View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by IUPUIScholarWorks

A Bayesian Adaptive Phase II Clinical Trial Design Accounting for Spatial Variation

Beibei Guo[1*] and Yong Zang[2,3]

[1]*Department of Experimental Statistics, Louisiana State University*

Baton Rouge, LA 70803, U.S.A.

∗ *email: beibeiguo@lsu.edu*

[2]*Department of Biostatistics, School of Medicine, Indiana University*

[3]*Center for Computational Biology and Bioinformatics, Indiana University*

Indianapolis, IN 46202, U.S.A.

For Peer Review

Abstract

Conventional phase II clinical trials evaluate the treatment effects under the assumption of patient homogeneity. However, due to inter-patient heterogeneity, the effect of a treatment may differ remarkably among subgroups of patients. Besides patient's individual characteristics such as age, gender, and biomarker status, a substantial amount of this heterogeneity could be due to the spatial variation across geographic regions because of unmeasured or unknown spatially varying environmental and social exposures. In this article, we propose a hierarchical Bayesian design for two-arm randomized phase II clinical trials that accounts for the spatial variation as well as patient's individual characteristics. We treat the treatment efficacy as an ordinal outcome and quantify the desirability of each possible category of the ordinal efficacy using a utility function. A cumulative probit mixed model is used to relate efficacy to patient-specific covariates and geographic region spatial effects. Spatial dependence between regions is induced through the conditional autoregressive priors on the spatial effects. A two-stage design is proposed to adaptively assign patients to desirable treatments and make treatment recommendations at the end of the trial according to each patient's

This is the author's manuscript of the article published in final edited form as:
Guo, B., & Zang, Y. (2018). A Bayesian adaptive phase II clinical trial design accounting for spatial variation. Statistical Methods in Medical Research, 0962280218797149. https://doi.org/10.1177/0962280218797149

spatial information and individual covariates. Simulation studies show that our proposed design has good operating characteristics and significantly outperforms an alternative phase II trial design that ignores the spatial variation.

KEY WORDS: personalized medicine; spatial variation; conditionally autoregressive; Bayesian adaptive design, phase II trial; Markov random field.

For Peer Review

1 Introduction

Increased understanding of the mechanistic heterogeneity of cancer has brought us to the era of precision medicine in clinical oncology. Under the paradigm of precision medicine, the effect of a treatment may differ significantly between patients due to inter-patient heterogeneity. Specifically, some drugs may benefit only 2% of the patients and there are even drugs that are harmful to certain ethnic groups [1]. Therefore, the conventional clinical trials that assume patient homogeneity are inappropriate to evaluate the individual level treatment effect. Indeed, to overcome the "one-size-fits-all" approach and to acknowledge patient heterogeneity, in January 2015, a new 215-million national Precision Medicine Initiative was launched [2]. Precision medicine provides an unbeatable opportunity for disease treatment and prevention that take into account each person's individual variability in genes, environment, and lifestyle.

Spatial variation is an important cause of patient heterogeneity in addition to clinical factors and biomarkers. Specifically, patients living in nearby areas are often exposed to similar physical and social environments such as air pollution, sunlight exposure, climatic factors [3], and neighborhood conditions [4, 5], so they tend to have similar spatial effects. On the other hand, patients living in areas that are far apart can be exposed to very different environments, so the spatial variation may be substantial. Spatial pattern has been observed in survival rate and incidence of various diseases such as cancer. For example, mortality rate and incidence of lung cancer are known to be associated with various environmental risk factors and vary across geographic regions [3, 6, 7, 8]. Henderson, Shimakura, and Gorst [9] demonstrated that there is good evidence that spatial dependence exists in survival rates of acute myeloid leukemia patients in northwest England. In a study of prevalence and morbidity of childhood asthma, Li and Ryan [10] showed that accounting for the spatial correlation improved the model. Therefore, it is reasonable to believe that the treatment effects may also vary substantially due to spatial variation, which is in turn determined by the living pattern of the patients. In fact, a reanalysis of the data from a clinical trial for the

For Peer Review

killed oral cholera vaccine did reveal that the spatial variation of the force of infection and vaccine coverage significantly affected the vaccine effectiveness [11]. Consequently, a clinical trial evaluating the personalized treatment effect should integrate the spatial variation into consideration, in addition to patient's individual covariates. Moreover, as the spatial effects are correlated between regions with the degree of correlation determined by the distance between regions, novel statistical methods are required to capture this spatial correlation.

Traditional phase II trials aim to examine the potential efficacy of a new treatment based on the assumption of patient homogeneity [12, 13, 14, 15, 16, 17]. However, in the presence of patient heterogeneity, this approach may draw incorrect conclusions. To account for patient heterogeneity, several phase II clinical trial designs have been proposed that take covariates such as patient prognostic covariates and disease subtypes into consideration. For example, Ibrahim, Ryan, and Chen [18] proposed to incorporate historical control information to adjust for covariate effects in trend tests for binary data. Thall, Sung, and Estey [19] described a statistical design that accommodates patient prognostic covariates as well as historical data in multicourse clinical trials. Thall et al. [20] presented a hierarchical Bayesian design for phase II clinical trials for diseases with multiple subtypes. Wathen et al. [21] proposed a design for single-arm phase II clinical trials that accounts for patient heterogeneity according to multiple prognostic subgroups. However, to the best of our knowledge, no phase II trial design has been developed to account for the spatial variation across geographic regions.

Our study was motivated by a phase II two-arm randomized clinical trial for Non-small cell lung cancer (NSCLC) patients, which is being conducted at the Indiana University Melvin and Bren Simon Cancer Center. A maximum of 100 patients will be enrolled and randomized to receive either the standard radiation therapy or a PI3K inhibitor plus radiation therapy. The outcome of interest is an ordinal variable indicating whether a patient reports partial or complete remission (PR/CR), stable disease (SD), or progressive disease (PD). The challenge of conducting this trial is that patients in the trial come from 12 different counties. Counties close to each other tend to share similar environments and living

For Peer Review

conditions while counties far apart could have quite distinct environments. Therefore, the spatial variation should be considered for trial conduction and data analysis.

In this article, we propose a Bayesian adaptive phase II clinical trial design that accommodates spatial variation and patient's individual characteristics. The objective of the proposed design is to determine whether the experimental treatment is superior compared with a standard treatment for each patient according to his/her spatial information and other individual covariates. To account for the spatial variation, we introduce spatially structured random effects with a conditional autoregressive prior distribution so that the random effects corresponding to geographic regions that are close to each other tend to have similar magnitude. These random effects can be interpreted as surrogates for unmeasured or unknown region-level exposure variables such as air pollution, which typically exhibit spatial pattern. In addition, rather than assuming a binary response outcome as in conventional phase II trials, we use ordinal efficacy outcomes. This is advantageous because the efficacy outcome is often scored as a multi-level ordinal variable such as PD, SD, PR, and CR in a real phase II trial, and dichotomizing the multi-level efficacy outcome into a binary variable looses information [22, 23]. This is especially true for molecularly targeted agents. The binary efficacy outcome in a traditional phase II trial typically defines CR/PR as a response, and PD/SD as a nonresponse. That is sensible for conventional cytotoxic agents that work by shrinking the tumor. However, due to a different functional mechanism, molecularly targeted agents function by stabilizing the tumor rather than shrinking it. So for these agents, SD is often regarded as a positive outcome because some targeted agents prolong survival by achieving durable SD without notable tumor progression. So although SD is less favorable than CR/PR, it should be distinguished from PD. We use a utility function elicited from physicians to quantify the desirability of the treatments assuming an ordinal efficacy outcome. Based on the proposed Bayesian spatial model and utility function, we develop a two-stage phase II clinical trial design with a personalized randomization scheme to guide treatment assignment and selection.

For Peer Review

The remainder of this article is organized as follows. In section 2, we present the probability model for the ordinal efficacy outcome and the prior distribution specification. Section 3 describes a two-stage design to assign patients to treatment arms based on utility and make final recommendations. Section 4 examines the operating characteristics of the proposed design through simulation studies, and concluding remarks are provided in Section 5.

2 Method

2.1 Probability models

Consider a two-arm phase II clinical trial comparing a new treatment with a standard treatment. Let Y be the ordinal efficacy outcome with $Y = 1, 2, \cdots, K$ representing increasingly desirable efficacy. Suppose the study area is subdivided into J geographic regions, which could be residential districts, clinical sites, cities, counties, states, etc, that are spatially arranged and potentially subject to spatial correlation. Let $\mathbf{X} = (X_1, \cdots, X_q)'$ denote a vector of q patient-specific covariates such as age, gender, and biomarker status. Suppose at an interim decision time, n_j patients have been enrolled in geographic region j $(j = 1, \cdots, J)$. The data from patient i $(i = 1, \cdots, n_j)$ in region j take the form $(Y_{ij}, \mathbf{X}_{ij}, Z_{ij})$, where Z_{ij} denotes the treatment indicator with $Z_{ij} = 1$ or 0 if he/she receives the experimental or the control arm, respectively. We model Y_{ij} using the cumulative probit mixed regression model. For $k = 1, \cdots, K-1$,

$$\Phi^{-1}\big(\Pr(Y_{ij} \leq k | \mathbf{X}_{ij}, Z_{ij}, \theta_j)\big) = \gamma_k - \beta_0 - \beta_1 Z_{ij} - \boldsymbol{\beta}_2 \mathbf{X}_{ij} - \boldsymbol{\beta}_3 Z_{ij} \mathbf{X}_{ij} - \theta_j \qquad (1)$$

where $-\infty = \gamma_0 < \gamma_1 < \cdots < \gamma_{K-1} < \gamma_K = \infty$ are boundaries, β_0 is an intercept, β_1 is the treatment main effect, $\boldsymbol{\beta}_2 = (\beta_{2,1}, \cdots, \beta_{2,q})$ is the vector of covariate main effects, $\boldsymbol{\beta}_3 = (\beta_{3,1}, \cdots, \beta_{3,q})$ is the vector of interaction effects between the treatment and covariates, and θ_j is the region-specific random effect for region j, which will be discussed in detail

https://mc.manuscriptcentral.com/smmr

later. To ensure that the parameters are identifiable, we impose one restriction $\gamma_1 = 0$. The cumulative probit model has a latent variable formulation. Specifically, there exists an unobserved latent variable

$$Y_{ij}^* \sim N(\beta_0 + \beta_1 Z_{ij} + \boldsymbol{\beta}_2 \mathbf{X}_{ij} + \boldsymbol{\beta}_3 Z_{ij} \mathbf{X}_{ij} + \theta_j, 1) \tag{2}$$

for patient i in region j such that $Y_{ij} = k$ if $\gamma_{k-1} < Y_{ij}^* \leq \gamma_k$. This latent variable approach facilitates the Bayesian inference under standard priors [24].

