

Bayesian Two-Stage Biomarker-Based Adaptive Design for Targeted Therapy Development

Xuemin Gu · Nan Chen · Caimiao Wei ·
Suyu Liu · Vassiliki A. Papadimitrakopoulou ·
Roy S. Herbst · J. Jack Lee

Received: 21 March 2012 / Accepted: 18 November 2014 / Published online: 4 December 2014
© International Chinese Statistical Association 2014

Abstract We propose a Bayesian two-stage biomarker-based adaptive randomization (AR) design for the development of targeted agents. The design has three main goals: (1) to test the treatment efficacy, (2) to identify prognostic and predictive markers for

Electronic supplementary material The online version of this article (doi:[10.1007/s12561-014-9124-2](https://doi.org/10.1007/s12561-014-9124-2)) contains supplementary material, which is available to authorized users.

X. Gu · N. Chen · C. Wei · S. Liu · J. J. Lee (✉)
Department of Biostatistics, The University of Texas MD Anderson Cancer Center,
1515 Holcombe Boulevard, Houston, TX 77030, USA
e-mail: jjlee@mdanderson.org

X. Gu
e-mail: xuemin.gu@bms.com

N. Chen
e-mail: nchen2@mdanderson.org

C. Wei
e-mail: caiwei@mdanderson.org

S. Liu
e-mail: syliu@mdanderson.org

Present address:

X. Gu
3551 Lawrenceville Road, Princeton, NJ 08540-4715, USA

V. A. Papadimitrakopoulou
Department of Thoracic, Head & Neck Medical Oncology, The University of Texas MD Anderson
Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA
e-mail: vpapadim@mdanderson.org

R. S. Herbst
Department of Internal Medicine, Yale School of Medicine, 333 Cedar Street, WWW221,
New Haven, CT 06520, USA
e-mail: roy.herbst@yale.edu

the targeted agents, and (3) to provide better treatment for patients enrolled in the trial. To treat patients better, both stages are guided by the Bayesian AR based on the individual patient's biomarker profiles. The AR in the first stage is based on a known marker. A Go/No-Go decision can be made in the first stage by testing the overall treatment effects. If a Go decision is made at the end of the first stage, a two-step Bayesian lasso strategy will be implemented to select additional prognostic or predictive biomarkers to refine the AR in the second stage. We use simulations to demonstrate the good operating characteristics of the design, including the control of per-comparison type I and type II errors, high probability in selecting important markers, and treating more patients with more effective treatments. Bayesian adaptive designs allow for continuous learning. The designs are particularly suitable for the development of multiple targeted agents in the quest of personalized medicine. By estimating treatment effects and identifying relevant biomarkers, the information acquired from the interim data can be used to guide the choice of treatment for each individual patient enrolled in the trial in real time to achieve a better outcome. The design is being implemented in the BATTLE-2 trial in lung cancer at the MD Anderson Cancer Center.

Keywords Adaptive design · Outcome-adaptive randomization · Bayesian Lasso · Predictive and prognostic biomarkers · Personalized medicine · Targeted therapy · Variable selection

1 Introduction

Clinical drug development is a long and costly process. One generally accepted estimation for the pre-approval cost of developing a new drug is 800 million US dollars [1]. When considering the expense associated with a product that fails to reach market, the cost major pharmaceutical companies must absorb per successful new drug launch may increase to 4–11 billion US dollars [2]. It often takes an average of 10 years in clinical development to obtain the approval of a new drug [3]; and the time is even longer if the duration of the preclinical work is included. Another problem in drug development is that the failure rate is extremely high. The success rate from first-in-man use to official registration is about 11 % across different disease areas [4]. For oncology, the success rate is even lower, at about 5 %. Furthermore, even approved drugs do not work for all patients. For the drugs that are already on the market, on average, only half of the people achieve a favorable response from whatever drugs they are taking [5]. These facts demonstrate the tough realities of the field of drug development. They also illustrate that the traditional methods for drug development do not work well and that we desperately need a new paradigm for improving the success rate of developing safe and effective drugs.

In recent decades, advances in biomedicine and genomics have fueled the development of targeted agents. Better knowledge of the mechanisms and biology underlying a disease has led to the development of agents targeted at the root cause of the disease, with the hope of offering more effective and less toxic therapies. Targeted agents, however, do not work for everyone and some may not work at all. Hence, the statistical challenges in the development of targeted agents requires not only the evaluation of

treatment efficacy but also the identification of prognostic (non-treatment specific) and predictive (treatment specific) biomarkers [6–9]. Typically, the overall treatment effect for targeted agents in the unselected overall patient population is low. But for patients who express a predictive marker (have a positive marker status), the matched targeted agent may offer high efficacy and result in low toxicity. Hence, the co-development of the targeted agents and their companion diagnostic markers is essential for the success of contemporary drug development [10–12]. In addition, upon the identification of each patient's marker profile, it is desirable to treat patients accordingly with the best available treatments in the clinical trial. These efforts make up the essential steps to achieving personalized medicine.

The targeted therapy development approach is especially suitable for finding effective treatments for cancer, an inherently heterogeneous disease [13]. Traditional cytotoxic chemotherapeutic agents kill both cancer cells and normal cells indiscriminately and have limited success [14]. In contrast, targeted agents directly tackle the abnormal biological pathways involved in the life cycle of cancer cells [15, 16]. One ultimate goal of targeted therapy is to provide customized treatment to individual patients based on their molecular profiles. Amidst all the current challenges, successful targeted agent development is well illustrated by examples such as the use of imatinib in treating BCR-ABL positive chronic myeloid leukemia, trastuzumab in treating HER-2 expressed breast cancer, and erlotinib in treating *EGFR* mutated lung cancer. In addition, the U.S. Food and Drug Administration approved two targeted agents with companion diagnostics in 2011, namely, vemurafenib in V600E *BRAF* mutated melanoma, and crizotinib in *EML-4-ALK* translocated lung cancer [17–21]. Although these successful examples are more exceptions than norms in oncology drug development, they unequivocally validate the concept of targeted agent development and pave the way for future discovery.

In our quest to find effective cancer therapies, we have designed, conducted, and completed a Bayesian adaptive randomization trial called Biomarker-Integrated Approaches of Targeted Therapy of Lung Cancer Elimination (BATTLE) [22, 23]. The trial involved testing four targeted agents with eleven pre-selected putative biomarkers. These markers were used to group patients into five marker groups. A hierarchical Bayes model was applied to model the primary endpoint of an 8-week disease control rate and to help assign patients to better treatments within each biomarker group. The trial was the first adaptive randomization trial in recurrent lung cancer with a mandatory pre-treatment biopsy requirement for biomarker analysis. We have demonstrated the feasibility of such a design and have identified several interesting findings to be validated in future trials [23].

One major limitation of the original BATTLE trial (BATTLE-1) was the pre-selection of biomarkers and bundling biomarkers into marker groups. It turned out that some of the selected markers had no predictive or prognostic values and the grouping of markers weakened their association with the treatment outcome. To rectify this problem, we have designed a Bayesian two-stage biomarker-based adaptive randomization (AR) design called BATTLE-2. The design has three main goals: (1) to test the treatment efficacy of a drug and drug combinations, (2) to identify prognostic and predictive markers for the targeted agents, and (3) to provide continually better treatment for patients enrolled in the trial. The relevant biomarkers are identified in

the first stage of the trial and validated in the second stage. Bayesian AR is applied in both stages to assign more patients to better treatments based on the individual patient's biomarker profiles. AR in the first stage is based on a known biomarker. A Go or No-Go decision can be made in the first stage by testing the overall treatment effects. If none of the experimental treatments show any promise in the overall and any marker groups, a No-Go decision is made and the trial can be stopped early. On the other hand, if a Go decision is made at the end of the first stage, a two-step Bayesian lasso strategy is implemented to select additional prognostic or predictive biomarkers based on the first stage data and other available information. Additional patients are enrolled in the second stage and the refined AR is applied to assign patients to treatments. Treatment effects, marker effects, and their interactions are estimated and tested using data acquired in both stages. Simulations are used to thoroughly study the operating characteristics of the design, including the control of per-comparison type I and type II errors, the probabilities of selecting important markers and noise markers, and the proportions of patient assigned to the different treatments.

Thanks to advancements in computational algorithms and computer speed, Bayesian methods have been invigorated in clinical trials over the past two decades. Compared to its frequentist counterpart, the Bayesian framework offers several attractive features in clinical trial designs [24,25]. One unique feature of the Bayesian method is that it is adaptive in nature. Bayesian adaptive design allows for continuous learning. The design is particularly suitable for the development of multiple targeted agents because much is unknown at the beginning of the trial, but information accumulates as the trial moves along [26]. By estimating the treatment effects, marker effects, and their interactions from the interim data, we can identify efficacious treatments and relevant biomarkers. As a result, we can apply response- or outcome-adaptive randomization to guide the choice of treatment for each individual patient enrolled in the trial in real time to achieve a better outcome. Despite some disagreement regarding its usefulness, adaptive randomization can provide better treatments for individual patients in the trial to maximize the overall outcome in the trial; hence, AR enhances individual ethics while equal randomization (ER) emphasized the group ethics [27–33]. The BATTLE-1 trial in lung cancer [23] and the I-SPY2 trial in breast cancer [34] are good examples of the application of a Bayesian adaptive design. Other Bayesian adaptive designs can be found in the literature [35–37].

Building on our experience from the BATTLE-1 trial, we have made several major improvements in the succeeding BATTLE-2 trial to develop biomarker-based targeted therapy in lung cancer. Although we have not developed new statistical methodology, we consider the statistical design novel in the following areas: (1) the construction of a two-stage design for biomarker selection in stage 1 and biomarker validation in stage 2, (2) the application of the two-step Bayesian lasso with group lasso to select important markers first followed by the adaptive lasso to refine the selection by further evaluating the markers' prognostic or predictive effect, and (3) the utilization of the outcome-adaptive randomization to assign more patients to more effective treatments throughout the trial based on the real-time update of marker and treatment information under the Bayesian framework.

In this paper, using the BATTLE-2 trial as an example, we describe the Bayesian two-stage biomarker-based adaptive design in detail. The concepts can be applied to

develop targeted agents in other therapeutic areas as well. Section 2 gives an overview of the BATTLE-2 trial design. Section 3 provides the statistical models and approaches. Section 4 lays out the implementation plans. Section 5 outlines the set-up of the simulation studies. Section 6 shows the simulation results. Section 7 concludes with a discussion.

