Published in final edited form as:

Clin Trials. 2013 ; 10(3): . doi:10.1177/1740774512470316.

Bayesian adaptive phase II screening design for combination trials

Chunyan Cai^a, Ying Yuan^b, and Valen E Johnson^c

^aBiostatistics/Epidemiology/Research Design Core, Center for Clinical and Translational Sciences, The University of Texas Health Science Center at Houston, Houston, TX, USA

^bDepartment of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

^cDepartment of Statistics, Texas A&M University, College Station, TX, USA

Abstract

Background—Trials of combination therapies for the treatment of cancer are playing an increasingly important role in the battle against this disease. To more efficiently handle the large number of combination therapies that must be tested, we propose a novel Bayesian phase II adaptive screening design to simultaneously select among possible treatment combinations involving multiple agents.

Methods—Our design is based on formulating the selection procedure as a Bayesian hypothesis testing problem in which the superiority of each treatment combination is equated to a single hypothesis. During the trial conduct, we use the current values of the posterior probabilities of all hypotheses to adaptively allocate patients to treatment combinations.

Results—Simulation studies show that the proposed design substantially outperforms the conventional multiarm balanced factorial trial design. The proposed design yields a significantly higher probability for selecting the best treatment while allocating substantially more patients to efficacious treatments.

Limitations—The proposed design is most appropriate for the trials combining multiple agents and screening out the efficacious combination to be further investigated.

Conclusions—The proposed Bayesian adaptive phase II screening design substantially outperformed the conventional complete factorial design. Our design allocates more patients to better treatments while providing higher power to identify the best treatment at the end of the trial.

Introduction

The use of combination therapies for cancer treatment can lead to treatment synergies that result in improved patient outcomes. The number of treatment combinations that must be tested is often quite large, however, which means that it is often not practical to conduct separate phase II trials on every possible combination of treatments. In this article, we describe a Bayesian adaptive trial design that facilitates the pooling of information obtained across treatment combinations by testing efficacy of all treatment combinations in a single trial. Important benefits of our trial designs include a reduction in the number of patients

Author for correspondence: Valen E Johnson, Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, USA. vjohnson@stat.tamu.edu.

Conflict of interest

There is no conflict of interest.

who must be recruited in order to evaluate each treatment combination, the assignment of a higher proportion of patients to efficacious treatments, faster patient accrual, and more rapid completion of trials.

To motivate our design methodology, we consider a recent drug-combination clinical trial conducted at MD Anderson Cancer Center to test the effectiveness of 16 combinations of four agents, A_1 , A_2 , A_3 and A_4 , in reducing the symptom burden experienced by patients with late-stage non-small-cell lung cancer who have received chemo-radiation therapies. The actual names of the drugs assessed in the trial are not specified here for reasons of confidentiality. The treatment combinations under investigation are enumerated in Table 1. The primary outcome variable for this trial was the area under the curve for five symptoms (pain, fatigue, drowsiness, sleep disturbance, and lack of appetite) measured daily using an interactive voice recording system for 10 days following the onset of radiation therapy. Each symptom was measured using the MD Anderson Symptom Inventory, which solicits ordinal ratings of symptoms on an 11-point scale ranging from 0 ('none at all') to 10 ('worst that can be imagined'). In contrast to typical cancer-treating agents, which are generally cytotoxic, the combinations of four agents tested in this trial were known to have minimal risks of toxicity, and thus, we focused herein on efficacy only. The goal of the trial is to identify the most efficacious combination to be further investigated in large trials.

A variety of screening designs have been proposed for trials comparing multiple independent or unique treatments. Among these, Thall *et al.* [1] proposed a two-stage phase II–III trial design to select the most promising treatment from k treatments in the first stage, and to compare the selected first-stage treatment with the standard of care in the second stage. Schaid *et al.* [2] presented a similar two-stage design that used survival data as an endpoint and that design allowed more than one treatment to be included in the second stage. Yao *et al.* [3] proposed a design to screen new treatments as a continuous process for identifying promising new therapeutic agents and determined the optimal sample size to be used with their design. Yao and Venkatraman [4], and Wang and Leung [5] extended that design to two-stage and fully sequential designs. Stout and Hardwick [6] developed a cost- and constraint-based decision-theoretic approach to the design of screening trials. Rossell *et al.* [7] proposed a screening design based on Bayesian decision theoretics that uses optimal linear boundaries. Ding *et al.* [8] developed a more systematic decision-making optimal phase II screening trial design using a utility function that accounts for sampling costs and possible future payoff. These aforementioned trial designs assume that the treatments under screening are independent (or unique) and assess their effects independently without borrowing information across the treatments. Therefore, these existing designs are not suitable to handle drug-combination trials, such as our motivating cancer trial, because in this case the treatments under comparison are interrelated, that is, the treatments are combinations obtained from the same set of agents.

We propose a Bayesian adaptive design to screen all combinations from a set of agents. Although each of the combinations is regarded as a treatment, we do not assume that they are independent or unique. We use information from all treatments to estimate the main effects and the interaction effects of the agents. That is, the data from the patient assigned to combinations are not only used to estimate the interaction effects but also used to estimate the main effects of each single agent. This is a key feature that distinguishes our design from the existing screening designs. Specifically, we model the main and synergistic effects of the treatment agents using a linear model, which facilitates borrowing information across the combinations. We cast the screening problem into a Bayesian hypothesis testing problem by constructing a series of hypotheses, each of which appoints one of the combinations as the most efficacious treatment. We utilize an encompassing prior with non-local prior constraints [9, 10] to accommodate the complex parameter constraints imposed by the

hypotheses. During the trial conduct, based on the observed data, we continuously update the posterior probabilities of the hypotheses and use them to adaptively allocate patients to effective combinations and select the best treatment. Extensive simulation studies show that compared to the standard (multiarm) balanced factorial design, the proposed design yields a significantly higher probability of selecting the best treatment. It also allocates more patients to efficacious treatments.