2.2 Prior specification

The region-specific random effects $\theta_j's$ can be interpreted as surrogates for unmeasured or unknown region-level covariates such as air pollution, neighborhood conditions, and social environments, which typically exhibit spatial pattern. If some of these region-level covariates could be measured, they should certainly be included in the regression model. However, these covariates are typically difficult or even impossible to measure [10]. Since neighboring regions tend to share similar environmental and social factors, it is reasonable to believe that random effects $\theta_j's$ corresponding to regions that are close to each other tend to have similar magnitude. Our approach aims to assign a prior distribution to $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_J)$ to achieve this spatial correlation.

We define two regions as neighbors if they share a common boundary. Different regions can have different numbers of neighbors. Consider the example illustrated in Figure 1 where there are 12 geographic regions. Region 1 has two neighbors, regions 2 and 5; region 2 has three neighbors, regions 1, 3, and 6; and region 6 has four neighbors, regions 2, 5, 7, and 10. We model the spatial correlation among regions by assigning $\boldsymbol{\theta}$ a Gaussian Markov random field prior distribution. We adopt the most popular intrinsic conditional autoregressive

For Peer Review

(CAR) models [25] with joint prior distribution of $\boldsymbol{\theta}$ given by:

$$p(\boldsymbol{\theta}|\sigma) \propto \exp\Big\{ -\frac{1}{2\sigma}\sum_{j\sim j'}(\theta_j - \theta_{j'})^2\Big\} \tag{3}$$

where $j \sim j'$ denotes that region j and region j' are neighbors. The hyperparameter $\sigma > 0$ controls the degree of smoothness. Small values of σ induce smoother realizations of the spatial effect. This prior is often denoted as $\boldsymbol{\theta} \sim \text{CAR}(\sigma)$. Under this prior, the distribution of each element θ_j for region j given all the other elements $\{\theta_{j^*}, j^* \neq j\}$ depends only on its neighbors. To see this, the conditional distribution of θ_j given all the other spatial effects follows a normal distribution

$$\theta_j|\{\sigma, \theta_{j^*}, j^* \neq j\} \sim N\Big(\frac{1}{\kappa_j}\sum_{j\sim j'}\theta_{j'}, \frac{\sigma}{\kappa_j}\Big) \tag{4}$$

where κ_j is the number of neighbors of region j. Since the conditional mean of θ_j is the average of the effects of its neighbors, neighboring $\theta_j's$ will tend to have similar values. Also the conditional variance $\frac{\sigma}{\kappa_j}$ is inversely proportional to the number of neighbors, which allows greater variability at regions with fewer neighbors and less variability at regions with more neighbors. This is appropriate for spatially correlated random effects because if a region has more neighbors, then there is more information about the value of its random effect from its neighbors. This CAR prior is improper because adding an arbitrary constant to each θ_j will not change the joint density (3). To address this issue, we add the constraint $\sum_{j=1}^{J}\theta_j = 0$ to guarantee identifiability and propriety of the prior distribution [26, 27, 5].

Now let's turn to the prior specification of the other parameters. We assign a conjugate Inverse Gamma prior to the hyperparameter σ. Since there may be little information about the spatial correlations between regions before we observe any data, we use a vague Inverse Gamma distribution $\sigma \sim \text{Inv-Gamma}(0.01, 0.01)$. To assign priors to the regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3)$, we follow a principle similar to that of Gelman et al. [28]. The basic idea is that a typical change in a covariate is unlikely to lead to a dramatic change

For Peer Review

in the probability of the response variable. Under our probit model, a change of 2.5 on the probit scale moves a probability of efficacy in one category from 0.1 to 0.89. It is typically reasonable to assume that the effect of a covariate is unlikely to be more dramatic than that. This is particularly true for clinical trials where the target efficacy probability is rarely outside that range. We first standardize each covariate. Binary covariates are scaled to have a mean of 0 and to differ by 1 in the two categories, and other covariates are scaled to have a mean of 0 and a standard deviation of 0.5. Each element of $\boldsymbol{\beta}$ is assigned an independent normal prior distribution $N(0, 1.25^2)$. Under this prior, a change in a covariate from one standard deviation below the mean to one standard deviation above the mean will result in a change mostly within the range of $(0, 2.5)$ on the probit scale. Finally, each element of the boundaries $\boldsymbol{\gamma} = \{\gamma_k, k = 2, \cdots, K-1\}$ has the interpretation of the increment in intercept when adding additional categories in the cumulative probability in the probit model. Hence, based on the same consideration as for $\boldsymbol{\beta}$, we assign each γ_k a uniform prior distribution U(0,2.65) so that *a priori*, there is a 95% probability that γ_k is between 0 and 2.5. In addition, we restrict the joint prior distribution of $\boldsymbol{\gamma}$ on the subspace $\gamma_2 < \cdots < \gamma_{K-1}$. For more details, please refer to Guo and Yuan [23].

2.3 Likelihood and Posterior

Letting $\Theta = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma, \boldsymbol{\theta}, \boldsymbol{Y}^*)$ denote all unobservables, where $\boldsymbol{Y}^* = \{Y_{ij}^*, i = 1, \cdots, n_j, j = 1, \cdots, J\}$ denote the set of all latent variables, and $\mathcal{D} = \big\{(Y_{ij}, \mathbf{X}_{ij}, Z_{ij}), i = 1, \cdots, n_j, j = 1, \cdots, J\big\}$ denote the data from the n_j patients in region j, for $j = 1, \cdots, J$, when an interim decision is to be made, the likelihood for Θ is given by

$$L(\Theta; \mathcal{D}) = \prod_{j=1}^{J} \prod_{i=1}^{n_j} \left[\Big\{ \sum_{k=1}^{K} 1(Y_{ij} = k) 1(\gamma_{k-1} < Y_{ij}^* \le \gamma_k) \Big\} \times \phi(Y_{ij}^*; \beta_0 + \beta_1 Z_{ij} + \boldsymbol{\beta}_2 \mathbf{X}_{ij} + \boldsymbol{\beta}_3 Z_{ij} \mathbf{X}_{ij} + \theta_j, 1) \right] \quad (5)$$

For Peer Review

where $1(A)$ is the indicator function that equals 1 if event A is true and 0 otherwise, and $\phi(\ ;\mu,\sigma^2)$ is the density of the normal distribution with mean μ and variance σ^2. The joint posterior distribution of Θ is

$$p(\Theta|D) \propto L(\Theta;\mathcal{D})p(\boldsymbol{\theta}|\sigma)p(\boldsymbol{\beta})p(\boldsymbol{\gamma})p(\sigma) \tag{6}$$

We use the Gibbs sampler [29] to draw posterior inference.

3 Decision Criteria

Conventional phase II trials with binary efficacy outcomes base decision rules on response rates, e.g., the probability of tumor response. For an ordinal efficacy outcome with three or more categories, one practical approach is to dichotomize efficacy into a binary variable (e.g., response if CR or PR and nonresponse if PD or SD) and then use conventional decision rules. However, this approach suffers from several drawbacks mentioned earlier such as loss of information and ignorance of the mechanisms of the treatment, e.g., SD, although not as desirable as CR/PR, is often regarded as a favorable outcome as many treatments prolong survival without much tumor shrinkage. To get a more informative and meaningful summary of the ordinal outcome, we propose to make decisions based on a utility function elicited from physicians that characterizes the importance/desirability of each category of efficacy. To illustrate the utility function, we assume the efficacy outcome is categorized as a 3-level ordinal variable, with $Y = 1, 2, 3$ representing PD, SD, and PR/CR, respectively. We elicit from physicians three utility values $\omega_1, \omega_2, \omega_3$ representing the desirability of the 3 categories $Y = 1, 2, 3$, respectively. The three utilities must satisfy $0 = \omega_1 < \omega_2 < \omega_3$. One example is the following

For patients in region j with covariates $\mathbf{X}$, we define the true utility for treatment arm

https://mc.manuscriptcentral.com/smmr
For Peer Review

	Y=1	Y=2	Y=3
ω	0	50	100

Z as

$$U_{\text{true}}(Z, \mathbf{X}, j) = \sum_{k=1}^{3} \omega_k \Pr(Y = k | Z, \mathbf{X}, j) \tag{7}$$

The experimental arm is considered superior to the control arm for patients in region j with covariates $\mathbf{X}$ if the true utility increases by a certain percentage δ, that is,

$$\Delta(\mathbf{X}, j) = \frac{U_{\text{true}}(Z = 1, \mathbf{X}, j) - U_{\text{true}}(Z = 0, \mathbf{X}, j)}{U_{\text{true}}(Z = 0, \mathbf{X}, j)} > \delta, \tag{8}$$

where δ is the minimum meaningful increment in utility and is pre-specified by the physicians. Note that $\Delta(\mathbf{X}, j)$ depends on covariates $\mathbf{X}$ and region j, so whether the experimental agent is promising depends on a patient's individual covariates and spatial information.

During the trial, given the interim data $\mathcal{D}_n$ collected from n patients, let

$$U_n(Z, \mathbf{X}, j) = \sum_{k=1}^{3} \omega_k \Pr(Y = k | Z, \mathbf{X}, j, \mathcal{D}_n) \tag{9}$$

denote the posterior estimate of the utility of treatment Z for patients with covariates $\mathbf{X}$ in region j, and let

$$\Delta_n(\mathbf{X}, j) = \frac{U_n(Z = 1, \mathbf{X}, j) - U_n(Z = 0, \mathbf{X}, j)}{U_n(Z = 0, \mathbf{X}, j)} \tag{10}$$

denote the posterior estimate of the corresponding percentage of increase in utility. Our decision criteria will be based on $\Delta_n(\mathbf{X}, j)$.

For a phase II trial that accounts for spatial variation and patient heterogeneity, decision making is difficult at the beginning of the trial, because estimates of the efficacy probabilities based on a small number of patients are highly unstable in the presence of patient heterogeneity and spatial effects. To alleviate this issue, we propose a two-stage design. Let the maximum sample size be N. In Stage I, we enroll N_1 patients and measure their covariates

For Peer Review

and residential regions. Patients in Stage I are equally randomized to the control and the experimental arms. The purpose of this stage is to collect preliminary data to facilitate the model fitting in stage II. In stage II, we continue to enroll the remaining $N_2 = N - N_1$ patients. Since whether the experimental treatment is promising for a patient depends on his/her unique individual covariates and residential region, we treat patients one by one in stage II and continuously update the posterior utilities to determine the treatment assignment for each patient. Specifically, assuming that $n-1$ patients have been enrolled in the trial, we assign a treatment to the nth patient with covariates $\mathbf{X}$ in region j as follows.

1. If $\Pr\big(\Delta_{n-1}(\mathbf{X}, j) > \delta\big) > c_1$, we assign this patient to the experimental arm.

2. If $\Pr\big(\Delta_{n-1}(\mathbf{X}, j) > \delta\big) < c_2$, we assign this patient to the control arm.

3. Otherwise, we use adaptive randomization to assign this patient to the experimental or the control arm with probabilities $\Pr\big(\Delta_{n-1}(\mathbf{X}, j) > \delta\big)$ and $\Pr\big(\Delta_{n-1}(\mathbf{X}, j) \leq \delta\big)$, respectively.

