subgroup) whereas the others are insensitive to the treatment (i.e. the insensitive subgroup). During the trial, the value of K will be updated in the light of accumulating data. As a result, the value of K may differ from one interim evaluation to another, depending on the observed data. Instead of using DIC to choose the value of K, an alternative approach is to treat K as an unknown parameter, and to estimate it together with the other parameters. This can be done by using the reversible jump Markov chain Monte Carlo algorithm (Green, 1995).

2.2. Prior specification and posterior estimation

We fit the proposed model by using a Markov chain Monte Carlo algorithm. We assign independent vague priors to the model parameters as follows:

$$\begin{aligned}
\gamma_{d(k)} &\overset{\text{IID}}{\sim} N(0,c), && d=0,\ldots,2, \quad k=1,\ldots,K,\\
\theta_{(k)} &\overset{\text{IID}}{\sim} N(g_k,c), && k=1,\ldots,K,\\
\sigma^2_{a(k)} &\overset{\text{IID}}{\sim} \text{IG}(a_a,b_a), && k=1,\ldots,K,\\
\beta &\sim N(0,c),\\
\sigma^2_v &\sim \text{IG}(a_v,b_v),\\
\sigma^2_w &\sim \text{IG}(a_w,b_w),\\
\sigma^2_\epsilon &\sim \text{IG}(a_\epsilon,b_\epsilon),\\
\tau^2_{(k)} &\overset{\text{IID}}{\sim} \text{IG}(a_\tau,b_\tau), && k=1,\ldots,K,\\
(\pi_1,\ldots,\pi_K) &\sim \text{Dir}(\lambda_1,\ldots,\lambda_K), && k=1,\ldots,K,
\end{aligned}$$

where $\text{IG}(a,b)$ denotes the inverse gamma distribution with a shape parameter a and a scale parameter b, and $\text{Dir}(\lambda_1,\ldots,\lambda_K)$ represents a Dirichlet distribution with probability parameters $(\lambda_1,\ldots,\lambda_K)$. We set $c=10^4$, $a_a=b_a=a_v=b_v=a_w=b_w=a_\epsilon=b_\epsilon=a_\tau=b_\tau=10^{-3}$ and $\lambda_1=\ldots=\lambda_K=2$ such that the priors are vague. We constrain $\theta_{(1)}<\ldots<\theta_{(k)}$ to avoid the potential computational issue of label switching and set $g_k=\log\{q'_k/(1-q'_k)\}$ with $q'_k=q_0+\{(k-1)/(K-1)\}(q_1-q_0)$, $k=1,\ldots,K$. We use the Gibbs sampler (Geman and Geman, 1984) to sample posteriors. The details of the posterior sampling can be found in Appendix A.

2.3. Trial design

The BLAST design has a total of M planned interim looks. Let $\mathcal{D}_m=\{(Z_{ij},Y_{ijl}),i=1,\ldots,I,j=1,\ldots,n_{i,m},l=1,\ldots,L\}$ denote the interim data at the mth look, where $n_{i,m}$ is the sample size for the ith cancer type at the mth interim look, $m=1,\ldots,M$. The BLAST design is described as follows.

Step 1: enrol $n_{i,1}$ patients with the ith cancer type for $i=1,\ldots,I$.
Step 2: given the mth interim data $\mathcal{D}_m$, $m=1,\ldots,M$, fit the proposed model.

(a) (*Futility stopping*) if $\Pr[p_i>\{(q_0+q_1)/2\}|\mathcal{D}_m]<Q_{\text{f}}$, suspend the accrual for the ith cancer type, where $(q_0+q_1)/2$ denotes the rate halfway between the null and target response rate, and Q_{f} is a probability cut-off for futility stopping.
(b) Otherwise, continue to enrol patients until the next interim analysis has been reached.

Step 3: once the maximum sample size has been reached or the accrual has been stopped for all cancer types due to futility, evaluate the treatment efficacy on the basis of the final data $\mathcal{D}$ as follows. If $\Pr(p_i>q_0|\mathcal{D})>Q$, declare that the treatment is effective for the ith cancer

type; otherwise, declare that the treatment is not effective for the ith cancer type, where Q is a probability cut-off.

In the design, the probability cut-offs Q_{f} and Q should be calibrated through simulation to achieve a desired type I error rate and power for each cancer type. This simulation-based calibration procedure is widely used in Bayesian clinical trial designs (Thall *et al.*, 1995; Yuan *et al.*, 2016).

3. Simulation study

3.1. Operating characteristics

We conducted extensive simulations to evaluate the operating characteristics of the BLAST design. Taking the setting of the motivating CDKN2A deficient solid tumour trial, the null response rate $q_0 = 0.2$ and the alternative response rate $q_1 = 0.3$. We considered six cancer types and two subgroups (i.e. sensitive subgroup and insensitive subgroup). Depending on its true response rate p_i, each of the cancer types belongs to either the sensitive subgroup or the insensitive subgroup. We constructed different scenarios by varying the subgroup membership for the cancer type. For example, in Table 1, scenario 1 represents the null case in which all cancer types belong to the insensitive subgroup (i.e. the drug is not effective for all patients). Scenario 2 represents the case in which cancer types 1–4 belong to the sensitive subgroup, whereas cancer types 5 and 6 belong to the insensitive subgroup. We simulated Y_{ij} from $\mathrm{Ber}(p_i)$, and eight repeated biomarker measures Z_{ijl}, equally spaced over the standardized timeframe [0, 1], based on model (3) with two different trajectory shapes (see Fig. A1 in the on-line supplementary materials): trajectory A,

$$\mu_{(k)}(t) = \begin{cases} 18 - 12\exp\{-6(t+0.05)\} & \text{if cancer type } i \in \text{sensitive subgroup,} \\ 9.11 & \text{if cancer type } i \in \text{insensitive subgroup;} \end{cases}$$

trajectory B,

$$\mu_{(k)}(t) = \begin{cases} 8.86 + \dfrac{8}{1+\exp\{-8(t-0.5)\}} & \text{if cancer type } i \in \text{sensitive subgroup,} \\ 4 + 5\exp(-t^{0.5}) & \text{if cancer type } i \in \text{insensitive subgroup.} \end{cases}$$

These two settings are chosen to represent the trajectories that we may encounter in practice. Specifically, in trajectory A, the biomarker measure increases and then plateaus in the sensitive subgroup, whereas it remains constant in the insensitive group. In trajectory B, for the sensitive subgroup, there is an incubation period during which the biomarker measure increases slowly before it increases more rapidly, whereas for the insensitive group the biomarker measure slowly decays over time to reflect the natural progression of the disease. We set $\beta = 1$, $\sigma_\epsilon^2 = 1.5$ and $\sigma_v^2 = \sigma_w^2 = 4$ for trajectory A, and $\sigma_v^2 = \sigma_w^2 = 7$ for trajectory B. Fig. A2 in the on-line supplementary materials shows an example of simulated biomarker data. To fit the biomarker data, we used four equally spaced knots in the penalized spline.

