Research Article

Statistics in Medicine

Received 28 May 2014, Accepted 31 December 2014 Published online 28 January 2015 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6442

Bridging continual reassessment method for phase I clinical trials in different ethnic populations

Suyu Liu,[a‡] Haitao Pan,[b,c‡] Jielai Xia,[c*] Qin Huang[d] and Ying Yuan[a,b*†]

Accumulating evidence shows that the conventional one-size-fits-all dose-finding paradigm is problematic when applied to different ethnic populations. Because of inter-ethnic heterogeneity, the dosage established in a landmark trial for a certain population may not be generalizable to a different ethnic population, and a follow-up bridge trial is often needed to find the maximum tolerated dose for the new population. We propose the bridging continual reassessment method (B-CRM) to facilitate dose finding for such follow-up bridge trials. The B-CRM borrows information from the landmark trial through a novel estimate of the dose-toxicity curve and accommodates the inter-ethnic heterogeneity using the Bayesian model averaging approach. Extensive simulation studies show that the B-CRM has desirable operating characteristics with a high probability to select the target dose. This article focuses on ethnic heterogeneity, but the proposed method can be directly used to handle other types of patient heterogeneity, for example, patient subgroups defined by prognostic factors or biomarkers. The software to implement the B-CRM design is available for free download at http://odin.mdacc.tmc.edu/~yyuan/**.**

Keywords: adaptive design; bridge studies; borrow strength; maximum tolerated dose; inter-ethnic heterogeneity; subgroups; dose finding

1. Introduction

The primary goal of a phase I clinical trial is to identify the maximum tolerated dose (MTD) of a new drug, which is defined as the dose with a toxicity probability that is closest to the target toxicity rate. A phase I clinical trial is important as it determines the MTD that will be further investigated in the subsequent phase II or III trials. Numerous statistical methods have been developed for phase I dose-finding studies, for example, the conventional '3 + 3' design [1], continual reassessment method (CRM) [2], decision-theoretic approach [3], random walk-based design [4], dose escalation method with overdose control [5], and the Bayesian optimal interval design [6], among others. Comprehensive reviews of dose-finding methods can be found in Chevret [7] and Ting [8].

Traditionally, phase I trials are conducted in a 'one-size-fits-all' fashion. That is, once the MTD of a drug is established in a landmark study based on a certain ethnic population (e.g., a Caucasian population), the results are directly extrapolated to other ethnic populations (e.g., an Asian population). Unfortunately, accumulating evidence shows that such a one-size-fits-all dose-finding paradigm is problematic, and ethnicity plays an important role in a patient's response to a drug [9]. The genetic and environmental differences among ethnic populations influence both the pharmacokinetics and pharmacodynamics of drugs [10]. As a result, different ethnic populations may have different MTDs. For example, a recent

[a] *Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, 77030, U.S.A.*
[b] *School of Statistics, Xi'an University of Finance and Economics, Xi'an, 710061, China*
[c] *Department of Health Statistics, The Fourth Military Medical University, Xi'an, 710032, China*
[d] *Office of Biostatistics, Center for Drug Evaluation, China Food and Drug Administration (CFDA), Beijing, 100038, China*
[*] *Correspondence to: Ying Yuan, Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, 77030, U.S.A., and Jielai Xia, Department of Health Statistics, The Fourth Military Medical University, Xi'an, 710032, China.*
[†] *E-mail: yyuan@mdanderson.org*
[‡] *Equal contribution*

study of sorafenib administered in patients with hepatocellular carcinoma (HCC) has found that the MTD of sorafenib is significantly lower among Asian patients than among non-Asian patients [11]. According to the manufacturer and the United States Food and Drug Administration (FDA), the recommended dose of sorafenib is 400 mg bid. That dose has been used in the pivotal phase III Sorafenib HCC Assessment Randomized Protocol trial, which involves a patient population drawn from Europe, North America, and South America [12]. However, the study of Barrera *et al.* [11] showed that Asian patients demonstrated poor tolerance to the manufacturer's recommended initial dose of the drug. Among a total of 36 Asian patients evaluated with this drug, 97% did not tolerate the FDA-indicated dose of 400 mg bid. Another example of inter-ethnic difference in drug tolerance is the administration of docetaxel, a broad spectrum taxane commonly used to treat solid tumors such as lung, breast, gastric, ovarian, and prostate tumors. The main side effects of docetaxel are myelosuppression and peripheral neuropathy. Different doses of docetaxel are administered in different geographic populations. In Caucasians, the common starting dose of docetaxel for the first-line treatment is 100 mg/m^2 [13]; whereas in countries with Asian populations such as China, the common starting dose is 70 to 75 mg/m^2 [14], and in Japan, the approved starting dose for docetaxel is 60 mg/m^2 [15]. Despite using the reduced doses, the incidence of febrile neutropenia was still higher among Asians than among Caucasians [14, 15]. Such inter-ethnic differences in docetaxel tolerance may arise from different clearance and exposure rates in Asians and Caucasians [16]. Goh *et al.* [17] reported that docetaxel clearance is approximately 40% lower, while the area under the curve for docetaxel (i.e., drug exposure) is approximately 25% higher in Asians than in Caucasians.

The inter-ethnic differences have been recognized by drug regulatory authorities. In 1999, the FDA published the guidance, 'Ethnic Factors in the Acceptability of Foreign Clinical Data' (i.e., E5 Guidance), which suggested the need to distinguish three ethnic categories (i.e., Asian, Black people, and Caucasian) in drug development. The guidance identified situations for which drugs could be ethnically sensitive and suggested the types of bridge studies that may be required to extrapolate clinical trial results from one region to another. In 2005, the FDA published another guidance and extended the race category further to include five minimum ethnic groups, namely, Caucasian, Black American/African American, Asian, American Indian/Alaskan Native, and Native Hawaiian/other Pacific Islander. In the same year, the FDA approved the first race-specific drug, isosorbide dinitrate and hydralazine hydrochloride (proprietary name: BiDil), for the treatment of congestive cardiac failure in black patients.

The goal of this article is to address the following bridge trial design question: given that a landmark phase I trial has been conducted in a landmark population and the corresponding MTD has been established, how do we design a follow-up trial (i.e., a bridge trial) to find the MTD for a new population? A straightforward approach is to ignore the early landmark trial and to conduct another independent phase I trial (e.g., using the CRM) to find the MTD for the new population. While this approach fully acknowledges the inter-ethnic heterogeneity, it is not efficient because the dose-toxicity relationships in different ethnic populations are expected to be closely related, even though there are some inter-ethnic differences. This is true because, after all, we are concerned with the same drug acting through the same biological mechanism in human beings. In other words, the dose-toxicity response observed in the landmark trial should inform the basic dose-toxicity behavior of the drug in the new population. Ignoring the data from the landmark trial is wasting useful information.

To address this issue, we propose the bridging CRM (B-CRM) design that utilizes the dose-toxicity data obtained from the landmark trial to achieve efficient dose finding in the follow-up trial while also acknowledging inter-ethnic heterogeneity. Specifically, we first estimate the dose-toxicity curve using the data from the landmark trial and then use that estimate to form a prior dose-toxicity curve, which is also known as the skeleton of the CRM, for the follow-up trial. To accommodate the inter-ethnic heterogeneity, we form multiple skeletons, by shifting the estimate of the dose-toxicity curve one dose level up or down, to represent a more conservative or aggressive dose response in the new population. We employ the Bayesian model averaging approach [28] to draw inference across multiple skeletons and adaptively make the decision of dose assignment and selection. This article focuses on ethnic heterogeneity, but the proposed method can be used to handle other types of patient heterogeneity, for example, patient subgroups defined by prognostic factors or biomarkers.