The remainder of this article is organized as follows. In section ‘Methods’, we describe our model, prior specification, and trial design. In section ‘Numerical studies’, we examine the operating characteristics of our design using simulation studies. We conclude with a brief discussion and comments in section ‘Conclusions’.

Methods

Probability model

We consider trials to evaluate the treatment effects of all possible combinations of k treatment agents, $A_1, A_2, \dots$, and A_k . We assume that each drug combination is assigned to one treatment arm, although it is straightforward to extend our design to trials where some combinations are excluded. Given k agents, there are $\binom{k}{r}$ different r -agent combinations, $r = 0, 1, \dots, k$, resulting in a total of $p = \sum_{r=0}^k \binom{k}{r} = 2^k$ combinations, including placebo group, to be evaluated. The goal of the trial is to identify the most efficacious treatment combination.

The outcome variable in the trial that motivates our research represents the mean change in the patient-reported symptom score. We assume that the outcome for the i th patient, y_i is continuous and follows a linear model of the form

$$y_i = \beta_0 + \beta_1 I_i(A_1) + \beta_2 I_i(A_2) + \dots + \beta_{1,2} I_i(A_1, A_2) + \dots + \beta_{1,2,\dots,k} I_i(A_1, A_2, \dots, A_k) + \varepsilon_i \quad (1)$$

where β_0 is the intercept of the linear model and $I_i(\cdot)$ is an indicator of whether patient i receives the given agents. For example, if patient i receives a combination of A_1 and A_2 , then $I_i(A_1) = I_i(A_2) = I_i(A_1, A_2) = 1$, whereas all the other indicator functions are then 0. Model (1) is flexible and accounts for the main and interaction effects of combining agents. Specifically, β_k represents the main treatment effect of A_k , $\beta_{k,k'}$ represents the two-way interaction or synergistic effect between A_k and $A_{k'}$ when $k \neq k'$, and so on. We assume that the residual ε_i follows a normal distribution with mean 0 and variance σ^2 . Binary and time-to-event outcomes can be modeled using a similar linear structure within a generalized linear model framework.

To cast the problem into a hypothesis testing framework, we define the null hypothesis H_0 to assert that no treatment is better than the placebo, and a series of alternative hypotheses $H_1, \dots, H_{p-1}$, where H_j asserts that the j th treatment combination is superior to all others. In our trial, for example, treatment j is superior to treatment k if it leads to a greater reduction in symptom burden. Let $\theta_0(\beta)$ denote the effect of the placebo and let $\theta_j(\beta)$, $j = 1, \dots, p-1$ denote the net treatment effect of the j th combination (or treatment arm). Under linear model (1), the treatment effect, $\theta_j(\beta)$, is a linear combination of the regression parameters, β s. For example, the treatment effect of the combination of A_1 and A_2 is given by $\theta_j(\beta) = \beta_0 + \beta_1 + \beta_2 + \beta_{1,2}$, and the treatment effect of the three-agent combination of A_1, A_2 , and A_3 is

$$\theta_j(\beta) = \beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_{1,2} + \beta_{1,3} + \beta_{2,3} + \beta_{1,2,3}.$$

To be consistent with the lung cancer trial, we assume that a smaller value of $\theta_j(\beta)$ (i.e., less symptom burden) represents a better response. Then the hypotheses can be formally expressed as

$$H_j: \theta_j(\beta) = \min(\theta_0(\beta), \dots, \theta_{p-1}(\beta)), \quad j=0, \dots, p-1$$

We let $\pi_j(\beta, \sigma^2)$ denote the prior distribution assigned to the unknown parameters β and σ^2 under H_j . Further discussion of the prior specification is provided in section ‘Delayed outcomes’; for the moment, we note that the domain of each prior is restricted to values of β that are consistent with the hypothesis under which they are defined [10]. Given these prior densities, the marginal density of the observed data $\mathbf{y}$ under H_j is

$$m_j(\mathbf{y}) = \iint f(\mathbf{y}|\beta, \sigma^2) \pi_j(\beta, \sigma^2) d\sigma^2 d\beta \quad (2)$$

and the Bayes factor [11, 12] of H_i to H_j is given by

$$B_{ij} = \frac{m_i(\mathbf{y})}{m_j(\mathbf{y})} \quad (3)$$

If $p(H_j)$ denotes the prior probability of H_j , then the posterior probability of H_j given the data $\mathbf{y}$ is

$$p(H_j|\mathbf{y}) = \frac{p(H_j)m_j(\mathbf{y})}{\sum_{i=0}^{p-1} p(H_i)m_i(\mathbf{y})} = \left[\sum_{i=0}^{p-1} \frac{p(H_i)}{p(H_j)} B_{ij} \right]^{-1} \quad (4)$$

If we assume that all hypotheses are equally likely a priori, then the posterior probability of H_j simplifies to

$$p(H_j|\mathbf{y}) = \frac{m_j(\mathbf{y})}{\sum_{i=0}^{p-1} m_i(\mathbf{y})} = \left[\sum_{i=0}^{p-1} B_{ij} \right]^{-1} \quad (5)$$

The value of $p(H_j|\mathbf{y})$ has a very intuitive probability interpretation – the probability that the j th combination is the best treatment conditional on the observed data. Meanwhile, the value of $p(H_0|\mathbf{y})$ is the probability that the placebo is the best treatment. Therefore, it provides a natural evidence-based mechanism to adaptively assign patients to efficacious combinations and select the most promising combination.