4. We repeat steps 1-3 until we reach the maximum sample size of N_2.

In the design algorithm, $0 \leq c_1, c_2 \leq 1$ are probability cutoffs with c_1 typically large (e.g., 0.9) and c_2 typically small (e.g., 0.1). These cutoffs will be tuned through simulation studies to achieve good design operating characteristics. Indeed, c_1 and c_2 can be viewed as counterparts of the early-stopping boundaries for superiority and futility, respectively, which are frequently used in conventional Bayesian adaptive designs. If $\Pr\big(\Delta_{n-1}(\mathbf{X}, j) > \delta\big) > c_1$, then there is sufficient evidence that the experimental arm is superior to the control arm for the nth patient with covariates $\mathbf{X}$ in region j, so he/she should be assigned to the experimental arm. The interpretation of c_2 is similar. However, unlike the conventional clinical trials that typically implement early-stopping rules by assuming patient homogeneity, the proposed design considers patient heterogeneity so the treatment assignment and decision rules are both tailored to each patient's individual profile $\mathbf{X}$ and residential region j and therefore cannot

For Peer Review

be generalized to the whole patient population. As a result, the proposed design should, in general, never early stop the trial for futility or superiority because each patient is unique.

After stage II is completed, for each combination of covariates $\mathbf{X}$ and region j, we claim the experimental treatment is promising if

$$\Pr\big(\Delta_N(\mathbf{X}, j) > \delta\big) > c \tag{11}$$

and claim the experimental arm is not promising otherwise, where c is a probability cutoff to be tuned through simulation.

4 Simulation

4.1 Simulation setting

We conducted simulation studies to assess the operating characteristics of the proposed design. The maximum sample size was $N = 100$, with stage I sample size $N_1 = 30$ and stage II sample size $N_2 = 70$. Let the covariate X be a binary biomarker with 50% positive marker prevalence. The efficacy outcome Y was taken to be a 3-level ordinal variable. To mimic our motivating trial, we assumed 12 geographic regions representing districts/counties/cities/states located as in Figure 1. We used the utility table provided in Section 3, i.e., $\omega_1 = 0$, $\omega_2 = 50$, and $\omega_3 = 100$. The minimum meaningful improvement of the utility was taken to be $\delta = 30\%$ so that the experimental agent was deemed promising relative to the standard treatment if the increase in utility exceeded 30%. We took the probability cutoffs $c_1 = 0.9$, $c_2 = 0.1$, and $c = 0.5$, which yielded desirable operating characteristics.

We constructed 8 scenarios to evaluate the performance of the proposed design. The true regression parameters for the 8 scenarios are provided in Table S1 in the Supplementary Materials. For each scenario, we considered two values of σ in the CAR model for $\boldsymbol{\theta}$, and

For Peer Review

two situations in terms of the relative population of the 12 regions. σ was set to 2 or 10 to induce different degrees of spatial variations among the regions. In the first setting of relative population, we assumed all 12 regions had the same populations so the relative population was $1 : 1 : \cdots : 1$. In the second setting, we assumed the relative population of the 12 regions was $1 : 2 : \cdots : 11 : 12$ to evaluate the performance of the proposed design with unequal populations of the regions. The probability of sampling a patient from a particular region was proportional to its relative population. Spatially structured random vectors $\boldsymbol{\theta}$ were generated from the CAR model following the technique described in Norton and Niu [5]. Conditional on the regional random effects $\boldsymbol{\theta}$, the efficacy outcomes were generated within each region under the probit model (1).

Under our simulation setting, there are 24 combinations of biomarker status and region. As $\theta_j's$ were randomly generated in each simulated trial, the true recommended treatment for a given combination of biomarker status and region could be different in different simulated trials even under the same scenario. To get a ballpark idea of the magnitude of the efficacy probabilities in each scenario, we present in Table 1 the true efficacy probabilities and true utility for each biomarker subgroup under each treatment when there is no spatial effect, that is, $\theta_j = 0$ for $j = 1, \cdots, 12$. Since the random effects $\theta_j's$ always have mean 0, these efficacy probabilities represent the "average state" of a scenario across the 12 geographic regions. With all $\theta_j's$ being 0, there is no spatial variation, so patients with the same biomarker status in the same treatment arm have the same efficacy probabilities and therefore the same utility. As shown by Table 1, the 8 scenarios were chosen to cover different cases of the "average" treatment effects across the 12 regions. In scenarios 1 and 2, the experimental treatment is "on average" not promising for marker negative group, and promising for marker positive group. In scenarios 3 and 4, the experimental treatment is "on average" promising for marker negative group, and not promising for marker positive group. In scenarios 5 and 6, the experimental treatment is "on average" not promising for both marker groups; and in scenarios 7 and 8, the experimental treatment is "on average" promising for both marker

For Peer Review

groups. Note that the experimental arm was considered superior to the control arm if the utility of that arm was at least 30% higher than that of the control arm.

To demonstrate the importance of accounting for the spatial variation in phase II trials, we compared our proposed design to a nonspatial design that incorporated the biomarker but ignored the spatial variation across the regions. That is, we used the same probit model (1), but with the spatial effects $\theta_j's$ dropped. To make a fair comparison, we used the same utility table and 2-stage strategy for both designs.

4.2 Simulation results

Under each scenario, we simulated 500 trials for each combination of σ and relative population of the J regions. As mentioned earlier, in each simulated trial under any given scenario, each of the 24 combinations of region and biomarker has its own recommended treatment due to the spatial random effects. Therefore, we report the simulation results in two ways. Table 2 shows the mean percent of correct selection (PCS) of the true recommended treatment across the 24 combinations of biomarker and region under each scenario, and Figure 2 shows the PCS of the true recommended treatment for each of the 24 combinations under each scenario. The complete results of the PCS for the 8 scenarios are provided in Tables S2 and S3 in the Supplementary Materials. The PCS measures the group ethics of the trial as it provides the performance of the designs to recommend treatments for future patients after the trial is completed. Table 3 shows the percentage of correct treatment assignment (PCTA) for the N_2 patients in stage II with adaptive randomization. This provides a measure of the individual ethics of the trial as it evaluates the performance of the designs to assign appropriate treatments to the patients enrolled in the trial.

The proposed spatial design significantly outperformed the nonspatial design in terms of both PCS and PCTA according to our simulation results. As shown in Table 2, the proposed spatial design yielded consistently higher mean PCS than the nonspatial design across all the simulation settings. In particular, since a larger value of σ induces larger

https://mc.manuscriptcentral.com/smmr
For Peer Review

spatial variations, the advantage of our proposed spatial design was more pronounced under $\sigma = 10$ than under $\sigma = 2$. For instance, in scenario 1 with equal populations, the spatial design yielded 7.3% higher mean PCS than the nonspatial design when $\sigma = 2$ (73.1% vs 65.8%), and 21.3% higher mean PCS when $\sigma = 10$ (77.8% vs 56.5%). Figure 2 depicts the PCS stratified by the region and biomarker subgroup. As shown by Figure 2, when $\sigma = 10$, our proposed spatial design yielded consistently higher PCS than the nonspatial design across the 24 combinations of region and biomarker status. For example, in scenario 1, with $\sigma = 10$ and equal populations, the proposed spatial model yielded $1\% - 15.8\%$ and $27.4\% - 39\%$ higher PCS than the nonspatial model across the 12 regions for marker negative and positive groups, respectively. In scenario 8, with $\sigma = 10$ and unequal populations, the proposed design resulted in $20.6\% - 34.8\%$ and $20.8\% - 29.6\%$ higher PCS across all regions under marker negative and positive groups, respectively. When $\sigma = 2$, the proposed design also gave higher PCS for almost all the combinations of region and marker group. For example, in scenarios 6, 7, and 8, PCS were consistently higher under the proposed design than under the nonspatial design for all combinations of region and biomarker status. In other scenarios, the proposed design yielded higher PCS for the majority of the 12 regions for each marker group. For example, in scenario 1, when the 12 regions had the same populations, the proposed design resulted in higher PCS for 11 out of the 12 regions for both marker groups, and slightly lower PCS for only one region, i.e., 62.4% vs 64.8% for region 1 under marker negative group and 79% vs 79.2% for region 6 under marker positive group.

Table 3 reports the PCTA under the two designs. Again, the proposed spatial design yielded higher PCTA than the nonspatial design under all settings in all scenarios. For example, in scenario 1, the proposed design resulted in 6.1% and 16.5% higher PCTA when $\sigma = 2$ and 10, respectively, under equal populations; and 5.1% and 16% higher PCTA, under unequal populations. To sum up, both the higher PCS and PCTA under the proposed design clearly demonstrate the importance of accounting for the spatial variation across geographic regions and the benefit of adopting the proposed spatial design.

For Peer Review

4.3 Sensitivity Analyses

We carried out sensitivity analyses to examine the robustness of our proposed spatial design to the prior distributions and the utility function. To simplify our presentation, only cases with equal populations were studied. In our original simulation studies, we used vague priors for all parameters. In the sensitivity studies, we let the prior distributions be even more non-informative. Specifically, we let the prior distribution of each element of $\boldsymbol{\beta}$ be $N(0, 5^2)$ so the prior standard deviation was twice of the original standard deviation. And we assigned γ_2 a Unif$(0, 5)$ prior distribution so the prior range was much wider than the original range. Under these new prior distributions, the results (see Tables 4 and 5 and Figure 3) are very similar to the original results reported in Figure 2 and Tables 2 and 3, suggesting that the proposed design is robust to the choice of the prior distributions.

To evaluate the sensitivity of the proposed design to the choice of the utility function, we considered the following 2 alternative utility functions: the first (second) new utility gives higher (lower) desirability scores to the outcome $Y = 2$. As shown in Figure 3 and Tables 4 and 5, the results under the two alternative utilities are generally similar to those reported in Figure 2 and Tables 2 and 3. The complete table of PCS is given in the Supplementary Materials (Table S4).

	Utility 1			Utility 2		
	Y=1	Y=2	Y=3	Y=1	Y=2	Y=3
ω	0	60	100	0	30	100

5 Discussion

We have described a Bayesian adaptive phase II clinical trial design which accounts for the spatial variation as well as patient heterogeneity. Our design uses the cumulative pro-

For Peer Review

bit regression to model the ordinal efficacy outcome and assigns an intrinsic conditional autoregressive prior distribution to the region-specific random effects to achieve spatial dependence. We propose a two-stage design to assign patients to desirable treatments based on their spatial information and individual covariates and make the personalized treatment recommendation for future patients after the trial is completed. Simulation results show that our proposed design has desirable operating characteristics and outperforms an alternative design that ignores the spatial variation.

