Testing the Effectiveness of Treatment for Cancers for which the Endpoint is Survival Using Bayesian Subgroup Analysis

by

Yueyang Han

B.Sc., University of Waterloo, 2022

Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

SIMON FRASER UNIVERSITY
Spring 2024

Declaration of Committee

Name:	**Yueyang Han**
Degree:	**Master of Science**
Thesis title:	**Testing the Effectiveness of Treatment for Cancers for which the Endpoint is Survival Using Bayesian Subgroup Analysis**
Committee:	**Chair: Wei (Becky) Lin** Lecturer, Statistics and Actuarial Science
	Haolun Shi Supervisor Assistant Professor, Statistics and Actuarial Science
	Jiguo Cao Committee Member Professor, Statistics and Actuarial Science
	Owen Ward Examiner Assistant Professor, Statistics and Actuarial Science

Abstract

We propose a basket trial design that tests the effectiveness of a new treatment for several types of cancers where the endpoint is the survival time. During the trial conduct, Bayesian subgroup analysis is conducted to classify the cancer types into different clusters according to both the survival time and the longitudinal biomarker measurements of the patient. Finally, we make Bayesian inferences to decide whether to stop recruiting patients for each cluster early and make conclusions about whether the treatment is effective for each cluster according to the estimated median survival time. The simulation study shows that our proposed method performs better than the independent approach and the Bayesian Hierarchical Modeling (BHM) method in most of the scenarios.

Keywords: Bayesian subgroup analysis; Longitudinal biomarkers; Phase I/II trials; Clinical trials for cancer

Acknowledgements

I would like to express my sincere gratitude to my supervisor Haolun Shi. His support and patience have been very crucial to my academic accomplishments during my master's study. He has a very solid and broad knowledge of statistical theory and its applications in pharmaceutical statistics, which helped me a lot when I conducted this research. I deeply thank him for helping me overcome the difficulties that I faced when I wrote this thesis. I will never forget his warmth and kindness.

I am also thankful to Dr. Jiguo Cao, Dr. Owen Ward, and Dr. Becky Lin for reading my thesis and providing insightful suggestions. My gratitude also goes to all other faculty members and staff in the department, especially those who taught me courses, which include Dr. Richard Lockhart, Dr. Boxin Tang, Dr. Liangliang Wang, Dr. Sonja Isberg, and Dr. Tom Loughin. Special thanks to Professor Ian Bercotivz who offered me a chance to work as a statistical consultant.

I would like to thank my fellow graduate students in the department who not only helped me academically but also made my life more colourful. Special thanks to Shifan Jia and Lu Tan who offered me support on collaborative course projects, and gave me other academic advice. Special thanks to Sidi Wu who asked insightful questions about this project.

I also appreciate my parents, my grandparents, my other family members, and my lifelong friends from Shanxi who bring me joyfulness and encourage me during this journey.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Traditional clinical trials for cancer treatments focus on the evaluation of the treatment for a specific cancer type. Nowadays, with the rapid development of drugs that deal with a certain molecular alteration, the focus of clinical trials for cancer treatments has shifted toward the evaluation of the treatment of a group of cancer types that share the same molecular alterations. This type of trial is called a basket trial, which has several advantages over traditional clinical trials. One of the advantages of evaluating the treatment for similar cancer types together is that the clinical trial can have a larger sample size and thus have a higher power. Another advantage of basket trial is that it allows the study of rare cancer types that have a very small sample size, which is inconvenient to study independently.

Some clinical trials assume that the treatment effects for all cancer types in the study are exchangeable, which is not usually the case because some cancer types might be more sensitive to the treatment than others. In this case, the exchangeability assumption often causes a higher type-I error rate if the cancer type is insensitive to the treatment. Several studies have considered classifying the cancer types into exchangeable clusters, so that information can be borrowed within each cluster. Chu and Yuan (2018 [CY18]) proposed a design in which the cancer types are grouped into one effective cluster and one ineffective cluster, and the cancer types that are grouped into the same cluster are considered to be exchangeable. Chen and Lee (2020 [CL20]) proposed a design in which the cancer types are classified into clusters, and the number of clusters changes dynamically during the trial. Fujikawa et al. (2020 [Fuj+20]) also proposed a basket trial that borrows information based on similarities across the subgroups. Zhou and Ji (2020 [ZJ20]) also proposed a basket trial which uses a formal Bayesian hypothesis testing procedure to test the efficacy of the treatment in subgroups of cancers, where the subgroups are classified according to the similarity of the probability that the alternative hypothesis is true. However, these four studies considered binary endpoints instead of continuous endpoints such as survival time.

Some studies have shown that biomarkers can be associated with the clinical outcomes of the patients. Samstein et al. (2019 [Sam+19]) showed that tumor mutational burden (TMB), which is a predictive biomarker, is associated with the overall survival (OS) of the patient. As a result, there

is a need to incorporate biomarker measurements when classifying the cancer types into subgroups with different efficacy. There are some clinical trial designs which incorporate biomarkers as classifiers which can help to classify cancer types into different subgroups. Takeda et al. (2022 [TLR22]) proposed a Bayesian subgroup design where the cancer types are classified into subgroups according to both the cancer type itself and a second classifier (biomarker). Liu et al. (2023 [LTR23]) also proposed a two-stage design where only patients whose biomarkers measured in the first stage are positive are enrolled in the second stage. These studies highlight the importance of biomarkers as a potential classifier for cancer types because higher values in some biomarkers can be associated with better clinical outcomes. Yin et al. (2021 [Yin+21]) proposed a method that combines finding the biomarker cutoff and testing the effectiveness of the treatment using Bayesian hierarchical modelling. However, these three designs only allow measuring the biomarker one time instead of allowing longitudinal biomarker measurements, and their endpoints are binary instead of continuous.

Longitudinal biomarkers are biomarkers that are collected multiple times over time during the clinical study, which can be used to track the progression and predict the outcome of the disease. Some clinical trials utilize longitudinal biomarker measurements to help predict the outcome of the disease. van Delft et al. (2022 [van+22]) conducted research in which serum tumor marker measurements, which are longitudinal biomarkers, are used to predict the immunotherapy non-response in patients with non-small cell lung cancer. However, this study did not consider classifying more than one cancer type into subgroups. Some clinical studies demonstrate longitudinal biomarkers are associated with a certain clinical outcome. Wu et al. (2017[Wu+17]) found that there is a longitudinal association between fasting blood glucose, which is a type of biomarker, and arterial stiffness risk in non-diabetic individuals. Paulo et al. (2020[Pau+20]) found that a "longitudinal increase of HbA1c was independently associated with higher rates of cardiovascular events in patients with type 2 diabetes and multivessel CAD", where HbA1c is a biomarker.

In some clinical trials, researchers are more interested in the progression-free survival time (PFS) instead of the binary indicator of whether the treatment is effective. Consequently, there is a need to develop a clinical trial method that evaluates the effectiveness of a treatment for cancers for which the endpoint is progression-free survival time (PFS) instead of a binary endpoint.

We propose the basket trial design to test the effectiveness of a specific new treatment for several types of cancers for which the endpoint is the survival time. Unlike traditional studies which treat each type of cancer separately, we use Bayesian subgroup analysis to first classify the cancer types into different clusters according to both the survival time and the biomarker measurements of the patients, and then estimate the parameters to find out whether the treatment is effective for each cluster of cancer types. We conclude that the treatment is effective for a cluster of cancer types if the estimated median survival time for this cluster is greater than the threshold that we desire.

In summary, our proposed clinical trial design has many advantages. First, we incorporate the longitudinal biomarker measurements along with the survival outcomes to help classify the cancer

types. Second, the endpoint in our proposed method is progression-free survival time, which is very uncommon in the clinical trial methods proposed by other researchers.

Our proposed method is motivated by the clinical trial which investigates the effectiveness of pan-HER kinase inhibitor neratinib for treating solid tumors harbouring HER2 and HER3 mutations. This trial is conducted by Hyman et al. (2018[Hym+18]), and it is registered at Clinicaltrials.gov. This is a basket trial that evaluates the efficacy of the drug on 21 cancer types, such as breast, lung, bladder and colorectal cancer. The patients were classified into different cohorts according to the tumor type and whether it is an HER2-mutant tumor or an HER3-mutant tumor. In total, 141 patients were enrolled in the trial. In this trial, the researchers measured tumor DNA and tumour-derived cell-free DNA in plasma as biomarkers, which can be used to help evaluate "how ERBB2/3 copy number and clonality as well as co-mutational pattern affected outcome" (2018[Hym+18]). The researchers are interested in the efficacy of the drug in each cohort. One of the secondary endpoints in this study is the progression-free survival time (PFS), which was estimated using the Kaplan-Meier method.

Chapter 2

Methods

2.1 Model

In this study, we consider I types of cancer. For each type of cancer, we recruit n_i patients. We measure the biomarker for each patient L-times, where the measurement times are denoted by t_1, $\cdots$, t_L. The biomarker measurement of the j-th patient in the i-th cancer type measured at time t_l is denoted by Z_{ijl}. We assume that we will measure the biomarker at the same time points for all patients. After we complete the biomarker measurements, we measure the survival time for each patient, denoted by t_{ij}. Let ω_{ij} be the censoring indicator for the j-th patient in the i-th cancer type, where $\omega_{ij} = 1$ if the patient is not censored, and $\omega_{ij} = 0$ if the patient is censored. We assume that the I types of cancer can be grouped into K clusters according to the biomarker measurements and the survival time. In this article, we consider the case where $K = 2$, which means that there is one effective cluster and one ineffective cluster. However, the method can be generalized for the cases where $K > 2$. The objective of the study is to determine whether the treatment is effective for each cluster of patients, which is reflected by the median survival time for each cluster of patients. The hypothesis test is

$$H_0 : \rho_i < q_0 \text{ versus } H_a : \rho_i > q_1,$$

where ρ_i is the median survival time for cluster i, q_0 is the median survival time cutoff under which the treatment is deemed ineffective, and q_1 is the median survival time cutoff over which the treatment is deemed effective.

Let π_{ik} denote the probability that cancer type i belongs to cluster k. Let C_i be an indicator of which cluster the cancer type belongs to. For example, $C_1 = 2$ means that the first cancer type belongs to the second cluster. We assume that C_i has the multinomial distribution:

$$C_i \sim \text{Multinomial}(\pi_{i1}, \ldots, \pi_{iK}).$$

Recall that Z_{ijl} is the biomarker measurement of the j-th patient in the i-th cancer type measured at time t_l. We assume the biomarker measures are structured as follows,

$$Z_{ijl}|(C_i = k) = \mu_k(t_l) + v_i + w_{ij} + \epsilon_{ijl},$$

which reflects the grouping structure of the model. Specifically, every cluster has a mean trajectory of the biomarker, which is denoted by $\mu_k(t_l)$. Every group within the cluster can have a mean trajectory that varies from the mean trajectory of the cluster, and the difference is denoted by v_i. Every patient within the group can have a mean trajectory that varies from the mean trajectory of the group, and the difference is denoted by w_{ij}. The error term of the j-th patient in the i-th cancer type measured at time t_l is denoted by ϵ_{ijl}. Here, $\mu_k(t_l)$ has the B-Spline structure.

$$\mu_k(t_l) = \gamma_{1(k)}B_{1,k}(t_l) + \gamma_{2(k)}B_{2,k}(t_l) + \ldots + \gamma_{S(k)}B_{S,k}(t_l),$$

where $B_{q,k}$ are the B-Spline basis functions, $\gamma_{q(k)}$ are the coefficients associated with the B-Spline basis function, and S is the number of B-Spline basis functions. In our simulation study, we set $S = 6$.