Trial design

We propose the following adaptive randomization scheme for the conduct of the trial. We assume that a total of N patients are available for testing, and that the first $m \times p$ patients are equally randomized into the p treatment arms using m replications of a complete factorial design, that is, m patients were randomized to each of the p arms. The advantage of using a factorial design is that it allows us to rapidly obtain preliminary estimates of the main treatment effects. Following the lead-in factorial phase of the design, subsequent patients are assigned to a treatment according to the posterior probability that each treatment is best. The resulting design can be described as follows:

1. Assign $m \times p$ patients to the p treatment arms using m replications of a factorial design.
2. For $i = m \times p + 1, \dots, N$, randomize the i th patient to the j th treatment arm with probability $p(H_j | \mathbf{y})$ $j = 0, \dots, p-1$, where $\mathbf{y} = (y_1, \dots, y_{i-1})'$ are the observed outcomes data from the first $i-1$ patients.
3. At the end of the trial, we select the combination j^* that has the highest posterior model probability, that is, $j^* = \operatorname{argmax}_j p(H_j | \mathbf{y}), j = 1, \dots, p-1$.

During the trial, we impose the following futility stopping rule: the trial is terminated for futility if

$$\max\{p(\theta_0 - \theta_j > \delta | \mathbf{y})\} < \alpha, \quad j = 1, \dots, p-1$$

where δ and α are the prespecified minimal effect size and threshold, respectively. That is, at any time during the trial, conditional on the observed data, if the probability of achieving an effect size of δ for the best treatment arm is below the threshold α , we terminate the trial. In practice, the value of δ can be elicited from investigators, and the values of design parameters m and α can be chosen by examining the operating characteristics of the trial in simulation studies.

Delayed outcomes

In general, outcome-dependent adaptive randomization, such as the one we have proposed, assumes that the outcome is quickly ascertainable so that when an incoming patient is ready for randomization, the previous patients have been assessed and their outcomes are completely observed. This assumption may not hold in practice. In many cases, the patient outcomes require a long follow-up time to be assessed (or the accrual is fast), so their outcomes are not available when a new patient is randomized. To address this delayed-outcome issue, one approach is to suspend accrual and wait until the outcomes of patients treated in the trial are fully observed. However, this approach is often not practical because it causes lengthy delays in a trial, wastes patient resources, and causes administrative problems. Alternatively, we propose to base our adaptive randomization scheme only on those patient outcomes that are available at the time that each new patient is randomized. Our simulation studies in section ‘Numerical studies’ show that with finite samples, this observed-data approach is competitive to the approach of suspending accrual.

Prior specification

We adopt the encompassing prior approach proposed by Klugkist *et al.* [9] and Klugkist and Hoijtink [13] to set the prior distributions on β and σ^2 under each hypothesis. In this approach, we first specify a prior distribution for the unconstrained model, and then based on that prior, we define prior densities under each hypothesis. More specifically, we begin

by assigning a noninformative prior to σ^2 of the form $\pi(\sigma^2) \propto 1/\sigma^2$. Given σ^2 and a hyperparameter g , we assume that β has a normal prior density of the form $\pi(\beta|\sigma^2) \sim N(\mathbf{0}, g\sigma^2\mathbf{I}_p)$, where $\mathbf{I}_p$ denotes a $p \times p$ identity matrix.

To modify the unconstrained priors for application to hypothesis $H_j, j = 0, \dots, p-1$, we restrict the domain of β under each hypothesis so that it is consistent with the assumptions of the given hypothesis [10]. That is, under hypothesis H_j , the domain of β is restricted to the value space satisfying the condition $\theta_j(\beta) = \min(\theta_0(\beta), \dots, \theta_{p-1}(\beta))$. This leads to the encompassing prior for H_j defined according to

$$\pi_j(\sigma^2) \propto \frac{1}{\sigma^2}, \quad \pi_j(\beta|\sigma^2) = \frac{1}{c_j} N(\mathbf{0}, g\sigma^2\mathbf{I}_p) I_{\min}(\theta_j(\beta)) \quad (6)$$

where $I_{\min}(\theta_j(\beta))$ denotes an indicator function of whether $\theta_j(\beta) = \min(\theta_0(\beta), \dots, \theta_{p-1}(\beta))$ and $c_j = \int N(\beta|\mathbf{0}, g\sigma^2\mathbf{I}_p) I_{\min}(\theta_j(\beta)) d\beta$

The prior densities used to define each hypothesis are thus nonlocal with respect to one another, which enables us to more rapidly exclude hypotheses that are inconsistent with the data [10].

Under the encompassing prior specification, it can be shown that the Bayes factor of H_i to H_j, B_{ij} , is given by

$$B_{ij} = \frac{m_i(\mathbf{y})}{m_j(\mathbf{y})} = \frac{r_i/c_i}{r_j/c_j}$$

where

$$r_j = \int t_\nu(\Sigma, \mu) I_{\min}(\theta_j(\beta)) d\beta$$

Here $t_\nu(\cdot)$ denotes a multivariate student distribution with degree of freedom $\nu = n$, scale matrix $\Sigma = V(\mathbf{y}^T\mathbf{y} - \mu^T V^{-1} \mu)/n$, and median $\mu = V\mathbf{X}^T\mathbf{y}$, where $V = ((1/g)\mathbf{I}_p + \mathbf{X}^T\mathbf{X})^{-1}$, $\mathbf{X}$ is the design matrix in model (1) and n is the number of patients who have completed the assessment during the course of the trial. The derivation of the Bayes factor, B_{ij} , is provided in Appendix A.