2 Overview of the BATTLE-2 Trial Design

The target population for the BATTLE-2 trial is patients with advanced stage, treatment-refractory, non-small cell lung cancer. The primary endpoint for the study is the 8-week disease control rate (DCR), which has been shown to be a good surrogate of overall survival based on the experience of SWOG [38] and our previous trial in this population, the BATTLE-1 [23]. The study schema is shown in Fig. 1 with four treatments. A total of 400 evaluable subjects with advanced stage non-small cell lung cancer will be enrolled over a 4-year period. With a conservative estimate that 10% of the subjects may have incomplete marker profiles due to limited numbers of tumor cells in biopsy samples, a total of 450 subjects will be enrolled. Patients with *EGFR* mutation or *EML4-ALK* translocation will be excluded as there are known effective treatments that target these aberrations. Patients with a prior history of having received erlotinib treatment will not be randomized into erlotinib-only arm. Treatment effects will be tested separately in the patient subgroups by whether or not patients had prior erlotinib exposure as well as in the overall patient groups.

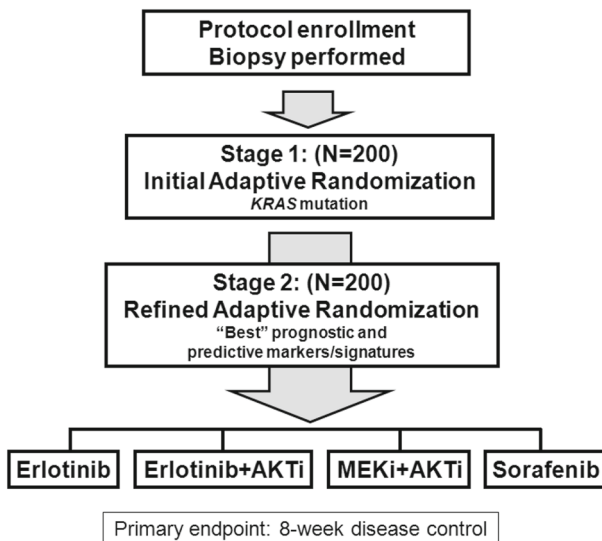


Fig. 1 Study schema for the BATTLE-2 trial. Four treatments will be studied: the EGFR inhibitor erlotinib (serving as a control group), erlotinib + an AKT-inhibitor (MK-2206), a MEK-inhibitor (AZD6244) + MK-2206, and sorafenib

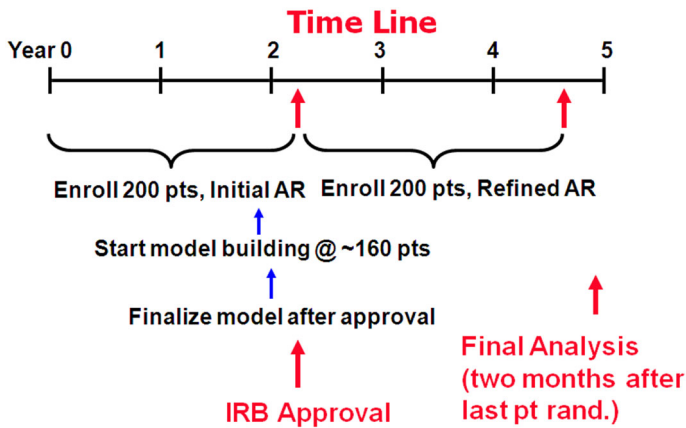


Fig. 2 Study timeline for the BATTLE-2 trial

As shown by the study timeline in Fig. 2, the study will be conducted in three defined steps:

- **Stage 1—Initial adaptive randomization (AR):** 200 patients will be adaptively randomized into one of the four treatment arms (three treatment arms for patients with a prior history of having received erlotinib) based on the available 8-week DCR data at the time and on the *KRAS* mutation status, which is generally accepted in the field as having relevance in predicting sensitivity or resistance to erlotinib and other targeted agents. An early futility stopping rule will be implemented after 70 patients are randomized. Starting from the 71th patient, the early stopping rule will be continuously evaluated until the end of the trial. If all the experimental arms (arms 2–4) do not show any improvement in efficacy over the control arm (arm 1) for the overall cohort and any of the biomarker groups (*KRAS* (–) or (+)) by prior erlotinib status, the trial will be stopped early. The treatment effect will also be estimated at the end of stage 1. If the trial is not stopped, it will proceed to stage 2.
- **Biomarker analysis at the end of stage 1—**Data from stage 1, as well as the patients' medical demographic variables, marker expressions, and treatment outcomes, supplemented by other up-to-date *in vitro* or *in vivo* data and information from the literature including the BATTLE-1 results, will be combined by the biostatistical and bioinformatics team to propose a refined predictive model. All biomarkers will be subject to a marker selection procedure. Examples of the putative markers are shown in Table 1. This biomarker analysis will select the best prognostic and/or predictive markers (either individual and/or composite markers) to be applied in stage 2.
- **Stage 2—Refined adaptive randomization:** Refined AR will be performed on the remaining 200 patients based on the model with all the markers selected from the previously mentioned analysis. Treatment effects, marker effects, and their interactions will be estimated and tested using data acquired in both stages.

Table 1 Molecular markers to be analyzed in stage 1 of the BATTLE-2 clinical trial using core needle biopsies

Markers	Tissue specimen	Biomarker
Established known markers	Diagnostic FFPE CNB	Mutation: <i>KRAS</i> (codons 12, 13 and 61 taken together)
High potential candidate markers	Research FFPE CNB	IHC protein: EGFR, p-AKT (Ser473), PTEN, HIF-1 α , and LKB1 Mutation: <i>PI3KCA</i> , <i>BRAF</i> , <i>AKT1</i> , <i>HRAS</i> , <i>NRAS</i> , <i>MAP2K1</i> (<i>MEK1</i>), <i>MET</i> , <i>CTNNB1</i> , <i>STK11</i> (<i>LKB1</i>) (<i>Sequenom</i> ®)
	Research Frozen CNB	Type of <i>KRAS</i> mutation (Cys/Val versus others), wild-type EGFR-erlotinib signature, epithelial–mesenchymal transition signature, sorefenib-sensitivity signature—Affymetrix
Discovery markers	Research Frozen CNB	mRNA profiling—Affymetrix miRNA expression (RT-qPCR): Hypoxia miRNA signature Protein profiling—RPPA
	Blood	EGFR-TKI SNPs

FFPE formalin-fixed and paraffin-embedded, *CNB* core needle biopsy, *IHC* immunohistochemistry, *FISH* fluorescence in situ hybridization, *p-AKT* phosphorylated-AKT, *RT-qPCR* reverse transcriptase-quantitative polymerase chain reaction, *RPPA* reverse phase protein array, *TKI* tyrosine kinase inhibitor, *SNP* single nucleotide polymorphism

3 Statistical Models and Approaches

The primary endpoint of the trial (the 8-week DCR) is modeled through a Bayesian logistic regression model. Bayesian outcome-adaptive randomization of patients is implemented in both stage 1 and stage 2 of the trial to treat patients better. The adaptive randomization of stage 1 is based on one known biomarker (denoted as M_1). At the end of stage 1, treatment effects in pre-defined patient populations are tested based on Bayesian posterior inferences. If a Go decision is made at the end of stage 1, a variable selection and model building procedure will be conducted to identify additional important biomarkers, which can be used in the adaptive randomization of patients in stage 2. Therefore, there are two different sets of Bayesian logistic regression: the first one using normal non-informative priors is for adaptive randomization and making Go/No-Go decisions, and the second one using shrinkage priors is for biomarker selection. All Bayesian computations are conducted in R by using BRugs, an R package (<http://www.r-project.org/>) based on OpenBugs (<http://www.openbugs.info/w/>).

3.1 Bayesian Models and Inferences on Treatment Effects and Marker Effects

A Bayesian logistic regression model is applied as the backbone of the statistical framework to model the probability of 8-week DCR, given treatments and markers. Without loss of generality, we assume M_1 is the pre-specified marker (*KRAS* mutation) and M_2 to M_K are the selected treatment-specific markers in the refined model. If we suppress the index of patients, the 8-week DCR, p_j , for patients on arm j can be written as follows,

$$\log \text{it}(p_j) = \mu_0 + \alpha_j T_j + \sum_{k=1}^K \beta_k M_k + \sum_{k=1}^K \gamma_{jk} T_k M_k + \left(\alpha'_j T_j Z + \sum_{k=1}^K \gamma'_{jk} T_k M_k Z \right) \quad (1)$$

where T_j is the treatment indicator for arm j , $j = 2, 3, 4$ (with arm 1 as the reference group), M_k is the marker status indicator for marker k , $k = 1, \dots, K$, and Z is the indicator for prior erlotinib treatment (i.e., erlotinib resistance). Therefore, μ_0 represents the mean 8-week DCR for patients on control treatment with negative markers, α is the treatment main effect, β is the biomarker main effect, and γ is the biomarker-treatment interaction term. In addition to the full model in Eq. (1), a reduced model is also used to make inference on the overall treatment effect while controlling for the main effect of biomarkers:

$$\log \text{it}(p_j) = \mu_0 + \alpha_j T_j + \sum_{k=1}^K \beta_k M_k + \alpha'_j T_j Z \quad (2)$$

The efficacy of the experimental treatments (arm 2, arm 3, or arm 4) will be compared with that of the erlotinib-only arm (arm 1). The treatment effects will be tested in two steps. First, we will test for the marginal treatment effect using the reduced model shown in Eq. (2). By comparing with the control arm, experimental treatment j ($j = 2, 3, 4$) will be claimed as having a significant marginal treatment effect if

$$\Pr(\alpha_j > 0) > \theta \quad (3)$$

in erlotinib-naïve patients, or if

$$\Pr(\alpha_j + \alpha'_j > 0) > \theta \quad (4)$$

in erlotinib-resistant patients, where θ is the threshold value for declaring statistical significance based on posterior inference. Next, we will test the treatment effect under the full model in Eq. (1), incorporating the treatment by marker interaction. A success will be claimed for treatment j ($j = 2, 3, 4$) in patients with a positive marker k tumor status if

$$\Pr(\alpha_j + \gamma_{jk} > 0) > \theta \quad (5)$$

for erlotinib-naïve patients, and if

$$\Pr(\alpha_j + \gamma_{jk} + \alpha'_j + \gamma'_{jk} > 0) > \theta \quad (6)$$

Table 2 True disease control rate by simple marker profiles

Marker profile	Marker indicator					Erlotinib-naïve $Z = 0$				Resistant $Z = 1$		
	M1	M2	M3	M4	M5	Arm1	Arm2	Arm3	Arm4	Arm2	Arm3	Arm4
(a) Null case												
0	0	0	0	0	0	0.3	0.3	0.3	0.3	0.2	0.2	0.2
1	1	0	0	0	0	0.1	0.1	0.1	0.1	0.01	0.01	0.01
2	0	1	0	0	0	0.3	0.3	0.3	0.3	0.2	0.2	0.2
3	0	0	1	0	0	0.3	0.3	0.3	0.3	0.2	0.2	0.2
4	0	0	0	1	0	0.3	0.3	0.3	0.3	0.2	0.2	0.2
5	0	0	0	0	1	0.3	0.3	0.3	0.3	0.2	0.2	0.2
(b) Alternative case												
0	0	0	0	0	0	0.3	0.3	0.3	0.3	0.2	0.2	0.2
1	1	0	0	0	0	0.1	0.5	0.8	0.6	0.4	0.7	0.5
2	0	1	0	0	0	0.6	0.6	0.6	0.6	0.5	0.5	0.5
3	0	0	1	0	0	0.3	0.8	0.3	0.3	0.7	0.2	0.2
4	0	0	0	1	0	0.3	0.3	0.8	0.3	0.2	0.7	0.2
5	0	0	0	0	1	0.3	0.3	0.3	0.8	0.2	0.2	0.7

Erlotinib-resistant patients are not assigned to arm 1

for erlotinib-resistant patients.