Some research has been done for bridging studies. Shih [18] proposed a method to determine whether a study is capable of bridging the foreign data to the new study. Lan *et al.* [19] proposed weighted Z-tests in which incorporate the prior observed data in bridging studies using weights. Gould *et al.* [20] developed a Bayesian predictive approach for designing and analyzing bridging studies that fully incorporate the information provided by the original trials. Gandhi *et al.* [21] proposed a Bayesian approach for inference from a bridging study with binary outcomes. Chow *et al.* [22] provided a review of statistical methods for

bridging studies. Most of the aforementioned works focus on statistical inference, and limited research has been done from the trial design perspective. Morita [23] proposed a phase I trial design that uses an informative prior to incorporate previous study information into the bridging study based on the CRM.

The remainder of the article is organized as follows. In Section 2, after a brief review of the original CRM, we propose a novel mixture estimate of the dose-toxicity curve using the landmark trial data. Based on this estimate, we present the procedure of using multiple skeletons to accommodate the inter-ethnic heterogeneity and that of using Bayesian model averaging to make the decision of dose assignment. In Section 3, we investigate the operating characteristics of the proposed B-CRM using simulation studies. In Section 4, we illustrate the proposed design using a phase I clinical trial for advance solid tumors and conclude with a brief discussion in Section 5.

2. Methods

2.1. Continual reassessment method

Let $(d_1, \ldots, d_J)$ denote a set of J prespecified doses for the drug under investigation. We assume that the dose-limiting toxicity (DLT) is recorded as a binary outcome, and the true dose toxicity monotonically increases with respect to the dose level. Let $(p_1, \ldots, p_J)$ be the prespecified toxicity probabilities of the J doses, $p_1 < \ldots < p_J$, which are also known as the skeleton. The CRM links the true toxicity probability at dose d_j, denoted as $\pi(d_j)$, with the prespecified prior toxicity probability p_j, using a working dose-toxicity model, such as

$$\pi(d_j) = p_j^{\exp(\alpha)} \tag{1}$$

for $j = 1, \ldots, J$, where α is an unknown parameter. To conduct the trial, the CRM continuously updates the estimate of the dose-toxicity model using the accrued information, and adaptively assigns incoming patients to the dose with an estimated toxicity probability closest to the prespecified target toxicity probability, ϕ. Once the maximum sample size is reached, the dose with a posterior toxicity probability closest to ϕ is selected as the MTD.

One important feature of the CRM is that the prior information on the dose-toxicity curve can be naturally incorporated into the model through the specification of the skeleton, which enhances the performance of the design. In the proposed B-CRM, we exploit this feature of the CRM to borrow the dose-toxicity information from the landmark trial for finding the MTD in the follow-up trial. In the next subsection, we first describe a method to estimate the dose-toxicity curve based on the dose-toxicity data generated by the landmark trial and then discuss how to incorporate such rich prior information into the follow-up trial by specifying multiple skeletons.

2.2. Estimation of dose-toxicity curve in landmark population

We assume that a landmark phase I trial has been conducted in a (landmark) population, with J_L prespecified doses, $b_1 < \ldots < b_{J_L}$. The trial identified dose b_{j^*} as the MTD and resulted in binomial data $\mathcal{D}_L = (x_j, m_j)$, where x_j is the number of patients who experienced toxicity, and m_j is the total number of patients treated at dose b_j, for $j = 1, \ldots, J_L$. Given $\mathcal{D}_L$, a straightforward way to estimate the dose-toxicity curve is to fit a probit model

$$\pi_j^{(P)} \equiv \pi^{(P)}(b_j) = \Phi(\beta_0 + \beta_1 b_j), \tag{2}$$

where the superscript in $\pi_j^{(P)}$ indicates it is a parametric estimate; $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution; β_0 and β_1 are intercept and slope parameters, respectively. We require $\beta_1 > 0$ such that toxicity monotonically increases with the dose. We adopt the probit model because of its intuitive toxicity tolerance interpretation [24]. The tolerance is defined as the dose intensity level below which toxicity does not occur and above which toxicity occurs. If we assume that the tolerance varies from subject to subject and is normally distributed, then the dose-toxicity curve follows a probit model of the form given by (2). This parametric approach is simple, but when the model is misspecified, the resulting estimate may be severely biased. Alternatively, we can nonparametrically estimate the toxicity probability of dose b_j using isotonic regression [25, 26],

$$\hat{\pi}_j^{(NP)} = \max_{0 \leqslant u \leqslant j} \min_{j \leqslant v \leqslant J_L} \frac{\sum_{k=u}^{v} x_k}{\sum_{k=u}^{v} m_k}.$$

This isotonic estimate satisfies the monotonic constraint $\pi_1^{(NP)} \leqslant \ldots \leqslant \pi_{J_L}^{(NP)}$, and can be easily obtained by applying the pooled–adjacent-violators algorithm (PAVA) [27] to the observed toxicity rate $\gamma_j = x_j/m_j$, $j = 1, \ldots, J_L$. Operatively, the PAVA replaces any adjacent γ_j's that violates the nondecreasing order by their (weighted) average so that the resulting estimates, $\hat{\pi}_j^{(NP)}$, become monotonic. Bhattacharya and Kong [26] showed that $\hat{\pi}_j^{(NP)}$ is consistent under mild conditions. The drawback of isotonic regression is that the resulting estimates can be highly variable, and it is difficult to estimate the toxicity probabilities for doses outside the observed dose range, $[b_1, b_{J_L}]$.

To inherit the merits of parametric regression and nonparametric isotonic regression, we propose a mixture (or weighted average) estimator of toxicity probabilities,

$$\hat{\pi}_j = w_j \hat{\pi}_j^{(P)} + (1 - w_j)\hat{\pi}_j^{(NP)}, \tag{3}$$

where weight w_i is chosen in a data-driven way such that if $\hat{\pi}_j^{(P)}$ is more accurate, more weight is assigned to the probit regression, and if $\hat{\pi}_j^{(NP)}$ is more accurate, more weight is assigned to the isotonic regression. We propose the following data-based weight,

$$w_j = \frac{\lambda_j}{\lambda_j + 1},$$

where

$$\lambda_j = \frac{\left(\hat{\pi}_j^{(P)}\right)^{x_j}\left(1 - \hat{\pi}_j^{(P)}\right)^{m_j - x_j}}{\left(\hat{\pi}_j^{(NP)}\right)^{x_j}\left(1 - \hat{\pi}_j^{(NP)}\right)^{m_j - x_j}} \tag{4}$$

is the (estimated) likelihood ratio evaluated at dose level j under the probit model and isotonic regression. Using the likelihood ratio as a weight, the parametric or nonparametric estimate that fits the data better will receive a higher weight. As a result, the proposed mixture estimator has the consistency property described in the following theorem (see the Appendix for the proof).

Theorem 1
The proposed mixture estimator in (3) is a consistent estimate of π_j.