We assume that $v_i \sim N(0, \sigma_v^2)$, $w_{ij} \sim N(0, \sigma_w^2)$, and $\epsilon_{ijl} \sim N(0, \sigma_\epsilon^2)$, where σ_v^2, σ_w^2, and σ_ϵ^2 are the variances of the corresponding Normal distributions. The survival function for the i-th patient in the j-th cancer type is assumed to be related to the cluster membership of the patient via a Cox-proportional hazard model with the baseline hazard following a Weibull distribution,

$$S_t(t_{ij}) = \exp\{-\lambda t_{ij}^r \exp(\theta_{(k)})\}.$$

where λ and r are the scale and shape parameters of the baseline Weibull distribution, and we specify $\theta_{(1)} = 0$ for the sake of identifiability. The corresponding probability density function of the survival time is

$$f_t(t_{ij}) = \exp\{-\lambda t_{ij}^r \exp(\theta_{(k)})\} r t_{ij}^{r-1} \lambda \exp(\theta_{(k)}).$$

We will use the Gibbs sampler to sample from posterior distributions of the parameters, which is common in Bayesian clinical trial designs.

2.2 Prior Distributions

We assign vague prior distributions to the parameters of the model. The prior distributions assigned to the parameters of the model are as follows.

$$(\pi_{i1}, \ldots, \pi_{iK}) \sim \text{Dirichlet}(2, \ldots, 2).$$

$$\gamma_{q(k)} \sim \text{N}(0, 10^4) \text{ for } q = 1, 2, \ldots, S \text{ and } k = 1, \ldots, K.$$

$$\sigma_v^2 \sim \text{IG}(10^{-3}, 10^{-3}).$$

$$\sigma_w^2 \sim \text{IG}(10^{-3}, 10^{-3}).$$

$$\sigma_\epsilon^2 \sim \text{IG}(10^{-3}, 10^{-3}).$$

$$\theta_2 \sim \text{Uniform}(-1, 1).$$

$$\lambda \sim \text{Gamma}(0.1, 0.1).$$

$$r \sim \text{Gamma}(0.1, 0.1).$$

In the distributions above, $\text{IG}(\alpha, \beta)$ denotes the inverse Gamma distribution with shape parameter α and scale parameter β.

2.3 Trial Design

This trial has M planned interim analyses. Let D_m be the observed data at the m-th interim analysis. Recall that ρ_i is the median survival time for cluster i.

If $P\{\rho_i > (q_0 + q_1)/2 | D_m\} < Q_f$, then stop recruiting patients for cancer types that belong to the i-th cluster and conclude that the treatment is ineffective for these types of cancers. Otherwise, continue to recruit patients for these cancer types. Here, Q_f is a probability cutoff; in this study, we set Q_f to be a small value, e.g., 0.05.

At the end of the study, declare that the treatment is effective if $P\{\rho_i > (q_0 + q_1)/2 | D_m\} > Q$ for the i-th cluster of patients. Otherwise, declare that the treatment is ineffective for the i-th cluster of patients. Here, Q is a probability cutoff; by default, we set $Q = 0.8$.

At each analysis, we determine the cluster membership of each group using the following rule. If the posterior probability that $C_i = 1$ is greater than 0.5, then the i-th cancer type belongs to the ineffective cluster. Otherwise, the i-th cancer type belongs to the effective cluster.

Chapter 3

Illustration

We illustrate how to implement our proposed design using a hypothetical clinical trial. Suppose we would like to evaluate whether a new drug is effective for each of the 12 types of cancers which share the same molecular aberration. Suppose the maximum number of patients in each cancer type is 50, and we have one planned interim analysis. In the first stage, we only recruit 30 patients for each cancer type. We first measure the biomarkers for each patient 20 times. After that, we record the observed survival time and censoring status of each patient.

After we collect the data, we fit the model and sample from the posterior distributions of the parameters. Recall that the interim stopping rule is that if $P\{\rho_i > (q_0 + q_1)/2|D_m\} < Q_f = 0.05$, then we stop recruiting patients for cancer types which belong to the i-th cluster, and conclude that the treatment is ineffective for these types of cancers. The posterior distribution of ρ_1 and ρ_2 are shown below.

The posterior probabilities are $P(C_1 = 1) = P(C_2 = 1) = P(C_3 = 1) = P(C_4 = 1) = P(C_5 = 1) = P(C_6 = 1) = 1$, and $P(C_7 = 1) = P(C_8 = 1) = P(C_9 = 1) = P(C_{10} = 1) = P(C_{11} = 1) = P(C_{12} = 1) = 0$, which means that cancer type 1, 2, 3, 4, 5, 6 belong to the first cluster, and cancer type 7, 8, 9, 10, 11, 12 belong to the second cluster. Since $P\{\rho_1 > (q_0 + q_1)/2|D_m\} = 0 < Q_f$, we declare that the treatment is ineffective for all the cancer types that are classified into the first cluster. Since $P\{\rho_2 > (q_0 + q_1)/2|D_m\} = 0.984 > Q_f$, we continue to recruit 20 more patients for each of the cancer types that are classified into the second cluster, and record the observed survival time and censoring status of each patient in these cancer types. In the second stage, we only use the data of the cancer types that are classified into the second cluster in the previous stage. After that, we fit the model again and estimate the posterior distribution of the parameters again. Recall that at the end of the study, we declare that the treatment is effective if $P\{\rho_i > (q_0 + q_1)/2|D_m\} > Q$ for the i-th cluster of patients. The posterior distribution of ρ_1 is shown below.

The posterior probabilities show that $P(C_7 = 1) = P(C_8 = 1) = P(C_9 = 1) = P(C_{10} = 1) = P(C_{11} = 1) = P(C_{12} = 1) = 1$, which means that cancer type 7, 8, 9, 10, 11, 12 are classified into

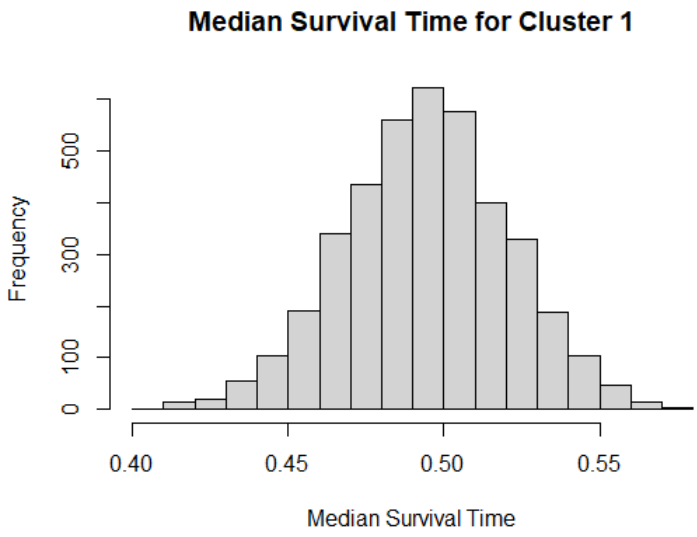

Figure 3.1: The posterior distribution of the median survival time for cluster 1 at the interim analysis.

the first cluster in this stage. Since $P\{\rho_1 > (q_0 + q_1)/2 | D_m\} = 0.999$, we declare that the treatment is effective for all these six cancer types.

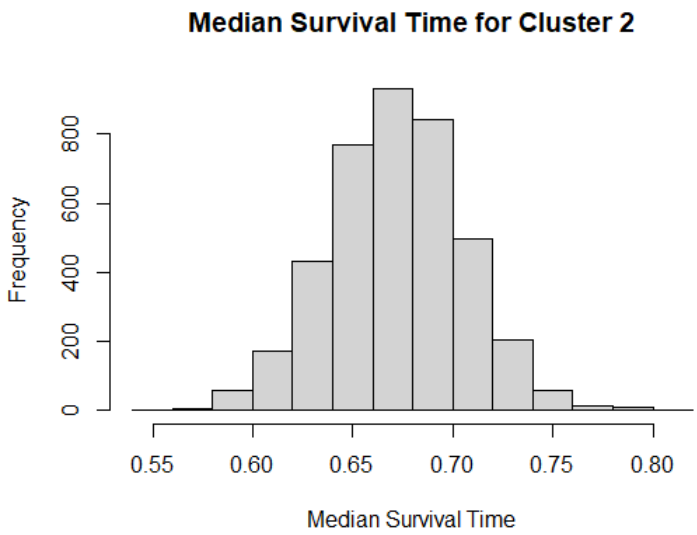

Figure 3.2: The posterior distribution of the median survival time for cluster 2 at the interim analysis.

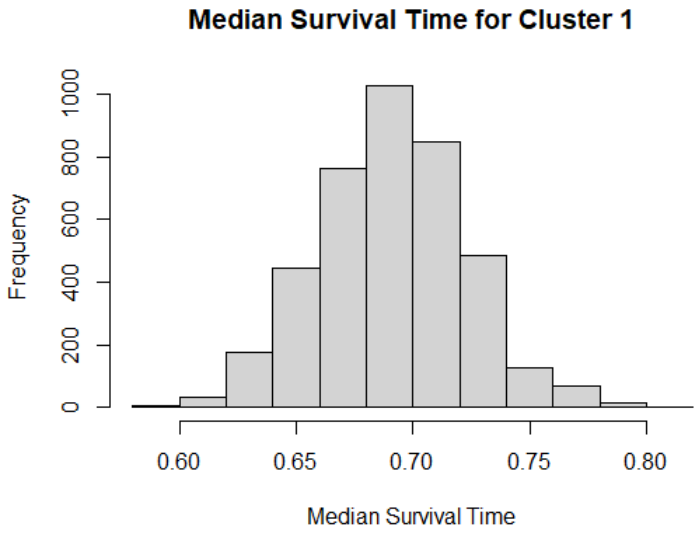

Figure 3.3: The posterior distribution of the median survival time for cluster 1 at the end of the trial.