Therefore for each experimental arm, the treatment effects will be tested in erlotinib-naïve patients, erlotinib-resistant patients, *KRAS* wild-type erlotinib-naïve patients, *KRAS* mutant erlotinib-naïve patients, *KRAS* wild-type erlotinib-resistant patients, and *KRAS* mutant erlotinib-resistant patients. By definition, if any of the above hypothesis tests erroneously reject the true null hypothesis of no treatment effect, a per-comparison error has been made. We choose θ to control the per-comparison-wise error rate numerically for all hypothesis tests for each treatment. In particular, we first conduct simulations under the null case using different values of θ . Then, the value of θ that attains the pre-specified per-comparison-wise rate among all experimental arms will be chosen. The power in the alternative case can be calculated based on the same value of θ selected from the null case.

Table 2 lists the proposed true DCRs of each treatment arm for both erlotinib-naïve and erlotinib-resistant patients by marker profile. Notice that the DCRs in erlotinib-resistant patients are expected to be lower than those in erlotinib-naïve patients. A common control (erlotinib-only treatment in erlotinib-naïve patients) is used for testing treatment efficacy in both naïve and resistant patients. Therefore, the DCRs in erlotinib-resistant patients are compared to a higher bar, which makes our modeling approach more conservative.

As shown in Table 2, the null case is the complete-null case associated with weak familiar-wise error rate control for testing treatment efficacy. The per-comparison-wise error rate control used in our simulation addresses the individual null hypothesis error rate.

3.2 Variable Selection and Model Building

Based on our model parameterization, the prognostic effect will be characterized by a large marker main effect term (β_k), and the predictive effect will be characterized by a large marker–treatment interaction term (γ_{jk} or $\gamma_{jk} + \gamma'_{jk}$). A marker will be deemed important if it has either a prognostic effect or a predictive effect. To facilitate the identification of prognostic and predictive effects, we apply the least absolute shrinkage and selection operator (lasso) method and implement the marker selection procedure through a Bayesian two-step lasso method. The Laplace prior is used in the Bayesian lasso for covariates effects. The Laplace prior can shrink the covariate estimation toward zero, and penalize smaller covariate effects more. Therefore, the Bayesian lasso can result in good estimation and variable selection simultaneously [39]. Specifically, the first step of variable selection is a group selection aimed at the identification of markers with either prognostic effects or predictive effects. The coefficients β_k , γ_{jk} , and γ'_{jk} are grouped together for inclusion or exclusion. This step intends to screen out markers without any effects. The second step is an individual selection for marker and its interactions with the treatments. In order to obtain more consistent estimates for the model parameters, we implement the idea of adaptive lasso [40] in a Bayesian framework to provide finer variable selection and estimation in both the group selection step and the individual selection step. Upon the conclusion of variable selection, the refined model will be used in adaptively randomizing patients in the second stage of the trial.

In the first step of variable selection, we group the main effect of the marker and the marker–treatment interactions into a vector and apply the Bayesian group lasso method to select important markers [41, 42]. In particular, let

$$\eta_k = (\beta_k, \gamma_{2k}, \gamma_{3k}, \gamma_{4k}, \gamma'_{2k}, \gamma'_{3k}, \gamma'_{4k})$$

be the row vector of covariates corresponding to the marker main effect, and to the marker–treatment interaction in erlotinib-naïve and resistant patients, respectively, for marker k . Thus for group lasso, the original Laplace prior for η_k should be

$$\pi(\eta_k | \lambda) \propto \exp(-\lambda \eta_k) \quad (7)$$

where $\|\eta_k\| = (\beta_k^2 + \gamma_{2k}^2 + \gamma_{3k}^2 + \gamma_{4k}^2 + \gamma'_{2k}{}^2 + \gamma'_{3k}{}^2 + \gamma'_{4k}{}^2)^{1/2}$. Based on the idea borrowed from adaptive lasso, the actual prior used for the Bayesian group lasso is equivalent to

$$\pi(\eta_k | \lambda) \propto \exp(-\lambda \|\eta_k\| / \|\hat{\eta}_k\|) \quad (8)$$

Here $\|\hat{\eta}_k\| = (\hat{\beta}_k^2 + \hat{\gamma}_{2k}^2 + \hat{\gamma}_{3k}^2 + \hat{\gamma}_{4k}^2 + \hat{\gamma}'_{2k}{}^2 + \hat{\gamma}'_{3k}{}^2 + \hat{\gamma}'_{4k}{}^2)^{1/2}$, which is estimated from a ridge regression with penalty parameter being chosen by 20-fold cross-validation.

Since the prior in Eq. (8) is not available in OpenBugs, we used a scale mixture of normal to produce the shrinkage prior for Bayesian group lasso. Based on the formulation of Bayesian group lasso in Kyung et al. [41], the priors used for the group variable η_k in Eq. (8) were

$$\begin{cases} \eta_k \sim N_{m_k}(0, \tau_k^2 I_k) \\ \tau_k^2 \sim \text{InvGamma}\left(\frac{m_k+1}{2}, \frac{\lambda_k^2}{2}\right) \\ \lambda^2 \sim \text{Gamma}(a, b) \end{cases} \tag{9}$$

where m_k is the dimension of η_k , and hyperparameters a and b are, respectively, 1 and 10 in our simulation. As compared with the formulation in Kyung et al. [41], there are three major differences in our implementation: (1) $\lambda_k = \lambda / \hat{\eta}_k$, which results in different penalties for different groups of variables, and the group with larger initial estimations in ridge regression will experience smaller penalties; (2) since a logistic regression is used, the priors in Eqs. (8) and (9) are not conditional on the variance of error term; (3) to follow the notation in OpenBugs, τ_k^2 is the precision of normal distribution, and an inverse gamma distribution is used for τ_k^2 . The derivation from prior (9) to (8) follows the procedures in Appendix of Kyung et al. [41]. Because of the three differences mentioned above, the derivation details are given in the Appendix 1.

Letting $\tilde{\eta}_k$ be a posterior sample of random vector η_k , and

$$\tilde{\eta}_k = (\tilde{\beta}_k \ \tilde{\gamma}_{2k} \ \tilde{\gamma}_{3k} \ \tilde{\gamma}_{4k} \ \tilde{\gamma}'_{2k} \ \tilde{\gamma}'_{3k} \ \tilde{\gamma}'_{4k})$$

be the posterior mean of η_k , we can compute the distance between the posterior sample and the zero vector: $T_k = (\tilde{\eta}_k - 0)^T W_k^{-1} (\tilde{\eta}_k - 0)$, where W_k^{-1} is the sample variance–covariance matrix. Let \tilde{T}_q be the q th empirical quantile of T_k , then for a given q we select the k th marker if $\tilde{\eta}_k^T W_k^{-1} \tilde{\eta}_k > \tilde{T}_q$. In other words, the larger the distance between $\tilde{\eta}_k$ and 0, the more likely the k th marker will be selected. In our simulation, a 30% quantile of T_k was used for the group selection step of variable selection. All empirical distance measures need to be normalized by m_k when different marker groups have different dimensions.

Letting Ω be the set of markers selected in the first step, the prior distribution for $\{\theta_k : k \in \Omega\}$ in the adaptive lasso of the second step of variable selection is

$$\pi(\theta_k | \lambda) \propto \exp\left(-\lambda \frac{|\theta_k|}{|\hat{\theta}_k^{\text{LS}}|}\right) \tag{10}$$

where θ_k is a generic representation of either the marker main effect or the marker–treatment interaction, $|\hat{\theta}_k^{\text{LS}}|$ is the least squares estimation of the parameter without regularization, and λ^2 follows a gamma distribution. Since the prior in (10) is readily available in OpenBugs, the scale mixture of normal priors [39,41] was not used. A variable will be selected if the 80% empirical posterior credible interval does not cover zero. The selections of the credible interval in this second step and the \tilde{T}_q in the first step can be adjusted to achieve a desirable false-positive rate and true-positive rate of the variable selection in the null case and alternative case separately. Based on the heredity principle [43,44], if a marker–treatment interaction term is selected, the main effect of the marker will automatically be included in the AR model, even though the main effect may not be significant in the final analysis.

3.3 Adaptive Randomization and Futility Early Stopping Rules

We apply response AR throughout the trial to treat patients more effectively in the trial based on the cumulative data. Due to the delay in observing the outcome, AR will start to kick in after the first 14 patients are enrolled and the first disease control status is measured.