Although $\left\{\hat{\pi}_j^{(P)}\right\}$ and $\left\{\hat{\pi}_j^{(NP)}\right\}$ are both monotonic, as weight w_j varies across doses, the mixture estimator $\{\hat{\pi}_j\}$ may occasionally violate the monotonicity assumption in finite samples. If that occurs, we can apply the PAVA algorithm to the $\hat{\pi}_j$'s to impose monotonicity. The transformed estimate will not take a form exactly the same as that given by (3), but it remains to be a consistent estimate of π_j. For an arbitrary dosage d between b_j and b_{j+1}, its toxicity probability can be estimated using linear interpolation

$$\hat{\pi}(d) = \hat{\pi}_j + \frac{d - b_j}{b_{j+1} - b_j}\left(\hat{\pi}_{j+1} - \hat{\pi}_j\right), \qquad b_j \leqslant d \leqslant b_{j+1}.$$

Other more sophisticated methods such as smoothing splines can also be used to extrapolate the $\hat{\pi}_j$'s and can obtain the estimate of $\pi(d)$. However, the simple linear interpolation is typically adequate because our goal here is not to pursue a precise estimate of the entire dose-response curve for the landmark population but to utilize the landmark trial data to provide a ballpark estimate of the toxicity probabilities at some prespecified doses (i.e., skeleton) for the follow-up population to facilitate the dose finding for the follow-up trial. Actually, as we never observe any data between b_j and b_{j+1}, all extrapolation methods are based on certain untestable model assumptions, and the observed data cannot inform us as to which extrapolation method is better. In addition, similar to the standard CRM, under the proposed B-CRM described in the succeeding texts, the accumulating data collected in the follow-up trial will quickly dominate the skeleton and will limit the influence of the skeleton.

2.3. *Bridging CRM*

We now consider how to design the follow-up trial to find the MTD for a new population, given that the estimate of the dose-toxicity curve has been obtained from the landmark trial as described previously.

We assume that a set of J doses, $d_1 < \ldots < d_J$, are under investigation in the follow-up trial. These doses do not have to be the same set of doses previously studied in the landmark trial (i.e., $b_1, \ldots, b_{J_L}$), and can be chosen based on the data collected from the landmark trial. For example, based on the dose-toxicity estimate (3) obtained from the landmark trial, we can choose the d_j's locally to the MTD. This can be done by taking the MTD identified in the landmark trial as the middle dose and then by determining other candidate doses by backsolving the dose-toxicity function such that the estimated toxicity probabilities of the other doses are equally spaced around the MTD. In general, no matter which approach is taken, the MTD identified in the landmark trial, that is, b_{j^*}, should be included as one of the investigational doses in the follow-up trial. Without a loss of generality, we assume that this dose corresponds to dose level t in the follow-up trial (i.e., $d_t = b_{j^*}$).

Because of inter-ethnic heterogeneity, we do not expect the dose-toxicity curve for the new population to be exactly the same as that of the landmark population. On the other hand, we also do not expect that these two curves dramatically differ from each other because they concern the same drug used to treat patients with the same type of disease. In practical use, the dose-toxicity curve of the new population should mostly resemble that of the landmark population, with some deviation. As a result, the MTD for the new population should be in the neighborhood of that of the landmark population (e.g., one dose level difference). Let $\hat{p}_j = \hat{\pi}(d_j) j = 1, \ldots, J$ denote the estimate of toxicity probability of d_j based on the landmark trial data. Under the CRM framework, we incorporate such prior information using three sets of skeletons:

$$\begin{aligned}
&\text{skeleton } 1: \quad p_j = \hat{p}_j \\
&\text{skeleton } 2: \quad p_j = \hat{p}_{j+1} \quad \text{for } j = 1, \ldots, J-1, \\
&\qquad\qquad\quad p_J = \frac{\hat{p}_J + 1}{2} \\
&\text{skeleton } 3: \quad p_j = \hat{p}_{j-1} \quad \text{for } j = 2, \ldots, J, \\
&\qquad\qquad\quad p_1 = \frac{\hat{p}_1}{2}.
\end{aligned}$$

That is, skeleton 1 represents that a priori the new population has the same toxicity profile as that of the landmark population; whereas skeletons 2 and 3 shift the dose-toxicity curve one level up and one level down, respectively. Under skeletons 1, 2, and 3, the MTD for the new population is a priori the same as one level lower or one level higher than that for the landmark population. For skeleton 2, when we shift the toxicity probabilities one level up, the toxicity probability of the highest dose level (i.e., p_J) will move out of the existing range, and thus we take p_J as the middle value between $\hat{p}_J$ and 1. Similarly, we set the toxicity probability of the lowest dose level (i.e., p_1) as the middle value between 0 and $\hat{p}_1$ in skeleton 3 when shifting the toxicity probability one level down. Following Yin and Yuan [28], we regard each skeleton as a CRM model and use the Bayesian model averaging (BMA) [29, 30] approach to estimate toxicity probabilities across multiple skeletons for adaptive dose assignment and selection.

Specifically, let $(M_1, \ldots, M_K)$ be the models corresponding to each set of prior guesses of the toxicity probabilities $\{(p_{11}, \ldots, p_{1J}), \ldots, (p_{K1}, \ldots, p_{KJ})\}$, where $K = 3$. Model M_k $(k = 1, \ldots, K)$ in the CRM is given by

$$\pi_{kj}(\alpha_k) = p_{kj}^{\exp(\alpha_k)}, \quad j = 1, \ldots, J,$$

which is based on the kth skeleton $(p_{k1}, \ldots, p_{kJ})$. Let $\text{pr}(M_k)$ be the prior probability that model M_k is the true model, that is, the probability that the kth skeleton $(p_{k1}, \ldots, p_{kJ})$ matches the true dose-toxicity curve. The value of $\text{pr}(M_k)$ should reflect the prior knowledge of whether the new population is likely to be less or more tolerable to the drug. For example, Asians are often expected to be less tolerable to certain drugs than Caucasians. Thus, if the landmark population is Caucasian, we may assign a high prior probability to skeleton 2 and a low prior probability to skeleton 3 when the new population is Asian. When there is no prior information regarding the relative tolerance between landmark and new populations, we can assign equal weights to the different skeletons by simply setting $\text{pr}(M_k) = 1/K$. Suppose at a certain stage of the trial, among n_j patients treated at dose level j, y_j patients have experienced DLT. Let $\mathcal{D} = \{(n_j, y_j), j = 1, \ldots, J\}$ denote the observed data, the likelihood function under model M_k is

$$L(\mathcal{D}|\alpha_k, M_k) = \prod_{j=1}^{J} \left\{ p_{kj}^{\exp(\alpha_k)} \right\}^{y_j} \left\{ 1 - p_{kj}^{\exp(\alpha_k)} \right\}^{n_j - y_j}.$$

The posterior model probability for M_k is given by

$$\text{pr}(M_k|\mathcal{D}) = \frac{L(\mathcal{D}|M_k)\text{pr}(M_k)}{\sum_{i=1}^{K} L(\mathcal{D}|M_i)\text{pr}(M_i)},$$

where $L(\mathcal{D}|M_k)$ is the marginal likelihood of model M_k.

$$L(\mathcal{D}|M_k) = \int L(\mathcal{D}|\alpha_k, M_k) f(\alpha_k|M_k) d\alpha_k.$$

α_k is the power parameter in the CRM associated with model M_k, and $f(\alpha_k|M_k)$ is the prior distribution of α_k under model M_k, for example, $f(\alpha_k|M_k) \sim N(0, 2)$. The BMA estimate for the toxicity probability at each dose level is given by

$$\bar{\pi}_j = \sum_{k=1}^{K} \hat{\pi}_{kj} \text{pr}(M_k|\mathcal{D}), \quad j = 1, \ldots, J, \tag{5}$$

where $\hat{\pi}_{kj}$ is the posterior mean of the toxicity probability of dose level j under model M_k, that is,

$$\hat{\pi}_{kj} = \int p_{kj}^{\exp(\alpha_k)} \frac{L(\mathcal{D}|\alpha_k, M_k) f(\alpha_k|M_k)}{\int L(\mathcal{D}|\alpha_k, M_k) f(\alpha_k|M_k) d\alpha_k} d\alpha_k.$$

Alternatively, we can use the model selection approach to estimate the toxicity probabilities and make the decision of dose assignment. That is, at each point of decision making for dose assignment, we select the model with the highest posterior probability, that is, model $k^* = \text{argmax}_{k\in(1,\ldots,K)}(\text{pr}(M_k|\mathcal{D}))$ and use that model to make inference and dose assignment. However, our numerical study shows that the BMA approach performs slightly better than the model selection approach (results not shown), thus we focus on the BMA approach hereafter.