Our spatial design can be extended to phase I dose-finding designs, multi-arm phase II designs, and large-scale phase III confirmatory trials. In particular, we anticipate a substantial benefit of accounting for the spatial variation in phase III trials because the spatial effect is generally expected to be more significant in a large-scale phase III trial involving multicenter, or even multinational collaboration. Further research in this area is warranted.

In our simulation studies, we included a limited number of individual covariates in our proposed spatial design. It is conceptually easy to incorporate more covariates, however, due to the typical small or moderate sample sizes of phase II trials, we do not recommend to include too many covariates in the model. For a moderate number of covariates, some dimension reduction or model selection technique could be used in conjunction with our method (e.g., see Guo and Yuan [23]).

Like many existing Bayesian adaptive clinical trial designs (e.g., [30, 31, 32]), our proposed design is based on adaptive randomization (AR). Compared with equal randomization (ER) that randomizes patients equally between treatments, AR uses the accumulating data to adapt the randomization probabilities so that on average more patients in the trial will receive the more effective treatment. Therefore, AR enhances the individual ethics of the trial and it is more desirable for patients enrolled in the trial. However, as a trade-off, AR sacrifices a little bit of power to detect a treatment effect. In addition, AR slightly increases the chance of a sample size imbalance in the wrong direction. To alleviate this problem, we equip the proposed design with a "burn-in" ER stage before applying AR, as suggested by

https://mc.manuscriptcentral.com/smmr
For Peer Review

Thall, Fox, and Wathen [33] and Zang and Lee [34]. Extension of the proposed design is warranted.

For Peer Review

References

[1] Schork, NJ. (2015) Personalized medicine: time for one-person trials. *Nature*, **520**, 609-611.

[2] Collins, FS., Varmus, H. (2015), A new initiative on precision medicine. *The New England Journal of Medicine*, 372, 793-795.

[3] Best, N., Arnold, R., Thomas, A., Waller, L., Conlon, E. (1999), Bayesian models for spatially correlated disease and exposure data. *Bayesian Statistics*, 6, 131-156.

[4] Culhane, J.F., Elo, I.T. (2005), Neighborhood context and reproductive health. *American Journal of Obstetrics and Gynecology*, 192, S22-S29.

[5] Norton, J., Niu, X. (2009), Intrinsically autoregressive spatiotemporal models with application to aggregated birth outcomes. *Journal of the American Statistical Association*, 104, 638-649.

[6] Alberg, A.J., Samet, J.M. (2003), Epidemiology of lung cancer. *CHEST*, 123, 21S-49S.

[7] Cordioli, M., Erspamer, L., Ranzi, A., Lauriola, P., Goldoni, C.A., Gatti, G., Razzini, G., Artioli, F. (2012), Environmental risk factors and lung cancer: a spatial analysis to define long-time exposure of population enrolled in a case-control study. *IX Simposio Internazionale di Ingegneria Sanitaria ed Ambientale (SIDISA)*

[8] Christian, W., Huang, B., Rinehart, J., Hopenhayn, C. (2011), Exploring geographic variation in lung cancer incidence in Kentucky using a spatial scan statistic: elevated risk in the appalachian coal-mining region. *Public Health Report*, 126, 789-796.

[9] Henderson, R., Shimakura, S., Gorst, D. (2002), Modeling spatial variation in Leukemia survival data. *Journal of the American Statistical Association*, 97, 965-972.

[10] Li, Y., Ryan, L. (2002), Modeling spatial survival data using semiparametric frailty models. *Biometrics*, 58, 287-297.

https://mc.manuscriptcentral.com/smmr
For Peer Review

[11] Ali, M., Emch M., Yunus M., Clemens J. (2009), Modeling spatial heterogeneity of disease risk and evaluation of the impact of vaccination. *Vaccine*, 27, 3724-3729.

[12] Gehan, EA (1961) The determination of the number of patients required in a preliminary and follow-up trials of a new chemotherapeutic agent *Journal of Chronic Diseases*, **13**, 346-353.

[13] Fleming. T. (1982) One-sample multiple testing procedure for phase II clinical trials. *Biometrics*, **38**, 243-151.

[14] Simon R. (1989) Optimal two-stage designs for phase II clinical trials. *Controlled Clinical Trials*, **10**, 1-10.

[15] Thall, P., Simon, R. (1994) Practical Bayesian guidelines for phase IIB clinical trials. *Biometrics*, **50**, 337-349.

[16] Chen, T. (1997), Optimal three-stage designs for phase II cancer clinical trials. *Statistics in Medicine*, 16, 2701-2711.

[17] Tan SB. Machin D. (2002) Bayesian two-stage designs for phase II clinical trials. *Statistics in Medicine*, **21**, 1991-2012.

[18] Ibrahim, J., Ryan, L., Chen, M. (1998), Using historical controls to adjust for covariates in trend tests for binary data. *Journal of the American Statistical Association*, 93, 1282-1293.

[19] Thall, P., Sung, H., Estey, E. (2002), Selecting therapeutic strategies based on efficacy and death in multicourse clinical trials. *Journal of the American Statistical Association*, 97, 29-39.

[20] Thall, P., Wathen, J., Bekele, B., Champlin, R., Baker, L., Benjamin, R. (2003), Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. *Statistics in Medicine*, 22, 763-780.

For Peer Review

[21] Wathen, J., Thall, P., Cook, J., Estey, E. (2008), Accounting for patient heterogeneity in phase II clinical trials. *Statistics in Medicine*, 27, 2802-2815.

[22] Houede, N., Thall, P., Nguyen, H., Paolett, X., Kramar, A. (2010), Utility-based optimization of combination therapy using ordinal toxicity and efficacy in phase I/II trials. *Biometrics*, 66, 532-540.

[23] Guo, B., Yuan, Y. (2016), Bayesian Phase I/II Biomarker-based Dose Finding for Precision Medicine with Molecularly Targeted Agents. *Journal of the American Statistical Association*, In Press.

[24] Albert, J., Chib, S. (1993), Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88, 669-679.

[25] Besag, J., York, J.C., Mollie, A. (1991), Bayesian image restoration, with two applications in spatial statistics (with discussion). *Ann. Inst. Statist. Math*, 43, 1-59.

[26] Assuncao, R., Potter, J., Cavenaghi, S. (2002), A Bayesian space varying parameter model applied to estimating fertility schedules. *Statistics in Medicine*, 21, 2057-2075.

[27] Banerjee, S., Carlin, BP, Gelfand, AE. (2004), Hierarchical modeling and analysis for spatial data. *Boca Raton: Chapman & Hall*.

[28] Gelman, A., Jakulin, A., Pittau, MG., Su, YS. (2008), A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2, 1360-1383.

[29] Gelfand, A.E., Smith, A.F.M. (1990), Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.

[30] Lee, J., Gu, X., Liu, S. (2010), Bayesian adaptive randomization designs for targeted agent development. *Clinical Trials*, 7, 584-596.

For Peer Review

[31] Yin, G., Chen, N., Lee, J. (2012), Phase II trial design with Bayesian adaptive randomization and predictive probability. *Journal of the Royal Statistical Society, Series C*, 61, 219-235.

[32] Yuan, Y., Guo, B., Munsell, M., Lu, K., Jazaeri, A. (2016), MIDAS: a practical Bayesian design for platform trials with molecularly targeted agents. *Statistics in Medicine*, 35, 3892-3906.

[33] Thall, P., Fox, P., Wathen, J. (2015), Statistical controversies in clinical research: scientific and ethical problems with adaptive randomization in comparative clinical trials. *Annals of Oncology*, 26, 1621-1628.

[34] Zang, Y., Lee, J. (2014), Adaptive clinical trial designs in oncology. *Chinese Clinical Oncology*, 3, 2304-3865.

For Peer Review

Table 1: True efficacy probabilities and utilities for the two biomarker subgroups ($X = 0/1$ for marker negative/positive group) under the two treatments ($Z = 0/1$ for control/experimental arm) when there is no spatial effect for the 8 scenarios. The last column "Δ" in each scenario shows the percentage of increase in utility for the experimental arm relative to the standard arm for each marker group. If $\Delta > 0.3$, the experimental arm is considered superior to the control. The bolded lines in each scenario represent the true recommended treatments for the two marker groups.

Z	X	$\Pr(Y=1)$	$\Pr(Y=2)$	$\Pr(Y=3)$	Utility	Δ	$\Pr(Y=1)$	$\Pr(Y=2)$	$\Pr(Y=3)$	Utility	Δ
		scenario 1					scenario 5				
0	0	**0.48**	**0.41**	**0.11**	**31.3**		**0.38**	**0.38**	**0.24**	**43.0**	
1	0	0.42	0.44	0.14	35.7	0.14	0.31	0.38	0.31	50.0	0.16
0	1	0.29	0.48	0.23	46.8		**0.21**	**0.37**	**0.42**	**60.4**	
1	1	**0.07**	**0.35**	**0.58**	**75.6**	0.62	0.13	0.32	0.56	71.7	0.19
		scenario 2					scenario 6				
0	0	**0.34**	**0.38**	**0.27**	**46.5**		**0.44**	**0.45**	**0.11**	**33.3**	
1	0	0.27	0.38	0.34	53.5	0.15	0.34	0.50	0.16	40.7	0.22
0	1	0.62	0.29	0.10	23.9		**0.13**	**0.47**	**0.40**	**63.8**	
1	1	**0.42**	**0.37**	**0.21**	**39.6**	0.65	0.07	0.39	0.54	73.7	0.15
		scenario 3					scenario 7				
0	0	0.31	0.58	0.12	40.3		0.69	0.21	0.10	20.3	
1	0	**0.07**	**0.51**	**0.42**	**67.7**	0.68	**0.54**	**0.28**	**0.18**	**32.2**	0.59
0	1	**0.16**	**0.60**	**0.24**	**54.2**		0.52	0.28	0.20	33.9	
1	1	0.10	0.56	0.34	62.4	0.15	**0.27**	**0.31**	**0.42**	**57.3**	0.69
		scenario 4					scenario 8				
0	0	0.50	0.34	0.16	32.9		0.46	0.33	0.21	37.6	
1	0	**0.31**	**0.38**	**0.31**	**50.0**	0.52	**0.21**	**0.33**	**0.46**	**62.4**	0.66
0	1	**0.66**	**0.26**	**0.08**	**21.3**		0.76	0.19	0.05	14.8	
1	1	0.60	0.30	0.11	25.3	0.19	**0.60**	**0.28**	**0.13**	**26.3**	0.77

Table 2: Mean percent of correct selection (PCS) of the true recommended treatment across the 24 combinations of region and biomarker status under the proposed spatial design and the nonspatial design.