In stage 1, the initial AR is based on the logistic regression model in Eq. (1) with $K = 1$ (*KRAS* mutation). In particular, for a given patient on arm j , we have

$$\log \text{it}(p_j) = \mu_0 + \alpha_j T_j + \beta_1 M_1 + \gamma_{j1} T_j M_1 + \alpha'_j T_j Z \quad (11)$$

The probability of a patient being assigned to the j th arm is proportional to the square root of

$$\Pr(p_j > p'_j \mid j' \in \{1, 2, 3, 4\} \mid j' \neq j) \quad (12)$$

which corresponds to the posterior probability of the j th arm having the largest DCR among all treatments. A normal prior with mean zero is used for each of the parameters in Eq. (1), and a large variance of 10 is used in the non-informative priors for all the parameters in the model. To ensure that the AR allows patients to be assigned to all treatment arms with reasonable probabilities in stage 1, we set the lower and upper bounds of the initial allocation probabilities (i.e., corresponding to the square root of the probabilities of each arm being the best) to be 0.2 and 0.8 to guard against extreme allocations. If the allocation probability of a certain arm is less than 0.2 or greater than 0.8, the original value of the allocation probability will be replaced by 0.2 or 0.8, and the real allocation rate will be determined by the normalized value of allocation probability by the sum of all allocation probabilities. For example, let us assume the initial allocation probabilities are: 0.01, 0.05, 0.1, and 0.84; since the first 3 numbers are less than 0.2 and the last one is greater than 0.8, these numbers will be automatically change to: 0.2, 0.2, 0.2, and 0.8; subsequently, the final allocation probabilities will become: 0.14, 0.14, 0.14, and 0.58.

Similarly, AR in stage 2 is performed based on comparing the DCRs among the treatments. At the time of a new patient being randomized, DCRs are computed from Eq. (1) after incorporating all treatment, marker, and outcome information from all available data at the time. The randomization probability is similar to Eq. (8), but we do not take the square root transformation to allow more patients to be randomized into better arms. Note that in both stage 1 and stage 2, erlotinib-naïve patients can be randomized to one of the four arms; whereas erlotinib-resistant patients can be randomized to only one of the three experimental arms (arms 2, 3, or 4) by design. Similarly, we set the range of the initial allocation probabilities in stage 2 to be the same as in stage 1 to guard against extreme allocations.

During AR, we also allow the trial to be stopped early for futility if all the experimental arms fail in all patients and in all marker groups after at least 70 patients have been treated in the trial. In particular, the trial will be stopped if

$$\Pr(\alpha_j + \gamma_{jk} + \delta > 0) > \theta_s \quad (13)$$

and

$$\Pr(\alpha_j + \gamma_{jk} + \alpha'_j + \gamma'_{jk} + \delta > 0) > \theta_s \quad (14)$$

for all combinations of j ($j = 2, 3, 4$) and for all markers of index k . However, the trial will not be stopped as long as one of the experimental arms is promising in a sub-population defined by any one of the markers included in the AR model. In our simulation, δ is set at -0.442 , corresponding to a 0.1 improvement in the DCR from 0.3; whereas θ_s is set at 0.4 to yield reasonable early stopping probabilities under various scenarios.

Tuning the design parameters (e.g., θ , θ_s , 30 % quantile of T_k , and 80 % empirical posterior credible interval does not cover zero, etc.) for choosing the threshold for decision rules was a time-consuming and tedious task. During the process, we did not do an exhaustive search for all these tuning parameters. Instead, we took the try-and-error approach to run various simulations at a few sets of tuning parameters, and chose the one that met our pre-specified statistical objectives including: 10 % per-comparison error rate and 80 % power for testing the efficacy of each experimental therapy versus control, at least 50 % early stopping probability in stage 1 under the null hypothesis, and 80 % power with a 10 % per-comparison error rate for identifying important prognostic and predictive markers. The tuning of the early stopping probability was more difficult than other decision rules, because the early stopping probability has a huge impact on the per-comparison error rate and power of the study. Intuitively, if the trial was stopped more often for futility, mean sample size will be reduced. The per-comparison error rate will decrease but the power of the study may suffer because of a smaller sample size. After the probability threshold value for the early stopping decision rule was chosen, the tuning of decision rules for treatment effect was relatively easy. For variable selection, the choice of the 30 % quantile of T_k and 80 % empirical posterior credible interval not covering zero results in desirable operating characteristics of high probability in selecting important markers and low probability in selecting noise markers. The two parameters are related. We performed the sensitivity analysis by trying various combinations of the thresholds until we found the desirable cutoffs which met our design objectives.

4 Trial Implementation Plans

This design creates a biopsy-based and biomarker-driven trial. All patients will be consented for a mandated biopsy before being randomized to the study's treatment arms. Biomarkers in the patients' samples will be analyzed to guide the treatment choice.

To facilitate trial conduct, a BATTLE-2 web-based database application for registration, biomarker analysis, randomization, and follow-up will be built. All information collected from the patients will be entered into the BATTLE-2 database. The biopsy and biomarker analysis results will be entered into the database as well. The adaptive randomization module is integrated to the database application through web-based services. Follow-up information, including the 8-week disease control status, will be entered into the database to facilitate the AR. To monitor the timeliness of measuring

the 8-week disease control status and to ensure timely data entry, we have developed an e-mail notification system to alert research nurses to schedule a patient visit when 6 weeks have passed since a patient was randomized. The system also tracks the time of the 8-week endpoint being recorded. E-mail alerts will be sent out starting from the time when an endpoint evaluation is two weeks overdue.

In both stage 1 and stage 2, the corresponding model parameters will be continuously updated during the trial using available data on 8-week disease control status. A patient will be excluded from analysis if the 8-week disease control status is not yet observed. However, since the trial is conducted seamlessly, information will continue to accrue over time and become available to be used for assigning treatment for the next patient whenever the information becomes available. For both adaptive randomization and data analysis, Bayesian logistic regression with non-informative normal prior distributions is used to allow the observed data to have a major influence on the model. This is different from the Bayesian logistic regression used in variable selection, where shrinkage priors are used. Given the observed data, the posterior distributions of the parameters are calculated using the Markov Chain Monte Carlo (MCMC) method. Bayesian AR will be performed based on the updated posterior probability of the 8-week DCR from the underlying logistic regression model at the time each new patient is randomized.

Concurrently throughout stage 1, many putative candidate and discovery biomarkers will be analyzed in the patients' tissue and blood samples (Table 1). Those data, as well as patients' medical demographic variables, marker expressions, and treatment outcomes, supplemented by other up-to-date *in vitro* or *in vivo* data and information from the literature, will be combined by the biostatistical and bioinformatics team to propose a refined predictive model. We will identify the "best-performing" markers from stage 1 to form the predictive model for the AR of the second stage. The approach will involve a variable selection and model building step using Bayesian logistic regression with shrinkage priors. The final model for the stage 2 AR will be reviewed and approved by the Internal and External Advisory Boards. Upon approval, the 200 patients in stage 2 will be adaptively randomized based on the refined model, which could include more than one treatment-specific marker. Also, the marker used in stage 1 could be dropped in stage 2. At the end of the trial, information collected from all patients enrolled in stages 1 and 2 will be combined to test the prognostic effect and predictive effect of putative markers selected in the final model. In addition, because the previous trial (BATTLE-1) also contained an erlotinib-only arm, predictive marker findings from BATTLE-1 for erlotinib will be validated in patients enrolled in BATTLE-2. Other findings from BATTLE-2 will need to be validated in future trials.

Biomarker selection will be based on statistical strength, biological plausibility, and practical considerations. The following specific steps will be taken: (a) filter the initial discovery markers to narrow them down to a small number of putative discovery markers based on available *in vitro* and *in vivo* data, data derived from clinical specimens from BATTLE-1, and information from existing literature and public data resources; (b) elicit and specify the prior probability of "predictiveness" of all candidate and discovery biomarkers based on available *in vitro* and *in vivo* data, data derived from clinical specimens related to patient outcomes from BATTLE-1, and application of

a critical filter from the existing literature and data resources; (c) construct the data likelihood based on patients' biomarker and outcome data in stage 1 of BATTLE-2; and (d) compute a posterior probability of marker predictiveness from the statistical model, applying variable selection and model-building procedures as described above. If more than two markers are identified for a given treatment, a composite marker will be formed by principal component analysis to choose the first few principal components that explain at least 80% of the variability to reduce the dimensionality of the parameters, hence, increase the model robustness. Biomarkers that emerge from that procedure will be ranked by applying the guidelines of posterior probability of marker predictiveness, biologic plausibility, and feasibility of Clinical Laboratory Improvement Amendments certification [10–12]. Upon the final selection of markers, we will build the logistic regression model to be used in stage 2.

5 Set-Up for Simulation Studies

All markers are binary with 20% of the markers positive for M_1 (*KRAS* mutation), and 50% of the markers positive for M_2 to M_{15} . Based on our experience in the BATTLE-1 trial, we assume that 40% of the patients had prior erlotinib treatment. Patients with prior exposure to erlotinib will not be assigned to the erlotinib-only arm.

The true 8-week DCRs under the null and the alternative hypotheses that were assumed in the simulation studies are listed in Table 2. Under the null hypothesis, the 8-week DCRs for erlotinib-naïve patients are assumed to be 30 or 10%, depending on the *KRAS* mutation status. For erlotinib-resistant patients, the corresponding 8-week DCRs are 20 or 1%. Under the alternative hypothesis, when predictive markers are identified for each of the combination treatments, the 8-week DCR is assumed to be 80% in erlotinib-naïve patients and 70% in erlotinib-resistant patients.

We assume that the true model is the same as model (1) for 5 important markers (M_1 , M_2 , M_3 , M_4 , and M_5). The true parameters of the model are solved based on the true response rates in Table 2, and then patient responses can be simulated from the true models in different scenarios.

Both the null and alternative scenarios listed in Table 2 are evaluated via simulation studies. For each treatment arm, the treatment effects will be tested in erlotinib-naïve patients, erlotinib-resistant patients, *KRAS* wild-type erlotinib-naïve patients, *KRAS* mutant erlotinib-naïve patients, *KRAS* wild-type erlotinib-resistant patients, and *KRAS* mutant erlotinib-resistant patients by using the models in Eqs. (1) and (2). The rejection regions of hypothesis testing are selected numerically based on the null scenario to control the error rate of erroneous rejection of the true null hypotheses.