Numerical studies

Operating characteristics

We evaluated the operating characteristics of the proposed Bayesian adaptive screening (BAS) trial design through extensive simulation studies. In the context of the lung cancer trial, we considered a total of 16 combinational treatments, including the placebo control, that result from four agents (Table 1). A total of 200 patients were available for enrollment (i.e., $N = 200$), and we performed $m = 2$ replications of the factorial design to obtain preliminary estimates of the treatment effects. The accrual rate was 12 patients per month, and it took 10 days to obtain the symptoms outcome. Because the accrual was fast, we faced the delayed-outcome problem, that is, when a new patient is accrued and ready for randomization, some patients treated in the trial may have not finished their 10-day assessment, and their outcomes are not available for calculating the randomization

probabilities for the new patient. To deal with this issue, we adopted the observed-data approach described previously and calculated the randomization probabilities based on observed data when the outcomes of some patients are not available. Because the observed-data approach supports continuous accrual, it took approximately 17 months to complete the trial. For this trial, the approach of suspending accrual apparently is not feasible because it would lead to an infeasibly long trial lasting at least 4.8 years. Although the accrual-suspension approach is not useful in practice, it provides a theoretical benchmark for comparison because it represents the optimal case that the complete data are available to determine treatment assignment. For convenience, we denote the BAS design based on the accrual-suspension approach as BAS_{susp} . We configured the simulation parameters, β , to generate 12 different efficacy scenarios. The simulation results of the selection percentage of each treatment under these 12 different efficacy scenarios are displayed in Table 2. The total selection percentage of target treatments and the total percentage of patients assigned to targets are summarized in Table 3. Under each scenario, the most efficacious combination was defined as the combination with the smallest value of symptom burden, $\theta(\beta)$. We set the residual variance $\sigma^2 = 130$ based on previous symptom report data. The two parameters involved in the futility stopping rule, α and δ , were set to 0.35 and 10, respectively. We also compared the proposed BAS design to a design based on 12 replications of the complete factorial design on the 16 treatments, randomly allocating the last 8 available patients to treatments. For the factorial design, the treatment with the lowest value of the least square estimate of $\theta(\beta)$ was selected as the best treatment at the end of the trial. This factorial design strategy has been employed in previous trial designs, such as Thall *et al.* [1] and Schaid *et al.* [2]. We carried out 2000 simulations for each scenario.

Scenarios 1–4 simulated the cases in which there was a single best treatment. In scenario 1, the best (or most efficacious) treatment was T_1 (i.e., single agent A_1), and the BAS design substantially outperformed the factorial design. The selection probability of the target treatment T_1 under the BAS design was 86.1%, while under the factorial design, it was 64.9%. In addition, compared to the factorial design, the BAS design allocated a significantly higher percentage of patients to the best treatment (6.3% vs. 39.3%, respectively). The performance of the BAS design was rather similar to that of the optimal BAS_{susp} design. The selection probability of the target treatment under the BAS design was only 1.1% lower than that of BAS_{susp} design, and the percentage of patients allocated to the best treatment were almost same in two designs (39.3% vs. 39.1%), suggesting that randomization based on observed data provided an efficient way to handle delayed outcomes. In scenario 2, the best treatment was T_5 , the combination of agents A_1 and A_2 . In this case, the selection probability of the BAS design was 21.5% higher than that of the factorial design, and the BAS design assigned 32.8% more patients to the best treatment. In scenarios 3 and 4, the three-drug combination T_{11} (i.e., combination of A_1 , A_2 , and A_3) and the four-drug combination T_{15} were defined as the optimal treatments. Comparisons under these scenarios were similar to those made under scenarios 1 and 2. The selection probabilities of the BAS design were more than 18% higher than those of the factorial design, and the percentages of patients assigned to the best treatment under the BAS design were more than 33% higher than those under the factorial design. Again, we observed that the performance of the BAS design was rather similar to that of the BAS_{susp} design.

Scenarios 5–8 were designed to evaluate the performance of the design when there were two best treatments that were equally effective. In scenario 5, T_1 and T_2 were the target treatments with the highest efficacy. The BAS design selected T_1 with a probability of 46.7% and T_2 with a probability of 47.5%, whereas the factorial design selected these two targets with probabilities of 41.0% and 43.1%, respectively. That is, the total selection probability of T_1 and T_2 under the BAS design was 10.1% higher than that under the factorial design. The percentage of patients assigned to the best treatment using the BAS

design was 38.8% higher than that using the factorial design. For scenarios 6–8, the BAS design again outperformed the factorial design, achieving substantially higher selection probabilities and assigning higher percentages of patients to the optimal treatments. Scenarios 9 and 10 had three optimal treatments, and scenario 11 had four target treatments. Under these scenarios, the performance of the proposed BAS design once again dominated that of the factorial design. Compared to the factorial design, the total selection probabilities of the target treatments under the BAS design were improved by 0.7%–5.9%, and the percentages of patients assigned to the best treatments were improved by 39.8%–44.6%. Scenario 12 represents the case in which the treatment effects of all combinations are the same as that of the placebo. Under this scenario, the BAS design and the factorial design terminated the trial due to futility, with respective probabilities of 85.2% and 90.5%.