Chapter 4

Simulation Study

We conduct a simulation study to evaluate the performance of the proposed design. We assume that the cancer types can be grouped into two clusters. The first cluster is the group where the treatment is ineffective, and the second cluster is the group where the treatment is effective. Let $q_0 = 0.5$, and $q_1 = 0.7$ in the hypothesis testing. In our simulation study, there are 12 cancer types and each of them might belong to either cluster 1 or cluster 2. We conduct 5 scenarios where the number of cancer types that belong to cluster 1 differs between each scenario. If a cancer type belongs to cluster 1 (ineffective cluster), then the median survival time we set for this cancer type is 0.5. By contrast, if a cancer type belongs to cluster 2 (effective cluster), then the median survival time we specify is 0.7. We set $\lambda = 1.96$, $r = 1.5$, $\theta_{(2)} = -0.505$ to satisfy these constraints.
For the ineffective group, we set $(\gamma_{1(1)}, \gamma_{2(1)}, \gamma_{3(1)}, \gamma_{4(1)}, \gamma_{5(1)}, \gamma_{6(1)}) = (1, 1, 1, 1, 1, 1)$, so that the mean value of the biomarker measurements stays the same across the observation period. By contrast, for the effective group, we specify $(\gamma_{1(2)}, \gamma_{2(2)}, \gamma_{3(2)}, \gamma_{4(2)}, \gamma_{5(2)}, \gamma_{6(2)}) = (1, 2, 3, 4, 5, 6)$, resulting in the mean biomarker measurements slowly increasing over time. Figure 6.1 shows the mean trajectories of the biomarker for the ineffective group and effective group in this simulation study. We simulate the biomarker measurements, the survival times, and the censoring times. The maximum number of patients for each cancer type is 50. There is one interim analysis conducted when the number of patients for each cancer type is 30. We compare the performance of our proposed model with the performance of the independent approach and the Bayesian Hierarchical Model (BHM) method. The independent approach uses a method similar to the proposed design without the clustering; it assumes that each cancer type has unique θ and r parameters in the survival function. The model specification and the Gibbs sampler of the BHM method can be found in a later section. The probability cutoffs Q_f and Q in the trial design are the same across the three methods so that the comparison would be meaningful. We conduct 500 simulations for each scenario.
Table 4.1 shows the results of the simulation study. In this table, "Early Stop" means that the trial

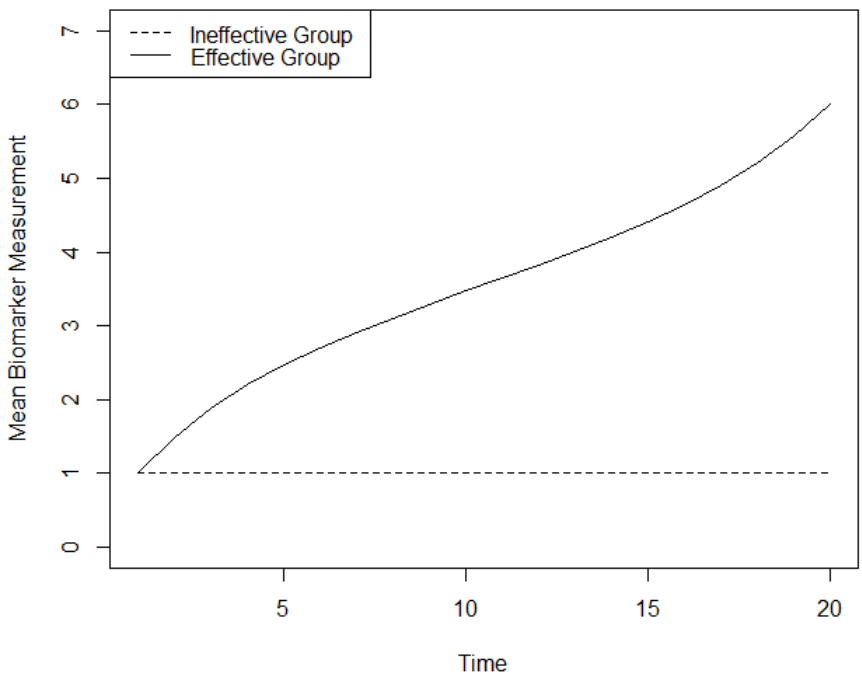

Figure 4.1: Mean biomarker measurement. The solid and dashed line represents the mean biomarker measurement for the effective group and the ineffective group, respectively.

is stopped for this cancer type after the interim analysis because the treatment is considered to be ineffective. If the trial is stopped early for a cancer type, then no new patients will be recruited for this cancer type. "Reject" means that the null hypothesis is rejected for this cancer type, and we conclude that the treatment is effective for this cancer type. "No-Reject" means that the trial is not early stopped and the null hypothesis is not rejected after interim 2. "Sample Size" means the average number of patients that are recruited in each cancer type, which are averaged over the cluster. In scenario A1, there are 12 effective cancer types and no ineffective cancer types. The early stopping rate for the effective group using our proposed model (0.0%) is lower than that using the independent approach (0.7%) or the BHM method (1.6%), which means that fewer patients who are actually in the effective group are wrongly stopped being recruited using our proposed model. The rejection rate for the effective group after interim 2 using our proposed model (99.6%) is higher than that using the independent approach (70.5%) or the BHM (73.6%) method, which means that more patients who are actually in the effective group are correctly classified into the effective group after the trial is finished using our proposed model. Our proposed method performs better than the other two methods in these two aspects because our proposed model borrows information across the cancer types that are classified into the same cluster.

			Early Stop %	Reject %	No-Reject %	Sample Size
A1	Proposed Model	Ineffective	-	-	-	-
		Effective	0.0	99.6	0.4	50.0
	Independent Approach	Ineffective	-	-	-	-
		Effective	0.7	70.5	28.9	49.9
	BHM	Ineffective	-	-	-	-
		Effective	1.6	73.6	24.8	49.7
A2	Proposed Model	Ineffective	67.8	9.8	22.4	36.4
		Effective	0.0	98.9	1.1	50.0
	Independent Approach	Ineffective	21.6	9.3	69.1	45.7
		Effective	0.7	68.0	31.2	49.9
	BHM	Ineffective	39.0	1.7	59.3	42.2
		Effective	1.8	71.2	27.0	49.6
A3	Proposed Model	Ineffective	85.4	1.4	13.2	32.9
		Effective	1.7	92.5	5.7	49.7
	Independent Approach	Ineffective	20.5	11.9	67.6	45.9
		Effective	0.7	63.3	36.0	49.9
	BHM	Ineffective	41.8	1.8	56.4	41.6
		Effective	1.9	69.7	28.4	49.6
A4	Proposed Model	Ineffective	99.0	0.0	1.0	30.2
		Effective	3.1	88.3	8.7	49.4
	Independent Approach	Ineffective	20.4	11.7	67.8	45.9
		Effective	0.7	49.1	50.2	49.9
	BHM	Ineffective	45.8	1.8	52.4	40.8
		Effective	2.5	67.1	30.4	49.5
A5	Proposed Model	Ineffective	96.4	0.1	3.6	30.7
		Effective	-	-	-	-
	Independent Approach	Ineffective	20.3	6.8	72.9	46.0
		Effective	-	-	-	-
	BHM	Ineffective	49.4	1.5	49.1	40.1
		Effective	-	-	-	-

Table 4.1: Results of the main simulation study which compares the performance of the proposed model with the independent approach and the BHM method. This table shows the early stopping rate, rejection rate, and sample size under five simulation scenarios.

Scenario A3 has 6 ineffective cancer types and 6 effective cancer types. Our proposed model results in a much higher early stopping rate (85.4%) for the ineffective group than the independent approach (20.5%) or the BHM method (41.8%), which means that more patients who are actually in the ineffective group are correctly stopped being recruited using our proposed model. This indicates that our proposed model results in a smaller sample size and is more likely to stop the trial early for the ineffective cluster. The rejection rate for the ineffective group after interim 2 using our proposed model (1.4%) is lower than that using the independent approach (11.9%) or the BHM method (1.8%), which means that our proposed model yields a lower type-I error rate than

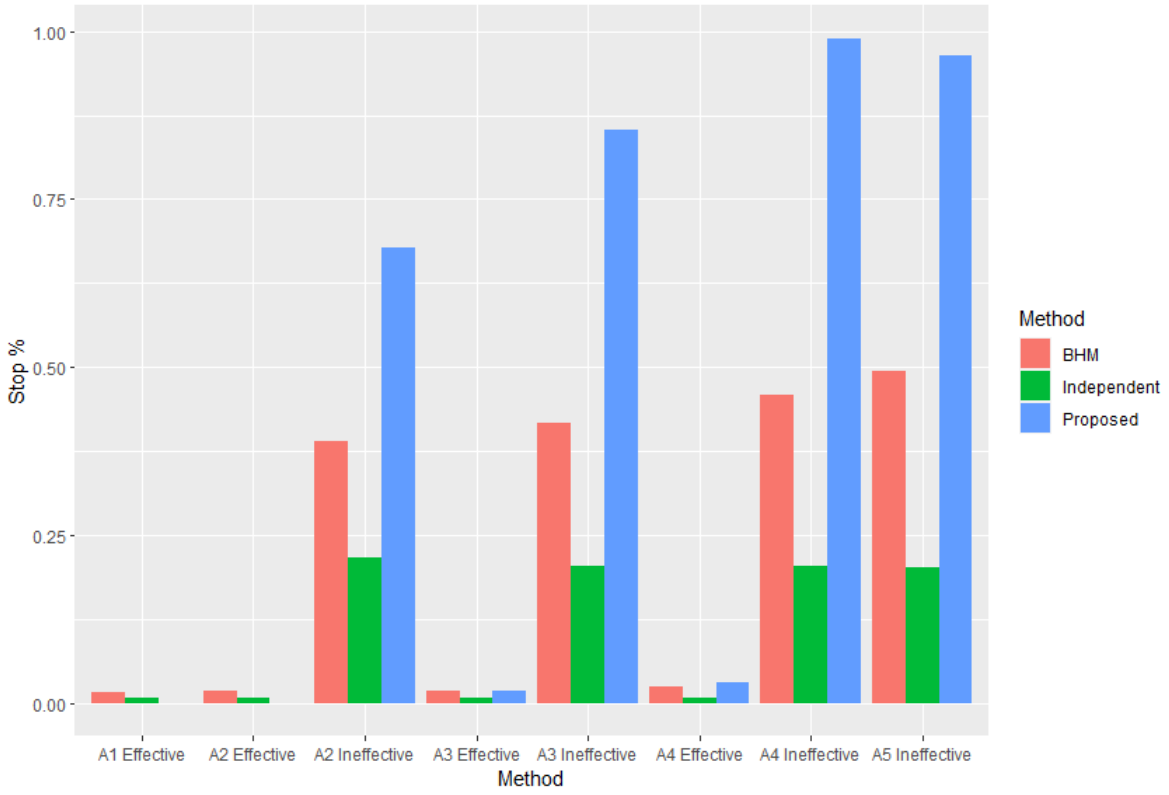

Figure 4.2: Early stopping rate in the main simulation study. The red, green, and blue bars represent the BHM method, the independent approach, and the proposed method, respectively. Scenarios A1 to A5 vary regarding the number of effective/ineffective treatments.

the BHM method and the independent approach. The early stopping rate for the effective group using our proposed model (1.7%) is only slightly higher than that using the independent approach (0.7%) and is lower than that using the BHM method (1.9%). The rejection rate for the effective group after interim 2 using our proposed model (92.5%) is higher than that using the independent approach (63.3%) or the BHM (69.7%) method, which indicates that our proposed method yields much higher power than the other two methods. Again, the reason why our proposed model performs better is that it borrows information within the subgroup.

In scenarios A2, A4, and A5, the number of ineffective cancer types is 3, 9, and 12, respectively, and the number of effective cancer types is 9, 3, and 0, respectively. In these three scenarios, our proposed design generally performs better than the other two methods in most aspects. However, there are some drawbacks to our proposed model. For instance, in scenario A4, the early stopping rate for the effective group using our proposed model (3.1%) is higher than that using the independent approach (0.7%) or the BHM (2.5%) method.