To demonstrate the improvement of BATTLE-2 design over BATTLE-1 design, we also simulated the clinical trial by using the design strategy of BATTLE-1 trial. The set-up for the BATTLE-1 strategy and the simulation results and are presented in Appendix 2 and Supplemental Tables 1 and 2. In addition to simulation results for BATTLE-1 design strategy, simulation results for equal randomization can also be found in the Supplemental Table 3. Bayesian inference is implemented in R by using a package called BRugs. The burn-in period for MCMC sampling is 1,000, and 25,000 data points are used in the posterior samples. Depending on the variable

Table 3 Power for testing treatment effect at the end of stage 1, early stopping probability during stage 1, and the mean sample sizes for stage 1

Treatment	Null ($\theta = 0.912$)						Overall treatment	Overall trial	Early stopping probability	Mean sample size
	Naïve (main)	Resist (main)	MI(-)	MI(+)	Resistant	MI(+)				
Arm 2	0.067	0.010	0.065	0.023	0.010	0.004	0.092	0.199	0.581	140
Arm 3	0.075	0.008	0.079	0.023	0.008	0.003	0.100			
Arm 4	0.063	0.011	0.062	0.023	0.012	0.001	0.088			
Alternative ($\theta = 0.912$)										
Arm 2	0.651	0.443	0.529	0.559	0.330	0.460	0.823	0.978	0.003	200
Arm 3	0.762	0.528	0.551	0.857	0.322	0.792	0.945			
Arm 4	0.668	0.456	0.527	0.662	0.315	0.567	0.842			

For each treatment, the treatment effect was first tested in erlotinib-naïve and erlotinib-resistant patients; and then treatment effects were tested by marker 1 status (KRAS mutation). Therefore, 6 hypothesis tests were conducted for each treatment. Columns 2–7 (left to right) in the table give the probabilities of rejecting the null hypothesis. The “Overall Treatment” column gives the probabilities of rejecting at least one of the six hypotheses for each experimental treatment. The value of θ used in hypothesis testing is 0.912, which means that the null hypothesis will be rejected if the 91.2 % quantile of the posterior distribution of the treatment effect does not include zero. The column “Overall Trial” gives the probabilities that at least one null hypothesis will be rejected for at least one treatment

selection results at the end of stage 1, the AR model for stage 2 may not be the same in different simulation runs. We choose the thinning parameter of the posterior sample to be twice the number of parameters in each model; and based on our observation the autocorrelation among the posterior samples of all the model parameters are small.

6 Simulation Results

We assume that the true model is the one given in Eq. (1) with the parameters calculated from the DCRs listed in Table 2. We carried out 2,000 simulation runs to study the operating characteristics of the design. The statistical power for testing the efficacy of the experimental treatments (arms 2, 3, and 4) versus the standard treatment (arm 1) at the end of stage 1 is given in Table 3. The powers to determine the treatment's main effect in erlotinib-naïve and erlotinib-resistant patients are shown in columns 2 and 3, respectively. The powers to determine the treatment effect in marker M1 subgroups are shown in columns 4–5 for erlotinib-naïve patients and in columns 6–7 for erlotinib-resistant patients, respectively. The powers for overall treatment effect are shown in column 8 and the power for overall trial effect (either arm 2, 3, or 4 is effective) is shown in column 9. Early stopping probabilities and mean sample sizes at the end of the first stage are shown in columns 10 and 11, respectively.

We chose a posterior probability cutoff value (θ) of 0.912 for declaring a positive treatment effect with the decision rules in (5) and (6). Using this cutoff, the per-comparison error rate in the null case is lower than 10% for each treatment. In our Bayesian modeling framework, a per-comparison error was an erroneous rejection of any one of the true null hypotheses of no treatment effect, and was defined by borrowing the idea from frequentist hypothesis testing. In the alternative case, with a maximum of 200 patients treated at the end of stage 1, we have 82, 95, and 84% power for testing the treatment effects for arms 2, 3, and 4, respectively. The overall power for the trial is 98% in the alternative case. The treatment-specific per-comparison error rate is controlled at or below 10% for treatments 2, 3, and 4, respectively, with an overall family-wise error rate of 20% in the null case. Although the null case was stopped early 58% of the time with an average sample size of 140, the alternative case was barely stopped early. Testing for the treatment effect at the end of stage 1 is important to make a “Go” or “No-Go” decision for determining whether the experimental treatments are warranted for further development or not. The statistical power for testing the treatment effect at the end of stage 2 will be sufficiently high after more patients have been enrolled; hence, it is not given here.

The marker selection probabilities for all simulation cases are shown in Table 4. A total of 15 markers were generated. Markers are selected by using the credible intervals of the posterior distribution of the model parameters. A 30% quantile value of the posterior distribution is used in the group selection. An 80% credible interval is used for the final marker selection, as described in Sect. 3.3. Markers with either a significant main effect (better treatment effect regardless of treatment arm) or significant interactions (associated with arms 2, 3, or 4) are defined as significant for the marginal marker effect. Under the null hypothesis, the probability of selection is 4% or lower in stage 1 and 3% or lower in stage 2 for all markers except marker 1

Table 4 Marker selection and final significance probabilities

	MK1	MK2	MK3	MK4	MK5	MK6	MK7	MK8	MK9	MK10	MK11	MK12	MK13	MK14	MK15
Null															
Stage 1															
Arm 1	0.275	0.015	0.020	0.010	0.015	0.014	0.012	0.011	0.011	0.013	0.016	0.015	0.012	0.014	0.016
Arm 2	0.005	0.010	0.012	0.010	0.009	0.008	0.006	0.011	0.005	0.010	0.007	0.012	0.010	0.008	0.007
Arm 3	0.004	0.012	0.007	0.010	0.008	0.005	0.006	0.009	0.008	0.006	0.008	0.005	0.010	0.012	0.008
Arm 4	0.004	0.009	0.008	0.009	0.013	0.005	0.010	0.010	0.007	0.006	0.010	0.007	0.008	0.008	0.006
Marginal sig.	0.280	0.038	0.038	0.032	0.038	0.029	0.027	0.034	0.026	0.030	0.033	0.036	0.033	0.035	0.031
Stage 2															
Arm 1	0.279	0.013	0.015	0.009	0.013	0.009	0.010	0.009	0.009	0.011	0.009	0.012	0.008	0.013	0.013
Arm 2	0.002	0.004	0.005	0.005	0.004	0.004	0.003	0.004	0.002	0.002	0.004	0.002	0.003	0.005	0.002
Arm 3	0.002	0.007	0.003	0.003	0.003	0.002	0.003	0.003	0.002	0.002	0.002	0.002	0.003	0.005	0.003
Arm 4	0.001	0.004	0.003	0.004	0.005	0.003	0.005	0.005	0.003	0.002	0.005	0.001	0.005	0.003	0.004
Marginal sig.	0.280	0.024	0.022	0.018	0.022	0.017	0.017	0.018	0.014	0.017	0.018	0.016	0.018	0.022	0.020
Alternative															
Stage 1															
Arm 1	0.266	0.789	0.094	0.096	0.091	0.065	0.059	0.060	0.057	0.051	0.059	0.067	0.052	0.058	0.063
Arm 2	0.082	0.132	0.873	0.066	0.076	0.038	0.031	0.031	0.040	0.032	0.037	0.033	0.032	0.028	0.038
Arm 3	0.501	0.129	0.070	0.849	0.069	0.033	0.033	0.037	0.029	0.032	0.039	0.035	0.029	0.035	0.035
Arm 4	0.168	0.122	0.077	0.073	0.871	0.036	0.033	0.031	0.038	0.038	0.042	0.033	0.034	0.030	0.023
Marginal sig.	0.790	0.915	0.899	0.875	0.891	0.145	0.133	0.134	0.133	0.131	0.150	0.146	0.121	0.131	0.134

Table 4 continued

	MK1	MK2	MK3	MK4	MK5	MK6	MK7	MK8	MK9	MK10	MK11	MK12	MK13	MK14	MK15
Stage 2															
Arm 1	0.419	0.900	0.135	0.129	0.132	0.045	0.035	0.045	0.039	0.031	0.046	0.042	0.040	0.039	0.045
Arm 2	0.034	0.057	0.863	0.035	0.038	0.015	0.016	0.013	0.014	0.012	0.015	0.012	0.015	0.010	0.016
Arm 3	0.401	0.047	0.033	0.836	0.035	0.011	0.016	0.016	0.008	0.014	0.016	0.012	0.013	0.015	0.014
Arm 4	0.096	0.048	0.031	0.035	0.860	0.016	0.012	0.013	0.015	0.016	0.017	0.019	0.015	0.012	0.010
Marginal sig.	0.763	0.915	0.895	0.866	0.888	0.076	0.072	0.078	0.068	0.065	0.080	0.076	0.071	0.072	0.077

Thirty percent and 80 % posterior credible intervals were separately used in the group selection step and the individual step of variable selection at the end of stage 1. Therefore, in the group selection step, a marker will be selected if the central 30 % credible interval does not cover zero; and in the individual selection step, the corresponding parameters will be selected if the central 80 % credible interval does not cover zero. Similarly, an 80 % posterior credible interval was used at the end of stage 2 to test the significance of model parameters. Markers with either significant main effect (in arm 1) or significant interactions (in arms 2, 3, or 4) are defined as significant for the marginal marker effect

Table 5 Patient allocation rate into one of the four treatments by prior treatment history of erlotinib and marker expression (positive or negative) for stage 1 and stage 2 under the null and alternative hypotheses