The maximum sample size for each cancer type is 25, with three interim analyses conducted when the sample size in each cancer type reaches 10, 15 and 20. We compared the BLAST design with the BHM approach (Berry *et al.*, 2013) and the independent approach. The BHM approach uses a conventional BHM similar to model (2) without the latent subgroup. The independent approach models Y independently in each cancer type by using a beta–binomial model, i.e. $Y_{ij} \sim \mathrm{Ber}(p_i)$ with a conjugate prior $p_i \sim \mathrm{beta}(0.03, 0.07)$. To make the comparison meaningful, we used the same interim stopping rule in the three designs, set $Q_{\mathrm{f}} = 0.05$ and calibrated Q for

Table 1. Simulation results of the independent, BHM and BLAST designs under biomarker trajectory setting A

Scenario	*Design*		*Results for the following cancer types:*						*Sample size*
			1	*2*	*3*	*4*	*5*	*6*	
A1		Response rate	0.2	0.2	0.2	0.2	0.2	0.2	
	Independent	% reject	9.9	10.1	10	10.1	10	9.9	132.9
		% stop	27.2	27.5	25.9	26.6	26.5	24.7	
	BHM	% reject	9.8	10.2	9.9	9.9	9.8	9.8	129.1
		% stop	30.4	30.1	30.7	28.8	30.6	29.2	
	BLAST	% reject	9.9	10.1	9.8	9.8	10.2	9.9	129.4
		% stop	31.2	30.8	29.1	31	29	29.7	
A2		Response rate	0.3	0.3	0.3	0.3	0.2	0.2	
	Independent	% reject	46.5	45.4	45.9	41.4	9.2	11.6	141.5
		% stop	5.6	7.3	6	5.4	27	26.2	
	BHM	% reject	69.6	68.6	72.2	70.8	45.8	42.3	147.2
		% stop	2.9	2.7	2.9	3.2	6.1	4.8	
	BLAST	% reject	90.7	91.1	92.3	91.4	11.3	11.9	140.5
		% stop	1	1	0.6	0.8	36.1	37	
A3		Response rate	0.3	0.3	0.2	0.2	0.2	0.2	
	Independent	% reject	45.4	43.4	10	9.4	10.6	10.2	137.3
		% stop	5.7	6.4	26.2	25.3	27	27.2	
	BHM	% reject	46.5	47.4	26.3	26.5	25.2	23.9	141.3
		% stop	7.8	7.3	13.6	14.7	15.8	14.3	
	BLAST	% reject	82.1	85.8	9.9	9.3	7.6	9	133.4
		% stop	2.1	1.2	34.6	34.9	32.4	33.3	
A4		Response rate	0.35	0.3	0.3	0.2	0.2	0.2	
	Independent	% reject	69	44.5	46.6	9.7	9.9	10.5	139.8
		% stop	2.2	6.4	6.8	25.6	27.1	25.2	
	BHM	% reject	74.9	62.8	66.6	39	36.4	36.4	146.0
		% stop	2.6	3.6	4	7.2	8.2	7.1	
	BLAST	% reject	94.7	89.1	91.4	8.6	9	7.6	137.6
		% stop	0.4	0.5	0.6	36.5	31.7	33.7	
A5		Response. rate	0.3	0.2	0.2	0.2	0.2	0.2	
	Independent	% reject	45	11.4	7.8	8.6	10.4	9.4	135.2
		% stop	6.2	27	25.5	25.7	26.7	25.3	
	BHM	% reject	35.8	15.9	18.7	17.8	15.7	16.2	135.9
		% stop	12.6	21.3	22.4	20.8	23.7	21.2	
	BLAST	% reject	71.7	11	10.2	10.3	10.1	11.2	130.0
		% stop	5.5	35	33.9	34.8	31.5	32.2	

each design to control the type I error rate to be 10% in each cancer type (see scenarios A1 and B1 in Tables 1 and 2). We conducted a total of 1000 simulations under each scenario.

Table 1 shows the simulation results when the biomarker follows trajectory A. Under each scenario, the first row reports the true response rate for the six cancer types, and the other rows report the percentage for rejecting H_0, the percentage for early stopping for each cancer type, and the average total sample size. Scenario A1 represents the null case in which the treatment is not effective for all cancer types and the percentage for rejecting H_0 is the type I error rate. As described above, the type I error rate of the three designs is controlled at 10% by calibrating the stopping probability cut-off Q. Compared with the independent design, the BHM and BLAST designs are more likely to stop the trial early, resulting in smaller total samples sizes. This is because the BHM and BLAST designs borrow information across cancer types and are thus more efficient than the independent design that performs interim analysis independently for each

Table 2. Simulation results of the independent, BHM and BLAST designs under biomarker trajectory setting B

Scenario	*Design*		*Results for the following cancer types:*						*Sample size*
			1	*2*	*3*	*4*	*5*	*6*	
B1		Response rate	0.2	0.2	0.2	0.2	0.2	0.2	
	Independent	% reject	9.9	10.1	10	10.1	10	9.9	132.9
		% stop	27.2	27.5	25.9	26.6	26.5	24.7	
	BHM	% reject	9.8	10.2	9.9	9.9	9.8	9.8	129.1
		% stop	30.4	30.1	30.7	28.8	30.6	29.2	
	BLAST	% reject	9.8	9.9	9.9	10.2	10.2	10	129.4
		% stop	31.2	31	29.3	31	29.1	29.9	
B2		Response rate	0.3	0.3	0.3	0.3	0.3	0.3	
	Independent	% reject	45.6	46.8	46.1	45.4	46.3	49.5	145.7
		% stop	5.9	6.5	7.3	6	5.8	5.8	
	BHM	% reject	87	85.3	87.8	86.6	85.5	85.3	149.2
		% stop	2.6	0.9	1	1.3	1	0.6	
	BLAST	% reject	87.1	89.5	89.7	89.4	88.1	88.9	149.4
		% stop	1.2	0.5	0.5	1	0.4	0.8	
B3		Response rate	0.2	0.2	0.35	0.3	0.3	0.35	
	Independent	% reject	10.8	9.4	64.1	44.8	47.1	67.1	142.0
		% stop	27.4	26.7	3.5	5.6	7.4	2.4	
	BHM	% reject	49.2	48.3	85.9	77.8	74.8	83.9	148.4
		% stop	2.9	3.8	1.5	1.2	2.1	0.9	
	BLAST	% reject	7.8	7.3	97.2	92.9	92.4	97	139.4
		% stop	43.2	41.6	0.4	0.3	0.2	0.3	
B4		Response rate	0.35	0.3	0.2	0.2	0.2	0.2	
	Independent	% reject	65.9	45.3	10.2	9	10	9.5	137.5
		% stop	2.6	7	26.9	25.6	28.5	24.8	
	BHM	% reject	62.7	52.4	29.6	29.1	27.2	26.1	143.3
		% stop	7.8	5	12	11.5	12.9	11.5	
	BLAST	% reject	91.5	85.9	8.3	8.7	6.7	8.2	133.8
		% stop	0.8	1.6	33.3	34.4	32	33.9	
B5		Response rate	0.3	0.3	0.3	0.3	0.3	0.2	
	Independent	% reject	46.8	46.9	44.5	46	45.7	9.9	143.7
		% stop	5.8	7.5	6.4	5.2	6.5	25	
	BHM	% reject	79	77.6	80.4	80.4	77.7	56	148.5
		% stop	12.6	1.6	1.8	1.8	2.3	2.6	
	BLAST	% reject	91.9	92.6	93.1	92.6	92.1	11.8	143.9
		% stop	0.8	0.6	0.5	0.8	0.5	41.1	