As demonstrated in the simulation study, the proposed BAS design achieved two important clinical goals simultaneously. First, it selected the best treatment arms with high probability. Second, it allocated more patients to the best treatments. This result seems somewhat surprising because the common notion is that these two goals are in conflict with each other. That is, it is often assumed that randomization schemes in which patients are allocated to effective treatments have less power to detect the best treatment at the end of the trial. This may be the case in comparisons of only two or three treatments, but in more complicated settings in which large numbers of treatments and treatment combinations are tested, the BAS offers significant gains in both power and patient allocation.

The success of our adaptive randomization scheme in accomplishing both goals simultaneously in high-dimension settings can be understood by noting that our design allocates more patients to the subset of treatments that are competitive. By reallocating patients away from ineffective treatments, we obtain higher power to distinguish between the top treatments. For example, in scenario 1, our design allocated 39.3 and 10.3 patients to the best and second best treatments (T_1 and T_{15}); in contrast, the factorial design allocated 6.3 patients to both T_1 and T_{15} . As a consequence, BAS had higher power to distinguish between T_1 and T_{15} .

Sensitivity analysis

The encompassing prior for β requires prespecification of the value for hyperparameter g . We conducted a sensitivity analysis to check the robustness of the design to the value of g . Specifically, we considered a tighter (or more informative) prior with $g = 5$ and a more diffused (or noninformative) prior with $g = 20$. Table 4 shows the results under scenarios 2, 4, 6, 8, and 10. Under each of these scenarios, the results with $g = 5$ or 20 were very similar to those reported in Table 2 (with $g = 10$), suggesting that the operating characteristics of the proposed design were not sensitive to the specification of g as long as it was reasonably diffuse. For example, in scenario 2, the selection probabilities of the target treatment, T_1 , were 87.3% and 87.1% under $g = 5$ and 20, respectively, which were very similar to that under $g = 10$ (86.6%). The percentages of patients assigned to T_1 were also very similar for $g = 5, 10$, and 20.

We conducted another sensitivity analysis to examine the performance of the proposed design when the outcome needs a longer assessment period to be evaluated. We assumed an assessment period of 60 days and an accrual rate of six patients per month. As shown in Table 5, the results were very similar to those reported in Table 2, in which the assessment period was 10 days with an accrual rate of 12 patients per month. This suggests that the proposed design is robust to the length of the assessment period and delayed outcomes.

Conclusion

We have proposed a Bayesian adaptive phase II screening design for trials combining multiple agents. Rather than testing each of the combinations independently, our design encompasses all the combinations of interest in a large screening trial. We model the main and synergistic effects of the treatment agents using a linear model and cast the screening problem into a Bayesian hypothesis testing problem. By using a factorial lead-in phase, we are able to quickly obtain preliminary estimates that each treatment combination is optimal, which enables us to quickly move into the adaptive phase of the algorithm. We utilize the encompassing prior with nonlocal constraints to accommodate the complex parameter constraints imposed by the hypotheses, and we continuously update the posterior probability that each treatment is best. Based on this posterior probability, we adaptively allocate patients to effective combinations and select the best treatment. The proposed design substantially outperformed a complete factorial design. Our design allocates more patients to better treatments, while providing higher power to identify the best treatment at the end of the trial.

The proposed design assumes that combined agents are administered simultaneously; therefore the order of the combination does not affect the treatment effect, for example, combination 'A + B' is equivalent to combination 'B + A'. However, if the combined agents are administered sequentially, the order of the treatment may impact the treatment effect. For example, the treatment effect of 'A + B' may be different from that of 'B + A'. This is particularly of concern if some of the agents in combination are therapeutical treatments, such as chemotherapy procedures or surgeries. In this case, we need to account for such an order effect in the dose–response model, for example, by including additional covariates indicating the order of the treatments.

Acknowledgments

We are grateful to the editor, associate editor, and two reviewers for their insightful comments, which substantially improved the article.

Funding

Valen Johnson's research was supported by Award Number R01 CA158113 and R01 CA026582 from the National Cancer Institute. Ying Yuan and Chunyan Cai's research was supported by Award Number R01 CA154591 from the National Cancer Institute. Yuan and Johnson were also supported by Core Grant P30 CA016672. Cai's research was also supported by the National Institutes of Health's Clinical and Translational Science Award grant (UL1 TR000371), awarded to the University of Texas Health Science Center at Houston in 2012 by the National Center for Clinical and Translational Sciences.

References

1. Thall PF, Simon R, Ellenberg SS. Two-stage selection and testing designs for comparative clinical trials. *Biometrika*. 1988; 75:303–10.
2. Schaid DJ, Wieand S, Therneau TM. Optimal two-stage screening designs for survival comparisons. *Biometrika*. 1990; 77:507–13.
3. Yao TJ, Begg CB, Livingston PO. Optimal sample size for a series of pilot trials of new agents. *Biometrics*. 1996; 52:992–1001. [PubMed: 8805764]
4. Yao TJ, Venkatraman ES. Optimal two-stage design for a series of pilot trials of new agents. *Biometrics*. 1998; 54:1183–89. [PubMed: 9750258]
5. Wang YG, Leung DHY. An optimal design for screening trials. *Biometrics*. 1998; 54:243–50. [PubMed: 9544519]
6. Stout QF, Hardwick J. Optimal screening designs with flexible cost and constraint structures. *J Stat Plan Infer*. 2005; 132:149–62.

