7-2020

BAYESIAN MODELING OF CENSORED DATA WITH APPLICATION TO META-ANALYSIS OF IMMUNOTHERAPY TRIALS

Xinyue Qi

BAYESIAN MODELING OF CENSORED DATA WITH APPLICATION TO META-ANALYSIS OF IMMUNOTHERAPY TRIALS

Final Doctoral Dissertation Submission

by

Xinyue Qi

Committee Chair/Academic Advisor: Ruosha Li, PhD

Dissertation Supervisor: Shouhao Zhou, PhD

Dissertation Co-supervisor: Christine B. Peterson, PhD

Minor Advisor: David R. Lairson, PhD

Breadth Advisor: Michael D. Swartz, PhD

July, 2020

BAYESIAN MODELING OF CENSORED DATA WITH APPLICATION TO META-ANALYSIS OF IMMUNOTHERAPY TRIALS

by

XINYUE QI, BS, MS

APPROVED:

Ruosha Li Digitally signed by Ruosha Li Date: 2020.06.12 16:27:59 -05'00'

RUOSHA LI, PHD

RUOSHA LI, PHD

Shouhao Zhou Digitally signed by Shouhao Zhou Date: 2020.06.11 16:43:22 -04'00'

SHOUHAO ZHOU, PHD

Christine Peterson Digitally signed by Christine Peterson Date: 2020.06.15 21:39:05 -05'00'

CHRISTINE B. PETERSON, PHD

David Lairson Digitally signed by David Lairson Date: 2020.06.16 08:27:49 -05'00'

DAVID R. LAIRSON, PHD

Michael D. Swartz Digitally signed by Michael D. Swartz Date: 2020.06.16 09:51:12 -05'00'

MICHAEL D. SWARTZ, PHD

RTZ, PHD

DEAN, THE UNIVERSITY OF TEXAS SCHOOL OF PUBLIC HEALTH

DEDICATION

To my family and friends

BAYESIAN MODELING OF CENSORED DATA WITH APPLICATION TO META-ANALYSIS OF IMMUNOTHERAPY TRIALS

by

XINYUE QI

BS, University of Illinois at Urbana-Champaign, 2015

MS, Duke University School of Medicine, 2017

Presented to the Faculty of The University of Texas

School of Public Health

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS

SCHOOL OF PUBLIC HEALTH

Houston, Texas

August, 2020

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere and deepest gratitude to my dissertation supervisors, Dr. Shouhao Zhou and Dr. Christine Peterson, for helping me to identify research topics, guiding me to find solutions when I encountered challenges during my research, and inspiring and encouraging me throughout my research. I am honored to have had the chance to meet and learn from both of these great mentors.

Second, I would like to express my profound appreciation to Dr. Ruosha Li, my academic advisor and committee chair, for her consistent help and support during the past three years. I am also grateful to my minor advisor, Dr. David Lairson, and breadth advisor, Dr. Michael Swartz, for all their suggestions and support as I completed the PhD program.

Third, I would like to give special thanks to Dr. Chan Shen for offering me the opportunity to work as a graduate research assistant at MD Anderson Cancer Center at the very beginning of the PhD program. In addition, I would like to thank my great collaborators from the Department of Lymphoma at MD Anderson for pushing our projects forward culminating in published research articles. I would also like to thank Lionel Santibanez, senior scientific editor at SPH, for reviewing the language and clarity of my dissertation.

Last but not least, I would like to thank my friends and peers from SPH and GSBS for providing me with generous assistance, companionship, and joy during so many long nights of research in the T. Boone Pickens Tower. Finally, I am deeply indebted to my parents for their unconditional love and support throughout my journey of studying abroad for so many years. I love you, mom and dad!

BAYESIAN MODELING OF CENSORED DATA WITH APPLICATION TO META-ANALYSIS OF IMMUNOTHERAPY TRIALS

Xinyue Qi, BS, MS, PhD

The University of Texas

School of Public Health, 2020

Dissertation Supervisor: Shouhao Zhou, PhD

Dissertation Co-supervisor: Christine B. Peterson, PhD

My dissertation builds on a systematic review of 125 clinical trials reporting on treatment-related adverse events (AEs) associated with PD-1/PD-L1 inhibitors published from 2010 to 2018. The motivating dataset contained the following study-level components extracted from each publication: trial name, number of treated patients, selected immunotherapy drug, dosing schedule, cancer type, number of AEs within each category, and the pre-specified criteria for AE reporting. The number of AEs were reported based upon all-grade (Grade 1-5) and Grade 3 or higher (Grade 3-5) severity. My overall objective was to increase our understanding of the toxicity profiles of five most common cancer immunotherapy drugs, and to evaluate AE incidence across subgroups in a meta-analysis setting. However, for assessing drug safety in clinical trials, a common challenge is that many published clinical studies do not report rare AEs. In particular, if the number of AEs observed is lower than a pre-specified cutoff value, these events may not always be reported in the publication (i.e., they are censored).

My doctoral dissertation research, thus, proposes an innovative statistical methodology for effectively handling censored rare AEs in the context of meta-analysis of immunotherapy trials. First, by deriving exact inference and robust estimates for the miss-

ing not at random data, we proposed a Bayesian multilevel regression model in the coarsened data framework to accommodate censored rare event data. We also demonstrated that if the censored information was ignored, the incidence probability of AEs would be overestimated. Second, to select the best Bayesian censored data model among a set of candidate models in the presence of complicated or high-dimensional features, we proposed an alternative strategy to implement Bayesian model selection for censored data analysis in Just Another Gibbs Sampling (JAGS). To generate deviance samples from a Bayesian model using JAGS, if censoring occurs, an existing function incorrectly calculates the value of deviance function because of the "wrong focus", i.e., the incorrect likelihood computed on the basis of model specification in JAGS. Therefore, we proposed a strategy to establish a simultaneous way to calculate the true value of deviance function in JAGS. The alternative strategy could be generalized to model other types of data and be applied to many other disciplines. Third, we developed a sparse Bayesian selection model with prior specifications on meta-analysis of censored rare AEs to perform selection of pairwise interactions between various study-level factors. Because the toxicity profiles of immunotherapy drugs may not be explained comprehensively by main effects of study-level factors, we identified the high-risk group by considering two-way interactions that impact the outcome of interest. Through simulation studies, we demonstrated that the proposed interaction selection method outperforms others in prediction accuracy and interaction identification in the presence of missing outcome data. Lastly, we also applied the proposed method to our real-world motivating dataset.

In sum, my dissertation work makes significant and innovative contributions to the field of applied statistics and cancer research.

Contents

List of Tables

List of Figures

Chapter 1

BACKGROUND

The Nobel Prize in Physiology or Medicine in 2018 was awarded to two cancer immunotherapy researchers who found alternative ways to activate the immune system in the body to attack cancer cells. One of them discovered that blocking programmed cell death protein 1 (PD-1, a protein on the surface of T cells) or programmed cell death ligand 1 (PD-L1, a protein on the cancer cells) enables the immune system to identify tumor cells and fight them back. This brilliant work inspired the development of the five most popular cancer immunotherapy drugs: Keytruda (Pembrolizumab) and Opdivo (Nivolumab), which inhibit PD-1, and Tecentriq (Atezolizumab), Bavencio (Avelumab) and Imfinzi (Durvalumab), which target PD-L1. Immunotherapy is an innovative and effective treatment approach to attack tumor cells for patients in different types of cancer.

However, it still poses a risk of adverse events (AEs) even if less toxic than other cancer treatments. The severity of AE can be ranked into five categories: Grade 1 (mild), Grade 2 (moderate), Grade 3 (severe), Grade 4 (life-threatening) and Grade 5 (death). There is also a "Grade 0", which is defined as absence of an AE. Since one study has limited information on study-level factors associated to AEs, in order to see the big picture of toxicity profiles of immunotherapy drugs, meta-analysis of relevant trials is

needed. In the past eight years, A number of published clinical trials have reported treatment-related AEs, providing an ideal resource for comprehensive analysis of AE incidences. The incidence/prevalence of an adverse event is defined as the proportion of subjects experiencing an adverse event in the study population [1]. The number of AE cases within each study is assumed to follow a binomial distribution.

In the reporting of clinical trials, AEs are typically summarized as a count of the observed number of incidents, which fall into many different categories. However, for categories with a small number of observed counts (typically less than 5% of the study sample size), trial reports may omit exact numbers for these rare events, and report only a cut-off. Therefore, quantitative modeling of sparse binomial event data in a meta-analysis setting has become a critical and challenging question in understanding the toxicity profiles of immunotherapy drugs.