cancer type. In scenario A2, cancer types 1–4 are sensitive to the treatment, whereas cancer types 5 and 6 are not sensitive to the treatment. The BLAST design outperforms the independent and BHM designs. As expected, the independent design maintains the type I error rate at 10% for the insensitive cancer types (i.e. cancer types 5 and 6) but has low power (i.e. 41.4–46.5%) to detect the treatment effects for sensitive cancer types (i.e. cancer types 1–4). In contrast, the proposed BLAST design yields substantially higher power (i.e. 90.7–92.3%) to detect the treatment effect in (sensitive) cancer types 1–4, whereas controlling the type I error rate reasonably close to 10% for (insensitive) cancer types 5–6. The BHM design fails to control the type I error rate, which is inflated to over 42% for (insensitive) cancer types 5 and 6. This is consistent with findings from previous research (Freidlin and Korn, 2013). In addition, compared with the independent and BHM designs, the BLAST design is more likely to stop treating the insensitive cancer types and less likely to stop treating the sensitive cancer types. The percentage of early stopping for

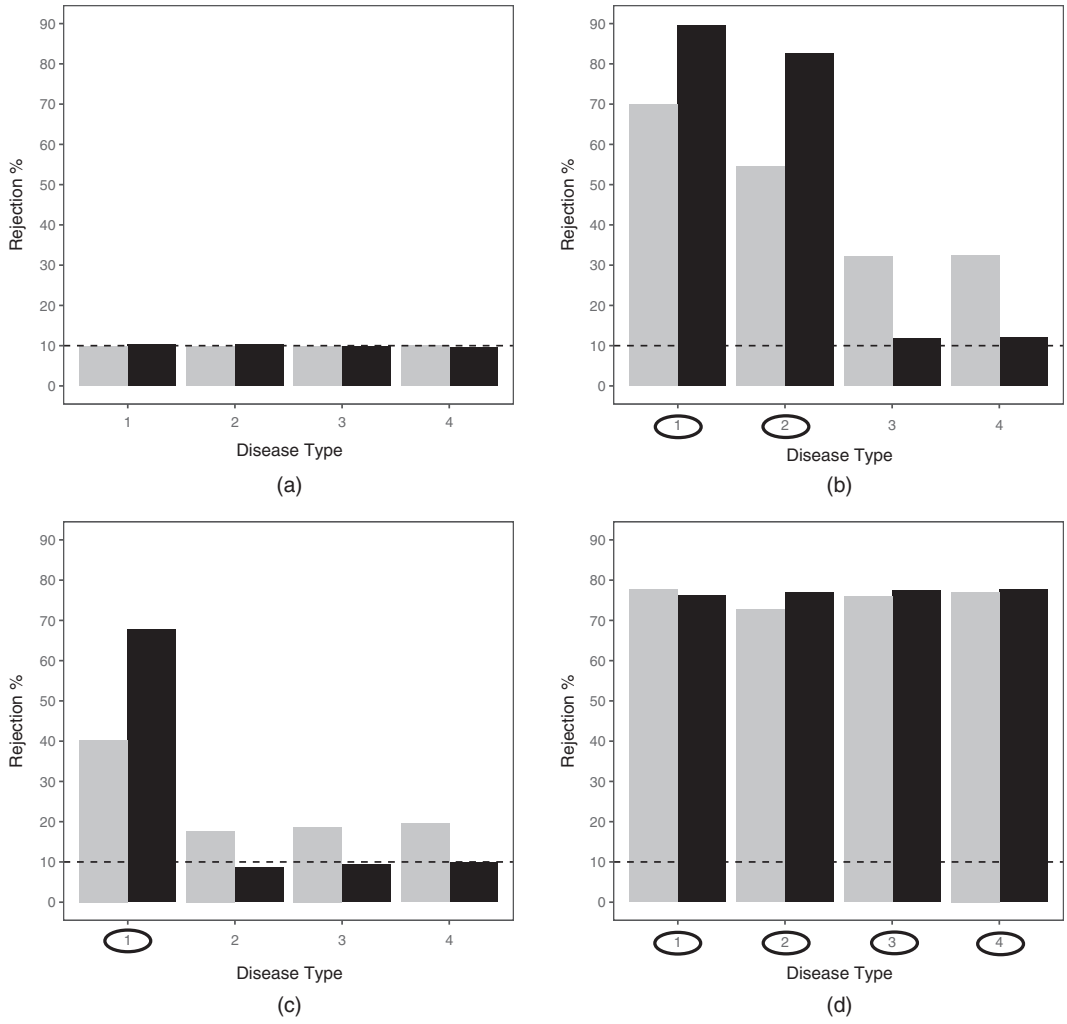

Fig. 1. Sensitivity analysis—comparing the rejection percentage of the BHM (▒) with that of the proposed BLAST design (■) when four cancer types are considered (on the x-axis, circled cancer types are responsive to the treatment, and uncircled cancer types are not responsive to the treatment; for example, scenario 1 serves as the calibration scenario, with all four cancer arms being non-responsive and having rejection percentage calibrated to 10%; scenario 2 has arms 1 and 2 as responsive, and others as non-responsive; scenario 3 has arm 1 as responsive, and arms 2–4 as non-responsive; scenario 4 has arms 1–4 as responsive): (a) scenario 1; (b) scenario 2; (c) scenario 3; (d) scenario 4

(insensitive) cancer types 5 and 6 is about 37% for the BLAST design, compared with 27% for the independent design and 6% for the BHM design. The percentage of incorrect early stopping for (sensitive) cancer types 1–4 is less than 1% in the BLAST design, compared with about 6% in the independent design and greater than 2.5% in the BHM design. As a result, the BLAST design tends to result in total sample sizes that are smaller than those under the independent and BHM designs. In scenario A3, there are two sensitive cancer types (i.e. cancer types 1 and 2) and four insensitive cancer types (i.e. cancer types 3–6). The BLAST design yielded high power (i.e. 82.1% and 85.8% for respective cancer types 1 and 2) and well-controlled type I error rates (i.e. 9.9%, 9.3%, 7.6% and 9% for respective cancer types 3, 4, 5 and 6), whereas the BHM design