	Equal Populations				Unequal Populations			
	$\sigma = 2$		$\sigma = 10$		$\sigma = 2$		$\sigma = 10$	
Scenario	Spatial	Nonspatial	Spatial	Nonspatial	Spatial	Nonspatial	Spatial	Nonspatial
1	0.731	0.658	0.778	0.565	0.708	0.636	0.764	0.554
2	0.709	0.628	0.744	0.545	0.688	0.629	0.723	0.536
3	0.782	0.711	0.818	0.569	0.772	0.701	0.806	0.566
4	0.670	0.572	0.707	0.518	0.625	0.547	0.708	0.529
5	0.724	0.613	0.745	0.577	0.709	0.621	0.745	0.568
6	0.717	0.630	0.749	0.572	0.711	0.622	0.725	0.566
7	0.698	0.614	0.750	0.500	0.719	0.592	0.743	0.493
8	0.746	0.634	0.768	0.496	0.739	0.581	0.765	0.499

For Peer Review

Table 3: Percent of correct treatment assignment (PCTA) for patients in the second stage of the trial under the proposed spatial design and the nonspatial design.

	Equal Populations				Unequal Populations			
	$\sigma = 2$		$\sigma = 10$		$\sigma = 2$		$\sigma = 10$	
Scenario	Spatial	Nonspatial	Spatial	Nonspatial	Spatial	Nonspatial	Spatial	Nonspatial
1	0.646	0.585	0.702	0.537	0.635	0.584	0.700	0.540
2	0.639	0.576	0.673	0.527	0.632	0.580	0.671	0.532
3	0.691	0.629	0.738	0.549	0.691	0.624	0.737	0.557
4	0.598	0.545	0.654	0.508	0.580	0.533	0.648	0.527
5	0.658	0.584	0.706	0.551	0.669	0.588	0.703	0.558
6	0.653	0.585	0.697	0.548	0.668	0.593	0.692	0.545
7	0.631	0.566	0.678	0.503	0.632	0.568	0.664	0.493
8	0.641	0.571	0.683	0.499	0.634	0.555	0.673	0.506

Table 4: Mean percent of correct selection (PCS) of the true recommended treatment across the 24 combinations of the region and biomarker status under the original simulation setting, alternative prior distributions, and two new utility functions.

	Original results		New Priors		New utility 1		New utility 2	
Scenario	$\sigma = 2$	$\sigma = 10$	$\sigma = 2$	$\sigma = 10$	$\sigma = 2$	$\sigma = 10$	$\sigma = 2$	$\sigma = 10$
1	0.731	0.778	0.734	0.777	0.703	0.770	0.732	0.775
2	0.709	0.744	0.713	0.720	0.709	0.738	0.683	0.720
3	0.782	0.818	0.815	0.798	0.802	0.818	0.772	0.800
4	0.670	0.707	0.623	0.702	0.639	0.716	0.638	0.710
5	0.724	0.745	0.695	0.738	0.734	0.749	0.669	0.721
6	0.717	0.749	0.671	0.727	0.742	0.757	0.655	0.711
7	0.698	0.750	0.711	0.751	0.713	0.758	0.754	0.728
8	0.746	0.768	0.709	0.748	0.765	0.779	0.750	0.788

For Peer Review

Table 5: Percent of correct treatment assignment (PCTA) for patients in the second stage of the trial under the original simulation setting, alternative prior distributions, and two new utility functions.

	Original results		New Priors		New utility 1		New utility 2	
Scenario	$\sigma = 2$	$\sigma = 10$	$\sigma = 2$	$\sigma = 10$	$\sigma = 2$	$\sigma = 10$	$\sigma = 2$	$\sigma = 10$
1	0.646	0.702	0.651	0.702	0.627	0.699	0.643	0.684
2	0.639	0.673	0.638	0.670	0.638	0.692	0.615	0.645
3	0.691	0.738	0.695	0.721	0.705	0.733	0.677	0.717
4	0.598	0.654	0.588	0.656	0.594	0.658	0.581	0.642
5	0.658	0.706	0.637	0.696	0.675	0.713	0.617	0.686
6	0.653	0.697	0.622	0.684	0.680	0.708	0.605	0.678
7	0.631	0.678	0.637	0.658	0.643	0.684	0.659	0.652
8	0.641	0.683	0.645	0.668	0.651	0.683	0.629	0.689

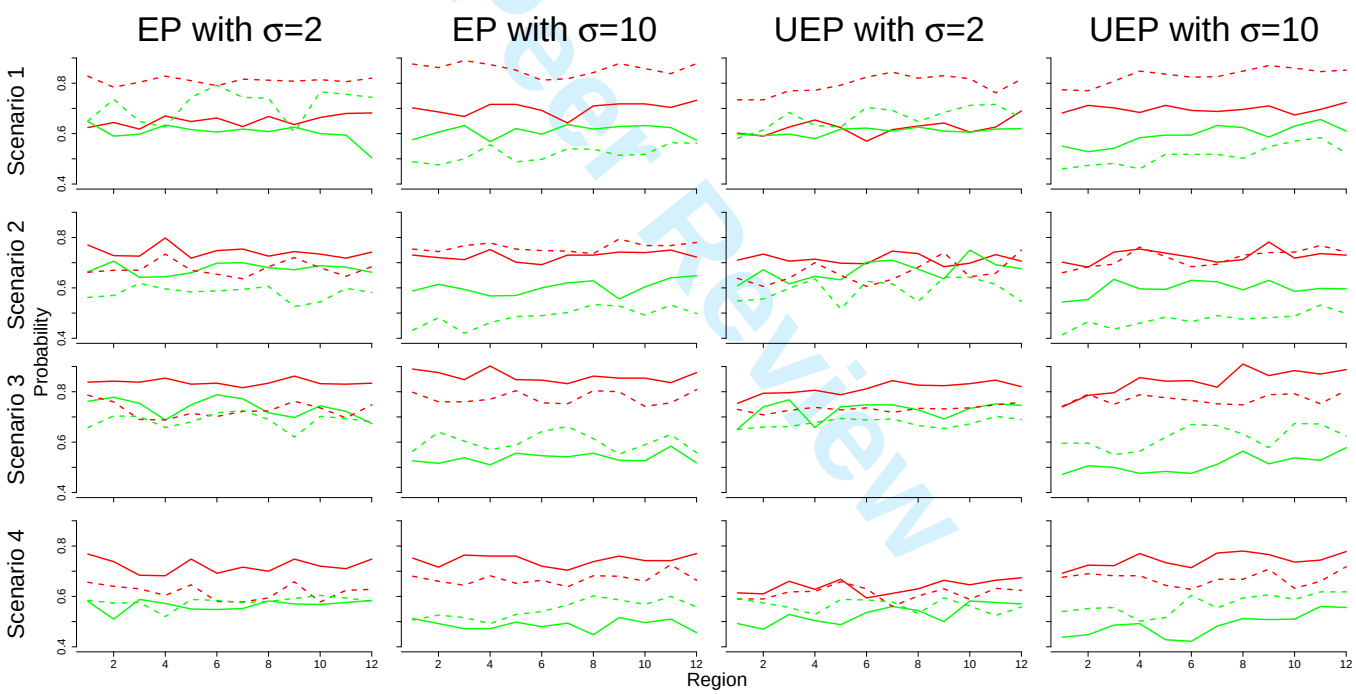
For Peer Review

Figure 1: An example of 12 spatially arranged geographic regions.

1	2	3	4
5	6	7	8
9	10	11	12

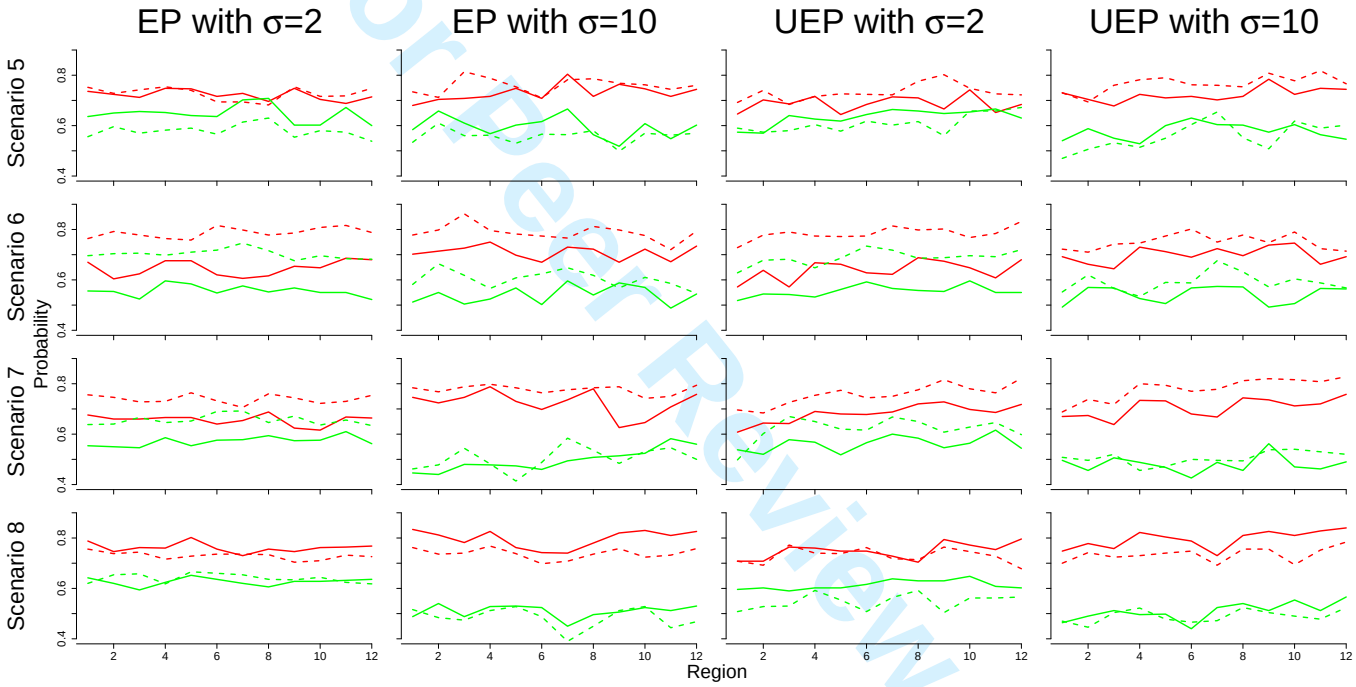
For Peer Review

Figure 2: Percent of correct selection (PCS) of the true recommended treatment for each of the 24 combinations of region and biomarker status for the 8 scenarios. The eight rows represent the 8 scenarios. In each row, the 4 plots from left to right represent equal populations (EP) and $\sigma = 2$, EP and $\sigma = 10$, unequal populations (UEP) with $\sigma = 2$, and UEP with $\sigma = 10$, respectively. In each plot, the red (green) solid line represents the PCS for the marker negative group under the proposed spatial model (nonspatial model); the red (green) broken line represents the PCS for the marker positive group under the proposed spatial model (nonspatial model).



For Peer Review

Figure 2 continued.