Motivating Example

The methodological work of my dissertation builds on a systematic review of clinical trials reporting PD-1 and PD-L1 inhibitors associated adverse events published from 2010 to 2018. The dataset contains the following study-level components: trial name, number of treated patients, selected immunotherapy drug, dosing schedule, cancer type, number of adverse events within each category, and the pre-specified criteria for AE reporting in the corresponding publication. Some trials may include multiple dose levels, others may treat a couple of cancer types. Here, the number of adverse events are reported based upon all-grade (Grade 1-5), and grade 3 or higher (Grade 3-5) severity. This motivating example illustrates how we encounter the missing outcome data in the meta-analysis of rare adverse events and why it is essential to take informative missingness into account. In meta-analysis to evaluate the rates of treatment-related adverse events (AEs) of PD-1/PD-L1 immune checkpoint inhibitors from 125 eligible trials, approximate 60%

of AEs on average for each study were not reported due to low incidence [2]. Some AEs were censored because they were less frequently observed based on the pre-determined censoring cutoffs by clinical professionals. In practice, those rare events may not always be reported. For example, left censoring occurs when some severe (Grade 3 or higher) AEs are not reported due to low incidence. Furthermore, right censoring occurs when some studies only report Grade 2 or higher AEs instead of all-grade AEs (if we aim to estimate the overall all-grade AEs in the meta-analysis). In the literature, meta-analyses either ignored the AEs with low incidence, or discarded the studies with missing outcome (AE) data, contributing to substantial publication selection bias. Therefore, those studies should not be ignored or treated as missing completely at random (MCAR) in the meta-analysis, otherwise, the overall inference on incidence rates would be biased. This type of missing data problem in the meta-analysis for rare events, however, has barely been addressed in the previous literature.

1.1 Literature Review

1.1.1 Meta-analysis of Rare Events

Meta-analysis synthesizes findings from a set of independent clinical studies and provides a more powerful analysis than from a single study [3]. However, traditional meta-analytical methods provide poor estimates of the true incidences of adverse events when events are rare [4]. When the event is rare, which are commonly encountered in routine practice, special attention should be paid statistically [5]. The existing method to model binary event is either an approximation method based on the normal distribution or an exact method based on the binomial distribution [6]. Estimating the probability of rare events using a normal approximation may also lead to biased results [7, 8]. There are a number of review articles for meta-analysis of rare events that focus on compar-

ing fixed-effects meta-analysis methods including inverse variance fixed effect method, Mantel-Haenszel, and Peto, with frequentist or Bayesian random-effect models [3, 9], but these articles do not handle any censored cases along with rare adverse events in the meta-analysis setting. There is a rich literature on statistical methods for meta-analysis of rare events, which includes Poisson random effects model to estimate relative risk between two treatment groups [10], new methods to estimate the treatment effect and heterogeneity parameter in the random-effect model [5], and a general exact meta-analysis approach to combine inferences from multiples studies in the rare event setting [11]. However, none of these papers mentioned missing data in the meta-analysis of rare event.

1.1.2 Missing Data in Meta-analysis

A common problem arises in the case of rare events, when the data was missing due to especially low incidence, which makes a meta-analysis of incidence rate more methodologically challenging. When we encounter missing outcome data in the meta-analysis setting, the first thing is to determine the mechanisms of missing data. The missingness of an outcome can be classified as missing completely at random (MCAR) if it does not depend on any other variables. Such scenario is generally not realistic in practice. For missing at random (MAR), the missingness does not depend on the actual unobserved outcome even if it is associated with observed data. Both MCAR and MAR are ignorable without leading to biased estimations. However, the missing data could also be nonignorable when it directly relies on the unobserved outcome. The data are then missing not at random (MNAR), also known as informatively missing because of relatedness [12]. For example, some adverse events are not reported due to being less frequently observed in a study on the basis of a pre-specified cutoff value. In the presence of informative missingness, if statistical analysis is only based on the likelihood of the observed data, then the inferences on parameters are biased. Furthermore, the missing data problem

can also be addressed under the framework of data coarsening. When the exact value of the data is not observed, but the unobserved true data lie in a subset of the sample space, such incomplete data is defined as coarse data [13]. A general model of incomplete data was proposed by Heitjan and Rubin to solve coarse data problems in the biomedical field [14]. If the data are coarsened at random (CAR), which is a generalization of MAR, as well as the model parameters and coarsening process are independent, suggesting the ignorability of the incompleteness mechanism, then the likelihood inference is still valid without considering the randomness in the coarsening [15].

There are relatively few articles dealing with the missing outcome data in the meta-analysis of clinical trials. Much of the work focuses on complete/available cases analysis by only including observed cases, imputation approaches (replacing with the mean or the last observed value), best-case/worst-case scenarios, and uncertainty interval for the summary estimates using adjusted weights [16]. These methods summarized above are capable of dealing with data that are MCAR or MAR, but they cannot directly give reliable and unbiased parameter estimations if the missing pattern is MNAR and/or the percentage of missingness is high. An early reference on meta-analysis of studies with varying proportions of missing data is Yuan and Little's paper [17]. Three methods were proposed to correct the bias caused by individual-level missing data in meta-analysis, but not for the context of study-level missing outcome data which is more frequent in the meta-analysis setting.

1.1.3 Bayesian Modeling of Censored Data and Implementation in JAGS

In the study-level meta-analysis of censored and rare events, we aggregate evidence across all relevant studies. For example, we aim to estimate the overall prevalence of adverse events related to the immunotherapy drugs. Such statistical inference has been increas-

ingly based on the evidence from meta-analysis of immuno-oncology clinical trials rather than one trial. However, existing statistical models for meta-analysis cannot account for the both censored and rare event data. A Bayesian hierarchical model is an appropriate approach to deal with censored rare event data in a meta-analysis setting since it naturally has advantages in describing the randomness and heterogeneity between studies as well as solving the missing data problems without multiple imputation techniques.The uncertainty associated with model parameters of interest can be described by prior probability distributions. Thus, to close the gap, we adopt the Bayesian framework to fit informatively censored rare event data within a Bayesian multilevel (hierarchical) logistic regression model framework, and to demonstrate how we can use this model to provide insights into the AE incidences of immunotherapy drugs in meta-analysis.

The major computational challenge of Bayesian inference is that, for all but very simple models, it is necessary to run simulations to obtain samples from the distribution of the parameters. For hierarchical models, it becomes more challenging due to integrations over plenty of unknown parameters. Markov Chain Monto Carlo (MCMC) algorithm, which approximate the posterior distribution of a model parameter by random sampling in a probabilistic space, is the most commonly used computational method to fit Bayesian models. The primary benefit of MCMC over other computational methods is that it allows uncertainty quantification, which is a key advantage of Bayesian modeling.

Just Another Gibbs Sampling (JAGS) is a software to generate posterior samples, which makes Bayesian hierarchical models easily to be implemented using MCMC simulation [18]. In particular, rjags is a R package that allows fitting JAGS model in R. The major advantage for fitting Bayesian models in JAGS using MCMC sampling is that we only need to specify likelihood functions and prior distributions in the model file without writing out the full conditional distributions for model parameters, especially when the closed form expressions are not available. In the presence of complicated features, how

to select the best one among a set of candidate models becomes extremely important. However, JAGS has the following limitations that prevent its much broader usage to perform deviance-based model selection. In the presence of censored data in the response variable, an existing function, known as dinterval distribution is commonly used to model censored data [19, 20]. There is an unsolved issue to calculate the correct likelihood and Bayesian deviance due to the "*wrong parameter focus*", i.e., the model specification for censored data in JAGS gives an incorrect likelihood [21].

As one of extensions to the proposed multilevel regression model, we aim to identify a model which is able to best describe the information in the data by comparing alternative models from a Bayesian perspective. Even if there is an existing function in R available to call the dic module and perform Bayesian model selection by directly extracting the deviance samples from a JAGS model to compute the posterior mean deviance, it fails to estimate the correct deviance in the presence of missing outcome data [21]. Such inconvenience of implementation in JAGS limits the usage of Bayesian methods to model censored data and perform model selection for censored data models. There is an unmet statistical computing demand to establish a simultaneous way to calculate the correct deviance function in JAGS when censored data occurs. In this work, our focus is on the algorithm of modeling censored data and its implementation in JAGS for Bayesian model selection. We will demonstrate in detail how to model censored data within the Bayesian framework using the alternative strategy in JAGS which assists in simultaneously calculating the correct deviance using posterior samples from MCMC simulations for further censored data model comparison.

1.1.4 Bayesian Variable Selection Methods

To better understand the toxicity profiles of immunotherapy drugs, in addition to including the marginal effects in a Bayesian hierarchical model, we identify the potential

two-way interactions between various study-level factors. The major statistical question is how to select the promising subsets of interaction terms by avoiding the overwhelming problem of computational burden posed by the high-dimensional set of interaction terms. By fitting a sparse linear model, we select only a subset of interaction terms which are non-zeros, which is beneficial both in terms of enabling scientific interpretation by identifying the truly important interactions, and in terms of statistical modeling, as the assumption of sparsity reduces the number of parameters to be estimated.