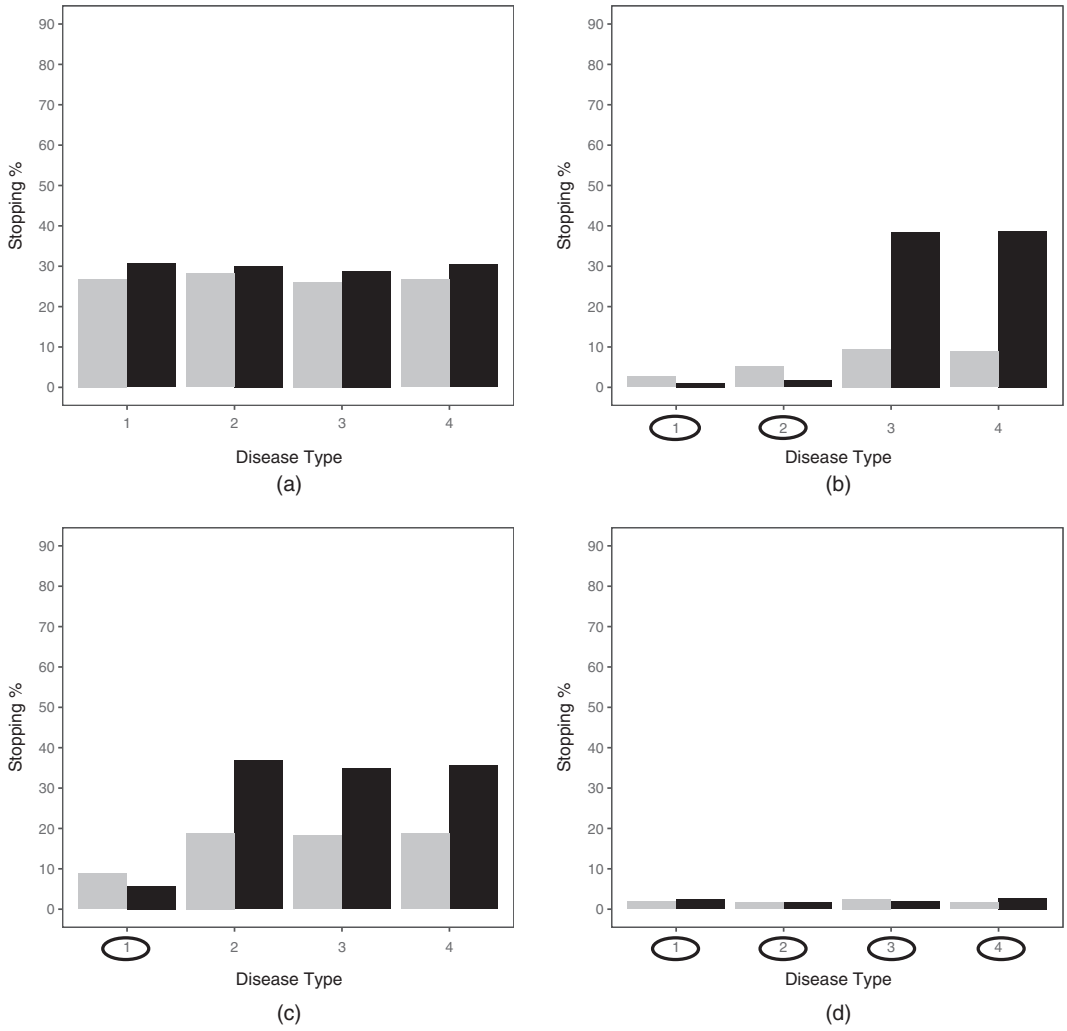

Fig. 2. Sensitivity analysis—comparing the early stopping percentage of the BHM (■) with that of the proposed BLAST design (■) when four cancer types are considered (on the *x*-axis, circled cancer types are responsive to the treatment, and uncircled cancer types are not responsive to the treatment): (a) scenario 1; (b) scenario 2; (c) scenario 3; (d) scenario 4

had lower power (i.e. 46.5% and 47.4% for respective cancer types 1 and 2) and substantially inflated type I error rates (i.e. 26.3%, 26.5%, 25.2% and 23.9% for respective cancer types 3–6). The independent approach yielded well-maintained type I error rates for the insensitive cancer types (10%, 9.4%, 10.6% and 10.2%), but low power for the sensitive cancer types (45.4% and 43.4%). Again, the BLAST design has smaller sample size, lower early stopping percentage for sensitive cancer types and higher stopping percentage for the insensitive cancer types than the other two designs. Similar results were observed in scenarios A4 and A5, which consider different (i.e. three and five respectively) numbers of insensitive cancer types. Table 2 shows the results under trajectory B. The results are generally similar to what is described above.

We also examined the estimates of response rates, including the absolute bias and mean-squared error MSE, under various designs. As shown in Figs A3 and A4 in the on-line