For Peer Review

Figure 3: Sensitivity analyses: percent of correct selection of the true recommended treatment under alternative prior distributions and utility functions. Each row represents a scenario. In each row, the 4 plots from left to right represent results from the original setting, alternative prior distribution, and the two new utility functions. In each row, the red (green) solid line represents marker negative (positive) group under $\sigma = 2$, and the red (green) broken line represents marker negative (positive) group under $\sigma = 10$.

For Peer Review

Figure 3 continued.

For Peer Review

Supplementary Materials for "A Bayesian Adaptive Phase II Clinical Trial Design Accounting for Spatial Variation"

Beibei Guo[1] and Yong Zang[2]

[1]*Department of Experimental Statistics, Louisiana State University*

Baton Rouge, LA 70803, U.S.A.

email: beibeiguo@lsu.edu

[2]*Department of Biostatistics, School of Medicine, Indiana University*

Indianapolis, IN 46202, U.S.A.

For Peer Review

Table S1: True model parameters for the 8 scenarios.

scenario	β_0	β_1	β_2	β_3	γ
1	0.1	0.2	0.5	0.8	1.3
2	0.4	0.2	-0.7	0.3	1.0
3	0.5	1.0	0.5	-0.7	1.7
4	0.0	0.5	-0.4	-0.4	1.0
5	0.3	0.2	0.5	0.2	1.0
6	0.2	0.3	1.0	0.1	1.4
7	-0.5	0.4	0.5	0.3	0.8
8	0.1	0.7	-0.8	-0.3	0.9

For Peer Review

Table S2: Percent of correct selection of the true recommended treatment for the 8 scenarios with equal relative population.

	$\sigma = 2$				$\sigma = 10$			
Region	$X = 0$ (Spatial)	$X = 0$ (Nonspatial)	$X = 1$ (Spatial)	$X = 1$ (NonSpatial)	$X = 0$ (Spatial)	$X = 0$ (Nonspatial)	$X = 1$ (Spatial)	$X = 1$ (NonSpatial)
				Scenario 1				
1	0.624	0.648	0.828	0.650	0.702	0.576	0.876	0.488
2	0.644	0.590	0.784	0.736	0.686	0.606	0.862	0.476
3	0.618	0.598	0.804	0.648	0.668	0.632	0.890	0.500
4	0.670	0.634	0.828	0.626	0.716	0.568	0.874	0.556
5	0.648	0.616	0.810	0.744	0.716	0.620	0.852	0.488
6	0.662	0.606	0.790	0.792	0.692	0.598	0.812	0.498
7	0.628	0.618	0.816	0.744	0.642	0.636	0.818	0.540
8	0.668	0.608	0.812	0.740	0.710	0.618	0.842	0.538
9	0.636	0.626	0.808	0.606	0.718	0.628	0.878	0.514
10	0.664	0.600	0.814	0.766	0.718	0.632	0.858	0.518
11	0.680	0.594	0.806	0.756	0.704	0.624	0.838	0.564
12	0.682	0.504	0.820	0.744	0.732	0.574	0.878	0.562
				Scenario 2				
1	0.770	0.664	0.662	0.562	0.730	0.588	0.754	0.432
2	0.728	0.706	0.670	0.570	0.720	0.614	0.744	0.482
3	0.726	0.642	0.670	0.618	0.712	0.594	0.768	0.420
4	0.798	0.644	0.734	0.596	0.752	0.568	0.778	0.462
5	0.718	0.660	0.670	0.584	0.702	0.570	0.754	0.486
6	0.748	0.698	0.654	0.588	0.692	0.600	0.748	0.490
7	0.754	0.700	0.636	0.594	0.730	0.620	0.746	0.502
8	0.726	0.680	0.684	0.606	0.730	0.628	0.736	0.534
9	0.744	0.672	0.720	0.526	0.742	0.556	0.794	0.528
10	0.734	0.688	0.680	0.544	0.740	0.604	0.768	0.492
11	0.718	0.682	0.646	0.598	0.750	0.640	0.768	0.532
12	0.742	0.662	0.684	0.582	0.722	0.648	0.780	0.498
				Scenario 3				
1	0.838	0.762	0.786	0.658	0.890	0.526	0.798	0.564
2	0.842	0.778	0.760	0.704	0.876	0.516	0.760	0.640
3	0.838	0.754	0.690	0.702	0.848	0.538	0.760	0.604
4	0.854	0.688	0.688	0.658	0.902	0.510	0.770	0.570
5	0.830	0.748	0.714	0.680	0.848	0.556	0.804	0.588
6	0.834	0.788	0.702	0.716	0.846	0.546	0.756	0.642
7	0.816	0.772	0.722	0.726	0.832	0.542	0.752	0.662
8	0.834	0.716	0.724	0.690	0.862	0.556	0.804	0.614
9	0.862	0.698	0.762	0.620	0.854	0.528	0.800	0.552
10	0.832	0.744	0.736	0.702	0.854	0.526	0.742	0.590
11	0.830	0.722	0.696	0.694	0.836	0.584	0.756	0.630
12	0.834	0.674	0.748	0.680	0.876	0.518	0.808	0.558
				Scenario 4				
1	0.768	0.580	0.656	0.584	0.752	0.512	0.680	0.506
2	0.738	0.510	0.640	0.574	0.716	0.492	0.660	0.526
3	0.684	0.588	0.630	0.576	0.764	0.472	0.644	0.516
4	0.682	0.572	0.604	0.520	0.760	0.472	0.682	0.492
5	0.748	0.550	0.646	0.588	0.760	0.498	0.652	0.528
6	0.692	0.548	0.580	0.584	0.720	0.480	0.664	0.540
7	0.716	0.552	0.578	0.576	0.704	0.494	0.638	0.566
8	0.700	0.582	0.594	0.580	0.738	0.448	0.682	0.602
9	0.748	0.570	0.658	0.592	0.760	0.516	0.680	0.586
10	0.720	0.568	0.578	0.604	0.742	0.496	0.660	0.568
11	0.710	0.576	0.624	0.594	0.742	0.510	0.726	0.600
12	0.748	0.584	0.628	0.580	0.770	0.456	0.664	0.558

For Peer Review

https://mc.manuscriptcentral.com/smmr

Table S2 continued.

	$\sigma = 2$				$\sigma = 10$			
Region	$X = 0$ (Spatial)	$X = 0$ (Nonspatial)	$X = 1$ (Spatial)	$X = 1$ (NonSpatial)	$X = 0$ (Spatial)	$X = 0$ (Nonspatial)	$X = 1$ (Spatial)	$X = 1$ (NonSpatial)
				Scenario 5				
1	0.736	0.636	0.752	0.556	0.680	0.584	0.734	0.534
2	0.724	0.650	0.728	0.596	0.704	0.658	0.712	0.610
3	0.712	0.656	0.742	0.570	0.708	0.610	0.814	0.560
4	0.748	0.652	0.754	0.582	0.716	0.568	0.788	0.562
5	0.746	0.640	0.740	0.590	0.748	0.602	0.754	0.530
6	0.716	0.636	0.694	0.566	0.708	0.616	0.710	0.566
7	0.728	0.702	0.694	0.614	0.804	0.666	0.782	0.564
8	0.696	0.708	0.682	0.630	0.716	0.564	0.786	0.580
9	0.748	0.602	0.754	0.554	0.764	0.518	0.768	0.498
10	0.704	0.602	0.716	0.580	0.746	0.608	0.762	0.570
11	0.688	0.672	0.718	0.574	0.716	0.548	0.744	0.562
12	0.714	0.600	0.748	0.538	0.744	0.602	0.760	0.570
				Scenario 6				
1	0.670	0.556	0.764	0.696	0.702	0.512	0.778	0.582
2	0.604	0.554	0.792	0.704	0.714	0.550	0.798	0.664
3	0.624	0.524	0.778	0.706	0.726	0.504	0.862	0.618
4	0.676	0.596	0.764	0.698	0.750	0.524	0.796	0.566
5	0.676	0.584	0.758	0.710	0.698	0.568	0.782	0.608
6	0.620	0.548	0.816	0.718	0.670	0.502	0.774	0.624
7	0.606	0.576	0.798	0.746	0.730	0.596	0.766	0.648
8	0.616	0.552	0.778	0.716	0.722	0.540	0.812	0.618
9	0.654	0.568	0.786	0.676	0.670	0.588	0.798	0.568
10	0.648	0.550	0.808	0.696	0.722	0.570	0.776	0.610
11	0.686	0.550	0.816	0.684	0.672	0.488	0.720	0.586
12	0.680	0.522	0.788	0.682	0.734	0.544	0.794	0.546
				Scenario 7				
1	0.676	0.554	0.756	0.638	0.746	0.446	0.784	0.462
2	0.660	0.550	0.746	0.640	0.724	0.440	0.768	0.478
3	0.660	0.546	0.728	0.666	0.746	0.480	0.788	0.544
4	0.666	0.586	0.730	0.646	0.788	0.478	0.798	0.482
5	0.666	0.554	0.764	0.652	0.730	0.474	0.784	0.414
6	0.640	0.576	0.732	0.690	0.698	0.460	0.764	0.488
7	0.654	0.578	0.706	0.692	0.736	0.494	0.776	0.584
8	0.688	0.594	0.760	0.646	0.780	0.508	0.784	0.536
9	0.624	0.574	0.744	0.672	0.626	0.514	0.788	0.484
10	0.616	0.576	0.722	0.636	0.646	0.524	0.742	0.532
11	0.668	0.610	0.730	0.656	0.708	0.582	0.750	0.546
12	0.664	0.562	0.754	0.634	0.758	0.560	0.794	0.500
				Scenario 8				
1	0.788	0.642	0.756	0.620	0.834	0.488	0.762	0.516
2	0.746	0.620	0.738	0.654	0.812	0.540	0.736	0.484
3	0.762	0.594	0.744	0.658	0.782	0.488	0.740	0.474
4	0.760	0.624	0.716	0.618	0.826	0.528	0.768	0.512
5	0.802	0.652	0.728	0.666	0.762	0.530	0.738	0.528
6	0.756	0.636	0.736	0.660	0.742	0.524	0.698	0.488
7	0.730	0.620	0.736	0.654	0.740	0.450	0.708	0.390
8	0.756	0.606	0.734	0.636	0.780	0.496	0.736	0.450
9	0.746	0.628	0.704	0.634	0.820	0.506	0.758	0.512
10	0.762	0.628	0.710	0.644	0.830	0.524	0.724	0.528
11	0.764	0.632	0.732	0.624	0.810	0.512	0.732	0.444
12	0.768	0.636	0.726	0.618	0.826	0.530	0.758	0.468

For Peer Review
https://mc.manuscriptcentral.com/smmr

Table S3: Percent of correct selection of the true recommended treatment for the 8 scenarios with unequal relative population.