There are a number of approaches in both the Bayesian and frequentist frameworks for achieving sparsity, with the most popular approach being the LASSO [22]. The major existing approaches to Bayesian variable selection in the context of regression modeling include indicator model selection, adaptive shrinkage, model space approach, stochastic search variable selection (SSVS) [23], and horseshoe. The two methods of indicator model selection simply set $\theta_j = I_j\beta_j$, where I_j is an auxiliary indicator variable for covariate j and β_j is jth regression coefficient. The Kuo & Mallick (K&M) method [24] assumes the priors of the indicator and effects are independent, indicating $P(I_j, \beta_j) = P(I_j)P(\beta_j)$, while Gibbs variable selection (GVS; [25]) assumes that prior distributions are conditionally independent of each other, suggesting $P(I_j, \beta_j) = P(I_j)P(\beta_j|I_j)$. Rather than using an auxiliary indicator variable,one may specify a normal prior as well as placing a hyper-prior on its variance to induce sparseness. For example, in the Laplacian shrinkage, a double exponential prior distribution is assigned to each model parameter [26]. In the model space approach, priors are placed on the number of selected covariates and corresponding coefficients using reversible jump MCMC (rjMCMC) technique [27]. The strength of model space approach is that only the selected variables are needed to be summed over leading to a smaller likelihood and a more convenient computation strategy. SSVS is a procedure to identify the promising subsets of predictors which have higher posterior probability using Gibbs sampling [28]. In a full Bayesian framework, a

widely used and parameter tuning-free approach to handle unknown sparsity is known as horseshoe prior, which has the following advantages over other procedures in terms of robustness, adaptivity, and tractability [29, 30]. To mitigate the computational challenge for high-dimensional data, Rovckova and George [31] proposed an alternative approach to stochastic search, known as Expectation-Maximization (EM) variable selection, which relies on the basis of the EM algorithm to quickly identify the promising subsets in the high-dimensional setting.

Each Bayesian variable selection method has certain strengths and weaknesses [23]; thus, none is optimal for all scenarios, but some are better for certain scenarios. The choice of method may depend heavily on the prior formulations. In general, rjMCMC is the fastest method in terms of computational speed, but it may have poorer mixing. The indicator model selection methods such as K&M and GVS are slower per iteration than the other methods. Even though it has a reasonable computational speed, the Laplacian method may perform poorly in mixing and separation. Overall, in terms of computational speed, efficiency of mixing, and performance of separation, the adaptive shrinkage approach using Jeffreys' prior can be considered as a benchmark. To handle large search space of variables, SSVS, as a popular search algorithm for high-dimensional data, is capable of randomly exploring a small portion of the whole model space. It has the ability to handle cases where there are more variables than the total number of observations, and it becomes more attractive when the speed is faster under the random effects model, and when the informative priors are assigned to improve the mixing performance and to achieve good separation.

1.2 Public Health Significance

Cancer, which is the second leading cause of death in the United States and worldwide, has a significant impact on public health. Since 2014, cancer immunotherapy (anti PD-1/PD-L1) drugs have been developed to treat different types of cancer. The adverse events (AEs) of these PD-1/PD-L1 inhibitors are often serious but rare, leading to sparse data [3]. It is critical to understand the toxicity profiles of immunotherapy drugs and evaluate the incidences of treatment-related AEs in the meta-analysis of published clinical trials, which may help guide clinical practice for clinicians in the future. Thus, conducting cancer-related public health research in the meta-analysis setting by aggregating medical information, such as drug safety data from a large number of published studies, is important, but it comes with certain statistical challenges.

A common statistical problem arise in the case of rare adverse events (AEs), which are encountered in routine practice, when the event data are missing (not reported) due to very low incidence, which makes a meta-analysis of AE incidence probability more methodologically challenging. To the best of our knowledge, this statistical problem on meta-analysis of censored rare AEs has not been addressed. Thus, in my dissertation work, we proposed a modified Bayesian multilevel logistic regression model to avoid the biased estimation on both observed and censored data, and derive the valid inference for better understanding the toxicity profiles of immunotherapy drugs. To perform censored model selection and implement in JAGS, we proposed an alternative strategy to model censored data in the Bayesian framework. We selected significant interaction terms among all candidates in the model and successfully identified the high toxicity subgroup among cancer patients treated with immunotherapy.

One significant impact of my dissertation work is that it shows the generalizability of the proposed strategy to model censored data and to conduct Bayesian model selection with implementation in JAGS. For example, the proposed strategy can be used to model

survival data, binary data, count data, and ranking data. In addition, the proposed strategy can be used to fit Bayesian hierarchical model for censored normal outcome [32], semiparametric accelerated failure time (AFT) models [33], illness-death model using Bayesian approach for semicompeting risks data [34], and Bayesian Thurstonian models for ranking data [35]. Furthermore, the proposed strategy can also be applied to many other fields, including survival analysis, behavioral science [36], environmental science [37], food science [38], as well as human health [39]. Finally, the proposed strategy can be extended to model truncated data, in particular, left-truncated right-censored observations that are common in survival analysis.

Another significant impact of my dissertation work in on drug adoption and public health policy-making. From the perspective of health economics, the unbiased estimation on AE incidence by severity can provide a reliable probability of each arm in the decision tree model to calculate the incremental cost-effectiveness ratios of immunotherapy drugs in the cost-effectiveness analysis. This information on AEs, in turn, may assist healthcare policymakers to reasonably budget for immunotherapy treatments for various types of cancer. Thus, my dissertation work makes significant and innovative contributions to the field of applied statistics, as well as cancer-related healthcare system.

1.3 Research Questions and Specific Aims

1.3.1 Aim 1: Bayesian Meta-analysis of Censored Rare Events with Stochastic Coarsening

Meta-analysis is a powerful tool for drug safety assessment by synthesizing findings from independent clinical trials. However, published clinical studies may or may not report all adverse events (AEs) if the observed number of AEs were fewer than a pre-specified

study-dependent cutoff, which can be considered as a special type of data coarsening [13] for rare events. To derive exact and robust inference, we investigate the stochastic nature of informative censoring and the conditions of ignorability for coarsened data mechanism. The proposed approach is illustrated using data from a recent meta-analysis of 125 clinical trials involving PD-1/PD-L1 inhibitors with respect to their toxicity profiles. We demonstrate that if the censored information is ignored, the incidence probability of adverse event is overestimated; this bias could have significant impact on immunotherapy drug adoption and public health policy.

1.3.2 Aim 2: Bayesian Censored Data Analysis in JAGS

Just Another Gibbs Sampling (JAGS) is a convenient tool to draw posterior samples using Markov Chain Monte Carlo for Bayesian modeling. However, the built-in function dinterval() to model censored data may limit its usage to perform likelihood based model comparison. The censored observations are ignored in the deviance monitor of the dic module in JAGS, such that an incorrect deviance of the model would be mistakenly reported at each iteration if censoring occurs. To establish an automatic approach to specify the correct deviance function in JAGS, we propose an alternative modeling strategy to implement Bayesian model selection for analysis of censored outcomes. The proposed approach is applicable to a broad spectrum of data types, including right-censored, left-censored and interval-censored survival data, and many different Bayesian model structures.

1.3.3 Aim 3: Bayesian Interaction Selection for Meta-analysis with Censored Rare Events.

In a meta-analysis of clinical trials reporting on adverse events (AEs) associated with PD-1/PD-L1 inhibitors, the toxicity profiles of immunotherapy drugs may not be explained comprehensively by main effects of study-level factors, such as AE categories, cancer types, and drug therapies. In the context of censored and rare AEs, as an extension of Bayesian modeling of censored rare events on the basis of Aim 1, we identify the potential two-way interactions between various study-level factors to better understand the toxicity profiles of immunotherapy drugs and, thus, guide clinical practice. In this work, we develop a sparse Bayesian approach with prior specifications to select pairwise interactions in a meta-analysis of censored and rare AEs, and compare its performance with that of other approaches. We demonstrate that the proposed approach outperforms other competitors in terms of prediction accuracy and interaction identification in simulation studies, and that it can be effectively applied to a recent meta-analysis on drug safety.