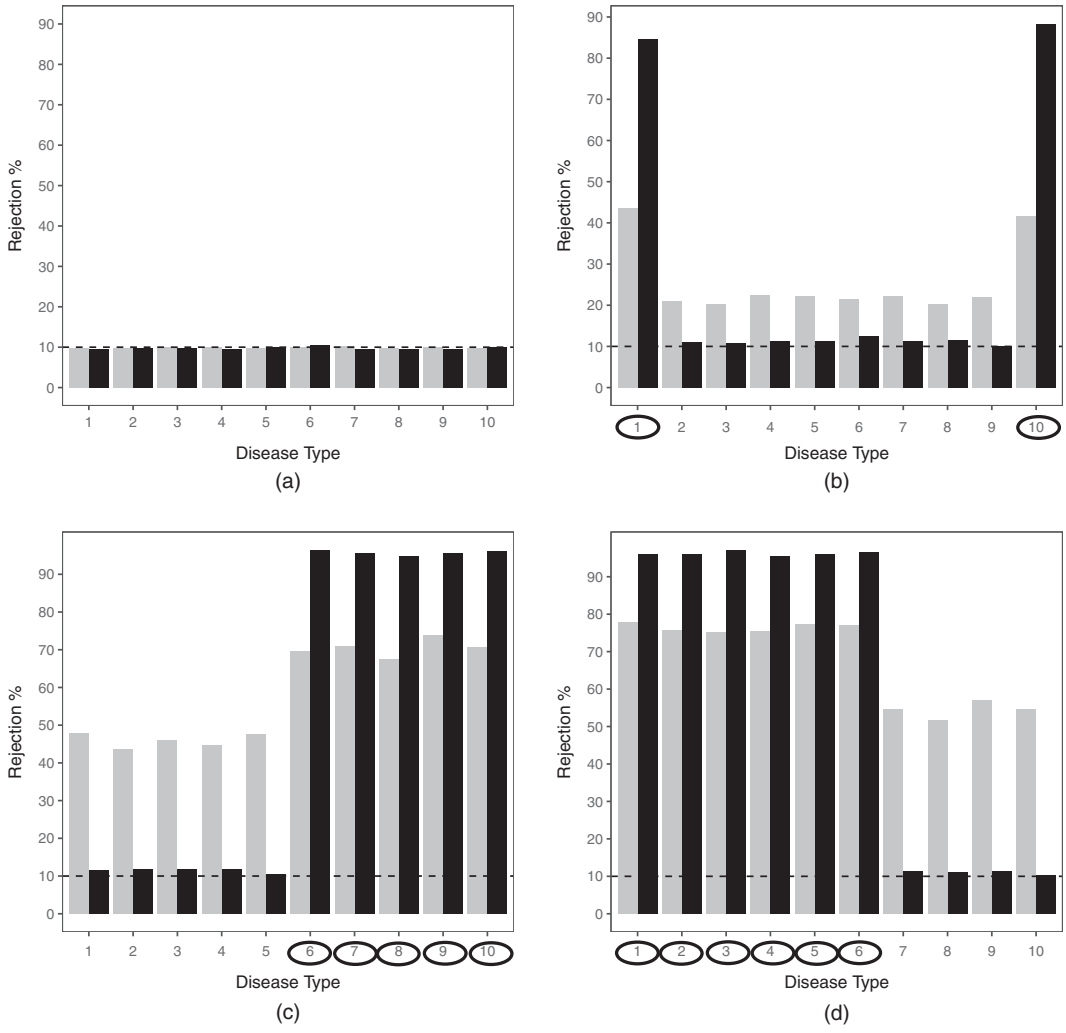

Fig. 3. Sensitivity analysis—comparing the rejection percentage of the BHM (▒) with that of the proposed BLAST design (■) when 10 cancer types are considered (on the *x*-axis, circled cancer types are responsive to the treatment, and uncircled cancer types are not responsive to the treatment; for example, scenario 1 serves as the calibration scenario, with all 10 cancer arms being non-responsive and having rejection percentage calibrated to 10%; scenario 2 has arms 1 and 10 as responsive and others as non-responsive; scenario 3 has arms 1–5 as non-responsive and arms 6–10 as responsive; scenario 4 has arms 1–6 as responsive and arms 4–10 as non-responsive): (a) scenario 1; (b) scenario 2; (c) scenario 3; (d) scenario 4

supplementary materials, the BLAST design generally has the best overall performance. The independent design has a larger MSE than the BHM and BLAST designs because it does not borrow information across cancer types. The BHM design tends to have larger absolute biases when the treatment effect is heterogeneous (i.e. scenarios A2–A5 and B3–B5) because it regards all the disease types as exchangeable and tends to shrink the estimate of treatment effect across them overly.

3.2. Sensitivity analysis

We evaluated the performance of the BLAST design when the basket trial contains four or

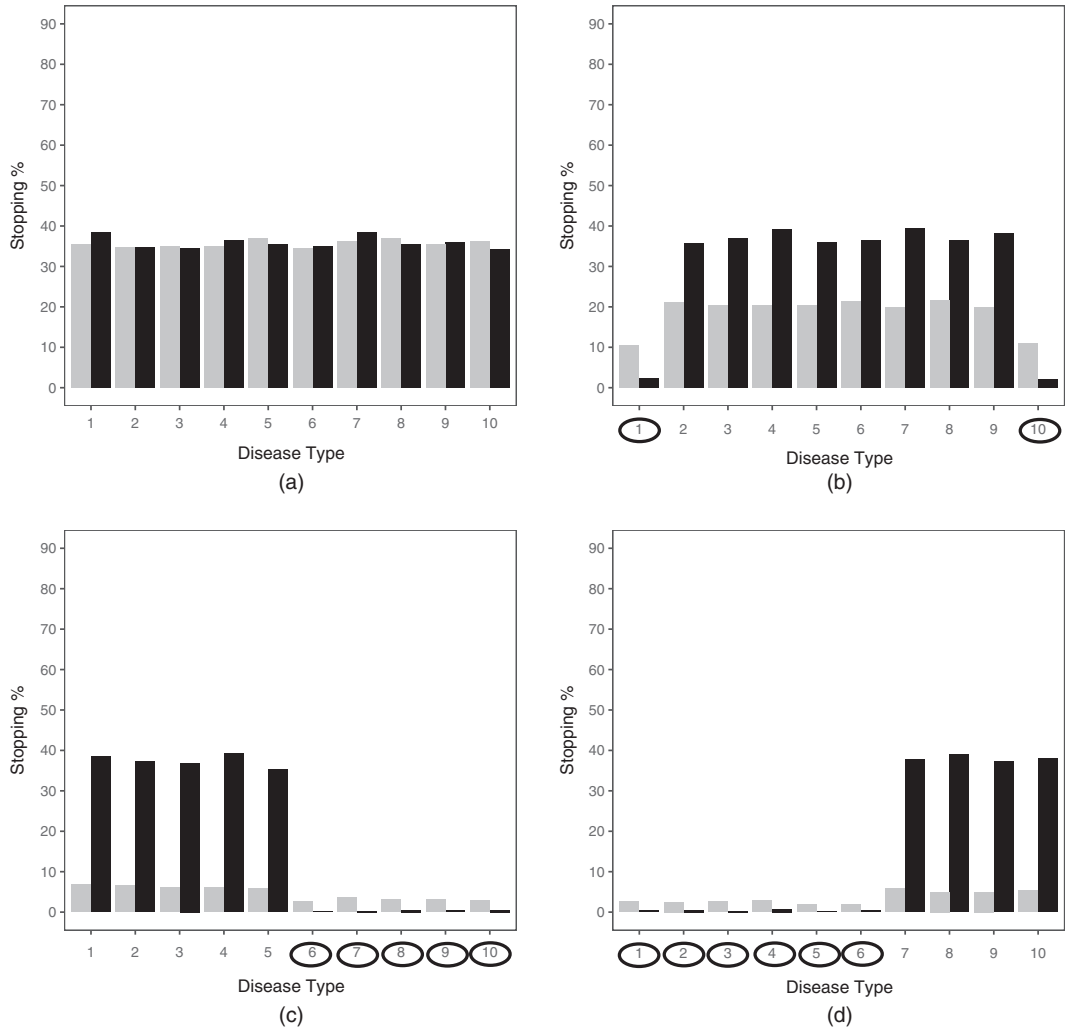

Fig. 4. Sensitivity analysis—comparing the early stopping percentage of the BHM (▒) with that of the proposed BLAST design (■) when 10 cancer types are considered (on the *x*-axis, circled cancer types are responsive to the treatment, and uncircled cancer types are not responsive to the treatment): (a) scenario 1; (b) scenario 2; (c) scenario 3; (d) scenario 4

10 cancer types. Figs 1–4 show the simulation results. We can see that, compared with the BHM design, the BLAST design yields substantially higher power to detect the treatment effect for sensitive cancer types and controls the type I error rate for the insensitive cancer types. Detailed results are provided in Tables A1 and A2 in the on-line supplementary materials.

We also studied the sensitivity of the BLAST design with respect to the priors. Table 3 shows the simulation results when setting hyperparameters $c = 10^2, a_a = b_a = a_v = b_v = a_w = b_w = a_\epsilon = b_\epsilon = 10^{-2}$, while keeping the other parameters unchanged. The results are generally similar to those reported in Table 1, suggesting that the choice of hyperparameters has little effect on the performance of the design.