	Resist	Naïve	MK1–	MK1+	MK2–	MK2+	MK3–	MK3+	MK4–	MK4+	MK5–
Null											
Stage 1											
Arm 1	0.000	0.188	0.092	0.193	0.112	0.113	0.113	0.112	0.113	0.112	0.114
Arm 2	0.337	0.271	0.304	0.272	0.297	0.298	0.297	0.297	0.297	0.297	0.297
Arm 3	0.332	0.270	0.303	0.261	0.296	0.294	0.292	0.298	0.297	0.293	0.295
Arm 4	0.331	0.271	0.301	0.274	0.295	0.295	0.298	0.293	0.293	0.298	0.294
Stage 2											
Arm 1	0.000	0.210	0.126	0.128	0.128	0.126	0.128	0.126	0.126	0.128	0.128
Arm 2	0.346	0.263	0.295	0.300	0.292	0.299	0.294	0.297	0.295	0.296	0.296
Arm 3	0.331	0.263	0.291	0.284	0.290	0.290	0.290	0.290	0.291	0.288	0.290
Arm 4	0.323	0.265	0.288	0.287	0.291	0.285	0.288	0.287	0.288	0.288	0.286
Alternative											
Stage 1											
Arm 1	0.000	0.143	0.088	0.077	0.087	0.086	0.086	0.086	0.085	0.087	0.085
Arm 2	0.317	0.275	0.300	0.258	0.291	0.293	0.292	0.291	0.291	0.292	0.294
Arm 3	0.349	0.301	0.306	0.373	0.320	0.320	0.320	0.320	0.319	0.321	0.319
Arm 4	0.334	0.281	0.305	0.291	0.303	0.301	0.302	0.303	0.304	0.300	0.302
Stage 2											
Arm 1	0.000	0.202	0.121	0.119	0.120	0.121	0.122	0.119	0.121	0.120	0.120
Arm 2	0.319	0.259	0.289	0.261	0.281	0.286	0.224	0.343	0.314	0.252	0.314
Arm 3	0.350	0.274	0.294	0.347	0.306	0.303	0.335	0.274	0.245	0.365	0.340
Arm 4	0.330	0.265	0.296	0.273	0.293	0.290	0.319	0.263	0.320	0.262	0.226
	MK5+	MK6–	MK6+	MK7–	MK7+	MK8–	MK8+	MK9–	MK9+	MK10–	MK10+
Null											
Stage 1											
Arm 1	0.111	0.113	0.113	0.113	0.113	0.113	0.112	0.114	0.112	0.112	0.113
Arm 2	0.298	0.296	0.299	0.298	0.297	0.295	0.299	0.298	0.296	0.299	0.295
Arm 3	0.295	0.296	0.294	0.295	0.295	0.296	0.294	0.294	0.296	0.294	0.296
Arm 4	0.296	0.296	0.294	0.295	0.296	0.296	0.295	0.295	0.297	0.294	0.296
Stage 2											
Arm 1	0.126	0.128	0.126	0.127	0.127	0.125	0.128	0.126	0.128	0.127	0.128
Arm 2	0.295	0.296	0.296	0.296	0.296	0.296	0.296	0.296	0.296	0.298	0.293
Arm 3	0.290	0.289	0.291	0.292	0.288	0.291	0.289	0.289	0.291	0.288	0.291
Arm 4	0.290	0.288	0.287	0.286	0.289	0.288	0.288	0.289	0.286	0.287	0.288
Alternative											
Stage 1											
Arm 1	0.087	0.086	0.086	0.086	0.086	0.086	0.087	0.086	0.086	0.087	0.085
Arm 2	0.289	0.292	0.291	0.291	0.292	0.292	0.291	0.292	0.291	0.290	0.293

Table 5 continued

	MK5+	MK6-	MK6+	MK7-	MK7+	MK8-	MK8+	MK9-	MK9+	MK10-	MK10+
Arm 3	0.321	0.319	0.321	0.320	0.319	0.320	0.319	0.320	0.320	0.321	0.319
Arm 4	0.303	0.304	0.301	0.302	0.302	0.302	0.303	0.301	0.303	0.302	0.303
Stage 2											
Arm 1	0.121	0.121	0.121	0.119	0.122	0.121	0.120	0.119	0.122	0.120	0.121
Arm 2	0.253	0.286	0.281	0.284	0.283	0.284	0.283	0.285	0.282	0.285	0.282
Arm 3	0.269	0.305	0.304	0.305	0.304	0.304	0.305	0.305	0.304	0.303	0.306
Arm 4	0.356	0.289	0.294	0.292	0.291	0.291	0.292	0.291	0.291	0.292	0.291

The allocation rates to arms 1–4 add up to 1 under each condition. The results for markers 11–15 are similar to those for markers 6–10; hence, they are omitted. Resist/Naïve stands for erlotinib-resistant or erlotinib-naïve; MK–/MK+ represents group that has negative/positive status for that marker

(*KRAS* mutation, for which selection probabilities are 28 % in both stages). In both the null and alternative cases, patients with a positive *KRAS* status will have a reduced response rate as compared with patients with the *KRAS* wild type, which translates into a small negative marker main effect (prognostic marker) in both scenarios. Therefore, higher selection probability of marker 1 is reasonable. Under the alternative hypothesis at the end of stage 2, the probabilities of selection for markers M_1 M_2 M_3 M_4 , and M_5 are about 76, 92, 90, 87, and 89 %, respectively. The probability of selecting an unimportant marker is under 15 % at the end of stage 1 and 8 % at the end of stage 2. As shown in Table 4, markers M_3 M_4 , and M_5 have higher selection probabilities for arms 2, 3, and 4, respectively, which is consistent with our model assumptions.

The patient allocation rates to one of the four treatment arms are summarized in Table 5 by corresponding columns separately for prior erlotinib treatment history and biomarker status. Under the null hypothesis, there is not much difference in patient allocation between the first and second stages. Erlotinib-naïve patients are assigned evenly to the four treatments; whereas erlotinib-resistant patients are assigned evenly to the three experimental treatments. There are also no differences in treatment allocation ratio between different marker groups. Under the alternative hypothesis, erlotinib-naïve patients are assigned with higher probability to arms 2, 3, and 4 because those treatments are more effective than the treatment of arm 1. Although the erlotinib-resistant patients are assigned about evenly to arms 2, 3, and 4, their allocation ratios differ depending on marker groups. In stage 1, only marker M_1 was used in the AR, thus, patient allocation ratios among the other markers are similar. In stage 2, markers M_2 M_3 M_4 , and M_5 have higher probabilities of being included in AR model. Therefore, the allocation of patient subgroups defined by M_2 M_3 M_4 , and M_5 will be different in the stage 2. For example, in stage 2, patients with a M_3+ status are assigned more frequently to arm 2, M_4+ are assigned more frequently to arm 3, and M_5+ are assigned more frequently to arm 4. These results indicate that the AR in both stage 1 and stage 2 has achieved its goal of treating more subjects with more effective treatments given patients’ marker profiles. With model refinement in stage 2, the false-positive rate of variable selection is further reduced; hence, patients are treated even more effectively than in stage 1.

7 Discussion

In the early phase of drug development, little information is available before the trial begins regarding treatment efficacy and the associated biomarkers. When sailing on such uncharted waters, it is essential to continuously learn and to make adjustments in response to learning throughout the course. Early drug development is all about learning. Adaptive designs allow for adjustments to the study conduct based on the interim data and provide a sensible and flexible approach to facilitate such learning.

We have developed a Bayesian two-stage biomarker-based adaptive randomization design for targeted agent development. The design allows for testing the treatment efficacy in stage 1 to make a Go or No-Go decision based on a futility early stopping rule. Putative prognostic and predictive biomarkers can be identified in stage 1 to help refining the adaptive randomization and assign more patients to more effective treatments in stage 2 based on each patient's specific marker profiles. Simulation studies show that the design has desirable operating characteristics in controlling the per-comparison error rate, achieving desirable statistical power for testing the treatment efficacy, providing high probability of selecting important markers and low probability of selecting noise-related markers, as well as assigning more patients to more effective treatments. Bayesian logistic regression provides a rich and flexible model for estimating treatment effect, marker effect, and their interactions. The Bayesian two-step lasso performs well for variable selection. The use of a group lasso in the first step is effective for identifying important markers and screening out non-contributing ones. The adaptive lasso in the second step can further refine variable selection to increase the statistical power while reducing false-positive selections. These results compare favorably with those of other variable selection methods that we have evaluated.

We implemented our Bayesian modeling in OpenBugs, where we only need to specify the model, the prior distributions, and the data likelihood. The full conditional distributions for Gibbs Sampling for logistic regression and standard Bayesian lasso can be found in the reference papers [39, 41], and also in the Appendices 3 and 4 for our modeling approach. Using OpenBugs will be slower in computation time as compared with writing Gibbs Sampler in fast low-level languages such as C/C++/Fortran. But using OpenBugs will save development time of programming. In our case, computation time is not an issue because we submit simulations in parallel such that each CPU will only need to run a few simulations. So the total computation time is the time of running a few simulations plus some overhead to combine all simulation results. Hence, we opted to use OpenBugs rather than developing our own Gibbs sampler. Using OpenBugs and R language have some additional benefits such as the standard documentation is available and validation is easier. Both issues are relevant when working with regulatory agencies, and are important for internal quality control.

The outcome-adaptive randomization is suitable for trials with relatively short-term endpoint such as the 8-week disease control status in the BATTLE-1 and BATTLE-2 trials. In order to adapt based on the interim data, the study endpoint needs to be recorded accurately and timely for each patient as the information accrues. Additional infrastructure support is required to ensure that the endpoint is objectively, accurately, and consistently evaluated and reported. A web-based centralized database with an automatic e-mail notification system can be constructed to facilitate timely data entry.

We open this trial at two centers (MD Anderson and Yale) and hope to demonstrate the feasibility of conducting such a study at multiple sites. The trial must use consistent patient eligibility criteria to ensure that patient characteristics do not drift over the course of the trial. Patient accrual also cannot be much faster than the accumulation of data. Otherwise, it will limit the value of adaptation.

Note that in this randomized phase II study, the efficacy endpoint is of the primary interest. We do not expect that the toxicity will be a main concern because the single agents and two-drug combinations in BATTLE-2 have been well studied. Hence, no formal statistical monitoring on toxicity will be conducted. However, toxicity will be monitored based on the CTCAE criteria as usual. Dose with holding or reduction will be implemented at the patient level whenever needed.

Other variable selection methods could have been used at the end of stage 1. The two-step Bayesian lasso method was used because of its good performance in simulations, and its consistence with the Bayesian adaptive design framework of the trial. For prediction purpose, a Bayesian model averaged (BMA) approach is more consistent with fundamental Bayesian idea. As a matter of fact, Barbieri and Berger [45] have proved that the best prediction model is the median probability model (which is often not the highest probability model) for mean squared error loss function. However, in our particular case, we did not use the BMA approach for the following reasons: (1) For targeted therapy development, it is more important to focus on just a few biomarkers that can differentiate treatment effects. These markers needed to be CLIA certified for clinical use to make treatment decisions. BMA is likely to select too many markers and is not practical in this regard; (2) We tried Bayesian model averaging using the BMA package in R. The performances of BMA were not as good as what we had by using Bayesian lasso.

Due to the complexity of the questions that need to be addressed by the trial, we were not able to provide the strict control of error rate analytically. Instead, the error rate of testing treatment effect was controlled numerically through simulation. Since prior information was incorporated into the null simulation case, our control of error rate was not even in the weak sense. Therefore in current regulatory environment, this design is not suitable for registration trial. However, in early phase study, the promise offered by our design to treat patient better within the trial and the flexibility of making decision are both desirable properties that can be leveraged on in the drug development process. Of course, as with all trials, the findings from our trial need to be validated in confirmatory trials [46].

There are several challenges in implementing the adaptive design including the computational demands, the lack of experience, and the lack of infrastructure, etc. However, many of these the road blocks can be easily overcome with better computing algorithms and hardware, building a web-based database for real-time data entry and update, and so on. With more such trials being conducted, the added complexity of conducting adaptive designs will become less and less an issue.