7. Rossell D, Müller P, Rosner GL. Screening designs for drug development. *Biostatistics*. 2007; 8:595–608. [PubMed: 17030589]
8. Ding M, Rosner GL, Müller P. Bayesian optimal design for phase II screening trials. *Biometrics*. 2008; 64:886–94. [PubMed: 18162113]
9. Klugkist I, Kato B, Hoijtink H. Bayesian model selection using encompassing priors. *Stat Neerl*. 2005; 59:57–69.
10. Johnson VE, Rossell D. On the use of non-local prior densities in Bayesian hypothesis tests. *J Roy Stat Soc B Met*. 2010; 72:143–70.
11. Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc*. 1995; 90:773–95.
12. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann Intern Med*. 1999; 130:1005–13. [PubMed: 10383350]
13. Klugkist I, Hoijtink H. The Bayes factor for inequality and about equality constrained models. *Comput Stat Data Anal*. 2007; 51:6367–79.

Appendix A. Derivation of Bayes factor under the encompassing prior

Under model (1) and the encompassing prior (6), the marginal density of data $\mathbf{y}$ under hypothesis H_j is given by

$$\begin{aligned} m_j &= \iint f(\mathbf{y}|\beta, \sigma^2) \pi_j(\beta|\sigma^2) \pi(\sigma^2) d\sigma^2 d\beta \\ &= \iint N(\mathbf{y}|\mathbf{X}\beta, \sigma^2 \mathbf{I}_n) \frac{1}{c_j} N(\beta|\mathbf{0}, g\sigma^2 \mathbf{I}_p) I_{\min}(\theta_j(\beta)) \frac{1}{\sigma^2} d\sigma^2 d\beta \end{aligned}$$

where

$$c_j = \int N(\beta|\mathbf{0}, g\sigma^2 \mathbf{I}_p) I_{\min}(\theta_j(\beta)) d\beta$$

Letting $\beta' = \beta/\sigma$, it follows that the normalizing constant

$$c_j = \int N(\beta'|\mathbf{0}, g\mathbf{I}_p) I_{\min}(\theta_j(\sigma\beta')) d\beta'$$

Recall that $\theta_j(\beta)$, $j = 0, \dots, p-1$, is a linear function of β , thus $I_{\min}(\theta_j(\sigma\beta')) = I_{\min}(\sigma\theta_j(\beta'))$. Since $\sigma > 0$, the order of $\{\sigma\theta_j(\beta')\}$ is the same as that of $\{\theta_j(\beta')\}$. Thus

$$I_{\min}(\theta_j(\sigma\beta')) = I_{\min}(\theta_j(\beta'))$$

and

$$c_j = \int N(\beta'|\mathbf{0}, g\mathbf{I}_p) I_{\min}(\theta_j(\beta')) d\beta'.$$

We can see that c_j is independent of σ^2 , which greatly simplifies the evaluation of the marginal density of $\mathbf{y}$. Then it follows that

$$\begin{aligned}
m_j &= \iint N(\mathbf{y}|\mathbf{X}\beta, \sigma^2 \mathbf{I}_n) \frac{1}{c_j} N(\beta|\mathbf{0}, g\sigma^2 \mathbf{I}_p) I_{\min}(\theta_j(\beta)) \frac{1}{\sigma^2} d\sigma^2 d\beta \\
&= \frac{1}{c_j} \iint N(\mathbf{y}|\mathbf{X}\beta, \sigma^2 \mathbf{I}_n) N(\beta|\mathbf{0}, g\sigma^2 \mathbf{I}_p) \frac{1}{\sigma^2} I_{\min}(\theta_j(\beta)) d\sigma^2 d\beta \\
&= \frac{\Gamma((n+p)/2)}{c_j g^{p/2} \pi^{(n+p)/2}} \int \left((\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) + \frac{1}{g} \beta^T \beta \right)^{-(n+p)/2} I_{\min}(\theta_j(\beta)) d\beta \\
&= \frac{\Gamma(n/2) \sqrt{|\mathbf{V}|}}{c_j g^{p/2} \pi^{n/2} (\mathbf{y}^T \mathbf{y} - \mu^T \mathbf{V}^{-1} \mu)^{n/2}} \int t_\nu(\Sigma, \mu) I_{\min}(\theta_j(\beta)) d\beta
\end{aligned}$$

Therefore, the Bayes factor of H_i to H_j , B_{ij} , is given by

$$B_{ij} = \frac{m_i(\mathbf{y})}{m_j(\mathbf{y})} = \frac{r_i/c_i}{r_j/c_j}$$

where

$$r_j = \int t_\nu(\Sigma, \mu) I_{\min}(\theta_j(\beta)) d\beta$$

Table 1

The 16 combinations of four agents (A_1 , A_2 , A_3 , and A_4) investigated in the lung cancer trial

Treatment	T_0	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}	T_{13}	T_{14}	T_{15}
A_1	0	1	0	0	0	1	1	1	0	0	0	1	1	1	0	1
A_2	0	0	1	0	0	1	0	0	1	1	0	1	1	0	1	1
A_3	0	0	0	1	0	0	1	0	1	0	1	1	0	1	1	1
A_4	0	0	0	0	1	0	0	1	0	1	1	0	1	1	1	1

Table 2

Operating characteristics of the proposed BAS, BAS_{susp}, and factorial designs. The efficacious treatments are boldfaced