Table 3. Sensitivity analysis of the BLAST design with different priors

Scenario		*Results for the following cancer types:*						*Sample size*
		1	*2*	*3*	*4*	*5*	*6*	
Trajectory A								
A1	Response rate	0.2	0.2	0.2	0.2	0.2	0.2	
	% reject	10.2	9.8	10.1	10.1	10	9.8	129.5
	% stop	31.7	30.3	29	31.5	28.6	29.2	
A2	Response rate	0.3	0.3	0.3	0.3	0.2	0.2	
	% reject	90.8	91.5	92.1	90.9	12.7	12.4	140.3
	% stop	1.3	0.9	0.9	0.9	36.7	38	
A3	Response rate	0.3	0.3	0.2	0.2	0.2	0.2	
	% reject	82.5	84.9	10.1	10	8.2	8.9	133.7
	% stop	2.1	1.2	32.9	35.7	32.3	32.4	
A4	Response rate	0.35	0.3	0.3	0.2	0.2	0.2	
	% reject	94.3	89.8	91	9.9	10.4	8.6	137.6
	% stop	0.4	0.6	0.7	36.2	32.1	34.5	
A5	Response rate	0.3	0.2	0.2	0.2	0.2	0.2	
	% reject	73.8	9.7	10.2	10.1	10	10.6	129.6
	% stop	5.4	35.1	33.1	36.1	32	33.1	
Trajectory B								
B1	Response rate	0.2	0.2	0.2	0.2	0.2	0.2	
	% reject	10	9.8	9.8	10	9.9	9.8	129.6
	% stop	31.7	30.3	29	31.5	28.6	29.2	
B2	Response rate	0.3	0.3	0.3	0.3	0.3	0.3	
	% reject	87.5	88.9	89.5	89.1	87.9	88.5	149.5
	% stop	1.1	0.5	0.6	1	0.5	0.7	
B3	Response rate	0.2	0.2	0.35	0.3	0.3	0.35	
	% reject	7.6	7.3	97	93.4	93.2	97.2	139.5
	% stop	42.8	41	0.4	0.2	0.1	0.2	
B4	Response rate	0.35	0.3	0.2	0.2	0.2	0.2	
	% reject	91.8	85.4	8.2	8.2	6.8	7.7	133.6
	% stop	0.8	1.6	34	35.1	32.4	33.3	
B5	Response rate	0.3	0.3	0.3	0.3	0.3	0.2	
	% reject	92.2	92.9	93.2	92.4	92.7	11.6	143.9
	% stop	0.7	0.6	0.5	0.6	0.3	41.2	

4. Conclusion

We proposed the BLAST design to evaluate the treatment effect of targeted agents in basket clinical trials. The BLAST design assumes that different cancer types belong to different latent subgroups. The cancer types within a subgroup have similar treatment effects. By jointly modelling the longitudinal biomarker measurements and treatment responses, the BLAST design simultaneously groups cancer types into different subgroups and makes Bayesian inference and go–no-go interim treatment decisions for each cancer type. The simulation shows that the BLAST design outperforms the BHM approach, with higher power and better controlled type I error rates.

One innovation of the BLAST design is that it leverages longitudinal biomarker measurements that are routinely taken in clinical trials to improve the efficiency of the basket trial. In some sense, this is necessary for achieving adaptive information borrowing, i.e. identifying cancer types (or baskets) with similar response and pooling information across them. We tried clustering the baskets on the basis of only tumour response Y_{ij} by using parametric methods (e.g. latent class

model) or non-parametric methods (e.g. K-means clustering). We found that it worked poorly because of small sample size and limited data information (i.e. a binary end point contains much less information than longitudinal measures). The resulting clusters are very unreliable, and there is a high probability that baskets are incorrectly clustered, leading to substantially inflated type I errors (the results are not shown).

The BLAST design aims to evaluate the treatment effect and makes the go–no-go decision for each of the cancer types, while allowing adaptive information borrowing across the cancer types with similar treatment effects by identifying and forming subgroups. The BLAST design can be easily adopted to make the go–no-go decision at the subgroup level if it is clinically desirable. In addition, we have focused on a single multitumour 'basket' that assesses a single targeted agent across multiple types of tumour sharing a specific genetic or molecular aberration. The BLAST design is also applicable to 'nested' basket trials, where multiple targeted therapies are evaluated simultaneously in a 'parent' basket trial, under which each agent forms a 'child' basket trial. In that case, our methodology can be implemented independently for each child basket trial, allowing different conclusions for different child basket trials.

Acknowledgements

We thank the reviewers and the Joint Editor for their very helpful comments that substantially improved this paper.

Appendix A

We employ the Gibbs sampler (Geman and Geman, 1984) as the posterior sampling strategy. We denote $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iJ})$, $\mathbf{Z}_{ij} = (Z_{ij}(1), \ldots, Z_{ij}(T))$ and $\mathbf{Z}_i = (\mathbf{Z}_{i1}, \ldots, \mathbf{Z}_{iJ})$. At each iteration, we conduct the following steps.

Step 1: update C_i. First, we reorganize model (3) as

$$\mathbf{Z}_i|\mathbf{X}_{i\alpha}, \mathbf{X}_{ib}, \mathbf{Y}_i, C_i = k \sim N(\mathbf{X}_{i\alpha}\alpha_{(k)} + \mathbf{X}_{ib}\mathbf{b}_i + \mathbf{Y}_i\beta, \sigma_\epsilon^2),$$

where $\boldsymbol{\alpha}_{(k)} = (\gamma_{0(k)}, \gamma_{1(k)}, \gamma_{2(k)}, a_{1(k)}, \ldots, a_{s(k)})^{\mathrm{T}}$, $\mathbf{b}_i = (v_i, w_{i1}, \ldots, w_{iJ})^{\mathrm{T}}$ and $\mathbf{X}_{i\alpha}$ and $\mathbf{X}_{ib}$ are design matrices of the ith arm associated with $\boldsymbol{\alpha}_{(k)}$ and the random-effect vector $\mathbf{b}_i$ respectively.

For $i = 1, \ldots, I$, draw C_i independently from the multinomial distribution with probability τ_{ik}, $k = 1, \ldots, K$, given by

$$\tau_{ik} = \frac{\pi_k N_{Q_i}(\omega_{i(k)}, \sigma_\epsilon^2 I_{Q_i})\mathrm{Bin}(n_i, p_{i(k)})}{\sum_{k=1}^{K} \pi_k N_{Q_i}(\omega_{i(k)}, \sigma_\epsilon^2 I_{Q_i})\mathrm{Bin}(n_i, p_{i(k)})}$$

where $\omega_{i(k)} = \mathbf{X}_{i\alpha}\alpha_{(k)} + \mathbf{X}_{ib}\mathbf{b}_i + \mathbf{Y}_i\beta$ and $N_{Q_i}(\omega, \sigma^2 I_{Q_i})$ is a Q_i-variate normal density function of $\mathbf{Z}_i$ with the mean vector ω and covariance matrix $\sigma^2 I_{Q_i}$, with $Q_i = Tn_i$ indexing the total number of biomarker measurements in arm i. $\mathrm{Bin}(n_i, p_{i(k)})$ denotes the binomial density function of $\mathbf{Y}_i$ with sample size n_i and probability $p_{i(k)}$, where $p_{i(k)} = \exp(\theta_{(k)} + u_i)/\{1 + \exp(\theta_{(k)} + u_i)\}$, with $u_i \sim N(0, \tau_{(k)}^2)$.