2.4. Dose-finding algorithm

The dose-finding algorithm for our B-CRM design is described as follows:

(1) Patients in the first cohort are treated at dose d_{t-1}, that is, the dose one level is lower than the MTD identified in the landmark trial (i.e., d_t). Note that we choose d_{t-1} rather than d_t as the starting dose to fit the physician's inclination to be conservative for safety purposes.
(2) At the current dose level, j^{curr}, based on the observed data, we calculate the BMA estimates for the toxicity probabilities, $\bar{\pi}_j$ $(j = 1, \ldots, J)$ and identify dose level j^{best} that has a toxicity probability closest to ϕ, that is,

$$j^{\text{best}} = \text{argmin}_{j\in(1,\ldots,J)} |\bar{\pi}_j - \phi|.$$

If $j^{\text{curr}} > j^{\text{best}}$, we de-escalate the dose level to $j^{\text{curr}} - 1$; if $j^{\text{curr}} < j^{\text{best}}$, we escalate the dose level to $j^{\text{curr}} + 1$; otherwise, the dose stays at the same level as j^{curr} for the next cohort of patients. Being conservative, we restrict the dose change to one level at a time.
(3) Once the maximum sample size is reached, the dose that has the toxicity probability closest to ϕ is selected as the MTD.

In addition, we impose the following safety stopping rule: if $\text{pr}(\pi_1 > \phi|\mathcal{D}) > 0.9$, the trial is terminated for safety. That is, if the lowest dose has a high probability of being overly toxic, we should stop the trial early for safety. The software to implement the proposed B-CRM design (written in R) can be found in supplementary materials and also is available for free download at http://odin.mdacc.tmc.edu/~yyuan/.

3. Simulation studies

We investigated the operating characteristics of the proposed B-CRM design through simulation studies. We considered six dose levels with a target toxicity probability of $\phi = 30\%$. The maximum sample size for the follow-up trial was 21 patients in cohorts of size 3. Suppose that the landmark trial has been done using a certain phase I trial design (e.g., the '3 + 3' design) that yielded the following data: the

Table I. Simulation study comparing the CRM, CRM using an IP-CRM and B-CRM. The underlined dose is the target dose.

			Dose level					
Scenario	Design		1	2	3	4	5	6
1		Pr(toxicity)	0.04	0.08	0.15	0.33	0.45	0.60
	CRM	selection(%)	0.0	2.3	23.0	**48.3**	22.6	3.4
		# of patients	0.1	2.0	7.1	7.0	3.9	0.9
	IP-CRM	selection(%)	0.0	0.3	25.3	**72.4**	2.0	0.0
		# of patients	0.0	0.4	7.9	12.4	0.4	0.0
	B-CRM	selection(%)	0.0	1.0	26.5	**56.1**	15.6	0.8
		# of patients	0.1	0.5	8.4	9.1	2.6	0.3
2		Pr(toxicity)	0.02	0.05	0.08	0.10	0.30	0.45
	CRM	selection(%)	0.0	0.1	1.4	15.5	**53.7**	29.2
		# of patients	0.0	0.8	4.1	4.5	6.5	5.0
	IP-CRM	selection(%)	0.0	0.0	3.3	59.9	**36.8**	0.0
		# of patients	0.0	0.1	4.4	12.2	4.3	0.0
	B-CRM	selection(%)	0.0	0.0	2.3	21.1	**59.5**	17.1
		# of patients	0.0	0.1	4.5	6.1	7.6	2.6
3		Pr(toxicity)	0.05	0.12	0.25	0.42	0.55	0.65
	CRM	selection(%)	0.6	11.4	**48.0**	31.9	6.9	0.2
		# of patients	0.4	4.0	8.9	5.2	2.1	0.3
	IP-CRM	selection(%)	0.1	6.1	**54.0**	39.1	0.7	0.0
		# of patients	0.0	1.8	11.3	7.7	0.1	0.0
	B-CRM	selection(%)	0.1	8.7	**56.3**	31.7	2.9	0.1
		# of patients	0.3	1.9	11.6	6.1	1.0	0.0
4		Pr(toxicity)	0.02	0.03	0.04	0.06	0.10	0.33
	CRM	selection(%)	0.0	0.0	0.1	1.1	18.8	**80.0**
		# of patients	0.0	0.4	3.5	3.2	4.5	9.3
	IP-CRM	selection(%)	0.0	0.0	0.4	40.2	59.4	**0.0**
		# of patients	0.0	0.0	3.5	10.4	7.1	0.0
	B-CRM	selection(%)	0.0	0.0	0.2	3.0	33.7	**63.1**
		# of patients	0.0	0.0	3.5	4.2	6.0	7.3
5		Pr(toxicity)	0.15	0.26	0.50	0.60	0.70	0.75
	CRM	selection(%)	16.8	**45.6**	21.4	1.7	0.2	0.0
		# of patients	3.3	7.8	6.4	0.9	0.2	0.0
	IP-CRM	selection(%)	9.7	**56.9**	31.1	1.9	0.0	0.0
		# of patients	1.7	8.4	9.4	1.4	0.0	0.0
	B-CRM	selection(%)	15.9	**58.6**	21.6	0.6	0.1	0.0
		# of patients	3.2	7.8	8.7	0.9	0.0	0.0
6		Pr(toxicity)	0.30	0.46	0.55	0.65	0.75	0.85
	CRM	selection(%)	**40.9**	19.0	3.5	0.3	0.0	0.0
		# of patients	6.3	5.4	4.1	0.5	0.1	0.0
	IP-CRM	selection(%)	**48.2**	35.0	9.2	0.8	0.0	0.0
		# of patients	5.7	7.6	6.7	0.7	0.0	0.0
	B-CRM	selection(%)	**48.6**	21.5	5.2	0.2	0.0	0.0
		# of patients	6.4	5.3	6.0	0.5	0.0	0.0

CRM, continual reassessment method; IP-CRM, informative prior continual reassessment method; B-CRM, bridging continual reassessment method.

number of patients treated at each dose $(m_1, \ldots, m_6) = (3, 3, 3, 6, 3, 0)$ and the number of patients who experienced dose-limiting toxicity at each dose $(x_1, \ldots, x_6) = (0, 0, 0, 1, 2, 0)$. Dose level 4 was identified as the MTD in the landmark trial. For the follow-up trial, we considered six toxicity scenarios that differ in the location of the true MTD. We note that the actual dosage of these six dose levels can be different from that in the landmark trial. In addition, in practice, the number of doses studied in the follow-up trial is not necessarily the same as that of the landmark trial. We compared the proposed B-CRM with two

methods: the conventional CRM (for a stand-alone trial), which does not actively borrow information from the landmark trial and the method proposed by Morita [23], which borrows information from the landmark trial using an informative prior under the CRM (referred to as the IP-CRM). For the IP-CRM method, a one-parameter logistic regression model was used,

$$\pi_j = \frac{\exp\left(\beta_0 + \beta_1 b_j\right)}{1 + \exp\left(\beta_0 + \beta_1 b_j\right)},$$

where $\beta_0 \equiv 3$, and dosage b_j was specified using 'backward fitting' [31] such that the prior estimates of the toxicity probabilities match the estimates from the landmark trial, which are obtained by fitting a logistic model to the landmark trial data. Following Morita [23], we assumed that β_1 follows a gamma prior $Ga(5, 5)$. For the conventional CRM, we chose the skeleton (0.12, 0.20, 0.30, 0.40, 0.50, 0.6) based on the method of Lee and Cheung [32], assuming an indifference interval of 0.1. We used the same starting dose (i.e., dose level 3) for all three designs. Strictly speaking, because this starting dose is chosen based on the landmark trial data, the CRM we considered here actually borrowed some information from the landmark trial.