Region Region	$\sigma = 2$ $X = 0$ (Spatial)	$X = 0$ (Nonspatial)	$X = 1$ (Spatial)	$X = 1$ (NonSpatial)	$\sigma = 10$ $X = 0$ (Spatial)	$X = 0$ (Nonspatial)	$X = 1$ (Spatial)	$X = 1$ (NonSpatial)
				Scenario 1				
1	0.602	0.598	0.734	0.582	0.682	0.550	0.774	0.460
2	0.590	0.592	0.734	0.614	0.712	0.528	0.770	0.474
3	0.626	0.598	0.770	0.684	0.702	0.542	0.808	0.482
4	0.654	0.580	0.772	0.632	0.684	0.584	0.848	0.460
5	0.624	0.618	0.792	0.626	0.712	0.594	0.836	0.518
6	0.570	0.622	0.824	0.704	0.692	0.594	0.824	0.518
7	0.616	0.610	0.844	0.692	0.688	0.632	0.826	0.518
8	0.630	0.626	0.820	0.648	0.696	0.624	0.848	0.502
9	0.642	0.610	0.830	0.684	0.710	0.586	0.870	0.548
10	0.606	0.606	0.818	0.712	0.674	0.630	0.860	0.570
11	0.626	0.618	0.762	0.716	0.696	0.656	0.846	0.584
12	0.690	0.620	0.816	0.662	0.724	0.610	0.852	0.522
				Scenario 2				
1	0.710	0.608	0.638	0.548	0.702	0.544	0.660	0.414
2	0.734	0.672	0.606	0.556	0.682	0.554	0.684	0.466
3	0.706	0.616	0.638	0.598	0.742	0.634	0.694	0.436
4	0.714	0.646	0.700	0.636	0.754	0.596	0.762	0.460
5	0.698	0.632	0.650	0.516	0.738	0.594	0.724	0.484
6	0.696	0.702	0.606	0.626	0.722	0.630	0.684	0.466
7	0.746	0.710	0.634	0.616	0.702	0.624	0.694	0.490
8	0.736	0.674	0.682	0.546	0.712	0.592	0.730	0.476
9	0.684	0.636	0.740	0.642	0.782	0.630	0.740	0.482
10	0.698	0.750	0.644	0.642	0.718	0.586	0.742	0.488
11	0.732	0.692	0.658	0.614	0.736	0.598	0.768	0.532
12	0.706	0.676	0.750	0.546	0.730	0.596	0.742	0.498
				Scenario 3				
1	0.754	0.652	0.730	0.650	0.740	0.472	0.742	0.596
2	0.794	0.740	0.708	0.660	0.786	0.506	0.790	0.596
3	0.796	0.768	0.726	0.662	0.796	0.500	0.750	0.550
4	0.806	0.658	0.738	0.678	0.856	0.476	0.788	0.564
5	0.788	0.740	0.728	0.694	0.842	0.484	0.776	0.620
6	0.812	0.748	0.736	0.688	0.844	0.476	0.766	0.670
7	0.844	0.748	0.718	0.692	0.818	0.512	0.752	0.666
8	0.826	0.728	0.734	0.666	0.910	0.564	0.748	0.632
9	0.824	0.692	0.732	0.654	0.864	0.514	0.788	0.578
10	0.832	0.734	0.736	0.672	0.884	0.538	0.792	0.674
11	0.846	0.752	0.748	0.702	0.870	0.528	0.752	0.672
12	0.820	0.746	0.758	0.690	0.888	0.578	0.806	0.624
				Scenario 4				
1	0.614	0.492	0.590	0.592	0.692	0.438	0.676	0.540
2	0.610	0.470	0.590	0.574	0.724	0.448	0.690	0.552
3	0.660	0.528	0.618	0.556	0.722	0.486	0.682	0.556
4	0.628	0.504	0.620	0.528	0.770	0.492	0.682	0.502
5	0.668	0.488	0.658	0.588	0.734	0.428	0.644	0.516
6	0.594	0.536	0.630	0.586	0.714	0.422	0.628	0.604
7	0.612	0.560	0.560	0.578	0.772	0.482	0.668	0.554
8	0.630	0.544	0.602	0.532	0.780	0.512	0.668	0.594
9	0.664	0.500	0.630	0.594	0.766	0.508	0.708	0.606
10	0.646	0.582	0.588	0.562	0.730	0.510	0.632	0.588
11	0.664	0.576	0.632	0.524	0.744	0.560	0.660	0.618
12	0.674	0.570	0.624	0.558	0.778	0.556	0.718	0.618

For Peer Review
https://mc.manuscriptcentral.com/smmr

Table S3 continued.

	$\sigma = 2$				$\sigma = 10$			
Region	$X = 0$ (Spatial)	$X = 0$ (Nonspatial)	$X = 1$ (Spatial)	$X = 1$ (NonSpatial)	$X = 0$ (Spatial)	$X = 0$ (Nonspatial)	$X = 1$ (Spatial)	$X = 1$ (NonSpatial)
				Scenario 5				
1	0.646	0.574	0.692	0.590	0.730	0.540	0.730	0.470
2	0.702	0.570	0.740	0.574	0.704	0.588	0.694	0.506
3	0.686	0.640	0.684	0.580	0.678	0.550	0.760	0.532
4	0.716	0.626	0.716	0.604	0.724	0.528	0.782	0.514
5	0.644	0.618	0.726	0.578	0.710	0.600	0.790	0.550
6	0.684	0.644	0.724	0.618	0.716	0.630	0.762	0.604
7	0.714	0.664	0.722	0.602	0.702	0.604	0.760	0.654
8	0.710	0.658	0.776	0.616	0.716	0.602	0.754	0.552
9	0.666	0.648	0.802	0.560	0.784	0.574	0.808	0.508
10	0.742	0.654	0.748	0.658	0.724	0.604	0.778	0.618
11	0.652	0.666	0.726	0.658	0.748	0.564	0.818	0.590
12	0.684	0.630	0.722	0.672	0.744	0.546	0.766	0.602
				Scenario 6				
1	0.572	0.518	0.728	0.628	0.692	0.492	0.722	0.552
2	0.638	0.544	0.778	0.678	0.662	0.570	0.710	0.620
3	0.572	0.542	0.790	0.682	0.644	0.568	0.742	0.566
4	0.668	0.532	0.774	0.648	0.730	0.526	0.746	0.534
5	0.662	0.562	0.772	0.686	0.712	0.506	0.774	0.590
6	0.628	0.592	0.774	0.734	0.690	0.568	0.802	0.588
7	0.622	0.566	0.814	0.718	0.724	0.574	0.750	0.676
8	0.688	0.558	0.798	0.686	0.696	0.572	0.778	0.632
9	0.674	0.554	0.802	0.688	0.738	0.492	0.748	0.572
10	0.648	0.596	0.768	0.696	0.746	0.506	0.790	0.604
11	0.608	0.550	0.784	0.692	0.662	0.566	0.724	0.588
12	0.680	0.550	0.832	0.720	0.692	0.564	0.714	0.568
				Scenario 7				
1	0.608	0.538	0.696	0.498	0.670	0.496	0.688	0.508
2	0.644	0.520	0.684	0.602	0.674	0.456	0.738	0.496
3	0.642	0.578	0.726	0.670	0.638	0.506	0.718	0.520
4	0.690	0.568	0.754	0.650	0.734	0.488	0.800	0.456
5	0.680	0.518	0.774	0.620	0.732	0.468	0.794	0.472
6	0.678	0.566	0.744	0.616	0.680	0.426	0.770	0.500
7	0.688	0.600	0.750	0.668	0.668	0.488	0.778	0.496
8	0.720	0.584	0.776	0.650	0.744	0.456	0.812	0.494
9	0.728	0.546	0.816	0.608	0.736	0.562	0.820	0.538
10	0.698	0.564	0.780	0.628	0.712	0.470	0.816	0.540
11	0.686	0.616	0.764	0.646	0.720	0.462	0.808	0.530
12	0.718	0.544	0.822	0.598	0.758	0.490	0.828	0.520
				Scenario 8				
1	0.708	0.596	0.708	0.508	0.748	0.464	0.700	0.470
2	0.708	0.602	0.692	0.528	0.778	0.490	0.742	0.446
3	0.764	0.590	0.772	0.530	0.758	0.512	0.724	0.504
4	0.760	0.602	0.740	0.592	0.822	0.496	0.730	0.522
5	0.748	0.602	0.738	0.554	0.804	0.498	0.740	0.478
6	0.748	0.616	0.762	0.508	0.788	0.440	0.748	0.466
7	0.728	0.638	0.718	0.564	0.730	0.524	0.692	0.472
8	0.704	0.630	0.714	0.592	0.810	0.540	0.756	0.524
9	0.794	0.630	0.764	0.504	0.826	0.512	0.756	0.504
10	0.772	0.648	0.746	0.562	0.810	0.554	0.694	0.490
11	0.754	0.608	0.728	0.562	0.828	0.512	0.752	0.478
12	0.796	0.602	0.678	0.566	0.840	0.566	0.784	0.524

For Peer Review

Table S4: Percent of correct selection of the true recommended treatment for the 8 scenarios under alternative prior distributions and utility functions.