Step 2: update π_k. Draw π_k, $k = 1, \ldots, K$, from the Dirichlet distribution,

$$\pi_k|\cdot \sim \mathrm{Dir}\left\{\sum_{i=1}^{I} I(C_i = 1) + \lambda_1, \ldots, \sum_{i=1}^{I} I(C_i = K) + \lambda_K\right\}.$$

Step 3: update $\theta_{(k)}$. There is no closed form for the posterior distribution of $\theta_{(k)}$, $k = 1, \ldots, K$, so we use the adaptive rejection Metropolis sampling algorithm to draw $\theta_{(k)}$ based on

$$\theta_{(k)}|\cdot \propto \prod_{i=1}^{I}\left\{\prod_{j=1}^{n_i}\frac{\exp(\theta_{(k)} + u_i)^{I(Y_{ij}=1)}}{1 + \exp(\theta_{(k)} + u_i)}\right\}\exp\left\{-\frac{(\theta_{(k)} - g_k)^2}{2c}\right\}.$$

We place an ordered constraint on $\theta_{(k)}$ such that $\theta_{(1)} < \ldots < \theta_{(k)}$ to avoid the label switching issue. We note that this constraint is consistent with the practical situation in which the non-responsive arms usually have lower efficacy measures than the responsive arms.

Step 4: update u_i. Sample u_i, $i = 1, \ldots, I$, via the adaptive rejection Metropolis sampling algorithm based on the full conditional

$$u_i|\cdot \propto \prod_{j=1}^{n_i} \frac{\exp(\theta_{(k)} + u_i)^{I(Y_{ij}=1)}}{1+\exp(\theta_{(k)} + u_i)} \exp\left(-\frac{u_i^2}{2\tau_{(k)}^2}\right).$$

Step 5: update $\tau_{(k)}^2$. Draw $\tau_{(k)}^2$, $k = 1, \ldots, K$, from the inverse gamma distribution

$$\tau_{(k)}^2|\cdot \sim \mathrm{IG}\left[a_\tau + \frac{1}{2}\sum_{i=1}^{I} I(C_i = k), b_\tau + \frac{1}{2}\sum_{i=1}^{I}\{u_i^2 I(C_i = k)\}\right].$$

Step 6: update $\boldsymbol{\alpha}_{(k)}$. We jointly update the penalized spline parameter vector $\boldsymbol{\alpha}_{(k)}$. Let $N_k = \Sigma_{i=1}^{I}\{n_i I(C_i = k)\}$ and $Q_k = TN_k$ denote the number of patients and the total number of biomarker measurements in the latent subgroup k respectively. Draw $\boldsymbol{\alpha}_{(k)}$ from its full conditional $f(\boldsymbol{\alpha}_{(k)}|\cdot) = N(\boldsymbol{\eta}_{\alpha(k)}, \mathbf{V}_{\alpha(k)})$, where

$$\mathbf{V}_{\alpha(k)} = (\boldsymbol{\Sigma}_{\boldsymbol{\alpha}(k)}^{-1} + \mathbf{X}_{\boldsymbol{\alpha}(k)}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\alpha}(k)}\sigma_\epsilon^{-2})^{-1},$$
$$\boldsymbol{\eta}_{\alpha(k)} = \sigma_\epsilon^{-2}\mathbf{V}_{\boldsymbol{\alpha}(k)}\mathbf{X}_{\boldsymbol{\alpha}(k)}^{\mathrm{T}}(\mathbf{Z}_{(k)} - \mathbf{X}_{\mathbf{b}(k)}\mathbf{b} - \mathbf{Y}_{(k)}\beta).$$

Here,

$$\boldsymbol{\Sigma}_{\boldsymbol{\alpha}(k)} = \begin{pmatrix} cI_3 & 0 \\ 0 & \sigma_{a(k)}^2 I_S \end{pmatrix}$$

represents the covariance matrix of $\boldsymbol{\alpha}_{(k)}$ for latent subgroup k, $\mathbf{X}_{\alpha(k)}$ is the $Q_k \times (3+S)$ design matrix for subgroup k, $\mathbf{Z}_{(k)}$ denotes the $Q_k \times 1$ biomarker measurement vector for subgroup k, $\mathbf{X}_{\mathbf{b}}(k)$ is the $Q_k \times (I+N)$ design matrix for subgroup k and $\mathbf{Y}_{(k)}$ denotes the $Q_k \times 1$ response outcome vector for subgroup k.

Step 7: update v_i and w_{ij}. We again jointly update the random-effects vector (v_i, w_{ij}), $i = 1, \ldots, I$, $j = 1, \ldots, J$, together. Let $N = \Sigma_{i=1}^{I} n_i$ and $Q = TN$ denote the total number of patients and the total number of biomarker measurements in the study respectively. Sample $\mathbf{b} = (\mathbf{v}, \mathbf{w})^{\mathrm{T}}$ from its full conditional $f(\mathbf{b}|\cdot) = N(\boldsymbol{\eta}_{\mathbf{b}}, \mathbf{V}_{\mathbf{b}})$, where

$$\mathbf{V}_{\mathbf{b}} = (\boldsymbol{\Sigma}_b^{-1} + \mathbf{X}_{\mathbf{b}}^{\mathrm{T}}\mathbf{X}_{\mathbf{b}}\sigma_\epsilon^{-2})^{-1},$$
$$\boldsymbol{\eta}_{\mathbf{b}} = \sigma_\epsilon^{-2}\mathbf{V}_{\mathbf{b}}\mathbf{X}_{\mathbf{b}}^{\mathrm{T}}(\mathbf{Z} - \mathbf{X}_\alpha\boldsymbol{\alpha}_{(k)} - \mathbf{Y}\beta).$$

Here,

$$\boldsymbol{\Sigma}_{\mathrm{b}} = \begin{pmatrix} \sigma_v^2 I_I & 0 \\ 0 & \sigma_w^2 I_N \end{pmatrix}$$

represents the covariance matrix of $\mathbf{b}$, $\mathbf{X}_{\mathbf{b}}$ denotes the $Q \times (I+N)$ design matrix associated with $\mathbf{b}$; $\mathbf{X}_\alpha$ represents the $Q \times (3+S)$ design matrix and $\mathbf{Y}$ is the $Q \times 1$ response outcome vector.