Chapter 2

JOURNAL ARTICLE 1

Bayesian Meta-analysis of Censored Rare Events with Stochastic Coarsening

2.1 Introduction

Over the past decade, a number of PD-1 and PD-L1 immune checkpoint inhibitors for cancer treatment have been approved by the Food and Drug Administration (FDA). These innovative drugs for immunotherapy, which enhance the ability of a patient's own immune system to attack tumor cells, have been shown to be efficacious in treating many types of cancer, and tend to be less toxic than traditional forms of cytotoxic chemotherapy. However, because they stimulate the immune system, immunotherapy drugs can lead to serious and even life-threatening side effects such as autoimmune-like disorders. To gain insight into the frequency of adverse events (AEs) associated with

PD-1 and PD-L1 immune checkpoint inhibitors, [2] conducted a systematic review of cancer therapy clinical trials published from 2010 to 2018. They selected trials in which patients were treated with one of the five single-agent immune checkpoint inhibitors, ultimately including 125 studies with a total of 20,218 patients. The primary outcome was the number of adverse events reported based upon all-grade (Grade 1-5) and grade 3 or higher (Grade 3-5) severity. To identify possible source of heterogeneity between studies, the following study-level information were also extracted: trial name, number of treated patients, selected immunotherapy drug, dosing schedule, cancer type, number of AEs within each category, and the pre-specified criteria for AE reporting.

Meta-analysis synthesizes findings from a set of independent clinical studies and provides a more powerful analysis than from a single study [3]. To analyze the anti-PD-1/PD-L1 AE data, special attention should be paid to rare events [5]. Standard methods to model binary events rely on either an approximation method based on the normal distribution or an exact method based on the binomial distribution [6]. Both approaches provide poor estimates of the true incidences of rare events [4]. For example, estimating the probability of rare events using a normal approximation may lead to significantly biased results [7, 8]. Reviews on statistical methods for meta-analysis of rare events were given in [3, 9], including a Poisson random effects model to estimate relative risk between two treatment groups [10], newer methods to estimate the treatment effect and heterogeneity parameter in the random-effect model [5], and a general exact meta-analysis approach to combine inferences from multiples studies in the rare event setting [11].

Nevertheless, the rare events data may be missing due to low incidence. In the anti-PD-1/PD-L1 AE data, approximately 60% of treatment-related AEs were not reported. Many AEs were missing because their observed incidences were lower than a pre-determined study-specific reporting cutoff (e.g. 3% or 5% of the study sample size). If statistical analysis was only based on the likelihood of the observed data, then the

inferences on incidence rates would be biased [40].

This type of missing data problem in the meta-analysis for rare events has barely been addressed. The limited literature on the missing outcome data in the meta-analysis of clinical trials either considered the missing data at individual participant level [17], or assumed that the data were missing completely at random (MCAR) or missing at random (MAR) [16]. However, as the missing data mechanism depends on the value of unobserved outcomes, the missing outcome cannot be simply ignored in the presence of missing not at random (MNAR) [12]. In literature, some meta-analyses either totally ignored the AEs with low incidence, or completely discarded the studies with missing AE data [41], contributing to substantial publication selection bias.

Furthermore, some AEs were still reported even though their incidences were lower than the reporting cutoff, which suggests a separate pattern for AE reporting. This pattern could be clinically recognized as 'must report' regardless of the clinical outcome. A set of AEs, for example, immune-related or cancer-type specific, could be of special interest for one clinical trial or disease group. Unfortunately, when an AE was reported with incidence larger than the cutoff, the actual pattern that it belongs to would not be identifiable. This creates additional technical difficulty to justify the missing data mechanism that can be ignored when we later explore the missing data problem in a coarsened data framework [13]. Therefore, it is essential to take both non-ignorable missingness and pattern non-identifiability into account in the meta-analysis of rare adverse events.

In this paper, we propose a Bayesian approach to handle rare event data in meta-analysis, with an aim to give feasible and reliable parameter estimations if the missing pattern is non-identifiable and/or the percentage of missingness is high. When data are MNAR, there are so many possible ways that the missing data can interact with the missing data mechanism. The rest of the article is organized as follows. To avoid making additional assumptions on missing pattern, in Section 2.2, we present a simplified case for

univariate outcome followed by adapting coarsened at random (CAR) to anti-PD-1/PD-L1 AE data. In Section 2.3, we explore a more complicated case for *multivariate* outcome in the *bivariate* coarsened data framework, and present a joint coarsening model when the coarsened data mechanism (CDM) is ignorable. In Section 2.4, we implement the proposed approach in Just Another Gibbs Sampler (JAGS; [42]) with a tailored presentation for simultaneous Bayesian model selection. In Section 2.5, we conduct numerical studies under distinct scenarios by comparing our Bayesian model with other conventional methods. In Section 2.6, we present the real data meta-analysis results demonstrating the advantage of our approach along with a sensitivity analysis. Lastly, some concluding remarks and discussion are given in Section 2.7.

2.2 Data Coarsening

Missing values are a form of data coarsening [40]. When the exact value of the data is not observed, but the unobserved true data lie in a subset of the sample space, such incomplete data can be defined as coarsened data, which can take the form of censoring, grouping, heaping or missing data. Heitjan and Rubin [13] generalized the seminal work by Rubin[43] for coarsened data and developed a general theory to identify when coarsening mechanism can be ignored when making likelihood-based inference about the parameters of interest. Jacobsen and Keiding [44] made a rigid justification in general sample space with the measure theory. Here we explore the informative missingness for the anti-PD-1/PD-L1 AE data in this framework.

To start from the simplified case, suppose that our primary focus is Grade 3 or higher(G35) AE, we treat G35 AE count as a binomial outcome and present it under the coarsened data framework in Section 2.2. Considering the inherent monotonicity between all-grade(G15) and G35 AE counts, and the potential correlation between the

coarsening mechanism corresponding to each AE severity group, in Section 2.3, we extend our univariate binomial distribution to more complicated multinomial distribution, and explore the multinomial outcome data {no AE, G12 AE, and G35 AE} will then be explored under the bivariate coarsening mechanism.

2.2.1 Notations

At the study level, we assume that the number of the jth adverse event Y_{ij} follows a binomial distribution given the total number of $n_{ij} = n_i$ patients recruited in the ith study. For simplicity in notation, the subscript ij is omitted in the absence of ambiguity. Y is a random variable representing the underlying truth, which can be observed incompletely, with the degree of incompleteness under the control of c and M. The threshold (cutoff) variable c is fully observed in each study, as specified in the published article of each clinical trial, such that the actual observation y_c may possibly be censored at c, dependent on the coarsening pattern indexed by M.

The variable M represents the nature of the coarsening mechanism [13], and determines the precision of reporting in the sense of which specific rule of mapping $Y \overset{M}{\longmapsto} Y_c$ to use in coarsening Y. M could be either fully observed in some cases, for instance, to label the units into subsets of Y_{obs} (indicated by $M = 0$) and Y_{mis} (indicated by $M = 1$) in a standard missing data example, when M can be modeled explicitly in *pattern mixture model* factorization; or stochastic in nature, when the coarsened observation Y_c do not necessarily imply known values of the pattern indicator $M = m$.

In reporting clinical studies, there are two general coarsening mechanisms: the number of adverse events Y could be either reported ($M = 0$; no coarsening) or censored only if below certain pre-specified cutoff value c ($M = 1$; censored). Denote $\Xi = \{0, 1, \cdots, n\}$, the sample space for a random variable Y, and $\Psi = \{0, 1\}$, the sample space for a random variable M. If the exact number of an adverse event was reported, it could be either

following the no coarsening mechanism $M = 0$ or following the censored mechanism $M = 1$ by being larger than c, creating a stochastic nature of the coarsening mechanism. $\mathbf{X}$ stands for a group of independent predictor variables which, for our application, include cancer type, AE subtype, drug and dosing schedule, and study used to model AE incidence. $\mathbf{Z}$ is a set of study-level covariates used to model coarsening mechanism.

2.2.2 Coarsening of Rare Events

Consider a general form of coarsening on the complete data (Y, n, c, x, M, z) when only variables Y and M are random, and n, c, x, and z are given and known. We suppose that, instead of observing Y and M directly, one only observes $Y_c = Y_c(Y, M, c)$, a coarse version of Y defined on the subspace of Ξ into which Y has fallen. The sample space of Y_c is a restricted subset of 2^{Ξ} with possible probability in mapping Y to Y_c, where 2^{Ξ} denotes the power set of Ξ. Given $Y = y$ and $M = m$, the conditional distribution which follows a deterministic rule, $r(y_c|y, m)$ for $(Y, M) \mapsto Y_c$, is a degenerate distribution:

$$r(y_c \mid y, m) = \begin{cases} 1, & \text{if } \begin{cases} m = 0, y_c = y. \\ y \leq c, m = 1, y_c = \{0, 1, \cdots, c\}. \\ y > c, m = 1, y_c = y. \end{cases} \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

Though not directly observed, the random variables Y and M can be inferred from the observed value $Y_c = y_c$.