Treatment effect				Selection percentage of each treatment											
θ_0	θ_1	θ_2	θ_3	BAS					BAS _{susp}			Factorial design			
θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}	θ_{12}	θ_{13}	θ_{14}	θ_{15}				
Scenario 1															
0	-25	-13	-12	0.0	86.1	0.0	0.0	0.0	87.2	0.0	0.0	0.0	64.9	0.0	0.0
-11	-20	-19	-18	0.0	2.5	1.4	0.5	0.0	2.4	1.0	0.2	0.0	8.9	4.4	2.6
-14	-15	-12	-10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.0	0.0
-16	-17	-14	-21	0.2	0.2	0.1	6.2	0.1	0.2	0.0	6.0	0.2	1.6	0.3	16.1
Scenario 2															
0	-20	-12	-13	0.0	2.8	0.0	0.0	0.0	2.8	0.0	0.1	0.0	10.0	0.0	0.0
-11	-25	-19	-18	0.0	86.6	1.4	0.5	0.0	85.4	1.5	0.5	0.0	65.1	4.2	3.1
-14	-15	-12	-10	0.0	0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.2	0.5	0.0	0.0
-16	-17	-14	-21	0.2	0.3	0.0	5.5	0.2	0.2	0.0	6.6	0.4	1.6	0.2	14.2
Scenario 3															
0	-10	-12	-13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1
-11	-20	-19	-18	0.0	3.3	1.2	0.6	0.0	2.5	1.1	0.7	0.0	8.6	4.5	2.9
-14	-15	-12	-25	0.0	0.0	0.0	86.0	0.0	0.0	0.0	87.5	0.1	0.2	0.1	67.6
-14	-17	-14	-21	0.1	0.1	0.2	6.2	0.0	0.2	0.0	5.7	0.2	1.3	0.1	13.5
Scenario 4															
0	-16	-12	-13	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.7	0.0	0.0
-11	-20	-19	-18	0.0	3.5	1.0	0.2	0.0	2.1	1.0	0.5	0.0	9.1	4.1	2.7
-14	-15	-12	-10	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.2	0.4	0.0	0.0
-16	-21	-14	-25	0.0	5.7	0.0	86.8	0.1	6.5	0.0	86.5	1.0	14.4	0.1	66.6
Scenario 5															
0	-25	-25	-13	0.0	46.7	47.5	0.0	0.0	46.1	48.8	0.0	0.0	41.0	43.1	0.0
-11	-20	-19	-18	0.0	1.0	0.2	0.2	0.0	0.6	0.3	0.0	0.0	4.1	2.0	1.4

Treatment effect												Selection percentage of each treatment															
θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}	BAS				BAS _{susp}				Factorial design							
θ_{12}	θ_{13}	θ_{14}	θ_{15}																								
-14	-15	-12	-10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
-16	-17	-14	-21	0.0	0.1	0.0	2.5	0.0	0.1	0.0	2.3	0.4	0.5	0.1	6.8												
Scenario 6																											
0	-25	-12	-13	0.0	46.2	0.0	0.0	0.0	47.0	0.0	0.0	0.0	43.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
-11	-25	-19	-18	0.0	49.1	0.2	0.1	0.0	48.8	0.2	0.2	0.2	43.0	1.8	1.1												
-14	-15	-12	-10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
-16	-17	-14	-21	0.1	0.0	0.0	1.7	0.0	0.0	0.0	1.5	0.2	0.9	0.0	8.3												
Scenario 7																											
0	-25	-12	-13	0.0	46.3	0.0	0.0	0.0	47.9	0.0	0.0	0.0	42.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
-11	-20	-19	-18	0.0	0.8	0.4	0.2	0.0	0.8	0.2	0.2	0.2	4.3	2.1	0.9												
-14	-15	-12	-25	0.0	0.0	0.0	48.5	0.0	0.0	0.0	46.8	0.0	0.0	0.0	41.3												
-16	-17	-14	-21	0.0	0.0	0.0	1.8	0.0	0.2	0.0	2.0	0.2	0.7	0.0	7.7												
Scenario 8																											
0	-25	-12	-13	0.0	47.5	0.0	0.0	0.0	49.4	0.0	0.0	0.0	43.8	0.0	0.2												
-11	-20	-19	-18	0.0	0.8	0.5	0.1	0.0	0.4	0.0	0.2	0.2	5.1	2.5	1.8												
-14	-15	-12	-10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0												
-16	-17	-14	-25	0.1	0.1	0.0	48.7	0.0	0.0	0.0	48.3	0.2	0.9	0.1	44.5												
Scenario 9																											
0	-25	-25	-13	0.0	31.3	31.5	0.0	0.0	33.4	30.9	0.0	0.0	31.1	32.0	0.0												
-11	-25	-19	-18	0.0	34.1	0.3	0.0	0.0	33.4	0.2	0.0	0.0	29.2	1.2	0.5												
-14	-15	-12	-10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0												
-16	-17	-14	-21	0.1	0.0	0.0	1.0	0.0	0.0	0.0	0.9	0.2	0.4	0.0	4.7												
Scenario 10																											
0	-25	-25	-25	0.0	30.9	33.7	32.2	0.0	30.6	34.1	31.3	0.0	30.4	30.9	29.6												
-11	-20	-19	-18	0.0	0.4	0.5	0.2	0.0	0.6	0.5	0.2	0.0	2.5	1.2	0.4												
-14	-15	-12	-10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0												