Table I shows the simulation results, including selection percentages and the number of patients treated at each dose, based on 1000 simulated trials. In scenario 1, the MTD is the fourth dose, the same dose level as the MTD of the landmark trial. Compared with the CRM, the B-CRM yielded a 7.8% higher percentage of correct selection (PCS) and assigned about two more patients to the MTD. In addition, the B-CRM was also 9.6% less likely to select the overly toxic doses (i.e., dose levels 5 and 6) than the CRM. The IP-CRM performed best with the highest PCS. However, as we will see, when the MTD of the follow-up trial differs from that of the landmark trial, the performance of the IP-CRM can be poor. Scenarios 2 and 3, respectively, present the cases in which the MTD in the follow-up trial is one level higher or lower than that identified in the landmark population. In these cases, the B-CRM outperformed the CRM with 6%–8% higher PCS and also lower probabilities of selecting the overly toxic doses. The IP-CRM showed large variation: it performed reasonably well in scenario 3 (PCS = 54.0%) but poorly in scenario 2 (PCS = 36.8%). In scenarios 4 and 5, the MTD in the follow-up trial is two levels different from that in the landmark trial. The B-CRM consistently performed well, with the PCS ranging from 58.6% to 63.1%. The CRM and IP-CRM were less stable. The CRM performed very well in scenario 4 but not as well in scenario 5, whereas the IP-CRM performed well in scenario 5 but very poorly in scenario 4. Scenario 6 has the first dose as the MTD, which is three levels different from the MTD identified in the landmark trial. In this case, the B-CRM and IP-CRM yielded similar PCS, but the B-CRM was 18.1% less likely to select doses above the MTD. The PCS of the CRM is about 8% lower than those of the B-CRM and IP-CRM. As a sensitivity analysis, we also investigated the operating characteristics of the designs given a different set of landmark trial data, that is, $(m_1, \ldots, m_6) = (3, 3, 6, 3, 0, 0)$ and $(x_1, \ldots, x_6) = (0, 0, 1, 2, 0, 0)$ for which dose level 3 was identified as the MTD of the landmark trial. The pattern of the results is generally similar to that given previously (Table II).

As we noted previously, in practice, we do not expect the dose-toxicity curve for the new population to dramatically differ from that for the landmark population because they concern the same drug that is used to treat patients with the same type of disease. The MTD for the new population should be in the neighborhood of that of the landmark population, that is, scenarios 1, 2, and 3 are more likely encountered in practice than other scenarios. In the case when there is strong prior knowledge that the MTD for the new population is much lower than the MTD for the landmark population, when specifying the investigational doses for the follow-up trial, we should choose more dose levels that are lower than the MTD identified in the landmark trial. Specifically, we can choose the dose that is most likely to be the MTD (for the new population) as dose level 4 (of the follow-up trial) and then add other doses. By doing so, we ensure that the MTD for the follow-up trial is still in the neighborhood of that of the landmark trial in terms of dose level (i.e., dose level 4), although they may be very different in terms of actual dosage. One advantage of using the power model (1) is that the actual dosages are not directly used in the model; rather, we use their associated toxicity probabilities (i.e., the skeleton). For example, in our simulation, we did not need to specify the actual dosages for the follow-up trials.

Table II. Sensitivity analysis of the CRM, IP-CRM, and B-CRM, given landmark trial data $(m_1, \ldots, m_6) = (3, 3, 6, 3, 0, 0)$ and $(x_1, \ldots, x_6) = (0, 0, 1, 2, 0, 0)$. The underlined dose is the target dose.

			Dose level					
Scenario	Design		1	2	3	4	5	6
1		Pr(toxicity)	0.05	0.12	0.35	0.42	0.55	0.65
	CRM	selection(%)	0.4	23.6	**46.9**	23.6	5.1	0.1
		# of patients	1.4	6.8	7.1	4.3	1.1	0.2
	IP-CRM	selection(%)	0.0	21.5	**76.2**	2.3	0.0	0.0
		# of patients	0.2	7.2	13.2	0.4	0.0	0.0
	B-CRM	selection(%)	0.0	22.3	**59.6**	16.5	1.4	0.0
		# of patients	0.4	7.7	9.6	2.9	0.3	0.0
2		Pr(toxicity)	0.04	0.08	0.15	0.26	0.45	0.60
	CRM	selection(%)	0.0	1.4	15.6	**47.9**	31.2	3.8
		# of patients	1.0	4.3	4.7	6.2	3.9	0.8
	IP-CRM	selection(%)	0.0	3.8	70.1	**25.3**	0.8	0.0
		# of patients	0.1	4.5	13.3	3.1	0.0	0.0
	B-CRM	selection(%)	0.0	3.0	23.7	**56.4**	16.0	0.9
		# of patients	0.2	4.7	6.6	7.1	2.2	0.2
3		Pr(toxicity)	0.15	0.26	0.50	0.60	0.70	0.75
	CRM	selection(%)	19.6	**50.5**	23.0	2.1	0.2	0.0
		# of patients	5.1	9.1	4.8	1.3	0.1	0.0
	IP-CRM	selection(%)	9.1	**63.4**	26.8	0.0	0.0	0.0
		# of patients	2.1	12.4	6.4	0.0	0.0	0.0
	B-CRM	selection(%)	10.0	**64.3**	22.4	1.1	0.0	0.0
		# of patients	2.5	12.3	5.2	0.5	0.0	0.0
4		Pr(toxicity)	0.02	0.05	0.08	0.10	0.30	0.45
	CRM	selection(%)	0.0	0.1	1.2	14.9	**50.3**	33.5
		# of patients	0.5	3.6	3.3	4.3	5.6	3.7
	IP-CRM	selection(%)	0.0	0.8	49.0	34.0	**16.2**	0.0
		# of patients	0.0	3.7	11.3	5.4	0.6	0.0
	B-CRM	selection(%)	0.0	0.4	5.6	33.3	**46.7**	14.0
		# of patients	0.1	3.7	4.7	5.7	5.4	1.4
5		Pr(toxicity)	0.30	0.46	0.55	0.65	0.75	0.85
	CRM	selection(%)	**48.1**	17.7	1.9	0.2	0.1	0.0
		# of patients	9.4	6.0	1.1	0.3	0.0	0.0
	IP-CRM	selection(%)	**52.9**	32.7	4.6	0.0	0.0	0.0
		# of patients	8.7	9.5	1.8	0.0	0.0	0.0
	B-CRM	selection(%)	**54.4**	22.3	3.1	0.3	0.0	0.0
		# of patients	8.9	8.0	1.5	0.1	0.0	0.0
6		Pr(toxicity)	0.02	0.03	0.04	0.06	0.10	0.25
	CRM	selection(%)	0.0	0.0	0.2	1.6	14.9	**83.3**
		# of patients	0.3	3.3	3.1	3.3	3.9	7.1
	IP-CRM	selection(%)	0.0	0.1	29.6	30.7	39.6	**0.0**
		# of patients	0.0	3.3	9.2	7.2	1.2	0.0
	B-CRM	selection(%)	0.0	0.0	1.0	14.3	22.7	**62.0**
		# of patients	0.0	3.4	3.8	4.6	4.5	4.7

CRM, continual reassessment method; IP-CRM, informative prior continual reassessment method; B-CRM, bridging continual reassessment method.