	New prior				New utility 1				New utility 2			
	$\sigma = 2$		$\sigma = 10$		$\sigma = 2$		$\sigma = 10$		$\sigma = 2$		$\sigma = 10$	
	X_0	X_1	X_0	X_1	X_0	X_1	X_0	X_1	X_0	X_1	X_0	X_1
	scenario 1											
1	0.692	0.788	0.720	0.900	0.682	0.810	0.776	0.888	0.646	0.830	0.734	0.896
2	0.672	0.798	0.682	0.850	0.624	0.776	0.778	0.814	0.612	0.844	0.674	0.868
3	0.660	0.782	0.666	0.844	0.640	0.734	0.690	0.826	0.628	0.812	0.652	0.860
4	0.690	0.770	0.702	0.904	0.636	0.760	0.594	0.880	0.646	0.856	0.652	0.832
5	0.628	0.838	0.688	0.856	0.628	0.768	0.720	0.824	0.596	0.848	0.694	0.880
6	0.656	0.788	0.688	0.842	0.622	0.744	0.724	0.850	0.598	0.848	0.670	0.808
7	0.644	0.820	0.704	0.826	0.612	0.748	0.582	0.872	0.600	0.852	0.724	0.864
8	0.698	0.818	0.676	0.864	0.640	0.756	0.726	0.832	0.602	0.832	0.722	0.870
9	0.650	0.832	0.684	0.876	0.660	0.818	0.704	0.888	0.652	0.854	0.706	0.912
10	0.692	0.792	0.698	0.846	0.604	0.770	0.698	0.830	0.636	0.838	0.692	0.842
11	0.644	0.790	0.720	0.856	0.616	0.754	0.652	0.834	0.624	0.836	0.650	0.858
12	0.652	0.810	0.680	0.870	0.652	0.822	0.662	0.824	0.628	0.848	0.666	0.880
	scenario 2											
1	0.746	0.740	0.758	0.748	0.724	0.718	0.760	0.826	0.680	0.706	0.754	0.762
2	0.732	0.682	0.720	0.704	0.732	0.700	0.696	0.774	0.644	0.696	0.686	0.722
3	0.714	0.658	0.736	0.738	0.704	0.650	0.706	0.750	0.680	0.706	0.714	0.730
4	0.764	0.642	0.724	0.710	0.738	0.682	0.744	0.774	0.706	0.738	0.750	0.742
5	0.712	0.734	0.722	0.682	0.744	0.704	0.736	0.780	0.676	0.750	0.688	0.736
6	0.720	0.724	0.714	0.658	0.714	0.642	0.668	0.682	0.618	0.708	0.684	0.684
7	0.716	0.690	0.704	0.702	0.750	0.652	0.710	0.716	0.624	0.666	0.674	0.728
8	0.736	0.688	0.752	0.726	0.710	0.680	0.690	0.728	0.668	0.690	0.636	0.752
9	0.692	0.728	0.682	0.698	0.744	0.712	0.818	0.700	0.654	0.696	0.740	0.728
10	0.722	0.688	0.708	0.740	0.732	0.674	0.818	0.674	0.646	0.680	0.750	0.726
11	0.752	0.698	0.728	0.718	0.750	0.676	0.790	0.722	0.678	0.714	0.690	0.724
12	0.724	0.698	0.750	0.746	0.748	0.724	0.710	0.750	0.676	0.698	0.714	0.762
	scenario 3											
1	0.876	0.736	0.862	0.754	0.820	0.800	0.870	0.786	0.886	0.652	0.860	0.746
2	0.884	0.748	0.848	0.726	0.840	0.742	0.852	0.784	0.896	0.648	0.884	0.728
3	0.848	0.802	0.854	0.752	0.810	0.752	0.860	0.770	0.902	0.606	0.886	0.714
4	0.786	0.786	0.866	0.736	0.810	0.846	0.838	0.782	0.898	0.618	0.880	0.736
5	0.836	0.742	0.882	0.734	0.820	0.792	0.804	0.804	0.890	0.664	0.880	0.726
6	0.882	0.738	0.834	0.730	0.776	0.806	0.824	0.800	0.928	0.628	0.864	0.692
7	0.850	0.768	0.830	0.734	0.814	0.824	0.824	0.720	0.908	0.636	0.842	0.728
8	0.918	0.830	0.828	0.746	0.798	0.794	0.868	0.798	0.916	0.638	0.902	0.706
9	0.832	0.764	0.868	0.718	0.812	0.782	0.836	0.800	0.882	0.718	0.904	0.760
10	0.844	0.772	0.850	0.724	0.768	0.800	0.862	0.830	0.918	0.650	0.862	0.742
11	0.808	0.768	0.862	0.768	0.804	0.814	0.848	0.826	0.880	0.624	0.828	0.724
12	0.896	0.856	0.888	0.760	0.838	0.788	0.850	0.790	0.902	0.644	0.872	0.730
	scenario 4											
1	0.714	0.608	0.740	0.562	0.696	0.642	0.784	0.698	0.754	0.642	0.772	0.718
2	0.686	0.584	0.740	0.644	0.628	0.644	0.762	0.684	0.682	0.614	0.750	0.676
3	0.622	0.562	0.754	0.590	0.652	0.626	0.740	0.636	0.672	0.590	0.742	0.628
4	0.716	0.538	0.816	0.714	0.658	0.656	0.774	0.678	0.684	0.586	0.764	0.606
5	0.700	0.580	0.748	0.644	0.670	0.608	0.792	0.678	0.706	0.632	0.784	0.680
6	0.594	0.584	0.804	0.562	0.642	0.550	0.726	0.630	0.632	0.596	0.782	0.614
7	0.646	0.590	0.762	0.662	0.594	0.604	0.754	0.660	0.684	0.554	0.746	0.630
8	0.642	0.584	0.738	0.612	0.682	0.620	0.786	0.662	0.664	0.564	0.772	0.590
9	0.714	0.592	0.770	0.646	0.696	0.602	0.768	0.694	0.676	0.600	0.846	0.720
10	0.634	0.566	0.756	0.646	0.654	0.600	0.752	0.662	0.696	0.546	0.782	0.626
11	0.658	0.554	0.794	0.684	0.632	0.614	0.762	0.658	0.668	0.552	0.746	0.668
12	0.706	0.588	0.772	0.678	0.698	0.670	0.748	0.688	0.720	0.596	0.762	0.644

For Peer Review

Table S4 continued.

	New prior				New utility 1				New utility 2			
	$\sigma = 2$		$\sigma = 10$		$\sigma = 2$		$\sigma = 10$		$\sigma = 2$		$\sigma = 10$	
	X_0	X_1	X_0	X_1	X_0	X_1	X_0	X_1	X_0	X_1	X_0	X_1
						scenario 5						
1	0.684	0.696	0.756	0.714	0.696	0.774	0.744	0.762	0.696	0.742	0.674	0.716
2	0.650	0.710	0.784	0.746	0.732	0.784	0.752	0.770	0.620	0.678	0.722	0.742
3	0.648	0.688	0.676	0.774	0.724	0.736	0.764	0.780	0.650	0.672	0.716	0.746
4	0.712	0.684	0.736	0.748	0.702	0.796	0.738	0.772	0.658	0.704	0.722	0.776
5	0.700	0.728	0.744	0.756	0.728	0.742	0.720	0.800	0.660	0.684	0.712	0.680
6	0.638	0.712	0.700	0.766	0.724	0.778	0.740	0.800	0.622	0.656	0.654	0.718
7	0.654	0.716	0.700	0.730	0.726	0.760	0.672	0.780	0.588	0.622	0.698	0.770
8	0.668	0.712	0.688	0.768	0.666	0.740	0.740	0.712	0.650	0.684	0.732	0.748
9	0.720	0.774	0.692	0.728	0.706	0.750	0.708	0.754	0.718	0.736	0.682	0.724
10	0.658	0.702	0.662	0.754	0.732	0.692	0.776	0.754	0.648	0.708	0.700	0.760
11	0.678	0.750	0.744	0.760	0.714	0.768	0.718	0.748	0.652	0.684	0.722	0.740
12	0.686	0.718	0.750	0.830	0.712	0.732	0.736	0.726	0.642	0.674	0.706	0.742
						scenario 6						
1	0.636	0.768	0.702	0.776	0.668	0.776	0.748	0.780	0.634	0.720	0.772	0.760
2	0.602	0.626	0.700	0.778	0.648	0.794	0.690	0.782	0.640	0.708	0.620	0.852
3	0.682	0.650	0.718	0.736	0.702	0.792	0.674	0.774	0.596	0.670	0.758	0.820
4	0.670	0.720	0.748	0.788	0.694	0.838	0.732	0.820	0.638	0.734	0.594	0.734
5	0.608	0.700	0.696	0.732	0.638	0.814	0.674	0.776	0.592	0.692	0.602	0.668
6	0.588	0.736	0.658	0.752	0.644	0.826	0.706	0.762	0.630	0.726	0.716	0.834
7	0.612	0.732	0.668	0.752	0.620	0.812	0.714	0.778	0.594	0.642	0.688	0.702
8	0.602	0.710	0.650	0.746	0.716	0.834	0.738	0.790	0.572	0.754	0.760	0.700
9	0.568	0.742	0.706	0.794	0.700	0.800	0.726	0.792	0.550	0.768	0.624	0.664
10	0.590	0.708	0.678	0.786	0.716	0.770	0.750	0.850	0.538	0.724	0.670	0.652
11	0.658	0.736	0.682	0.728	0.662	0.872	0.730	0.816	0.572	0.728	0.694	0.648
12	0.652	0.806	0.708	0.764	0.672	0.808	0.752	0.824	0.578	0.724	0.708	0.824
						scenario 7						
1	0.656	0.766	0.740	0.810	0.714	0.716	0.696	0.848	0.704	0.832	0.680	0.756
2	0.660	0.720	0.730	0.840	0.638	0.758	0.714	0.820	0.712	0.802	0.686	0.748
3	0.660	0.732	0.722	0.832	0.678	0.776	0.690	0.784	0.736	0.780	0.730	0.718
4	0.716	0.786	0.722	0.828	0.690	0.656	0.738	0.814	0.728	0.784	0.700	0.730
5	0.614	0.750	0.698	0.822	0.692	0.732	0.730	0.774	0.722	0.790	0.726	0.818
6	0.674	0.748	0.712	0.812	0.700	0.708	0.690	0.750	0.688	0.802	0.682	0.742
7	0.686	0.776	0.684	0.780	0.740	0.738	0.676	0.814	0.704	0.758	0.670	0.718
8	0.690	0.806	0.712	0.776	0.732	0.764	0.732	0.822	0.708	0.784	0.688	0.774
9	0.646	0.742	0.722	0.770	0.652	0.714	0.702	0.832	0.736	0.786	0.682	0.780
10	0.634	0.734	0.722	0.746	0.706	0.714	0.692	0.800	0.730	0.786	0.708	0.738
11	0.670	0.744	0.690	0.724	0.680	0.680	0.698	0.816	0.698	0.804	0.682	0.756
12	0.666	0.776	0.662	0.770	0.746	0.792	0.730	0.820	0.720	0.808	0.742	0.810
						scenario 8						
1	0.678	0.716	0.824	0.712	0.722	0.770	0.806	0.732	0.786	0.670	0.844	0.758
2	0.776	0.680	0.796	0.700	0.782	0.768	0.798	0.724	0.782	0.724	0.838	0.766
3	0.684	0.674	0.790	0.688	0.768	0.730	0.780	0.726	0.838	0.698	0.824	0.770
4	0.734	0.726	0.828	0.684	0.806	0.772	0.858	0.750	0.780	0.696	0.810	0.754
5	0.734	0.726	0.856	0.692	0.750	0.734	0.820	0.746	0.772	0.706	0.830	0.774
6	0.760	0.718	0.788	0.694	0.776	0.754	0.776	0.732	0.778	0.736	0.790	0.750
7	0.702	0.712	0.774	0.654	0.804	0.804	0.816	0.744	0.764	0.734	0.798	0.744
8	0.692	0.710	0.770	0.692	0.802	0.752	0.820	0.746	0.740	0.720	0.810	0.788
9	0.726	0.680	0.862	0.734	0.760	0.790	0.834	0.762	0.768	0.760	0.816	0.782
10	0.722	0.716	0.778	0.660	0.732	0.742	0.840	0.746	0.806	0.738	0.812	0.734
11	0.708	0.680	0.812	0.654	0.766	0.714	0.804	0.726	0.810	0.758	0.764	0.736
12	0.670	0.690	0.840	0.670	0.784	0.768	0.838	0.776	0.736	0.688	0.838	0.782

For Peer Review