The complete coarsened-data likelihood, which correctly accounts for the coarsening

of Y and the stochastic nature of the coarsening mechanism M, can be written as,

$$\begin{aligned}\mathcal{L}_C(\theta \mid y_c) \propto f_1(y_c \mid \theta) &= \int_{\Psi}\int_{\Xi} f_{Y,M}(y, m \mid n, x, z, \theta, \gamma) r(y_c \mid y, m, c) dy dm \\ &= \prod_{I(y_c > c)} \text{Bin}(y_c | n, x, \theta) \prod_{I(y_c \in \{0,\ldots,c\})} \text{Bin}(y_c \mid n, x, \theta) \text{Pr}(m = 0 \mid y = y_c, z, \gamma) \\ &\quad \prod_{I(y_c = \{0,\ldots,c\})} \sum_{k=0}^{c} \text{Bin}(k \mid n, x, \theta) \text{Pr}(m = 1 \mid y = k, z, \gamma),\end{aligned} \tag{2.2}$$

where the subscript C implies both complete and correct, $f_{Y,M}(y, m|n, x, z, \theta, \gamma)$ is the joint distribution of Y and M and the integration is with respect to the underlying counting measure. Here, we assume the parameters describing the measurement process (θ) are functionally independent of those describing the coarsening mechanism (γ), suggesting $Y \perp\!\!\!\perp M \mid \theta, \gamma$, thus, we can present the joint probability density function of Y and M in the *selection model* factorization

$$f_{Y,M}(y, m \mid n, x, z, \theta, \gamma) = f_Y(y|n, x, \theta) \cdot f_{M|Y}(m|y, z, \gamma).$$

The second equation in (2.2) then holds because the subspace $r(y_c|y, m) = 1$ in (2.1) can be written equivalently as

$$\{y > c, y_c = y\} \cup \{y \leq c, m = 0, y_c = y\} \cup \{y \leq c, m = 1, y_c = \{0, 1, \cdots, c\}\}.$$

The general form of the likelihood $\mathcal{L}_C$ in (2.2) is subject to the specific choice of $f_{M|Y}$. Modeling the mechanism can be challenging, and parameters can often be poorly identified [45]. If the coarsening mechanism M is ignorable, the observed random variable Y_c becomes only a function of Y and model specification can then be significantly simplified.

The degenerate conditional distribution $r_{ign}(y_c \mid y, \theta)$ maps $Y \mapsto Y_c$:

$$r_{ign}(y_c \mid y) = \begin{cases} 1, & \text{if } \begin{cases} y_c = y. \\ y \leq c, y_c = \{0, 1, \cdots, c\}. \end{cases} \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

The corresponding coarsened data likelihood is

$$\begin{aligned} \mathcal{L}_{ign}(\theta \mid y_c) \propto f_2(y_c \mid \theta) &= \int_{\Xi} f_Y(y \mid n, x, \theta) r_{ign}(y_c \mid y, c) dy \\ &= \prod_{I(y_c \in \{0,\ldots,n\})} \text{Bin}(y_c \mid n, \theta) \prod_{I(y_c = \{0,\ldots,c\})} \sum_{k=0}^{c} \text{Bin}(k \mid n, \theta). \end{aligned} \tag{2.4}$$

This likelihood $\mathcal{L}_{ign}$ is appealing because it allows inference for θ using only the density of interest f_Y and the observed data y_c. However, note that M is not observed when the number of an adverse event $Y > c$. Hence $\mathcal{L}_{ign}$ could incorrectly specify the likelihood for statistical inference especially when the coarsening process is stochastic [15].

2.2.3 Ignorability

A question of primary interest concerns whether the coarsening process in (2.2) is ignorable. In the coarsening data framework [13], Heitjen and Rubin suggested a sufficient condition for coarsened at random (CAR) based on the conditional distribution of Y_c given Y. If (a) the data are coarsened at random (CAR), and (b) the model parameters θ in data process and γ in the coarsening process are *a priori* independent, then the coarsening mechanism is ignorable without affecting the statistical inference.

When the coarsening process, M, is stochastic, it is difficult to verify the condition of CAR, which, in turn, makes it challenging to identify the coarsened data mechanism for anti-PD-1/PD-L1 AE data. Given a deterministic rule in (2.1), we propose the following

Lemma for stochastic coarsening process M with a deterministic rule r.

Lemma 1. *The data y are coarsened at random (CAR) if the observed data y_c are coarsened by a deterministic rule $r(y_c \mid y, m, \theta, \gamma) = r(y_c \mid y, m)$, and given any z and γ, $f_{M|Y}(m \mid y, z, \gamma)$ takes the same value for all $y \in y_c$.*

The proof of *Lemma 1* is given in *Appendix 2.8*. Compared with *Theorem 1* in [13], *Lemma 1* simplifies the sufficient condition for ignorable missingness when a deterministic mapping rule r is present, making it much more intuitive and easier to verify.

For Bayesian inference, if the data are CAR, and the model parameter θ and coarsening process γ are *a priori* independent, then the complete coarsened-data likelihood $\mathcal{L}_C$ is proportional to the coarse data likelihood $\mathcal{L}_{ign}$ and the posterior distribution of θ based on $\mathcal{L}_{ign}$ equals the correct posterior distribution based on $\mathcal{L}_C$. Therefore, the likelihood do not need to account for the stochastic nature of the coarsening mechanism, and θ and γ are *a posteriori* independent.

2.3 Joint Modeling under Coarsened Data Mechanism

Returning to our motivating data application, from the clinical perspective, grade 3 or higher adverse events (Grade 3-5 AEs) contain partial information in all-grade (Grade1-5) AEs, that is, the number of Grade 3-5 AEs never exceeds the number of Grade 1-5 AEs in each study. Furthermore, the coarsening mechanisms for Grade 3-5 AEs and Grade 1-5 AEs can be correlated. Such inherent monotonicity of G35 and G15 counts, and correlation between data coarsening mechanisms should both be considered in a joint coarsening model setting. Therefore, instead of modeling our Grade 1-5/Grade 3-5 AE data separately, we present a joint model for censored rare events based upon no-

grade (G0), grade 1-2 (G12) and grade 3-5 (G35) under the bivariate data coarsening mechanism. To be specific, the proposed joint model for coarsened data can handle both left-censored G35 AEs and right-censored G0 AEs (when G15 AEs are left-censored).

2.3.1 Extending Univariate to Multivariate

Given all-grade (G15) AE, and grade 3 or higher (G35) AE count of the jth AE in the ith study in the motivating example, the grade 1-2 (G12) AE is obtained by $Y_{ij}^{12} = Y_{ij}^{15} - Y_{ij}^{35}$ and no-grade (G0) AE is $Y_{ij}^{0} = n_i - Y_{ij}^{15}$. Suppose our random outcome vector $\mathbf{Y}_{ij} = \mathbf{Y} = (Y^0, Y^{12}, Y^{35})$ is a length-k vector of jth AE in the ith study, where $k = 3$, e.g. a realization of random vector, $\mathbf{y}_{ij} = \mathbf{y} = (y^0, y^{12}, y^{35})$. Then, we can model these data by assuming a multinomial distribution: $\mathbf{Y} \sim \text{Multinom}(n; \theta^0, \theta^{12}, \theta^{35})$, where $n = n_i = \sum_{k=1}^{3} y^{(k)}$ is the total number of patients in the ith study, and $\theta^{(\cdot)}$ is the probability of incidence at each severity level (G0/G12/G35) with condition $\sum_{k=1}^{3} \theta^{(k)} = 1$.

2.3.2 Bivariate Coarsening Mechanisms

To explore the potential correlation between data coarsening mechanisms, rather than using univariate Bernoulli distribution to model each of coarsening mechanism, M, separately in Section 2.2, we now consider a bivariate Bernoulli random vector (M_1, M_2), in which, $M_1 = \{0, 1\}$ is the data coarsening mechanism for Y^0 and $M_2 = \{0, 1\}$ is the data coarsening mechanism for Y^{35}. Therefore, (M_1, M_2) takes values from a set of pairs {(0,0), (0,1), (1,0), (1,1)} in the Cartesian product space, suggesting that $M_1 \times M_2 = \{(m_1, m_2) \mid m_1 \in M_1, m_2 \in M_2\} = \{0, 1\} \times \{0, 1\} = \{0, 1\}^2$.

In contrast of the coarse version of Y, $Y_c = Y_c(Y, M, c)$ in Section 2.2, now, $\mathbf{Y}_c$ becomes a function of $Y^0, Y^{12}, Y^{35}, M_1, M_2, c^0, c^{35}$, where $c^0 = n - c^{15} - 1$ and c^{35} are cutoff value corresponding to G0 and G35, respectively. Specifically, no-grade AE, Y^0, is right-censored if $Y^0 > c^0$, and Grade 3 or higher AE, Y^{35}, is left-censored if $Y^{35} \leq c^{35}$.

Given a multivariate vector of outcomes and a bivariate vector of coarsening mechanisms, the conditional distribution then follows a more complicated deterministic rule, $r(y_c^0, y_c^{35} \mid y^0, y^{35}, m_1, m_2)$. Such deterministic rule in (2.9) for $(Y^0, Y^{35}) \xmapsto{M_1,M_2} (Y_c^0, Y_c^{35})$ is free of parameters, resulting in a degenerate distribution provided in *Appendix* 2.8.