4. Application

A multi-center phase I study was recently conducted to find the MTD of BKM120 in adult patients with advanced solid tumors [33]. BKM120 is a potent, highly specific oral inhibitor of the intracellular phosphatidylinositol-3-kinase (PI3K) pathway, which regulates cellular functions such as cell proliferation, growth, survival, and apoptosis. Selective inhibition of the PI3K pathway provides a promising therapeutic approach to treat cancer. A total of six doses were investigated, that is, 12.5, 25, 50, 80, 100,

Table III. The number of DLTs at six doses in the phase I trial of BKM120 for patients with advanced solid tumors.

	Dose (mg)					
	12.5	25	50	80	100	150
Number of patients	1	2	5	6	17	4
Number of DLTs	0	0	0	1	4	2

DLT, dose-limiting toxicity.

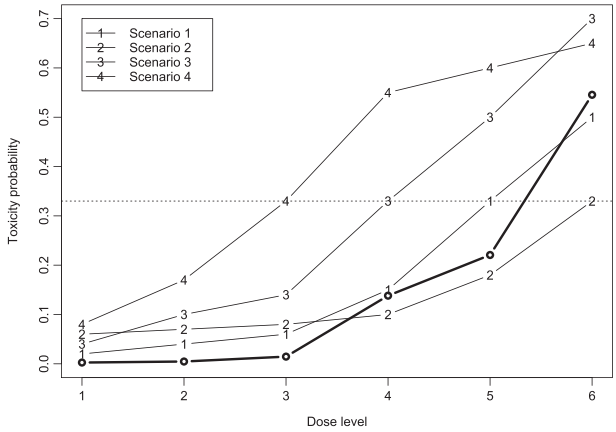

Figure 1. Four dose-toxicity curves for the new population and the estimate of the dose-toxicity curve for the landmark population (represented by the thick line). The horizontal dotted line indicates the target toxicity probability.

or 150 mg. DLTs were evaluated during the first treatment cycle (28 days). The main DLTs were defined as any grade 3 or higher hematologic or nonhematologic toxicity according to the common terminology criteria for adverse events version 3.0. The MTD was defined as the highest dose of BKM120 yielding a DLT rate not higher than 33%. A total of 35 patients were treated in the trial. The resulting dose-toxicity data are shown in Table III, with a dose of 100 mg selected as the MTD. As the patients in this trial came entirely from the USA, Canada, the Netherlands, and Spain, the identified MTD may not be applicable to Asian populations. Our collaborators at the Fourth Military Medical University in China, are interested in conducting a follow-up phase I bridge trial to establish the MTD of BKM120 in Chinese patients.

We applied the proposed B-CRM to design the follow-up trial, in which we considered the same six doses that were evaluated in the landmark trial. The maximum sample size was 24 patients. Based on the dose-toxicity data from the landmark trial (Table III), we estimated the toxicity probabilities of the doses using the proposed mixture estimator, yielding $(\hat{\pi}_1, \ldots, \hat{\pi}_6) = (0.002, 0.004, 0.014, 0.137, 0.220, 0.546)$. Accordingly, we constructed three skeletons:

$$\begin{aligned}
\text{skeleton}\ \ 1: &\ \left(p_1, \ldots, p_6\right) = (0.002, 0.004, 0.014, 0.137, 0.220, 0.546)\\
\text{skeleton}\ \ 2: &\ \left(p_1, \ldots, p_6\right) = (0.004, 0.014, 0.137, 0.220, 0.546, 0.773)\\
\text{skeleton}\ \ 3: &\ \left(p_1, \ldots, p_6\right) = (0.001, 0.002, 0.004, 0.014, 0.137, 0.220),
\end{aligned}$$

where skeletons 1–3 represent that the MTD for Chinese patients is a priori the same as one level lower or one level higher than that for non-Asian patients. To evaluate the operating characteristics of the B-CRM for this trial, we considered four scenarios that differ in both the location of the MTD and the shape of the true dose-toxicity curve for the follow-up trial (Figure 1). We simulated 1000 trials under each scenario. For the purpose of comparison, we also applied the CRM and IP-CRM. The results (Table IV) show that compared to the CRM and IP-CRM, the proposed B-CRM has the most reliable operating characteristics. It selected the true MTD consistently with high probabilities (62.5–68.5%) and assigned the majority of the patients (i.e., about 10 or more) to the MTD. The CRM and IP-CRM performed well in some scenarios (e.g., scenario 2 for the CRM and scenario 4 for the IP-CRM) but worse in other scenarios (e.g., scenarios 1 and 4 for the CRM and scenario 2 for the IP-CRM).

Table IV. Operating characteristics of the B-CRM, CRM and IP-CRM for the phase I BKM120 dose-finding trial. The underlined dose is the target dose.

			Dose (mg)					
Scenario	Method		12.5	25	50	80	100	150
1		Pr(toxicity)	0.02	0.04	0.06	0.15	0.33	0.50
	CRM	selection (%)	0.0	0.0	0.5	17.6	**56.1**	25.8
		# patients	0.0	0.1	1.8	7.2	8.9	6.0
	IP-CRM	selection (%)	0.0	0.0	2.2	27.9	**61.0**	8.9
		# patients	3.0	3.0	3.9	6.2	6.7	1.2
	B-CRM	selection (%)	0.0	0.0	0.1	18.0	**65.3**	16.6
		# patients	0.0	0.0	0.8	6.0	11.9	5.2
2		Pr(toxicity)	0.06	0.07	0.08	0.10	0.18	0.33
	CRM	selection (%)	0.0	0.0	0.1	1.7	22.9	**75.3**
		# patients	0.0	0.0	1.1	4.4	5.9	12.5
	IP-CRM	selection (%)	0.0	0.3	7.8	14.7	50.4	**26.8**
		# patients	3.1	3.3	4.9	5.0	5.5	2.3
	B-CRM	selection (%)	0.0	0.0	0.1	2.5	28.9	**68.5**
		# patients	0.0	0.0	0.4	2.6	7.5	13.4
3		Pr(toxicity)	0.04	0.10	0.14	0.33	0.50	0.70
	CRM	selection (%)	0.0	0.6	17.0	**58.7**	23.0	0.7
		# patients	0.1	0.9	6.1	11.0	5.0	1.0
	IP-CRM	selection (%)	0.2	0.5	25.8	**59.2**	14.2	0.1
		# patients	3.1	3.5	6.8	8.0	2.5	0.1
	B-CRM	selection (%)	0.0	0.3	15.2	**62.5**	21.8	0.2
		# patients	0.1	0.5	4.6	12.0	5.9	0.9
4		Pr(toxicity)	0.08	0.17	0.33	0.55	0.60	0.65
	CRM	selection (%)	1.4	23.5	**56.8**	16.4	0.6	0.0
		# patients	0.9	5.5	10.3	6.1	0.8	0.2
	IP-CRM	selection (%)	1.9	14.5	**71.4**	12.0	0.2	0.0
		# patients	3.5	5.7	11.1	3.4	0.3	0.0
	B-CRM	selection (%)	0.6	16.1	**64.5**	15.2	2.9	0.5
		# patients	0.9	4.1	10.8	5.6	1.9	0.7

5. Conclusion

We have proposed the B-CRM to find the MTD of a drug for a new ethnic population, given that a landmark trial has been conducted to establish the MTD in a landmark population. Our method borrows dose-toxicity information from the landmark trial and also accounts for inter-ethnic differences. We propose a novel mixture estimator to estimate the dose-toxicity curve using the data yielded by the landmark trial. Based on the resulting estimate, we form multiple skeletons to borrow information from the landmark trial and also accommodate the inter-ethnic heterogeneity. We use the Bayesian model averaging approach to make inference across multiple skeletons and make the decision of dose assignment. Simulation studies show that the proposed method yields higher MTD selection percentages and also assigns more patients to the MTD than the conventional dose-finding method, which does not borrow information across trials.