In summary, our proposed model outperforms the independent approach or the BHM method in most cases in terms of the higher early stopping rate for the ineffective group, lower early stopping

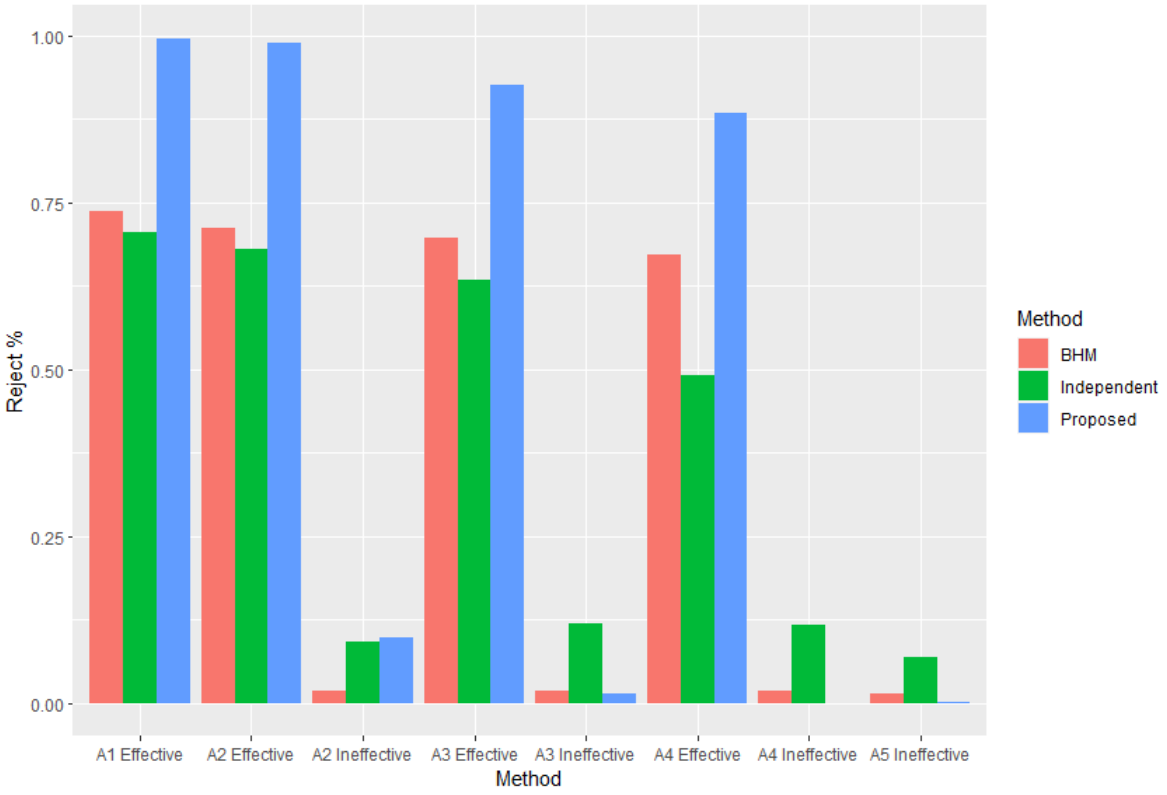

Figure 4.3: Rejection rate, which is the proportion of trials where the null hypothesis is rejected, in the main simulation study. The red, green, and blue bars represent the BHM method, the independent approach, and the proposed method, respectively. Scenarios A1 to A5 vary regarding the number of effective/ineffective treatments.

rate for the effective group, higher rejection rate for the effective group, and lower rejection rate for the ineffective group.

Chapter 5

Sensitivity Analysis

We conduct three sensitivity analyses to evaluate the robustness of the proposed model. In the first sensitivity analysis, we change the prior distribution of θ_2 from Uniform(-1,1) to Uniform(-2,0), while there are still 12 cancer types so that we can evaluate whether the model performs well when the prior distribution of a certain parameter is changed.

Table 5.1 shows the results of the simulation study of the first sensitivity analysis. In scenario B1, there are 12 effective cancer types and no ineffective cancer types. The early stopping rate for the effective group using our proposed model (0.0%) is lower than that using the independent approach (0.7%) or the BHM method (1.6%). The rejection rate for the effective group after interim 2 using our proposed model (99.5%) is higher than that using the independent approach (70.5%) or the BHM (73.6%) method. In both of these aspects, our proposed model outperforms the other two methods.

In scenario B2, there are 3 ineffective cancer types and 9 effective cancer types. The early stopping rate for the ineffective group using our proposed model (58.0%) is higher than that using the independent approach (21.6%) or the BHM method (39.0%). The early stopping rate for the effective group using our proposed model (0.0%) is lower than that using the independent approach (0.7%) or the BHM method (1.8%). The rejection rate for the effective group after interim 2 using our proposed model (97.2%) is higher than that using the independent approach (68.0%) or the BHM (71.2%) method. However, the rejection rate for the ineffective group after interim 2 using our proposed model (23.0%) is higher than that using the independent approach (9.3%) or the BHM (1.7%) method, which is a drawback to our proposed model.

In scenarios B3, B4, and B5, the number of ineffective cancer types is 6, 9, and 12, respectively, and the number of effective cancer types is 6, 3, and 0, respectively. In these three scenarios, our proposed design generally performs better than the BHM method and the independent approach in most of the cases, although there are some cases in which our proposed model performs worse than the other two methods. This simulation study shows that the performance of our proposed

			Early Stop %	Reject %	No-Reject %	Sample Size
B1	Proposed Model	Ineffective	-	-	-	-
		Effective	0.0	99.5	0.5	50.0
	Independent Approach	Ineffective	-	-	-	-
		Effective	0.7	70.5	28.9	49.9
	BHM	Ineffective	-	-	-	-
		Effective	1.6	73.6	24.8	49.7
B2	Proposed Model	Ineffective	58.0	23.0	19.0	38.4
		Effective	0.0	97.2	2.8	50.0
	Independent Approach	Ineffective	21.6	9.3	69.1	45.7
		Effective	0.7	68.0	31.2	49.9
	BHM	Ineffective	39.0	1.7	59.3	42.2
		Effective	1.8	71.2	27.0	49.6
B3	Proposed Model	Ineffective	59.2	3.0	37.8	38.2
		Effective	7.5	64.5	27.9	48.5
	Independent Approach	Ineffective	20.5	11.9	67.6	45.9
		Effective	0.7	63.3	36.0	49.9
	BHM	Ineffective	41.8	1.8	56.4	41.6
		Effective	1.9	69.7	28.4	49.6
B4	Proposed Model	Ineffective	96.8	0.0	3.2	30.6
		Effective	10.6	81.7	7.7	47.9
	Independent Approach	Ineffective	20.4	11.7	67.8	45.9
		Effective	0.7	49.1	50.2	49.9
	BHM	Ineffective	45.8	1.8	52.4	40.8
		Effective	2.5	67.1	30.4	49.5
B5	Proposed Model	Ineffective	97.3	0.0	2.7	30.5
		Effective	-	-	-	-
	Independent Approach	Ineffective	20.3	6.8	72.9	46.0
		Effective	-	-	-	-
	BHM	Ineffective	49.4	1.5	49.1	40.1
		Effective	-	-	-	-

Table 5.1: Simulation results of the sensitivity analysis when the prior distribution of one hyperparameter is changed. This table shows the early stopping rate, rejection rate, and sample size under five simulation scenarios. Scenarios B1 to B5 vary regarding the number of effective/ineffective treatments.

method is generally superior even if we change the prior distribution of one hyperparameter.

In the second sensitivity analysis, we increase the number of cancer types from 12 to 18, while keeping the prior distribution of θ_2 to be Uniform(-1,1), so that we can evaluate whether the model performs well when there are more cancer types.

Table 5.2 shows the results of the simulation study of the second sensitivity analysis. In scenarios

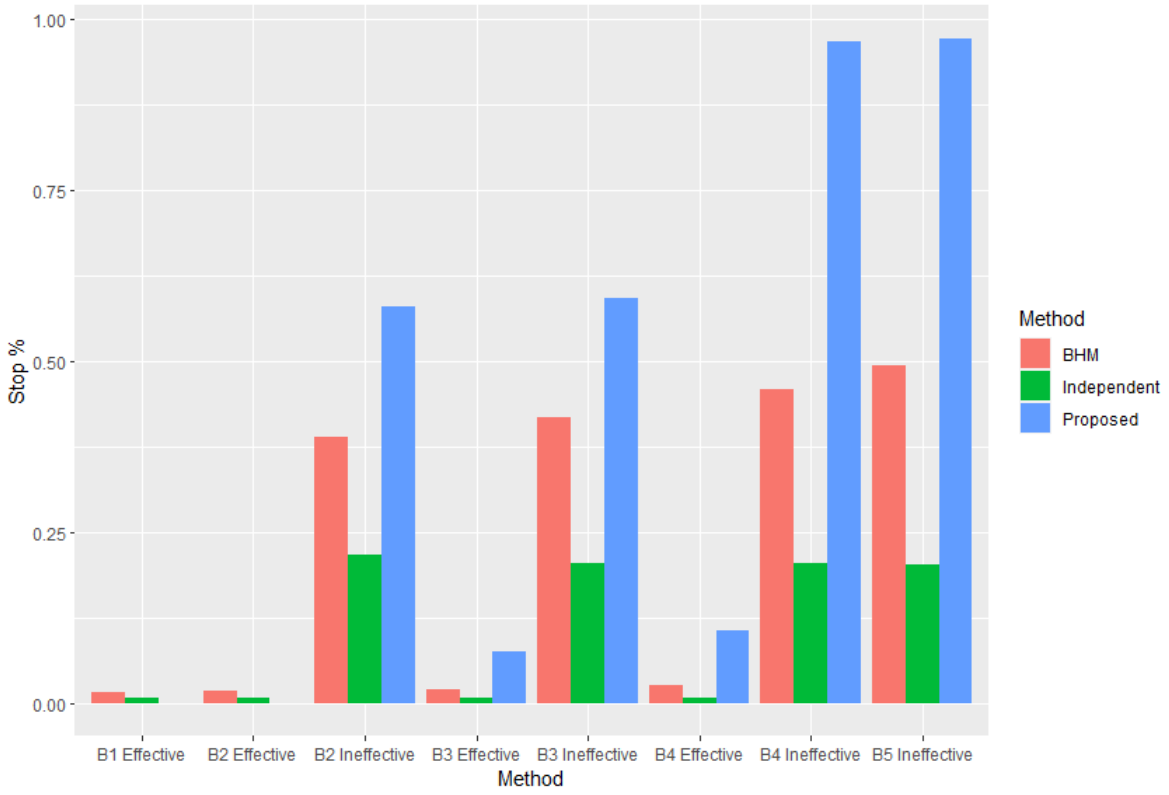

Figure 5.1: Early stopping rate when the prior distribution of one hyperparameter is changed. The red, green, and blue bars represent the BHM method, the independent approach, and the proposed method, respectively. Scenarios B1 to B5 vary regarding the number of effective/ineffective treatments.

C1, C2, and C3, the number of ineffective cancer types is 5, 9, and 18, respectively, and the number of effective cancer types is 13, 9, and 0, respectively. In these three scenarios, our proposed model performs better than the other two methods in general in terms of the early stopping rate and the rejection rate. For example, in scenario C2, the early stopping rate for the ineffective group using our proposed model (89.4%) is higher than that using the independent approach (20.1%) or the BHM method (43.4%). The rejection rate for the ineffective group after interim 2 using our proposed model (0.4%) is lower than that using the independent approach (12.2%) or the BHM method (1.9%). The rejection rate for the effective group after interim 2 using our proposed model (93.0%) is higher than that using the independent approach (64.0%) or the BHM (69.0%) method. The only aspect in this scenario where our proposed model performs worse is that the early stopping rate for the effective group using our proposed model (1.8%) is higher than that using the independent approach (0.5%) and is only slightly lower than that using the BHM (2.2%) method. This simulation study shows that our proposed method performs greatly even if we increase the number of cancer types.