To identify CAR for multinomial outcome data, we extend our *Lemma 1* in Section 2.2 to *Lemma 2* for bivariate stochastic coarsening process, (M_1, M_2), with a deterministic rule in (2.9). The *Lemma 2* and its proof are given in *Appendix 2.8.* Therefore, for any observed y_c^0 and y_c^{35}, if $f_{\mathbf{M}|\mathbf{Y}}(m_1, m_2 \mid y^0, y^{35}, z, \gamma)$ takes the same value for all $y^0 \in y_c^0, y^{35} \in y_c^{35}$, then the complete coarsened-data likelihood $\mathcal{L}_C$ in (2.10) is proportional to the coarse data likelihood $\mathcal{L}_{ign}$ in (2.6) and the posterior distribution of θ based on $\mathcal{L}_{ign}$ in n(2.6) equals the correct posterior distribution based on $\mathcal{L}_C$ in (2.10). For Bayesian inference, if the data are CAR, and the model parameter θ and coarsening process γ are *a priori* independent, then the likelihood do not need to account for the stochastic nature of the coarsening mechanism, and θ and γ are *a posteriori* independent.

If the coarsening mechanism vector (M_1, M_2) is ignorable, the observed random vector (Y_c^0, Y_c^{35}) would be a function of Y^0, Y^{35} only. The degenerate conditional distribution $r_{ign}(y_c^0, y_c^{35} \mid y^0, y^{35}, \theta)$ maps $(Y^0, Y^{35}) \mapsto (Y_c^0, Y_c^{35})$:

$$r_{ign}(y_c^0, y_c^{35} \mid y^0, y^{35}) = \begin{cases} 1, & \text{if } \begin{cases} y_c^0 = y^0, y_c^{35} = y^{35}. \\ y_c^0 > c^0, y_c^0 = \{c^0+1, \cdots, n - y_c^{35}\}, y_c^{35} = y^{35}. \\ y_c^0 = y^0, y_c^{35} \le c^{35}, y_c^{35} = \{0, 1, \cdots, c^{35}\}. \\ y^0 > c^0, y_c^{35} \le c^{35}, y_c^0 = \{c^0+1, \cdots, n\}, y_c^{35} = \{0, 1, \cdots, c^{35}\}, y_c^0 + y_c^{35} \le n. \end{cases} \\ 0, & \text{otherwise.} \end{cases} \tag{2.5}$$

To summarize, we extended the binomial outcomes in the univariate coarsening framework to the multinomial outcomes in the bivariate coarsening framework by specifying the deterministic rules for constructing both complete likelihood and ignored like-

lihood. The proposed rule in (2.5) by ignoring the coarsening mechanisms helps us to model the multinomial outcomes jointly in the next section.

2.3.3 Coarsening Model of Multinomial Outcomes

Let the probability of AE incidence at each severity category be $\theta_1 = \theta^0, \theta_2 = \theta^{12}$, and $\theta_3 = \theta^{35}$ in the multinomial outcome setting. To fit a multinomial logistic regression model for $k = 3$ outcome category, we need $k - 1 = 2$ logit functions comparing G35 to G0, and comparing G12 to G0, and denote the ratio of the probability of each severity category (G12 or G35) relative to a baseline category (G0) as $r_1 = \theta_2/\theta_1$, $r_2 = \theta_3/\theta_1$. The two logit functions are as follows:

$$\log\left(\frac{\theta_2}{\theta_1}\right) = \log(r_1) = \mu^{G_{12}} + \beta + \alpha^{G_{12}} + \eta + \zeta,$$

and

$$\log\left(\frac{\theta_3}{\theta_1}\right) = \log(r_2) = \mu^{G_{35}} + \beta + \alpha^{G_{35}} + \eta + \zeta,$$

where $\mu^{G_{12}}$ and $\mu^{G_{35}}$ are overall mean of log odds of G12 to G0, and G35 to G0, respectively. Here, two functions share the same parameters consisting of study-level effects (β), drug-dose effects (η) and effects of cancer type (ζ) except the effects of AE subtype (α). Therefore, the incidence probability of each severity category can be rewritten as,

$$\theta_1 = \frac{1}{1 + r_1 + r_2}, \qquad \theta_2 = \frac{r_1}{1 + r_1 + r_2}, \qquad \theta_3 = \frac{r_2}{1 + r_1 + r_2},$$

where the odds ratios of AE outcome at G35 and at G12 being compared to the no-grade AE are,

$$r_1 = \exp(\mu^{G_{12}} + \beta + \alpha^{G_{12}} + \eta + \zeta)$$

$$r_2 = \exp(\mu^{G_{35}} + \beta + \alpha^{G_{35}} + \eta + \zeta)$$

To estimate the incidence probabilities defined above, we can construct the coarse data likelihood of joint multinomial model by ignoring the coarsening mechanisms as

$$\begin{aligned}
\mathcal{L}_{ign}(\boldsymbol{\theta} \mid y_c^0, y_c^{35}) \propto f(y_c^0, y_c^{35} \mid \boldsymbol{\theta}) &= \int_{\Xi} f_{\mathbf{Y}}(y^0, y^{35} \mid n, \mathbf{x}, \boldsymbol{\theta}) r_{ign}(y_c^0, y_c^{35} \mid y^0, y^{35}, c^0, c^{35}) d\mathbf{y} \\
&= \prod_{y_c^0, y_c^{35} \in \{0,\ldots,n\}} \text{Multi}(y^0 = y_c^0, y^{35} = y_c^{35} | n, \mathbf{x}, \boldsymbol{\theta}) \\
&\prod_{y_c^0 = \{c^0+1,\ldots,n-y_c^{35}\}, y_c^{35} \in \{0,\ldots,n\}} \sum_{k=c^0+1}^{n-y^{35}} \text{Multi}(k, y_c^{35} \mid n, \mathbf{x}, \boldsymbol{\theta}) \\
&\prod_{y_c^0 \in \{0,\ldots,n\}, y_c^{35} = \{0,\ldots,c^{35}\}} \sum_{k=0}^{c^{35}} \text{Multi}(y_c^0, k | n, \mathbf{x}, \boldsymbol{\theta}) \\
&\prod_{y_c^0 = \{c^0+1,\ldots,n\}, y_c^{35} = \{0,\ldots,c^{35}\}, y_c^0 + y_c^{35} \leq n\}} \sum_{k_1=c^0+1}^{n-k_2} \sum_{k_2=0}^{c^{35}} \text{Multi}(k_1, k_2 \mid n, \mathbf{x}, \boldsymbol{\theta})
\end{aligned} \tag{2.6}$$

where c^0 and c^{35} are cutoffs corresponding to the study-level criteria to report no-grade and Grade 3-5 AEs. Note that $\text{Multi}(y^0, y^{35}) = \text{Multi}(y^0, y^{12} = n - y^0 - y^{35}, y^{35})$ and $c^0 = n - c^{15} - 1$.

The prior specifications for parameters in the model are briefly summarized below. We place a non-informative normal prior with a large variance on $\mu^{G_{12}}, \mu^{G_{35}}$, respectively.

$$\mu^{G_{12}}, \mu^{G_{35}} \sim N(0, \sigma_\mu^2), \quad \text{where } \sigma_\mu = 100$$

The main effects follow normal distributions with mean 0 and variance $\sigma_{(\cdot)}^2$,

$$\beta \sim \text{N}(0, \sigma_\beta^2), \quad \alpha^{G_{12}}, \alpha^{G_{35}} \sim \text{N}(0, \sigma_\alpha^2), \quad \eta \sim \text{N}(0, \sigma_\eta^2), \quad \zeta \sim \text{N}(0, \sigma_\zeta^2),$$

Following the recommendation in [46], we assign weakly-informative half-Cauchy prior

distributions to the standard deviation parameters,

$$\sigma_\beta, \sigma_\alpha, \sigma_\eta, \sigma_\zeta \sim C^+(0, A),$$

where scale parameter $A = 25$.

2.4 Model Implementation and Selection using JAGS

In order to perform inference on the model specified above, we apply Just Another Gibbs Sampling (JAGS) to generate samples from the posterior distribution. JAGS makes Bayesian hierarchical models easy to implement using MCMC simulation [18] and fit in R. In the presence of censored data in the response variable, an existing function, known as dinterval distribution is commonly used to model censored data [19, 20]. However, such model specification for censored data in JAGS calculates an incorrect likelihood [21], which also hinders us to obtain the true deviances of candidate models directly from JAGS for model assessment.