This article focuses on ethnic heterogeneity, but the proposed method can be used to handle other types of patient heterogeneity. For example, based on certain prognostic factors or biomarkers, patients often can be divided into several subgroups that have different levels of sensitivity to a drug. In this case, we can first use the conventional method to find the MTD in one subgroup and then employ the proposed B-CRM to find the MTD in other subgroups. In some situations, certain modifications are needed for the proposed method to accommodate different types of prior information. For example, suppose a drug has been tested in adults, but we are interested in finding the MTD of that drug for children. Because children are typically more susceptible to toxicity, and the MTD for children is most likely lower than that for adults, we may want to modify our three elicited skeletons such that the MTDs of the three skeletons are the same as one level lower and two levels lower than the MTD of the landmark population. That is, we

replace the skeleton that is one level higher with a skeleton that is two levels lower than the MTD of the landmark population.

In the proposed B-CRM, we use the point estimate (i.e., posterior mean) of toxicity probability to determine the dose escalation and deescalation (i.e., step 2 of the dose-finding algorithm described in Section 2.4). Ishizuka and Ohashi [34] and Neuenschwander *et al.* [35] pointed out that the use of the point estimate may cause the tendency of aggressively allocating patients to toxic doses, at least under the logistic dose-toxicity model. To address this issue, Neuenschwander *et al.* [35] proposed to divide the toxicity probability into four intervals (i.e., under-dosing, targeted toxicity, excessive toxicity, and unacceptable toxicity) and used the posterior probabilities of these intervals for the decision making in the CRM. The same strategy can be readily adopted here to enhance the performance of the B-CRM.

Appendix A: Proof of Theorem 1

Let π_j denotes the true toxicity probability of dose j. Because the nonparametric estimate is generally consistent (Bhattacharya and Kong, 2007), that is, $\hat{\pi}_j^{(NP)} \rightarrow \pi_j$, the consistency of $\hat{\pi}$ depends on the property of the parametric estimate $\hat{\pi}_j^{(P)}$. When the probit model is correctly specified, $\hat{\pi}_j^{(P)}$ is a consistent estimate of π_j. Therefore, $\hat{\pi}_j$ is consistent because

$$\hat{\pi}_j = w_i\hat{\pi}_j^{(P)} + (1 - w_i)\hat{\pi}_j^{(NP)} \rightarrow w_i\pi_j + (1 - w_j)\pi_j = \pi_j.$$

We now show that $\hat{\pi}_j$ is still consistent when the probit model is misspecified. In this case, the parametric estimate $\hat{\pi}_j^{(P)}$ is generally not consistent with $\hat{\pi}_j^{(P)} \rightarrow \pi_j^*$, where π_j^* is a constant not equal to π_j. Define y_{ij} as the binary toxicity indicator for the ith subject treated at dose j. The likelihood ratio of the two binomial distributions, as shown in (4), can be rewritten using the Bernoulli density as follows:

$$\lambda_j = \frac{\prod_{i=1}^{m_j} f\left(y_{ij}|\hat{\pi}_j^{(P)}\right)}{\prod_{i=1}^{m_j} f\left(y_{ij}|\hat{\pi}_j^{(NP)}\right)},$$

where $f\left(y_{ij}|\hat{\pi}_j^{(P)}\right) = \left(\hat{\pi}_j^{(P)}\right)^{y_{ij}}\left(1 - \hat{\pi}_j^{(P)}\right)^{1-y_{ij}}$ and $f\left(y_{ij}|\hat{\pi}_j^{(NP)}\right) = \left(\hat{\pi}_j^{(NP)}\right)^{y_{ij}}\left(1 - \hat{\pi}_j^{(NP)}\right)^{1-y_{ij}}$. Thus,

$$\log(\lambda_j) = \sum_{i=1}^{m_j} \log \frac{f\left(y_{ij}|\hat{\pi}_j^{(P)}\right)}{f\left(y_{ij}|\hat{\pi}_j^{(NP)}\right)},$$

where each term in the summation has a mean of

$$\begin{aligned} E\left(\log\left(\frac{f\left(y_{ij}|\hat{\pi}_j^{(P)}\right)}{f\left(y_{ij}|\hat{\pi}_j^{(NP)}\right)}\right)\right) &\rightarrow E\left(\log\left(\frac{f\left(y_{ij}|\pi_j^*\right)}{f(y_{ij}|\pi_j)}\right)\right) \\ &= E\left(\log\left(\frac{f(y_{ij})}{f(y_{ij}|\pi_j)}\right) - \log\left(\frac{f(y_{ij})}{f\left(y_{ij}|\pi_j^*\right)}\right)\right) \\ &= H(\pi_j) - H\left(\pi_j^*\right), \end{aligned}$$

with $H(\theta)$ denoting the Kullback–Leibler information of the form

$$\begin{aligned} H(\theta) &= E\left(\log\left(\frac{f(y_{ij})}{f(y_{ij}|\theta)}\right)\right) \\ &= \int \log\left(\frac{f(y_{ij})}{f(y_{ij}|\theta)}\right) f(y_{ij})\, \mathrm{d}y_{ij} \end{aligned}$$

Based on Jensen's inequality, $H(\theta)$ is minimized at the true parameter value, $\theta = \pi_j$, with a minimum value of 0. Therefore,

$$E\left(\log\left(\frac{f\left(y_{ij}|\hat{\pi}_j^{(P)}\right)}{f\left(y_{ij}|\hat{\pi}_j^{(NP)}\right)}\right)\right) < 0.$$

That is, $\log(\lambda_j)$ is the sum of m_j independent identically distributed (iid) random variables with negative mean. By the law of large numbers, $\log(\lambda_j) \to -\infty$ as $m_j \to \infty$. As a result, $w_j = \lambda_j/(1+\lambda_j) \to 0$. So we have

$$\hat{\pi}_j = w_j\hat{\pi}_j^{(P)} + (1 - w_j)\hat{\pi}_j^{(NP)} \to 0\pi_j^* + (1 - 0)\pi_j = \pi_j.$$

Acknowledgements

The authors thank the associate editor and two reviewers for very insightful and constructive comments that substantially improved the article. Pan's research was partially supported by Research Grant 81302513 from the National Science Foundation of China and China Postdoctoral Science Foundation funded project of 2014M562601. Xia's research was partially supported by Research Grant 81273176 from the National Science Foundation of China. Yuan's research was partially supported by grants CA154591, CA016672, and 5P50CA098258 from the National Cancer Institute.