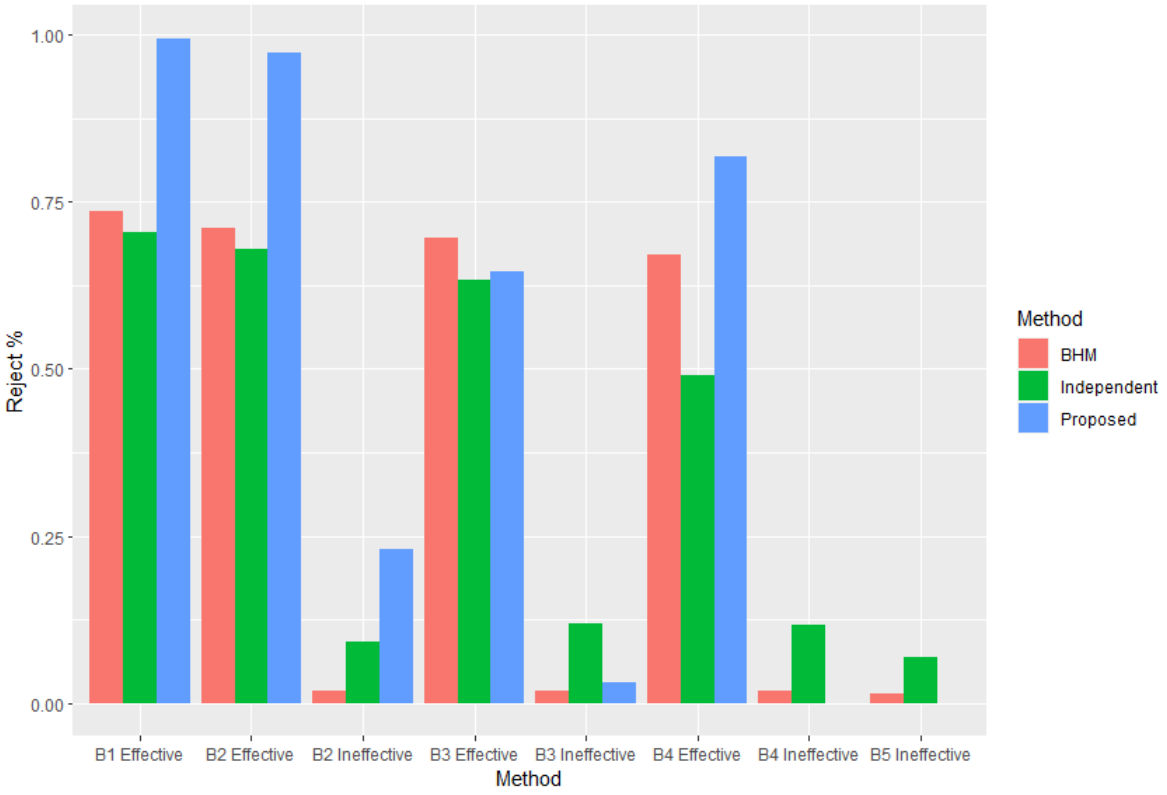

Figure 5.2: Rejection rate, which is the proportion of trials where the null hypothesis is rejected when the prior distribution of one hyperparameter is changed. The red, green, and blue bars represent the BHM method, the independent approach, and the proposed method, respectively. Scenarios B1 to B5 vary regarding the number of effective/ineffective treatments.

In the third sensitivity analysis, we increase the number of cancer types from 12 to 18 and change the prior distribution of θ_2 from Uniform(-1,1) to Uniform(-2,0), so that we can evaluate the performance of the model when there are more cancer types and the prior distribution of one hyperparameter is changed.

Table 5.3 shows the results of the simulation study of the third sensitivity analysis. In scenario D1, D2, and D3, the number of ineffective cancer types is 5, 9, and 18, respectively, and the number of effective cancer types is 13, 9, and 0, respectively. In most cases, our proposed model outperforms the other two methods in terms of the early stopping rate and rejection rate. For example, in scenario D1, The early stopping rate for the ineffective group using our proposed model (75.4%) is higher than that using the independent approach (19.3%) or the BHM method (39.5%). The early stopping rate for the effective group using our proposed model (0.0%) is lower than that using the independent approach (0.5%) or the BHM method (2.0%). The rejection rate for the effective

			Early Stop %	Reject %	No-Reject %	Sample Size
C1	Proposed Model	Ineffective	82.2	4.8	13.0	33.6
		Effective	0.0	98.4	1.6	50.0
	Independent Approach	Ineffective	19.3	10.5	70.2	46.1
		Effective	0.5	68.8	30.7	49.9
	BHM	Ineffective	39.5	1.8	58.7	42.1
		Effective	2.0	70.8	27.3	49.6
C2	Proposed Model	Ineffective	89.4	0.4	10.2	32.1
		Effective	1.8	93.0	5.2	49.6
	Independent Approach	Ineffective	20.1	12.2	67.7	46.0
		Effective	0.5	64.0	35.5	49.9
	BHM	Ineffective	43.4	1.9	54.7	41.3
		Effective	2.2	69.0	28.8	49.6
C3	Proposed Model	Ineffective	97.3	0.0	2.7	30.5
		Effective	-	-	-	-
	Independent Approach	Ineffective	20.3	7.5	72.2	45.9
		Effective	-	-	-	-
	BHM	Ineffective	51.7	1.6	46.7	39.7
		Effective	-	-	-	-

Table 5.2: Simulation results of the sensitivity analysis when the number of cancer types increases. This table shows the early stopping rate, rejection rate, and sample size under three simulation scenarios.

group after interim 2 using our proposed model (96.8%) is higher than that using the independent approach (68.8%) or the BHM (70.8%) method. The only drawback to our proposed model in this scenario is that the rejection rate for the ineffective group after interim 2 using our proposed model (10.2%) is only slightly lower than that using the independent approach (10.5%) and is higher than that using the BHM (1.8%) method.

In summary, our proposed model still performs better in most of the scenarios, which means that the model is not very sensitive to the number of cancer types or the prior distribution of the hyperparameter.

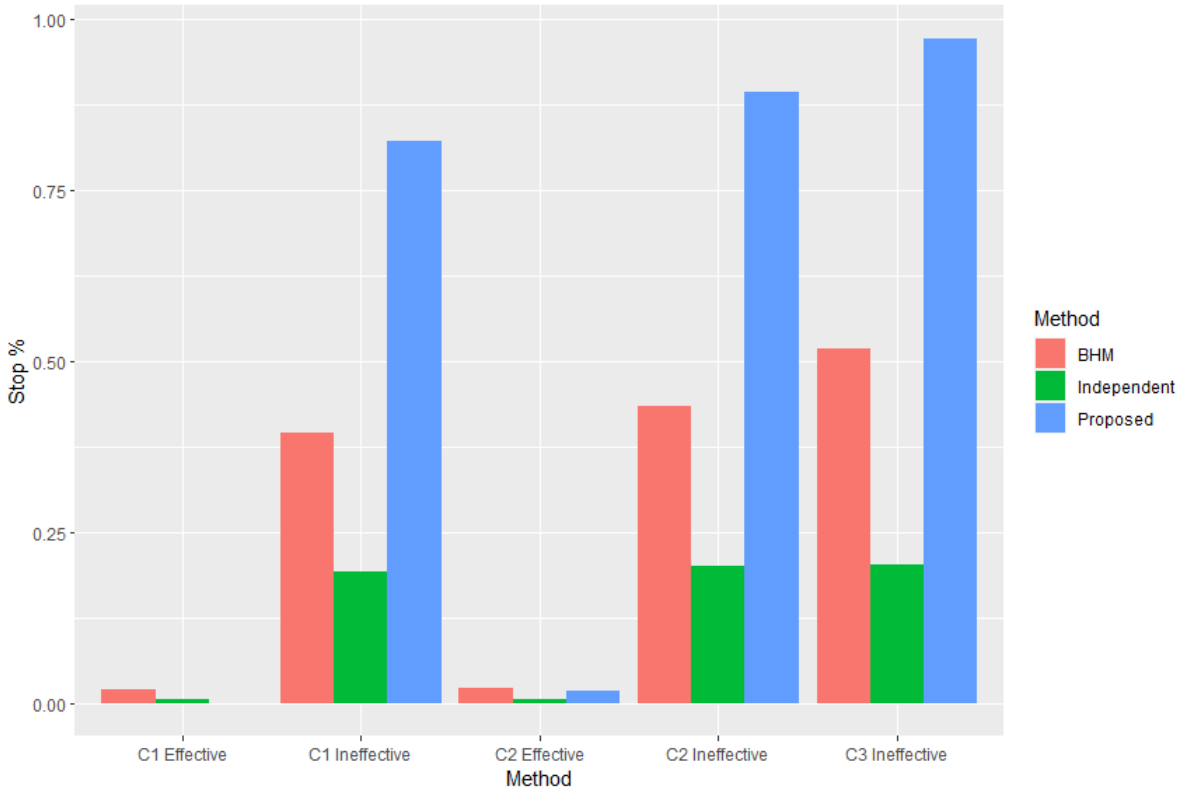

Figure 5.3: Early stopping rate when the number of cancer types increases. The red, green, and blue bars represent the BHM method, the independent approach, and the proposed method, respectively. Scenarios C1 to C3 vary regarding the number of effective/ineffective treatments.

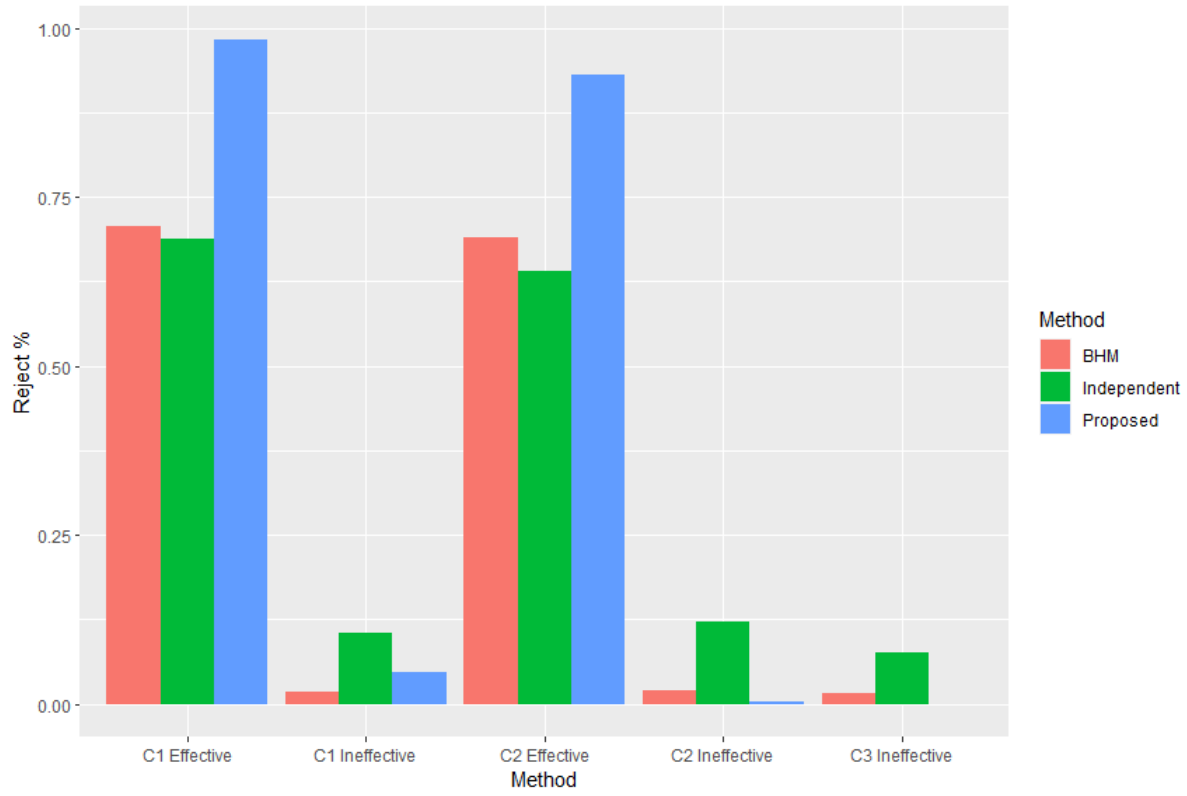

Figure 5.4: Rejection rate, which is the proportion of trials where the null hypothesis is rejected when the number of cancer types increases. The red, green, and blue bars represent the BHM method, the independent approach, and the proposed method, respectively. Scenarios C1 to C3 vary regarding the number of effective/ineffective treatments.