Therefore, we aim to overcome the difficulty of modeling censored data based on the ideas of data augmentation. In our alternative modeling strategy, we first introduce a censoring status variable, W, which follows a Bernoulli distribution, with $W = 1$ indicating left-censoring, and $W = 0$ indicating right-censoring. When a binomial outcome, Y, is left-censored at a cutoff, c, with study-level sample size (n), covariate (x) and AE incidence (θ), the likelihood can be described by the cumulative distribution $F_Y(c) = F_Y(c|n, x, \theta)$. If we assume that $\Pr(W = 1) = F_Y(c) = \sum_{k=0}^{c} \binom{n}{k} \theta^k (1-\theta)^{n-k}$, then for left-censored data ($w = 1$), the probability mass function of W is $f_W(w; \Pr(W = 1)) = \Pr(W = 1)^w (1 - \Pr(W = 1))^{1-w} = F_Y(c)$.

Proposition 1. *The likelihood generated from this alternative JAGS model using the cumulative binomial distribution for censored data is identical to the exact likelihood.*

The proof of *Proposition 1* can be found in *Appendix 2.8*. Our strategy creates the right focus of model parameters and produces the true likelihood for the censored data. Meanwhile, it is beneficial for us to identify the best model using deviance information criterion (DIC; [47]) for both coarsening mechanism in the sensitivity analysis conducted later in Section 2.6.2 and for our event data in the presence of complicated study-level features. Most importantly, the proposed algorithm in JAGS properly draws posterior samples, as well as simultaneously computes the correct deviance for model selection.

In the multinomial setting in Section 2.3, the distribution of Y^{35} conditional on $Y^0 = y^0$ follows a binomial distribution with sample size, $n^* = n - y^0$, and incidence probability parameter, $\theta^* = \frac{\theta_3}{\theta_2+\theta_3}$, denoted as $Y^{35} \mid Y^0 \sim \text{Bin}(n - y^0, \frac{\theta_3}{\theta_2+\theta_3})$. Meanwhile, the conditional distribution of Y^0 given $Y^{35} = y^{35}$ is denoted as $Y^0 \mid Y^{35} \sim \text{Bin}(n - y^{35}, \frac{\theta_1}{\theta_1+\theta_2})$. The multinomial outcomes, $\mathbf{Y} = (Y^0, Y^{12}, Y^{35})$ can be partly observed and indicated by a vector of censoring status variable $\mathbf{W} = (W_1, W_2, W_3)$. For example, the data $\mathbf{y} = (y^0, \mathsf{NA}, \mathsf{NA})$ shows only grade 0 (G0) AE is observed, while grade 3 or higher (G35) AE is *left*-censored at a cutoff, c^{35}, with sample size $(n - y^0)$ and AE incidence $(\frac{\theta_3}{\theta_2+\theta_3})$. In that case, the likelihood of Y^{35} is described by a cumulative binomial distribution, $\Pr(W_3 = 1) = \Pr(Y^{35} \leq c^{35}) = F_{Y^{35}}(c^{35}) = \sum_{k=0}^{c^{35}} \binom{n-y^0}{k} \left(\frac{\theta_3}{\theta_2+\theta_3}\right)^k \left(1 - \frac{\theta_3}{\theta_2+\theta_3}\right)^{n-y^0-k}$. Similarly, the data $\mathbf{y} = (\mathsf{NA}, \mathsf{NA}, y^{35})$ indicates only G35 AE is observed, while G0 AE is *right*-censored at a cutoff, c^0, with sample size $(n - y^{35})$ and AE incidence $(\frac{\theta_1}{\theta_1+\theta_2})$, suggesting the likelihood of Y^0 is given by $\Pr(W_1 = 0) = \Pr(Y^0 > c^0) = 1 - F_{Y^0}(c^0) = 1 - \sum_{k=0}^{c^0} \binom{n-y^{35}}{k} \left(\frac{\theta_1}{\theta_1+\theta_2}\right)^k \left(1 - \frac{\theta_1}{\theta_1+\theta_2}\right)^{n-y^{35}-k}$.

2.5 Simulation

In this section, we conduct a simulation study to assess the performance of the proposed Bayesian model in estimating the incidence rates and odds ratios (ORs) in the meta-

analysis of rare adverse events (AEs) with censored information, as well as compare it with that of other existing methods. According to a recent meta-analysis of treatment-related AEs of PD-1/PD-L1 inhibitors in clinical trials [2], AEs are either rare or censored if their number is not reported due to very low incidence. When this scenario arises in medical research, the common solution is to use observed (partial) data while ignoring the censored information [41], which results in the overestimation of the incidence probability of AEs because not all of the available data are included in the analysis.

2.5.1 Settings

To assess the performance of the proposed model that incorporates both observed and censored data, we consider four scenarios: (1) no censoring; (2) low percentage (40%) of censoring; (3) high percentage (80%) of censoring; and (4) mixed percentage of censoring, which suggests no censoring for Drug 1, 40% for Drug 2, and 80% for Drug 3. In Scenario 1, the number of AEs for all studies are fully observed. In the other scenarios (Scenarios 2–4), which include censored observations, data with low incidence are informatively censored to mimic real-world cases, where low and zero events are often censored. Therefore, in Scenario 2, we treat the 40% of AE data with low incidence as censored data and the 60% of AE data that have a relatively higher incidence as observed data. Similarly, in Scenario 3, in order to stress test the robustness of estimation in a more extreme case of censoring, 80% of AE data with low incidence are treated as censored and the remaining 20% are treated as observed. Lastly, in Scenario 4, which is more comprehensive, all studies corresponding to Drug 1, the top 60% of studies corresponding to Drug 2, and the top 20% of studies corresponding to Drug 3 are treated as observed data, and the remaining studies for each drug are treated as censored data. Such an unbalanced case of censoring for different drugs can illustrate the real performance of odds ratio (OR) estimation, when similar biased effects in incidence estimation can no longer be canceled

out in OR estimation.

We compare the proposed model, Bayesian method of censored data (BMCD), with four other methods: the pooled estimation method after continuity correction (PEM), the normal approximation method (NAM), the logistic regression method (LRM), and the normal approximation method with robust variance estimator (RVE). In PEM [48], we pool observations by drug and add 0.5 correction to those studies with zero observations to avoid undefined OR of pairwise comparison. The 95% confidence intervals (CIs) for drug effects are calculated by the exact binomial test. We exponentiate the confidence limits of the logarithm of OR to obtain the 95% CIs of OR [49]. In NAM, as a standard method in practice [50], we use a normal likelihood procedure to estimate the incidence rate by taking the inverse logit of the observed logit incidence [6] of each drug weighted by its within-drug variance. In LRM, we estimate the drug effects by an exact method through fitting a generalized linear model with logit link. In addition, we compare the performance of NAM with and without robust variance estimators [51]. Therefore, in RVE, instead of Fisher information, we implement the sandwich estimator of variance into NAM to improve the robustness of the statistical inference on incidence rates and ORs.

The total number of studies for each drug is fixed at 10 to reflect the typical number of studies in a meta-analysis. Our outcome of interest, the number of AEs for each study, is generated from a binomial distribution with number of patients ($n = 100$) and probability of events ($d_1 = 0.025, d_2 = 0.025$, and $d_3 = 0.013$, respectively). The probability of incidence is determined by the range of incidence rate for the main dose of the corresponding drug to mimic the real-world data example in the next section. Based on the selected incidence probabilities, the true OR between Drug 2 and Drug 1 is 1.0, and the true OR between Drug 3 and Drug 1 (or Drug 2) is 0.5. We assess the coverage probability of 95% CIs, point estimations with associated standard errors, mean absolute

deviations, and root mean squared errors of all six parameters of interest in the four scenarios.

2.5.2 Simulation Results

The results are based on 10,000 simulated data sets. For each method, we repeated the same data generation procedure in order to be able to compare results across methods. Figure 2.1 gives boxplots for point estimations with corresponding standard errors of incidence rates and odds ratios (ORs) by scenario and method. Coverage probabilities (CPs) of six parameters of interest by scenario and method are displayed in bar charts in Figure 2.2. In Table 2.1, performance in terms of both mean absolute deviations (MADs) and root mean square errors (RMSEs) of incidence rates and ORs based on the five methods are shown for the four scenarios.

When there is no censoring (Scenario 1), the proposed method (BMCD) has CPs, MADs, and RMSEs on incidence rates and ORs that are almost identical to those of the PEM and LRM. Of the five methods compared, the PEM can be considered the gold standard/benchmark for both interval and point estimations. Our results indicate that the BMCD is not inferior to the PEM. They also indicate that the CP for each drug obtained from the NAM appeared to be unstable on estimating incidence rates of rare events compared with the other methods. The performance of the RVE is even worse compared with that of the NAM because the model was properly specified. The point estimations of incidence rates in both NAM and RVE are overestimated in Scenario 1. This finding is consistent with arguments mentioned in the normal approximation for rare events [7] and biased results of estimation for rare events using normal approximation [8].