The successful use of Bayesian designs has been demonstrated under many settings [47, 48]. The proposed Bayesian two-stage biomarker-based adaptive design offers an additional efficient and flexible procedure for drug development. It can be used to identify effective treatments and relevant biomarkers, and to match effective treatments

with patients’ biomarker profiles to best treat patients in the trial. It offers a rational approach toward the goal of personalized medicine.

Acknowledgments The authors thank Ms. LeeAnn Chastain for editorial assistance. The work was supported in part by grants CA016672 and CA155196 from the National Cancer Institute. The clinical trial was supported in part by Merck Research Laboratories and Bayer HealthCare. We also would like to thank two anonymous reviewers, the associate editor, and the editor for their thorough review and constructive critiques. Our manuscript has been improved by providing answers in addressing these critics.

Conflict of interest The authors did not have any conflict of interest related to this work to disclose.

Appendix 1

The prior for η_k in (8) can be obtained by integrating out τ_k^2 :

$$\begin{aligned} & \int_0^\infty \left(\frac{\tau_k^2}{2\pi}\right)^{\frac{m_k}{2}} \exp\left[-\frac{\eta_k}{2} \tau_k^2\right] \frac{\left(\frac{\lambda_k^2}{2}\right)^{\frac{m_k+1}{2}}}{\Gamma\left(\frac{m_k+1}{2}\right)} (\tau_k^2)^{-\frac{m_k+1}{2}-1} \exp\left[-\frac{\lambda_k^2}{2\tau_k^2}\right] d\tau_k^2 \\ &= \int_0^\infty \left(\frac{1}{2\pi}\right)^{\frac{m_k}{2}} \frac{\left(\frac{\lambda_k^2}{2}\right)^{\frac{m_k+1}{2}}}{\Gamma\left(\frac{m_k+1}{2}\right)} \left(\frac{1}{\tau_k^2}\right)^{\frac{1}{2}} \exp\left[-\frac{\eta_k}{2} \left(\tau_k^2 - \frac{\lambda_k}{\eta_k}\right)^2 - \lambda_k \eta_k\right] d\tau_k^2 \\ &= \int_0^\infty \left(\frac{1}{2\pi}\right)^{\frac{m_k}{2}} \frac{\left(\frac{\lambda_k^2}{2}\right)^{\frac{m_k+1}{2}}}{\Gamma\left(\frac{m_k+1}{2}\right)} \left(\frac{\lambda_k^2}{\tau_k^2}\right)^{\frac{1}{2}} \exp\left[-\frac{\lambda_k^2 \left(\tau_k^2 - \frac{\lambda_k}{\eta_k}\right)^2}{2\frac{\lambda_k^2}{\eta_k} \tau_k^2}\right] \exp[-\lambda_k \eta_k] d\tau_k^2 \\ &= \left(\frac{1}{2\pi}\right)^{\frac{m_k-1}{2}} \frac{\left(\frac{1}{2}\right)^{\frac{m_k+1}{2}}}{\Gamma\left(\frac{m_k+1}{2}\right)} \lambda_k^{m_k} \exp[-\lambda_k \eta_k] \int_0^\infty \left(\frac{\lambda_k^2}{2\pi(\tau_k^2)^3}\right)^{\frac{1}{2}} \\ & \quad \times \exp\left[-\frac{\lambda_k^2 \left(\tau_k^2 - \frac{\lambda_k}{\eta_k}\right)^2}{2\frac{\lambda_k^2}{\eta_k} \tau_k^2}\right] d\tau_k^2 \end{aligned}$$

The pdf of inverted Gaussian distribution is

$$\left(\frac{\gamma}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left[-\frac{\gamma(x - \mu)^2}{2\mu x}\right]$$

When letting $\tau_k^2 = x$, $\frac{\lambda_k}{\eta_k} = \mu$, $\lambda_k^2 = \gamma$, and integrating out τ_k^2 , we get the prior in (8).

Appendix 2

We used the same true scenario in Table 2 to simulate true patient response. However, instead of randomizing patients based on individual biomarker information (BATTLE-2 design strategy), patients were randomized based on biomarker group information (BATTLE-1 design strategy). Following BATTLE-1 design strategy, 3 biomarkers with decreasing priorities were used to group patients into 4 groups. A patient will be in group 1 if the first marker (highest priority) is positive; otherwise, the patient will be in group 2 if the second marker is positive; otherwise, the patient will be in group 3 if the third marker is positive; and the patient will be in group 4 if all 3 biomarkers are negative.

For the null case with the BATTLE-1 design strategy, the 3 markers used to group patients were marker 3, 4, and 5. Three different alternative cases were simulated using BATTLE-1 design strategy to represent 3 different situations. In Alternative 1 case, we assumed that the 3 true predictive markers were correctly identified, and marker 3, 4 and 5 were used to group patients. Unless specified explicitly, we always assumed that the marker with lower index has higher priority. In Alternative 2 case, we used marker 3, 6, and 7 to group patients, which means that one true predictive marker has been assigned the highest priority by chance. In Alternative 3 case, we used 3 randomly chosen markers to group patients, and this is equivalent to grouping patients with no credible marker information. The same hypothesis testings for the Go/No-Go decisions as in the simulation for BATTLE-2 design were conducted at the end of stage 1, and the same interim futility analyses were implemented. To make the results comparable to BATTLE-2 design, *KRAS* mutation and prior erlotinib history, in addition to the marker group information, were always included in the adaptive randomization. The same models with *KRAS* mutation and prior erlotinib history in (1) and (2) were used in hypothesis testing, and the marker group information was only used in adaptive randomization. Therefore, the differences in simulation results of BATTLE-1 and 2 design strategies were due to differences in randomization strategies and the added variable selection in BATTLE-2:

- (1) BATTLE-1 design strategy: Markers were used to group patients, and group information was used in adaptive randomization in BATTLE-1 design;
- (2) BATTLE-2 design strategy: Individual biomarkers were subjected to variable selection and, upon being selected, markers were directly used in adaptive randomization in BATTLE-2 design.

For comparison purpose, we simulated the clinical trials by following the design strategy of BATTLE-1 trial, and the results for hypothesis testing at the end of stage 1 are shown in Supplemental Table 1. The patient responses were simulated from the same true scenarios in Table 2. The same decision criteria and modeling approaches of BATTLE-2 design were also used for BATTLE-1 simulation. The error rates were controlled at the same rate of 10 % for each experimental treatment. Therefore, the differences in operating characteristics between BATTLE-1 and BATTLE-2 simulations can be attributed to the different strategies of incorporating biomarker information in adaptive randomization. The simulation results for Alternative 1 case in Supplemental

Table 1 show that, in the best scenario where the predictive markers were correctly identified before the start of the clinical trial, BATTLE-1 design has higher power on the treatment (arm 2) associated with the marker having the highest priority, and has lower power on the remaining 2 experimental treatments (arm 3 and 4). In the worst scenario where none of the true predictive markers were correctly identified before the start of the trial (Alternative 3 case), BATTLE-2 trial design has higher power on all experimental treatments. In practice, Alternative 3 case is more likely to happen than Alternative 1 case. Because the purpose of clinical trial is to study experimental treatment and the biomarkers associated the treatment. It is unlikely that we are going to have perfect biomarker information at the beginning of clinical trial.

The patient allocation results for BATTLE-1 design strategy are shown in Supplemental Table 2. As compared with the results in Table 5 for BATTLE-2 design, the predictive marker with the highest priority can allocate more patients to the beneficial treatment arm in the most optimal situation (Alternative 1 and 2 cases) where the predictive marker has been correctly identified before the trial. In all other cases, the BATTLE-1 design strategy failed to learn from ongoing trial to treat patients better.

If credible informative biomarkers are available, using pre-selected biomarkers will be more powerful [49]. Our simulation comparison with BATTLE-1 design strategy also demonstrated that. However, in the exploratory stage of study when no credible biomarker information is available, our simulations showed that combining biomarkers into biomarker groups may lead to higher error rate in biomarker selection.

We also simulated the first stage of the trial using an equal randomization. The results are shown in the Supplemental Table 3. Under the null hypothesis, when the overall error rate for the trial is still controlled at 20 %, the average error rate for each experimental treatment arm is about 1–2 % lower (0.077–0.091 under ER and 0.088–0.100 for AR). When the alternative hypothesis is true, the increase in average power for each experimental treatment arm can be as high as 15 % (0.97 to 0.99 for ER and 0.82 to 0.95 for AR). However, the increase of the overall power of the trial is only 2 % (0.999 for ER and 0.978 for AR). With AR and early stopping rule, it can result in a sample size saving of 60 patients under the null hypothesis. AR also resulted in more patients to be treated with more effective treatments.

Appendix 3

The likelihood function of logistic model can be expressed as:

$$\prod_{i=1}^n \left[\left(\frac{\exp(X_i \theta)}{1 + \exp(X_i \theta)} \right)^{y_i} \left(\frac{1}{1 + \exp(X_i \theta)} \right)^{1-y_i} \right] = \frac{\exp(\sum_{i \in \Omega_R} X_i \theta)}{(1 + \exp(X_i \theta))^n}$$

where i is the index for individual patient, n is the total number of patients, Ω_R is the set of index for all responders, X_i is a generic representation of covariate, and θ is a generic representation of model parameter. Without loss of generality, we assume that both X_i and θ are scales.

For Bayesian adaptive lasso, the prior for θ given λ is

$$\pi(\theta|\lambda) \propto \exp\left(-\lambda \frac{|\theta|}{\hat{\theta}}\right)$$

where $\hat{\theta}$ is a square root n consistent initial estimate of θ . The square of λ follows a gamma distribution:

$$\pi(\lambda^2) = \frac{\lambda^{2(a-1)} e^{-\frac{\lambda^2}{b}}}{b^a \Gamma(a)}$$

where a and b are hyper parameters.

The conditional posterior distributions of θ and λ are

$$\theta|\lambda \propto \frac{\exp(\sum_{i \in \Omega_R} X_i \theta)}{(1 + \exp(X_i \theta))^n} \cdot \exp\left(-\lambda \frac{|\theta|}{\hat{\theta}}\right)$$

$$\lambda|\theta \propto \lambda^{2(a-1)} e^{-\frac{\lambda^2}{b}} \cdot \exp\left(-\lambda \frac{|\theta|}{\hat{\theta}}\right)$$

It is easy to show that all above conditional distributions are log-concave. Therefore, the adaptive rejection sampling implemented in OpenBugs can be conveniently utilized to generate the posterior distribution of model parameters for Bayesian adaptive lasso.