			Early Stop %	Reject %	No-Reject %	Sample Size
D1	Proposed Model	Ineffective	75.4	10.2	14.4	34.9
		Effective	0.0	96.8	3.2	50.0
	Independent Approach	Ineffective	19.3	10.5	70.2	46.1
		Effective	0.5	68.8	30.7	49.9
	BHM	Ineffective	39.5	1.8	58.7	42.1
		Effective	2.0	70.8	27.3	49.6
D2	Proposed Model	Ineffective	73.4	1.4	25.2	35.3
		Effective	6.4	77.5	16.1	48.7
	Independent Approach	Ineffective	20.1	12.2	67.7	46.0
		Effective	0.5	64.0	35.5	49.9
	BHM	Ineffective	43.4	1.9	54.7	41.3
		Effective	2.2	69.0	28.8	49.6
D3	Proposed Model	Ineffective	97.9	0.0	2.1	30.4
		Effective	-	-	-	-
	Independent Approach	Ineffective	20.3	7.5	72.2	45.9
		Effective	-	-	-	-
	BHM	Ineffective	51.7	1.6	46.7	39.7
		Effective	-	-	-	-

Table 5.3: Simulation results of the sensitivity analysis when the number of cancer types increases, and the prior distribution of one hyperparameter is changed. This table shows the early stopping rate, rejection rate, and sample size under three simulation scenarios.

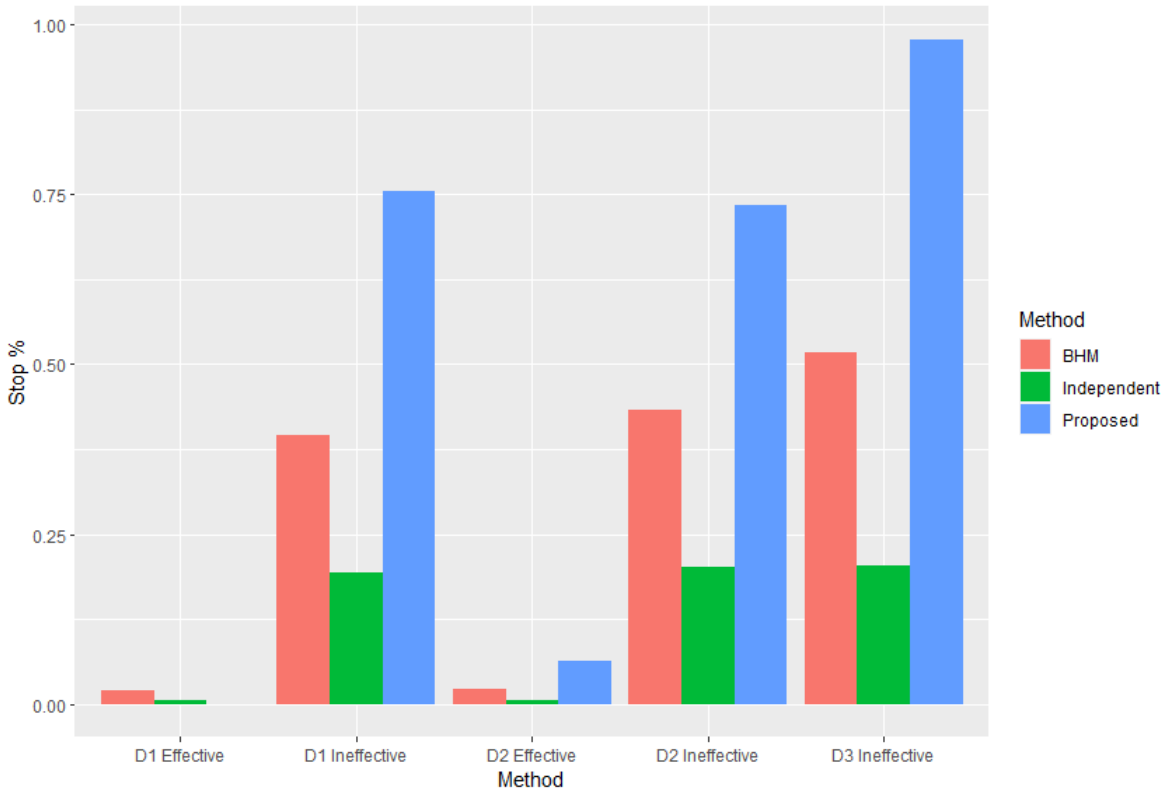

Figure 5.5: Early stopping rate when the number of cancer types increases, and the prior distribution of one hyperparameter is changed. The red, green, and blue bars represent the BHM method, the independent approach, and the proposed method, respectively. Scenarios D1 to D3 vary regarding the number of effective/ineffective treatments.

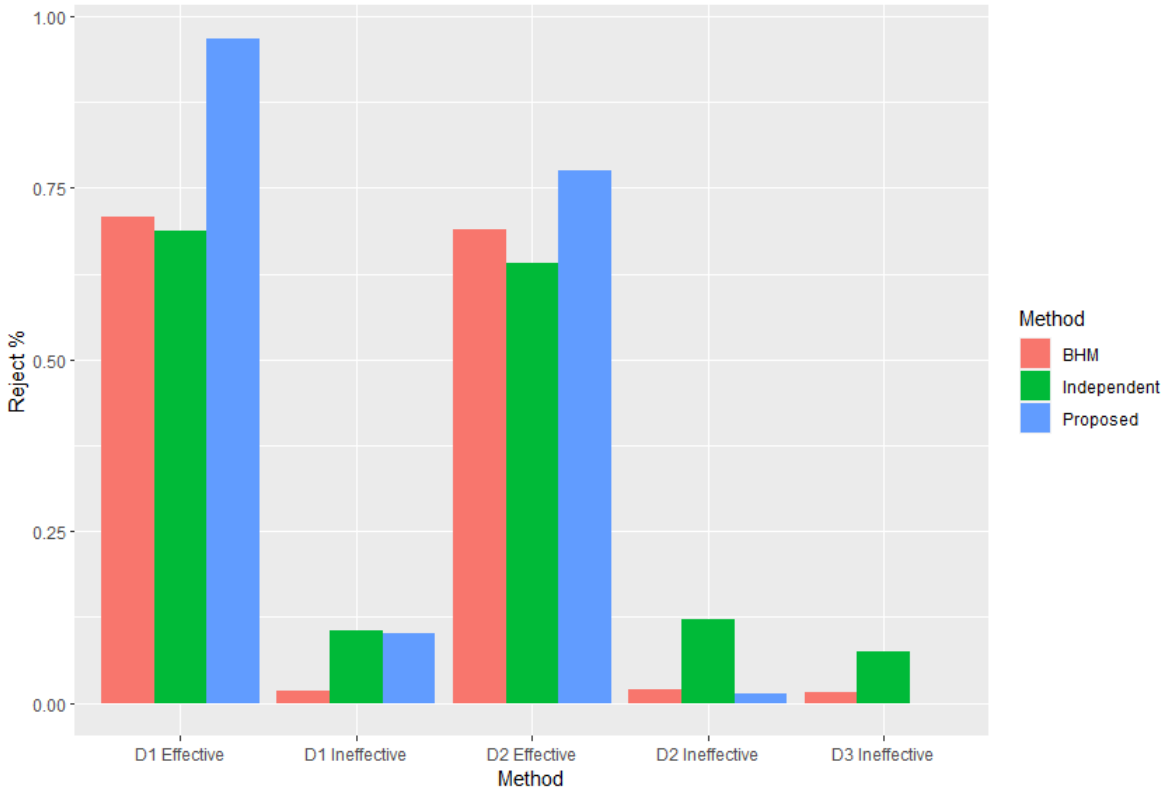

Figure 5.6: Rejection rate, which is the proportion of trials where the null hypothesis is rejected, when the number of cancer types increases, and the prior distribution of one hyperparameter is changed. The red, green, and blue bars represent the BHM method, the independent approach, and the proposed method, respectively. Scenarios D1 to D3 vary regarding the number of effective/ineffective treatments.

Chapter 6

Conclusion and Discussion

In conclusion, we propose a basket trial method to test the effectiveness of a treatment for several types of cancers for which the endpoint is the survival time, where the cancer types are classified according to the survival time and the biomarker measurements. We conduct Bayesian inference and make interim decisions about whether to stop recruiting for each cluster of patients. The simulation study shows that our proposed method performs better than the independent approach and the BHM method in most of the scenarios. The main advantage of our proposed method is that it incorporates the longitudinal biomarker measurements and the continuous endpoint.

There are some limitations of the proposed design. First, the biomarkers in our proposed design must be measured multiple times, so that the method is not suitable for clinical trials in which biomarkers can only be measured one time. In some clinical trials, researchers only measure biomarkers one time because it is expensive or time-consuming to measure the biomarkers. Second, our proposed design only considers one continuous endpoint (progression-free survival time) as the primary endpoint; it does not incorporate other secondary endpoints into the study.

Some potential research topics can extend our proposed method. For example, we can include a toxicity indicator as a secondary endpoint, which can be used in the classification of cancer types. We can also measure the immune response of the patients, and use that additional information to help classify the cancer types into subgroups.

Chapter 7

Gibbs Sampler for the Proposed Model

Step 1: We update C_i.
We denote $\boldsymbol{Z_{ij}} = (Z_{ij1}, ..., Z_{ijL})$, and $\boldsymbol{Z_i} = (\boldsymbol{Z_{i1}}, ..., \boldsymbol{Z_{iJ}})$. For $i = 1, ..., I, C_i \sim$ Multinomial$(\tau_{i1}, ..., \tau_{iK})$, where

$$\tau_{ik} = \frac{\pi_{ik} \mathrm{N}_{Q_i}(\boldsymbol{u_{i(k)}}, \sigma_\epsilon^2 \boldsymbol{I_{Q_i}}) \prod_j [\{f_t(t_{ij})\}^{\omega_{ij}} \{S_t(t_{ij})\}^{1-\omega_{ij}}]}{\sum_{i=1}^K \pi_{ik} \mathrm{N}_{Q_i}(\boldsymbol{u_{i(k)}}, \sigma_\epsilon^2 \boldsymbol{I_i}) \prod_j [\{f_t(t_{ij})\}^{\omega_{ij}} \{S_t(t_{ij})\}^{1-\omega_{ij}}]},$$

$$\text{where } \boldsymbol{u_{i(k)}} = \boldsymbol{X_{i\alpha}} \boldsymbol{\alpha_{(k)}} + \boldsymbol{X_{ib}} \boldsymbol{b_i}, Q_i = Ln_i,$$

$$\boldsymbol{\alpha_{(k)}} = (\gamma_{1(k)}, \gamma_{2(k)}, \cdots, \gamma_{S(k)})^\top,$$

$$\boldsymbol{b_i} = (v_i, w_{i1}, ..., w_{iJ})^\top.$$

Here, $\mathrm{N}_{Q_i}(w, \sigma^2 I_{Q_i})$ is a Q_i-variate normal density function of Z_i, $\boldsymbol{X_{i\alpha}}$ is a Q_i-by-S design matrix of the i-th arm associated with $\alpha_{(k)}$, and $\boldsymbol{X_{ib}}$ is a Q_i-by-$(1 + J)$ design matrix of the i-th arm associated with the random-effect vector $\boldsymbol{b_i}$.
This formula helps to classify cancer types into subgroups based on the prior probability π, the biomarker measurements (Z), and the survival outcomes $f_t(t_{ij})$ and $S_t(t_{ij})$. We calculate the product of the likelihoods from the three kinds of information for each subgroup k and then normalize it by dividing the product of the likelihoods by the sum of the product of the likelihoods over k. If τ_{ik} is higher for some subgroup k than for other subgroups, then this means that based on the information, it is more likely that the i-th cancer type belongs to the k-th subgroup.
Step 2: Next, we derive the full conditional distribution of π_i as follows. Since the prior distribution of π_i is $(\pi_{i1}, \ldots, \pi_{iK}) \sim$ Dirichlet$(2, \ldots, 2)$, and the likelihood of C_i is $C_i|\pi_i \sim$ Multinomial$(\pi_{i1}, \ldots, \pi_{iK})$, we get the posterior distribution based on the conjugacy of Multinomial/Dirichlet distributions.