Step 8: update β. Sample β from its full conditional $f(\beta|\cdot) = N(\eta_\beta, V_\beta)$, where

$$V_\beta = (c^{-1} + \sigma_\epsilon^{-2}\sum_i\sum_j Y_{ij})^{-1},$$
$$\eta_\beta = \sigma_\epsilon^{-2}V_\beta\mathbf{Y}^{\mathrm{T}}(\mathbf{Z} - \mathbf{X}_\alpha\alpha_{(k)} - \mathbf{X}_{\mathbf{b}}\mathbf{b}).$$

Step 9: update $\sigma_{a(k)}^2$, σ_v^2, σ_w^2 and σ_ϵ^2. Draw the variance parameters sequentially from their full conditionals:

$$\sigma^2_{a(k)}|\cdot \sim \text{IG}\left(a_a + \frac{S}{2}, b_a + \frac{1}{2}\sum_{s=1}^{S} a^2_{s(k)}\right),$$

$$\sigma^2_{v}|\cdot \sim \text{IG}\left(a_v + \frac{I}{2}, b_v + \frac{1}{2}\sum_{i=1}^{I} v^2_{i}\right),$$

$$\sigma^2_{w}|\cdot \sim \text{IG}\left(a_w + \frac{N}{2}, b_w + \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{n_i} w^2_{ij}\right),$$

$$\sigma^2_{\epsilon}|\cdot \sim \text{IG}\left\{a_\epsilon + \frac{Q}{2}, b_\epsilon + \frac{1}{2}(\mathbf{Z} - \mathbf{X}_\alpha \boldsymbol{\alpha}_{(k)} - \mathbf{X}_\mathbf{b}\mathbf{b} - \mathbf{Y}\beta)^{\mathrm{T}}(\mathbf{Z} - \mathbf{X}_\alpha \boldsymbol{\alpha}_{(k)} - \mathbf{X}_\mathbf{b}\mathbf{b} - \mathbf{Y}\beta)\right\}.$$

References

Berry, D. A. (2015) The brave new world of clinical cancer research: adaptive biomarker-driven trials integrating clinical practice with clinical research. *Molec. Oncol.*, **9**, 951–959.

Berry, S. M., Broglio, K. R., Groshen, S. and Berry, D. A. (2013) Bayesian hierarchical modeling of patient subpopulations: efficient designs of phase II oncology clinical trials. *Clin. Trials*, **10**, 720–734.

Chu, Y. and Yuan, Y. (2017) A Bayesian basket trial design using a calibrated Bayesian hierarchical model. *Clin. Trials*, to be published.

Cunanan, K. M., Iasonos, A., Shen, R., Begg, C. and Gonen, M. (2017) An efficient basket trial design. *Statist. Med.*, **36**, 1568–1579.

Eilers, P. H. and Marx, B. D. (1996) Flexible smoothing with B-splines and penalties. *Statist. Sci.*, **11**, 89–102.

Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and Its Applications*. Boca Raton: CRC Press.

Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. (2008) *Longitudinal Data Analysis*. Boca Raton: CRC Press.

Flaherty, K. T., Puzanov, I., Kim, K. B., Ribas, A., McArthur, G. A., Sosman, J. A., O'Dwyer, P. J., Lee, R. J., Grippo, J. F., Nolop, K. and Chapman, P. B. (2010) Inhibition of mutated, activated BRAF in metastatic melanoma. *New Engl. J. Med.*, **363**, 809–819.

Fleming, G. F., Sill, M. W., Darcy, K. M., McMeekin, D. S., Thigpen, J. T., Adler, L. M., Berek, J. S., Chapman, J. A., DiSilvestro, P. A., Horowitz, I. R. and Fiorica, J. V. (2010) Phase II trial of trastuzumab in women with advanced or recurrent, HER2-positive endometrial carcinoma: a Gynecologic Oncology Group study. *Gyn. Oncol.*, **116**, 15–20.

Freidlin, B. and Korn, E. L. (2013) Borrowing information across subgroups in phase II trials: is it useful? *Clin. Cancer Res.*, **19**, 1326–1334.

Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., Tosolini, M., Camus, M., Berger, A., Wind, P. and Zinzindohoué, F. (2006) Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science*, **313**, 1960–1964.

Gatzemeier, U., Groth, G., Butts, C., Van Zandwijk, N., Shepherd, F., Ardizzoni, A., Barton, C., Ghahramani, P. and Hirsh, V. (2004) Randomized phase II trial of gemcitabine-cisplatin with or without trastuzumab in HER2-positive non-small-cell lung cancer. *Ann. Oncol.*, **15**, 19–27.

Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattn Anal. Mach. Intell.*, **6**, 721–741.

Green, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732.

Green, P. J. and Silverman, B. W. (1993) *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach*. New York: Chapman and Hall.

Li, Y., Lin, X. and Müller, P. (2010) Bayesian inference in semiparametric mixed models for longitudinal data. *Biometrics*, **66**, 70–78.

Okada, H. and Mak, T. W. (2004) Pathways of apoptotic and non-apoptotic death in tumour cells. *Nat. Rev. Cancer*, **4**, 592–603.

Pardoll, D. M. (2012) The blockade of immune checkpoints in cancer immunotherapy. *Nat. Rev. Cancer*, **12**, 252–264.

Prahallad, A., Sun, C., Huang, S., Di Nicolantonio, F., Salazar, R., Zecchin, D., Beijersbergen, R. L., Bardelli, A. and Bernards, R. (2012) Unresponsiveness of colon cancer to BRAF (V600E) inhibition through feedback activation of EGFR. *Nature*, **483**, 100–103.

Redig, A. J. and Jänne, P. A. (2015) Basket trials and the evolution of clinical trial design in an era of genomic medicine. *J. Clin. Oncol.*, **33**, 975–977.

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003) *Semiparametric Regression*. Cambridge: Cambridge University Press.

Simon, R. and Roychowdhury, S. (2013) Implementing personalized cancer genomics in clinical trials. *Nat. Rev. Drug Discvry*, **12**, 358–369.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *J. R. Statist. Soc.* B, **64**, 583–639.
Thall, P. F., Simon, R. M. and Estey, E. H. (1995) Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes. *Statist. Med.*, **14**, 357–379.
Thall, P. F., Wathen, J. K., Bekele, B. N., Champlin, R. E., Baker, L. H. and Benjamin, R. S. (2003) Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. *Statist. Med.*, **22**, 763–780.
Tiacci, E., Trifonov, V., Schiavoni, G., Holmes, A., Kern, W., Martelli, M. P., Pucciarini, A., Bigerna, B., Pacini, R., Wells, V. A. and Sportoletti, P. (2011) BRAF mutations in hairy-cell leukemia. *New Engl. J. Med.*, **364**, 2305–2315.
Yuan, Y., Nguyen, H. Q. and Thall, P. F. (2016) *Bayesian Designs for Phase I-II Clinical Trials*. London: Chapman and Hall.
Zuiverloon, T. C., Nieuweboer, A. J., Vékony, H., Kirkels, W. J., Bangma, C. H. and Zwarthoff, E. C. (2012) Markers predicting response to bacillus Calmette-Guérin immunotherapy in high-risk bladder cancer patients: a systematic review. *Eur. Urol.*, **61**, 128–145.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'BLAST: Bayesian latent subgroup design for basket trials supplemental materials'.