Appendix 4

For group parameter η in Bayesian adaptive group lasso, the prior distribution is

$$\pi(\eta | \lambda) \propto \exp(-\lambda |\eta| / \|\hat{\eta}\|)$$

where $\hat{\eta}$ is an initial estimate of η . In our simulation, we used the gamma mixture of normals in Eq. (9). After τ_k in Eq. (9) is integrated out, Appendix 1 shows that these two formulations are equivalent while conditional on λ .

The same gamma prior distribution is assumed for the square of λ :

$$\pi(\lambda^2) = \frac{\lambda^{2(a-1)} e^{-\frac{\lambda^2}{b}}}{b^a \Gamma(a)}$$

The likelihood function of logistic model can be expressed as:

$$\prod_{i=1}^n \left[\left(\frac{\exp(X_i^T \eta)}{1 + \exp(X_i^T \eta)} \right)^{y_i} \left(\frac{1}{1 + \exp(X_i^T \eta)} \right)^{1-y_i} \right] = \frac{\exp(\sum_{i \in \Omega_R} X_i^T \eta)}{(1 + \exp(X_i^T \eta))^n}$$

Both X_i and η are column vectors. Without loss of generality, we assume that there is only one group of variables in the model. Therefore, the posterior distribution of model parameters can be sampled from

$$\eta, \tau | X, Y \propto \frac{\exp(\sum_{i \in \Omega_R} X_i^T \eta)}{(1 + \exp(X_i^T \eta))^n} \tau^m \exp\left[-\frac{\eta^2}{2} \tau^2\right] \lambda^{m+1} \tau^{-m-3} \exp\left[-\frac{\lambda^2}{2 \hat{\eta}^2 \tau^2}\right] \lambda^{2(a-1)} \exp\left[-\frac{\lambda^2}{b}\right]$$

where Y is a vector of observed response outcomes, m is the dimension of η , and $\eta^2 = \eta_1^2 + \dots + \eta_m^2$. The full conditional distributions are

$$\begin{aligned} \eta | \tau, \lambda, X, Y &\propto \frac{\exp(\sum_{i \in \Omega_R} X_i^T \eta)}{(1 + \exp(X_i^T \eta))^n} \exp\left[-\frac{\eta^2}{2} \tau^2\right] \\ \tau | \eta, \lambda, X, Y &\propto \tau^{-3} \exp\left[-\frac{\lambda^2}{2 \hat{\eta}^2 \tau^2}\right] \exp\left[-\frac{\eta^2}{2} \tau^2\right] \\ \lambda | \eta, \tau, X, Y &\propto \lambda^{2a+m-1} \exp\left[-\lambda^2 \left(\frac{1}{b} + \frac{1}{2 \hat{\eta}^2 \tau^2}\right)\right] \end{aligned}$$

References

- DiMasi JA, Hansen RW, Grabowski HG (2003) The price of innovation: new estimates of drug development costs. *J Health Econ* 22(2):151–185
- Herper M (2012) The truly staggering cost of inventing new drugs. Forbes Website. <http://www.forbes.com/sites/matthewherper/2012/02/10/the-truly-staggering-cost-of-inventing-new-drugs/>
- CSDD Outlook (2009) Background: a methodology for counting cost for pharmaceutical R&D. Tufts Center for the study of Drug Development, Boston
- Kola I, Landis J (2004) Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 3:711–716
- Spear BB, Heath-Chiozzi M, Huff J (2001) Clinical application of pharmacogenetics. *Trends Mol Med* 7:201–204
- Simon R (2010) Clinical trial designs for evaluating the medical utility of prognostic and predictive biomarkers in oncology. *Pers Med* 7(1):33–47
- Mandrekar SJ, Sargent DJ (2009) Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. *J Clin Oncol* 27(24):4027–4034
- Mandrekar SJ, Sargent DJ (2001) Design of clinical trials for biomarker research in oncology. *Clin Invest* 1(12):1629–1636
- Wang SJ (2007) Biomarker as a classifier in pharmacogenomics clinical trials: a tribute to 30th anniversary of PSI. *Pharm Stat* 6(4):283–296
- Scher HI, Nasso SF, Rubin EH, Simon R (2011) Adaptive clinical trial designs for simultaneous testing of matched diagnostics and therapeutics. *Clin Cancer Res* 17(21):6634–6640
- Goozner M (2012) Drug approvals 2011: focus on companion diagnostics. *J Natl Cancer Inst* 104(2):84–86
- Poste G, Carbone DP, Parkinson DR, Verweij J, Hewitt SM, Jessup JM (2012) Leveling the playing field: bringing development of biomarkers and molecular diagnostics up to the standards for drug development. *Clin Cancer Res* 18(6):1515–1523
- Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144(5):646–674
- Reck M, Gatzemeier U (2004) Chemotherapy in stage-IV NSCLC. *Lung Cancer* 45:s217–222

15. Schiller JH, Harrington D, Belani CP, Langer C, Sandler A, Krook J, Zhu J, Johnson DH for the Eastern Cooperative Oncology Group (2002) Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. *N Engl J Med* 346:92–98
16. Silvestri G, Rivera M (2005) Targeted therapy for the treatment of advanced non-small cell lung cancer: a review of the epidermal growth factor receptor antagonists. *Chest* 183:29–42
17. McClellan M, Benner J, Schilsky R et al (2011) An accelerated pathway for targeted cancer therapies. *Nat Rev Drug Discov* 10:79–80
18. Chabner BA (2011) Early accelerated approval for highly targeted cancer drugs. *N Engl J Med* 364(12):1087–1089
19. Bates SE, Amiri-Kordestani L, Giaccone G (2012) Drug development: portals of discovery. *Clin Cancer Res* 18(1):23–32
20. Sharma MR, Schilsky RL (2012) Role of randomized phase III trials in an era of effective targeted therapies. *Nat Rev Clin Oncol* 9(4):208–214
21. Rubin EH, Gilliland DG (2012) Drug development and clinical trials—the path to an approved cancer drug. *Nat Rev Clin Oncol* 9(4):215–222
22. Zhou X, Liu SY, Kim ES, Herbst RS, Lee JJ (2008) Bayesian adaptive design for targeted therapy development in lung cancer—a step toward personalized medicine. *Clin Trials* 5:181–193
23. Kim ES, Herbst RS, Wistuba II, Lee JJ, Blumenschein GR, Tsao A, Stewart DJ, Hicks ME, Erasmus J, Gupta S, Alden CM, Liu S, Tang X, Khuri FR, Tran HT, Johnson BE, Heymach JV, Mao L, Fossella F, Kies MS, Papadimitrakopoulou V, Davis SE, Lippman SM, Hong WK (2011) The BATTLE trial: Personalizing therapy for lung cancer. *Cancer Discov* 1(1):44–53
24. Berry DA (2005) Introduction to Bayesian methods III: use and interpretation of Bayesian tools in design and analysis. *Clin Trials* 2:295–300; discussion 301–304, 364–378
25. Berry DA (2006) A guide to drug discovery: Bayesian clinical trials. *Nat Rev Drug Discov* 5:27–36
26. Berry DA (2011) Adaptive clinical trials in oncology. *Nat Rev Clin Oncol* 9(4):199–207
27. Hu F, Rosenberger WF (2003) Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. *J Am Stat Assoc* 98:671–678
28. Hu F, Rosenberger WF (2006) The theory of response-adaptive randomization in clinical trials. Wiley, Hoboken
29. Thall PF (2002) Ethical issues in oncology biostatistics. *Stat Methods Med Res* 11:429–448
30. Berry DA (2004) Bayesian statistics and the efficiency and ethics of clinical trials. *Stat Sci* 19:175–187
31. Korn EL, Freidlin B (2010) Outcome-adaptive randomization: is it useful? *J Clin Oncol* 21:100–120
32. Berry DA (2010) Adaptive clinical trials: the promise and the caution. *J Clin Oncol* 21:606–609
33. Lee JJ, Chen N, Yin G (2012) Worth adapting? Revisiting the usefulness of outcome-adaptive randomization. *Clin Cancer Res* 18(17):4498–4507
34. Barker AD et al (2009) I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharm Ther* 86:97–100
35. Berry SM, Carlin BP, Lee JJ, Müller P (2010) Bayesian adaptive methods for clinical trials. Chapman & Hall, Boca Raton
36. Lee JJ, Gu X, Liu S (2010) Bayesian adaptive randomization designs for targeted agent development. *Clin Trials* 7(5):584–596
37. Eickhoff JC, Kim K, Beach J, Kolesar JM, Gee JR (2010) A Bayesian adaptive design with biomarkers for targeted therapies. *Clin Trials* 7:546–556
38. Lara PN, Natale R, Crowley J et al (2009) Phase III trial of irinotecan/cisplatin Compared with etoposide/cisplatin in extensive-stage small-cell lung cancer: clinical and pharmacogenomic results from SWOG S0124. *J Clin Oncol* 27:2530–2535
39. Park T, Casella G (2008) The Bayesian lasso. *J Am Stat Assoc* 103(482):681–686
40. Zou H (2006) The adaptive lasso and its oracle properties. *J Am Stat Assoc* 101(476):1418–1429
41. Kyung M, Gill J, Ghosh M (2010) Penalized regression, standard errors, and Bayesian lassos. *Bayesian Anal* 5:369–412
42. Meier L, van de Geer S, Bühlmann P (2008) The group lasso for logistic regression. *J R Stat Soc B* 70:53–71
43. Chipman H (1996) Bayesian variable selection with related predictors. *Can J Stat* 24:17–36
44. McCullagh P, Nelder J (1989) Generalized linear models. Chapman and Hall, Boca Raton
45. Barbieri MM, Berger J (2004) Optimal predictive model selection. *Ann Stat* 32(3):870–897

46. Brannath W, Zuber E, Branson M, Bretz F, Gallo P, Posch M, Racine-Poon A (2009) Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. *Stat Med* 28(10):1445–1463
47. Biswas S, Liu DD, Lee JJ, Berry DA (2009) Bayesian clinical trials at the University of Texas MD Anderson Cancer Center. *Clin Trials* 6(3):205–216
48. Lee JJ, Chu CT (2012) Bayesian clinical trials in action. *Stat Med* 31(25):2955–2972
49. Wang S-J, Li M-C (2013) Impacts of predictive genomic classifier performance on subpopulation-specific treatment effects assessment. *Stat Biosci*. doi:[10.1007/s12561-013-9092-y](https://doi.org/10.1007/s12561-013-9092-y)