$$\pi_i|. \sim \mathrm{Dir}\{I(C_i = 1) + 2, ..., I(C_i = K) + 2\},$$

where $I(\cdot)$ is the indicator function.
Step 3: Next, we derive the full conditional distribution of σ_v^2, σ_w^2, and σ_ϵ^2. Since $v_i \sim N(0, \sigma_v^2)$, $w_{ij} \sim N(0, \sigma_w^2)$, and $\epsilon_{ijl} \sim N(0, \sigma_\epsilon^2)$, where σ_v^2, σ_w^2, and σ_ϵ^2 are the variances of the corresponding

Normal distributions, and the prior distributions are $\sigma_v^2 \sim \text{IG}(10^{-3}, 10^{-3})$, $\sigma_w^2 \sim \text{IG}(10^{-3}, 10^{-3})$, and $\sigma_\epsilon^2 \sim \text{IG}(10^{-3}, 10^{-3})$, we are calculating the posterior distribution of the variance of the Normal distribution where the mean is known and the prior distribution of the variance follows Inverse Gamma distribution. As a result, we get the posterior distributions.

$$\begin{gathered}\sigma_v^2|. \sim \text{IG}(10^{-3} + 0.5I, 10^{-3} + 0.5\sum_{i=1}^{I} v_i^2),\\ \sigma_w^2|. \sim \text{IG}(10^{-3} + 0.5N, 10^{-3} + 0.5\sum_{i=1}^{I}\sum_{j=1}^{n_j} w_{ij}^2),\\ \sigma_\epsilon^2|. \sim \text{IG}(10^{-3} + 0.5LN, 10^{-3} + 0.5\sum_{i=1}^{I}\sum_{j=1}^{n_j}\sum_{l=1}^{L} \epsilon_{ijl}^2).\end{gathered}$$

Step 4: We update $\boldsymbol{b} = (v_1, \cdots, v_I, w_1, \cdots, w_{IJ})^\top$, where $N = \sum_{i=1}^{i} n_i$, $\boldsymbol{X_\alpha}$ is a Q-by-S design matrix associated with $\boldsymbol{\alpha}$, and $\boldsymbol{X_b}$ is a Q-by-$(I+N)$ design matrix associated with the random-effect vector $\boldsymbol{b}$. The values of $\gamma_{q(k)}$ for each observation (each measurement for each patient) are stored in $\boldsymbol{X_\alpha}$, where each row corresponds to one observation in the dataset, and each column corresponds to one coefficient associated with the B-Spline basis function. The values of v_i and w_{ij} for each observation (each measurement for each patient) are stored in $\boldsymbol{X_b}$, where each row corresponds to one observation in the dataset, and each column corresponds to one parameter in $\boldsymbol{b}$.
Since $Z_{ijl}|(C_i = k) = \mu_k(t_l) + v_i + w_{ij} + \epsilon_{ijl}$, we can use the matrix notations introduced in the previous section to get that $\boldsymbol{Z} - \boldsymbol{X_\alpha\alpha} \sim \text{N}(\boldsymbol{X_b b}, \text{diag}(\sigma_\epsilon^2, \cdots, \sigma_\epsilon^2))$. When we update $\boldsymbol{b}$ in this step, we are calculating the posterior distribution of the mean of a multivariate Normal random variable where the variance is known. Let $\boldsymbol{\Sigma_b}$ denote the variance matrix of the prior distribution of $\boldsymbol{b}$. Since the prior distribution of $\boldsymbol{b}$ is $\text{N}(0, \boldsymbol{\Sigma_b})$, we get the prior distribution of $\boldsymbol{X_b b}$ is $\text{N}(0, \boldsymbol{X_b \Sigma_b X_b^\top})$. As a result, the posterior distribution of $\boldsymbol{X_b b}$ is also a multivariate Normal distribution $\text{N}(\text{E}(\boldsymbol{X_b b}), \text{Var}(\boldsymbol{X_b b}))$, where

$$\begin{aligned}\text{Var}(\boldsymbol{X_b b}) &= [(\boldsymbol{X_b \Sigma_b X_b^\top})^{-1} + \text{diag}(\sigma_\epsilon^{-2}, \cdots, \sigma_\epsilon^{-2})]^{-1},\\ \text{E}(\boldsymbol{X_b b}) &= \text{Var}(\boldsymbol{X_b b})\text{diag}(\sigma_\epsilon^{-2}, \cdots, \sigma_\epsilon^{-2})(\boldsymbol{Z} - \boldsymbol{X_\alpha\alpha}).\end{aligned}$$

As a result, the posterior distribution of $\boldsymbol{b}$ is $f(\boldsymbol{b}|.) \sim \text{N}(\boldsymbol{\eta_b}, \boldsymbol{V_b})$, where

$$\begin{aligned}\boldsymbol{V_b} = \text{Var}(\boldsymbol{X_b^{-1} X_b b}) = \boldsymbol{X_b^{-1}}\text{Var}(\boldsymbol{X_b b})(\boldsymbol{X_b^{-1}})^\top = [(\boldsymbol{X_b^\top}\text{Var}(\boldsymbol{X_b b})^{-1}\boldsymbol{X_b}]^{-1} = (\boldsymbol{\Sigma_b^{-1}} + \boldsymbol{X_b^\top X_b}\sigma_\epsilon^{-2})^{-1},\\ \boldsymbol{\eta_b} = \boldsymbol{X_b^{-1}}\text{E}(\boldsymbol{X_b b}) = \boldsymbol{X_b^{-1} X_b}\text{Var}(\boldsymbol{b})\boldsymbol{X_b^\top}\sigma_\epsilon^{-2}(\boldsymbol{Z} - \boldsymbol{X_\alpha\alpha}) = \sigma_\epsilon^{-2}\boldsymbol{V_b X_b^\top}(\boldsymbol{Z} - \boldsymbol{X_\alpha\alpha}).\end{aligned}$$

In summary, the full conditional distribution of $\boldsymbol{b}$ is as follows,

$$\begin{gathered}f(\boldsymbol{b}|.) \sim \text{N}(\boldsymbol{\eta_b}, \boldsymbol{V_b}), \text{ where}\\ \boldsymbol{V_b} = (\boldsymbol{\Sigma_b^{-1}} + \boldsymbol{X_b^\top X_b}\sigma_\epsilon^{-2})^{-1},\\ \boldsymbol{\eta_b} = \sigma_\epsilon^{-2}\boldsymbol{V_b X_b^\top}(\boldsymbol{Z} - \boldsymbol{X_\alpha\alpha}),\\ \boldsymbol{\Sigma_b} = \begin{pmatrix}\sigma_v^2\boldsymbol{I_I} & 0\\ 0 & \sigma_w^2\boldsymbol{I_N}\end{pmatrix}.\end{gathered}$$

Step 5: We update $\boldsymbol{\alpha_{(k)}} = (\gamma_{1(k)}, \gamma_{2(k)}, \ldots, \gamma_{S(k)})^\top$. Let Q_k denote the number of biomarker measurements in cluster k. Here, $\boldsymbol{X_{\alpha(k)}}$ is a Q_k-by-S design matrix associated with $\boldsymbol{\alpha}$ for cluster k,

and $\boldsymbol{X_{b(k)}}$ is a Q_k-by-$(I+N)$ design matrix associated with the random-effect vector for cluster k. The values of $\gamma_{q(k)}$ for each observation (each measurement for each patient) are stored in $\boldsymbol{X_{\alpha(k)}}$ for each cluster k, where each row corresponds to one observation in the dataset for cluster k, and each column corresponds to one coefficient associated with the B-Spline basis function. The values of v_i and w_{ij} for each observation (each measurement for each patient) are stored in $\boldsymbol{X_{b(k)}}$ for each cluster k, where each row corresponds to one observation in the dataset for cluster k, and each column corresponds to one parameter in $\boldsymbol{b}$.

Since $Z_{ijl}|(C_i = k) = \mu_k(t_l) + v_i + w_{ij} + \epsilon_{ijl}$, we can use the matrix notations introduced in the previous section to get that $\boldsymbol{Z} - \boldsymbol{X_{b(k)}}\boldsymbol{b} \sim \mathrm{N}(\boldsymbol{X_{\alpha(k)}}\boldsymbol{\alpha_k}, \mathrm{diag}(\sigma_\epsilon^2, \cdots, \sigma_\epsilon^2))$. When we update $\boldsymbol{\alpha_k}$ in this step, we are calculating the posterior distribution of the mean of a multivariate Normal random variable where the variance is known. Let $\boldsymbol{\Sigma_{\alpha_k}}$ denote the variance matrix of the prior distribution of $\boldsymbol{\alpha_k}$. Since the prior distribution of $\boldsymbol{\alpha_k}$ is $\mathrm{N}(0, \boldsymbol{\Sigma_{\alpha_k}})$, we can derive the posterior distribution of $\boldsymbol{\alpha_k}$ using exactly the same method in step 4.

$$\begin{aligned}\boldsymbol{V_{\alpha(k)}} &= (\boldsymbol{\Sigma^{-1}_{\alpha(k)}} + \boldsymbol{X^\top_{\alpha(k)}}\boldsymbol{X_{\alpha(k)}}\sigma_\epsilon^{-2})^{-1},\\ \boldsymbol{\eta_{\alpha(k)}} &= \sigma_\epsilon^{-2}\boldsymbol{V_{\alpha(k)}}\boldsymbol{X^\top_{\alpha(k)}}(\boldsymbol{Z_{(k)}} - \boldsymbol{X_{b(k)}}\boldsymbol{b}).\end{aligned}$$

In summary, the full conditional distribution of $\boldsymbol{\alpha_{(k)}}$ is as follows,

$$\begin{aligned}f(\boldsymbol{\alpha_{(k)}}|.) &\sim \mathrm{N}(\boldsymbol{\eta_{\alpha(k)}}, \boldsymbol{V_{\alpha(k)}}), \text{ where}\\ \boldsymbol{V_{\alpha(k)}} &= (\boldsymbol{\Sigma^{-1}_{\alpha(k)}} + \boldsymbol{X^\top_{\alpha(k)}}\boldsymbol{X_{\alpha(k)}}\sigma_\epsilon^{-2})^{-1},\\ \boldsymbol{\eta_{\alpha(k)}} &= \sigma_\epsilon^{-2}\boldsymbol{V_{\alpha(k)}}\boldsymbol{X^\top_{\alpha(k)}}(\boldsymbol{Z_{(k)}} - \boldsymbol{X_{b(k)}}b),\\ \boldsymbol{\Sigma_{\alpha(k)}} &= 10^4\boldsymbol{I_S}.\end{aligned}$$

Step 6: We update θ_2 using Metropolis sampling. Since the prior distribution of θ_2 is $\theta_2 \sim \mathrm{Uniform}(-1, 1)$, and the likelihood which involves θ_2 is $\prod_{ij}[\{f_t(t_{ij})\}^{\omega_{ij}}\{S_t(t_{ij})\}^{1-\omega_{ij}}]$, we get the posterior distribution by multiplying the prior and the likelihood.

$$f(\theta_2|.) \propto f_{unif}(\theta_2, -1, 1)\prod_{ij}[\{f_t(t_{ij})\}^{\omega_{ij}}\{S_t(t_{ij})\}^{1-\omega_{ij}}],$$

where $f_t(t_{ij})$ is the probability density function of PFS of the i-th patient in the j-th cancer type, and $S_t(t_{ij})$ is the survival function of PFS of the i-th patient in the j-th cancer type. Here, $f_{unif}(x, a, b)$ is the probability density function of a Uniform(a,b) random variable.