References

1. Storer BE. Design and analysis of phase I clinical trials. *Biometrics* 1989; **45**:925–937.
2. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics* 1990; **46**:33–48.
3. Whitehead J, Brunier H. Bayesian decision procedures for dose determining experiments. *Statistics in Medicine* 1995; **14**:885–893.
4. Durham SD, Flournoy N, Rosenberger WF. A random walk rule for phase I clinical trials. *Biometrics* 1997; **53**:745–760.
5. Babb J, Rogatko A, Zacks S. Cancer phase I clinical trials: efficient dose escalation with overdose control. *Statistics in Medicine* 1998; **17**:1103–1120.
6. Liu S, Yuan Y. Bayesian optimal interval designs for phase I clinical trials. *Journal of the Royal Statistical Society: Series C* 2014. DOI: 10.1111/rssc.12089 (Early View).
7. Chevret S. *Statistical Methods for Dose-Finding Experiments*. John Wiley and Sons Ltd: England, 2006.
8. Ting N. *Dose Finding in Drug Development*. Springer: Cambridge, MA, 2006.
9. Huang S, Temple R. Is this the drug or dose for you? Impact and consideration of ethnic factors in global drug development, regulatory review, and clinical practice. *Nature* 2008; **84**:287–294.
10. Yasuda SU, Zhang L, Huang S. The role of ethnicity in variability in response to drugs: focus on clinical pharmacology studies. *Nature* 2008; **84**:417–423.
11. Barrera K, Imagawa DK, Yamamoto M, Katz M, Sanati H, Nguyen A, Nguyen L, Ho J, Gebhardt K, Hoang T. Sorafenib for treatment of hepatocellular carcinoma in the Asian–American population: Does one size fit all. *2009 ASCO Gastrointestinal Cancers Symposium (Abstract #198)*, 2009.
12. Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc JF, de Oliveira AC, Santoro A, Raoul JL, Forner A, Schwartz M, Porta C, Zeuzem S, Bolondi L, Greten TF, Galle PR, Seitz JF, Borbath I, Häussinger D, Giannaris T, Shan M, Moscovici M, Voliotis D, Bruix J; SHARP Investigators Study Group. Sorafenib in advanced hepatocellular carcinoma. *The New England Journal of Medicine* 2008; **359**:378–390.
13. Joensuu H, Kellokumpu-Lehtinen PL, Bono P, Alanko T, Kataja V, Asola R, Utriainen T, Kokko R, Hemminki A, Tarkkanen M, Turpeenniemi-Hujanen T, Jyrkkiö S, Flander M, Helle L, Ingalsuo S, Johansson K, Jääskeläinen AS, Pajunen M, Rauhala M, Kaleva-Kerola J, Salminen T, Leinonen M, Elomaa I, Isola J; FinHer Study Investigators. Adjuvant docetaxel or vinorelbine with or without trastuzumab for breast cancer. *New England Journal of Medicine* 2006; **354**:809–820.
14. Perng RP, Shih JF, Chen YM, Chou KC, Lee YC, Tsai CM. A phase II study of single agent docetaxel chemotherapy for non-small cell lung cancer. *Japanese Journal of Clinical Oncology* 2000; **30**:429–434.
15. Maruyama R, Nishiwaki Y, Tamura T, Yamamoto N, Tsuboi M, Nakagawa K, Shinkai T, Negoro S, Imamura F, Eguchi K, Takeda K, Inoue A, Tomii K, Harada M, Masuda N, Jiang H, Itoh Y, Ichinose Y, Saijo N, Fukuoka M. Phase III study, V-15-32, of gefitinib versus docetaxel in previously treated Japanese patients with non-small-cell lung cancer. *Journal of Clinical Oncology* 2008; **26**:4244–4252.
16. Bruno R, Hille D, Riva A, Vivier N, ten Bokkel Huinnink WW, van Oosterom AT, Kaye SB, Verweij J, Fossella FV, Valero V, Rigas JR, Seidman AD, Chevallier B, Fumoleau P, Burris HA, Ravdin PM, Sheiner LB. Population pharmacokinetics/pharmacodynamics of docetaxel in phase II studies in patients with cancer. *Journal of Clinical Oncology* 1998; **16**:187–196.

17. Goh BC, Lee SC, Wang LZ, Fan L, Guo JY, Lamba J, Schuetz E, Lim R, Lim HL, Ong AB, Lee HS. Explaining interindividual variability of docetaxel pharmacokinetics and pharmacodynamics in Asians through phenotyping and genotyping strategies. *Journal of Clinical Oncology* 2002; **20**:3683–3690.
18. Shih WJ. Clinical trials for drug registration in Asian-Pacific countries: proposal for a new paradigm from a statistical perspective. *Controlled Clinical Trials* 2001; **22**:357–366.
19. Lan GKK, Soo YW, Siu CT, Wang M. The use of weighted Z-tests in medical research. *Journal of Biopharmaceutical Statistics* 2005; **15**:625–639.
20. Gould AL, Jin T, Zhang L, Wang W. A predictive Bayesian approach to the design and analysis of bridging studies. *Journal of Biopharmaceutical Statistics* 2012; **22**:916–934.
21. Gandhi M, Mukherjee B, Biswas D. A Bayesian approach for inference from a bridging study with binary outcomes. *Journal of Biopharmaceutical Statistics* 2012; **22**:935–951.
22. Chow S, Chieh C, Liu J, Hsiao C. Statistical methods for bridging studies. *Journal of Biopharmaceutical Statistics* 2012; **22**:903–915.
23. Morita S. Application of the continual reassessment method to a phase I dose-finding trial in Japanese patients: east meets west. *Statistics in Medicine* 2010; **30**:2090–2097.
24. Agresti A. *Categorical Data Analysis* 2nd edition. Wiley: New York, 2002.
25. Ayer M, Brunk HD, Ewing GM, Reid WT, Silverman E. An empirical distribution function for sampling with incomplete information. *The Annals of Mathematical Statistics* 1955; **26**:641–647.
26. Bhattacharya R, Kong M. Consistency and asymptotic normality of the estimated effective doses in bioassay. *Journal of Statistical Planning and Inference* 2007; **137**:643–658.
27. Barlow RE, Bartholomew DJ, Bremner JM, Brunk HD. *Statistical Inference under Order Restrictions: The Theory and Application of Isotonic Regression*. Wiley: New York, 1972.
28. Yin G, Yuan Y. Bayesian model averaging continual reassessment method in phase I clinical trials. *Journal of the American Statistical Association* 2009; **104**:954–968.
29. Madigan D, Raftery AE. Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association* 1994; **89**:1535–1546.
30. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial. *Statistical Science* 1999; **14**: 382–401.
31. Garrett-Mayer E. The continual reassessment method for dose-finding studies: a tutorial. *Clinical Trials* 2006; **3**:57–71.
32. Lee S, Cheung Y. Model calibration in the continual reassessment method. *Clinical Trials* 2009; **6**:227–238.
33. Bendell JC, Rodon J, Burris HA, Jonge M, Verweij J, Birle D, Demanse D, Buck SS, Ru QC, Peters M, Goldbrunner M, Baselga J. Phase I, dose-escalation study of BKM120, an oral pan-class I PI3K inhibitor, in patients with advanced solid tumors. *Journal Of Clinical Oncology* 2012; **30**:282–290.
34. Ishizuka N, Ohashi Y. The continual reassessment method and its applications: a Bayesian methodology for phase I cancer clinical trials. *Statistics in Medicine* 2001; **20**:2661–2681.
35. Neuenschwander B, Branson M, Gsponer T. Critical aspects of the Bayesian approach to phase I cancer trials. *Statistics in Medicine* 2008; **27**:2420–2439.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.