Treatment effect										Selection percentage of each treatment									
θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}	BAS		BAS _{susp}			Factorial design		
θ_{12}	θ_{13}	θ_{14}	θ_{15}																
-16	-17	-14	-21	0.0	0.0	0.0	1.1	0.0	0.0	0.0	1.1	0.0	0.0	0.0	1.1	0.1	0.3	0.0	4.2
Scenario 11																			
0	-25	-12	-13	0.0	23.5	0.0	0.0	0.0	26.3	0.0	0.0	0.0	0.0	0.0	25.9	0.0	0.0	0.0	0.0
-11	-25	-19	-18	0.0	25.7	0.2	0.2	0.0	25.4	0.2	0.0	0.0	0.0	0.0	22.7	0.8	0.8	0.6	0.6
-14	-15	-12	-25	0.0	0.0	0.0	0.0	24.7	0.0	0.0	0.0	24.6	0.0	0.0	24.6	0.0	0.0	0.0	25.2
-16	-17	-14	-25	0.0	0.0	0.0	0.0	24.6	0.0	0.0	0.1	22.7	0.0	0.2	0.0	0.2	0.0	0.0	24.0
Scenario 12																			
0	0	0	0	0.0	0.9	1.1	0.6	0.0	0.8	1.1	0.9	0.0	0.8	0.5	0.4				
0	0	0	0	1.0	0.8	1.5	1.4	1.0	0.8	0.8	1.0	0.4	1.0	0.4	0.8				
0	0	0	0	1.0	0.8	0.8	1.1	1.0	1.0	1.2	0.8	0.8	0.5	1.0	0.6				
0	0	0	0	0.8	1.1	0.8	1.1	1.1	1.1	1.2	1.1	0.7	0.4	0.4	0.8				

BAS: Bayesian adaptive screening; BAS_{susp}: Bayesian adaptive screening design based on the accrual-suspension approach.

Summary of the simulation results, including the total selection percentage of target treatments and percentage of patients treated at the target treatments

Table 3

Scenario	1	2	3	4	5	6	7	8	9	10	11	12
	Total selection percentage of target treatments											
BAS	86.1	86.6	86.0	86.8	94.2	95.3	94.8	96.2	96.9	96.8	98.5	85.2 ^a
BAS _{Susp}	87.2	85.4	87.5	86.5	94.9	95.8	94.7	97.7	97.7	96.0	99.0	85.1 ^a
Factorial design	64.9	65.1	67.6	66.6	84.1	86.9	83.4	88.3	92.3	90.9	97.8	90.5 ^a
	Total percentage of patients treated at target treatments											
BAS	39.3	39.0	39.2	39.8	51.3	54.7	51.2	55.9	61.5	58.6	69.6	
BAS _{Susp}	39.1	38.7	40.4	40.0	51.8	54.6	51.2	56.8	61.6	58.5	69.5	
Factorial design	6.3	6.2	6.2	6.3	12.5	12.5	12.5	12.5	18.8	18.8	25.0	

BAS: Bayesian adaptive screening; BAS_{Susp}: Bayesian adaptive screening design based on the accrual-suspension approach.

^aThe percentage of trials terminated due to futility.

Table 4

Sensitivity analysis with different values of g under scenarios 2, 4, 6, 8, and 10 for the proposed Bayesian adaptive screening (BAS) design. The efficacious treatments are boldfaced

Scenario	$g = 5$			$g = 20$		
	0.0	2.6	0.0	0.0	2.8	0.0
2	0.0	87.3	1.1	0.5	87.1	1.2
	0.0	0.0	0.0	0.0	0.0	0.0
	0.1	0.2	0.0	5.9	0.2	0.3
4	0.0	0.1	0.0	0.0	0.1	0.0
	0.0	2.5	0.7	0.2	0.0	2.9
	0.0	0.0	0.0	0.0	0.1	0.0
6	0.2	5.0	0.0	88.2	0.2	5.9
	0.0	48.4	0.0	0.0	45.1	0.0
	0.0	46.5	0.4	0.2	0.0	50.5
8	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.1	0.0	1.8	0.0	0.0
	0.0	49.5	0.0	0.0	48.6	0.0
10	0.0	0.8	0.4	0.2	0.0	0.6
	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.2	0.0	47.0	0.0	0.0
	0.0	32.2	32.1	32.1	0.0	32.9
	0.0	0.4	0.4	0.2	0.0	0.5
	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	1.2	0.0	0.0

Table 5

Sensitivity analysis for the proposed BAS design with an assessment period of 60 days and an accrual rate of 6 patients per month. The efficacious treatments are boldfaced

Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5		Scenario 6		Scenario 7		Scenario 8		Scenario 9		Scenario 10		Scenario 11		Scenario 12																									
0.0	86.5	0.0	0.0	2.8	0.0	0.0	0.0	0.0	0.0	48.5	45.6	0.0	0.0	47.9	0.0	0.0	47.2	0.0	0.0	31.9	31.9	0.0	0.0	32.3	32.0	31.8	0.0	24.2	0.0	0.1	0.0	1.0	1.0	1.2													
0.0	3.2	1.1	0.5	0.0	85.3	1.4	0.6	0.0	2.7	1.3	0.2	0.0	2.8	1.1	0.9	0.0	0.7	0.3	0.4	0.0	47.0	0.3	0.4	0.0	0.8	0.6	0.4	0.0	0.8	0.2	0.2	0.0	0.6	0.4	0.1	0.0	24.6	0.4	0.2	0.9	1.1	1.2	1.1				
0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	25.0	1.2	1.0	1.1	1.1				
0.2	0.3	0.0	5.3	0.1	0.3	0.2	5.7	0.0	0.2	0.2	6.6	0.2	6.2	0.0	84.8	0.0	0.2	0.0	1.7	0.0	0.0	0.0	1.6	0.0	0.0	0.0	2.6	0.1	0.1	0.0	47.8	0.0	0.2	0.0	1.3	0.0	0.0	0.0	1.4	0.0	0.1	0.0	24.5	0.9	1.0	1.1	1.1