Step 7: We update λ using Metropolis sampling. Since the prior distribution of λ is $\lambda \sim \mathrm{Gamma}(0.1, 0.1)$, and the likelihood which involves λ is $\prod_{ij}[\{f_t(t_{ij})\}^{\omega_{ij}}\{S_t(t_{ij})\}^{1-\omega_{ij}}]$, we get the posterior distribution by multiplying the prior and the likelihood.

$$f(\lambda|.) \propto f_{gamma}(\lambda, 0.1, 10)\prod_{ij}[\{f_t(t_{ij})\}^{\omega_{ij}}\{S_t(t_{ij})\}^{1-\omega_{ij}}],$$

where $f_{gamma}(x, \alpha, \beta)$ is the probability density function of a Gamma random variable where the shape is α and the rate is β.

Step 8: We update r using Metropolis sampling. Since the prior distribution of r is $r \sim \mathrm{Gamma}(0.1, 0.1)$, and the likelihood which involves r is $\prod_{ij}[\{f_t(t_{ij})\}^{\omega_{ij}}\{S_t(t_{ij})\}^{1-\omega_{ij}}]$, we get the posterior distribution by multiplying the prior and the likelihood.

$$f(r|.) \propto f_{gamma}(r, 0.1, 10) \prod_{ij} [\{f_t(t_{ij})\}^{\omega_{ij}} \{S_t(t_{ij})\}^{1-\omega_{ij}}],$$

where $f_{gamma}(x, \alpha, \beta)$ is the probability density function of a Gamma random variable where the shape is α and the rate is β.

Chapter 8

Model Specification and Gibbs Sampler for the BHM

The structure of the Bayesian Hierarchical Model is as follows. The survival function for the i-th patient in the j-th cancer type is related to the cancer type of the patient in the following way, where θ_i, λ_i, and r_i are the parameters corresponding to each cancer type.

$$S_t(t_{ij}) = \exp\{-\lambda_i t_{ij}^{r_i} \exp(\theta_i)\}.$$

We assume the parameters have the distributions, where μ_θ, σ_θ^2, α_λ, β_λ, α_r, and β_r are the hyper-parameters.

$$\begin{aligned}
\theta_i &\sim \mathrm{N}(\mu_\theta, \sigma_\theta^2), \\
\lambda_i &\sim \mathrm{Gamma}(\alpha_\lambda, \beta_\lambda), \\
r_i &\sim \mathrm{Gamma}(\alpha_r, \beta_r).
\end{aligned}$$

We assume the hyperparameters have the prior distributions.

$$\begin{aligned}
\mu_\theta &\sim \mathrm{N}(0, 10), \\
\sigma_\theta^2 &\sim \mathrm{Unif}(0, 10), \\
\alpha_\lambda &\sim \mathrm{Unif}(0, 1), \\
\beta_\lambda &\sim \mathrm{Unif}(0, 1), \\
\alpha_r &\sim \mathrm{Unif}(0, 1), \\
\beta_r &\sim \mathrm{Unif}(0, 1).
\end{aligned}$$

The Gibbs sampler is as follows.
Step 1: We update θ_i, λ_i, and r_i using Metropolis sampling

$$\begin{aligned}
f(\theta_i|.) &\propto f_N(\theta_i, \mu_\theta, \sigma_\theta^2) \textstyle\prod_j [\{f_t(t_{ij})\}^{\omega_{ij}} \{S_t(t_{ij})\}^{1-\omega_{ij}}], \\
f(\lambda_i|.) &\propto f_{gamma}(\lambda_i, \alpha_\lambda, \beta_\lambda) \textstyle\prod_j [\{f_t(t_{ij})\}^{\omega_{ij}} \{S_t(t_{ij})\}^{1-\omega_{ij}}], \\
f(r_i|.) &\propto f_{gamma}(r_i, \alpha_r, \beta_r) \textstyle\prod_j [\{f_t(t_{ij})\}^{\omega_{ij}} \{S_t(t_{ij})\}^{1-\omega_{ij}}].
\end{aligned}$$

where $f_N(x, \mu, \sigma)$ is the probability density function of a Normal(μ, σ) random variable where the mean is μ and the standard deviation is σ, and $f_{gamma}(x, \alpha, \beta)$ is the probability density function of a Gamma random variable where the shape is α and the rate is β. Here, $f_t(t_{ij})$ is the probability density function of PFS of the i-th patient in the j-th cancer type, and $S_t(t_{ij})$ is the survival function of PFS of the i-th patient in the j-th cancer type.
Step 2: We update μ_θ, σ^2_θ, α_λ, β_λ, α_r, and β_r using Metropolis sampling

$$f(\mu_\theta|.) \propto f_N(\mu_\theta, 0, 10)\{\prod_i f_N(\theta_i, \mu_\theta, \sigma^2_\theta)\},\ f(\sigma^2_\theta|.) \propto f_{unif}(\sigma^2_\theta, 0, 10)\{\prod_i f_N(\theta_i, \mu_\theta, \sigma^2_\theta)\},$$
$$f(\alpha_\lambda|.) \propto f_{unif}(\alpha_\lambda, 0, 1)\{\prod_i f_{gamma}(\lambda_i, \alpha_\lambda, \beta_\lambda)\},\ f(\beta_\lambda|.) \propto f_{unif}(\beta_\lambda, 0, 1)\{\prod_i f_{gamma}(\lambda_i, \alpha_\lambda, \beta_\lambda)\},$$
$$f(\alpha_r|.) \propto f_{unif}(\alpha_r, 0, 1)\{\prod_i f_{gamma}(r_i, \alpha_r, \beta_r)\},\ f(\beta_r|.) \propto f_{unif}(\beta_r, 0, 1)\{\prod_i f_{gamma}(r_i, \alpha_r, \beta_r)\},$$

where $f_N(x, \mu, \sigma^2)$ is the probability density function of a Normal(μ, σ^2) random variable where the mean is μ and the standard deviation is σ, $f_{unif}(x, a, b)$ is the probability density function of a Uniform(a,b) random variable, and $f_{gamma}(x, \alpha, \beta)$ is the probability density function of a Gamma random variable where the shape is α and the rate is β.

Bibliography

[CL20] Nan Chen and J Jack Lee. "Bayesian cluster hierarchical model for subgroup borrowing in the design and analysis of basket trials with binary endpoints". In: *Statistical Methods in Medical Research* 29.9 (2020). PMID: 32178585, pp. 2717–2732. DOI: 10.1177/0962280220910186. eprint: https://doi.org/10.1177/0962280220910186. URL: https://doi.org/10.1177/0962280220910186.

[CY18] Yiyi Chu and Ying Yuan. "BLAST: Bayesian latent subgroup design for basket trials accounting for patient heterogeneity". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 67.3 (2018), pp. 723–740. DOI: https://doi.org/10.1111/rssc.12255. eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/rssc.12255. URL: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssc.12255.

[Fuj+20] Kei Fujikawa et al. "A Bayesian basket trial design that borrows information across strata based on the similarity between the posterior distributions of the response probability". In: *Biometrical Journal* 62.2 (2020), pp. 330–338. DOI: https://doi.org/10.1002/bimj.201800404. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bimj.201800404. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.201800404.

[Hym+18] David M. Hyman et al. "HER kinase inhibition in patients with HER2- and HER3-mutant cancers". In: *Nature* 554.7691 (Feb. 2018), pp. 189–194. ISSN: 1476-4687. DOI: 10.1038/nature25475. URL: https://doi.org/10.1038/nature25475.

[LTR23] Shufang Liu, Kentaro Takeda, and Alan Rong. "An adaptive biomarker basket design in phase II oncology trials". In: *Pharmaceutical Statistics* 22.1 (2023), pp. 128–142. DOI: https://doi.org/10.1002/pst.2264. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pst.2264. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/pst.2264.

[Pau+20] Cury R. Paulo et al. "Association of Longitudinal Values of Glycated Hemoglobin With Cardiovascular Events in Patients With Type 2 Diabetes and Multivessel Coronary Artery Disease". English. In: *JAMA Network Open* 3.1 (2020). URL: http://proxy.

lib.sfu.ca/login?url=https://www.proquest.com/scholarly-journals/association-longitudinal-values-glycated/docview/2668181506/se-2.

[Sam+19] Robert M. Samstein et al. "Tumor mutational load predicts survival after immunotherapy across multiple cancer types". In: *Nature Genetics* 51.2 (Feb. 2019), pp. 202–206. ISSN: 1546-1718. DOI: 10.1038/s41588-018-0312-8. URL: https://doi.org/10.1038/s41588-018-0312-8.

[TLR22] Kentaro Takeda, Shufang Liu, and Alan Rong. "Constrained hierarchical Bayesian model for latent subgroups in basket trials with two classifiers". In: *Statistics in Medicine* 41.2 (2022), pp. 298–309. DOI: https://doi.org/10.1002/sim.9237. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.9237. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.9237.

[van+22] Frederik A. van Delft et al. "Modeling strategies to analyse longitudinal biomarker data: An illustration on predicting immunotherapy non-response in non-small cell lung cancer". In: *Heliyon* 8.10 (2022), e10932. ISSN: 2405-8440. DOI: https://doi.org/10.1016/j.heliyon.2022.e10932. URL: https://www.sciencedirect.com/science/article/pii/S2405844022022204.

[Wu+17] Yuntao Wu et al. "Longitudinal fasting blood glucose patterns and arterial stiffness risk in a population without diabetes". English. In: *PLoS One* 12.11 (Nov. 2017). URL: http://proxy.lib.sfu.ca/login?url=https://www.proquest.com/scholarly-journals/longitudinal-fasting-blood-glucose-patterns/docview/1966423335/se-2.

[Yin+21] Guosheng Yin et al. "Bayesian Hierarchical Modeling and Biomarker Cutoff Identification in Basket Trials". In: *Statistics in Biopharmaceutical Research* 13.2 (2021), pp. 248–258. DOI: 10.1080/19466315.2020.1811146. eprint: https://doi.org/10.1080/19466315.2020.1811146. URL: https://doi.org/10.1080/19466315.2020.1811146.

[ZJ20] Tianjian Zhou and Yuan Ji. "RoBoT: a robust Bayesian hypothesis testing method for basket trials". In: *Biostatistics* 22.4 (Feb. 2020), pp. 897–912. ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxaa005. eprint: https://academic.oup.com/biostatistics/article-pdf/22/4/897/40579672/kxaa005.pdf. URL: https://doi.org/10.1093/biostatistics/kxaa005.