When 40% of data are censored (Scenario 2), the proposed method (BMCD) performs better than the others in estimating incidence rates; its performance in Scenario 2

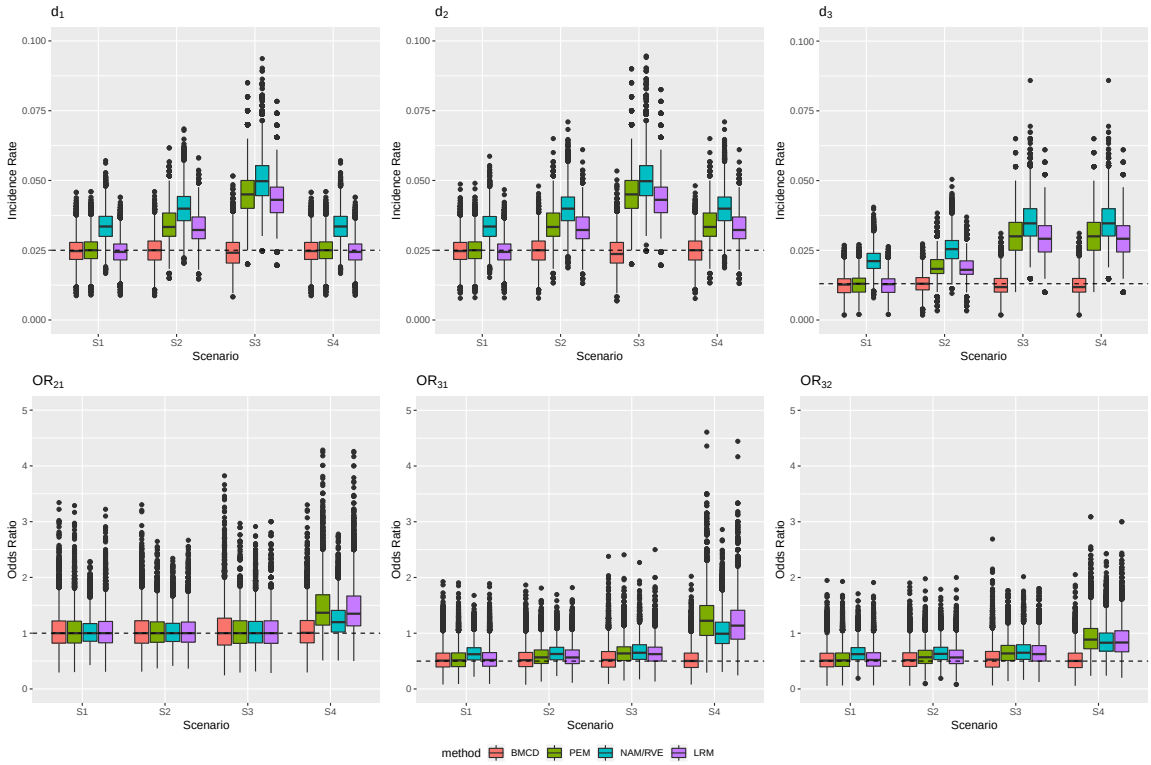

Figure 2.1: Point estimations (PEs) with standard errors (SEs) of drug effects (incidence rates of drugs; d) and odds ratios (ORs) for five methods, Bayesian method of censored data (BMCD), pooled estimation method after continuity correction (PEM), normal approximate method (NAM), logistic regression model (LRM), as well as normal approximate method with robust variance estimation (RVE) under four scenarios: (S1) 0% censoring; (S2) 40% censoring; (S3) 80% censoring; and (S4) mixed censoring.

is as good as it is in Scenario 1. Because censored observations are ignored under other four methods (PEM, NAM, LRM, and RVE), it is unsurprising that the point estimations of incidence rates are overestimated and that the CPs in Scenario 2 are much lower than those in Scenario 1. In contrast, the performance of BMCD in Scenario 2 is almost identical to its performance in Scenario 1 for both interval and point estimations.

In a more extreme scenario where 80% of data are censored (Scenario 3), the proposed method (BMCD) performs well, with little information loss compared with Scenarios 1 and 2. However, all other estimators of drug effects led to inferior CP due to increased percentage of censoring. The point estimations obtained from PEM, LRM,

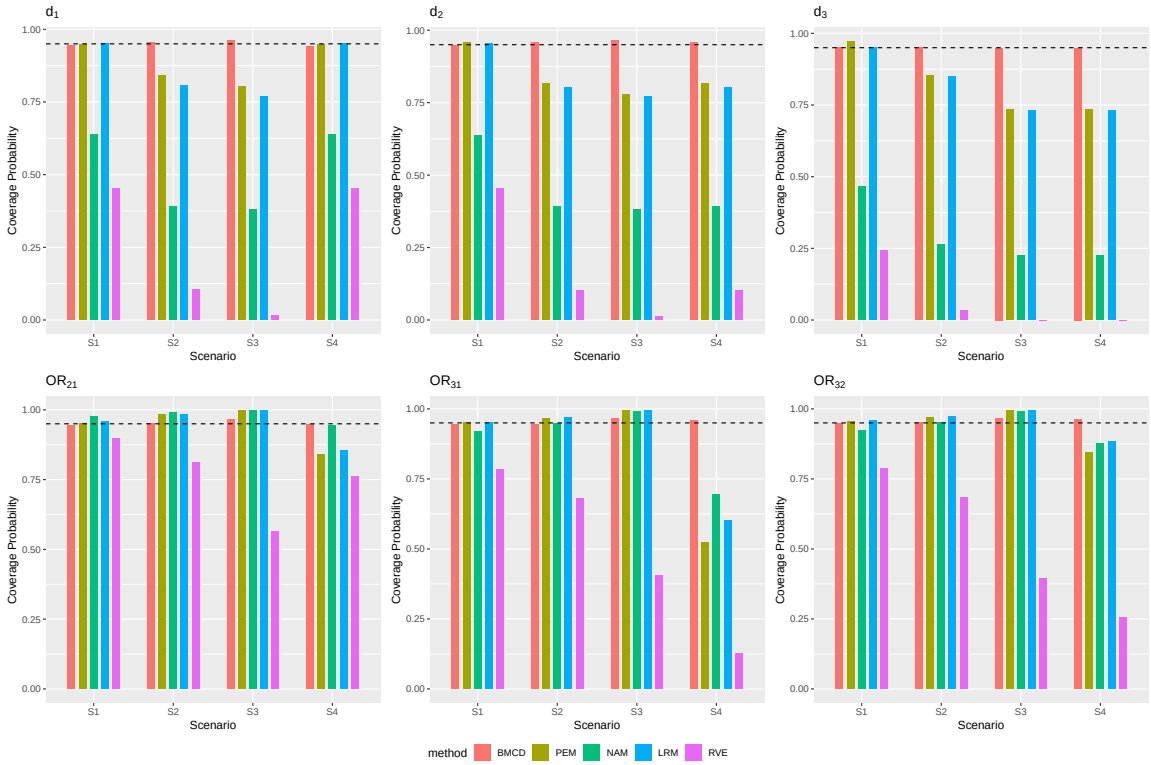

Figure 2.2: Coverage probabilities (CPs) of drug effects (d) and odds ratios (ORs) for five methods, Bayesian method of censored data (BMCD), pooled estimation method after continuity correction (PEM), normal approximate method (NAM), logistic regression model (LRM), as well as normal approximate method with robust variance estimation (RVE) under four scenarios: (S1) 0% censoring; (S2) 40% censoring; (S3) 80% censoring; and (S4) mixed censoring.

NAM, and RVE in Scenario 3 are more biased than those obtained from these methods in Scenario 2. Based on the MAD and RMSE, there were larger deviations from true values of incidence rates compared with those in Scenario 2. Overall, the BMCD yields not only more stable and superior coverage, but also unbiased estimator of incidence rates and ORs in all three scenarios.

Keeping the censoring pattern fixed as in Scenario 2 (40% missing) and Scenario 3 (80% missing) across drugs results in the unbiased estimations on ORs even if the point estimations of incidence are overestimated. Therefore, other than fixed censoring in Scenarios 2 and 3, mixed censoring (0%/40%/80%; Scenario 4) is designed to show that

Scenario	Parameter	True Value	% of Missing	Mean Absolute Deviation				Root-mean-squared Error			
				BMCD	PEM	LRM	NAM/RVE	BMCD	PEM	LRM	NAM/RVE
S1	d_1	0.025	0%	0.004	0.004	0.004	0.009	0.005	0.005	0.005	0.010
	d_2	0.025	0%	0.004	0.004	0.004	0.009	0.005	0.005	0.005	0.010
	d_3	0.013	0%	0.003	0.003	0.003	0.008	0.004	0.004	0.003	0.009
	OR_{21}	1.000		0.245	0.242	0.235	0.190	0.326	0.322	0.313	0.246
	OR_{31}	0.500		0.152	0.151	0.150	0.170	0.201	0.200	0.200	0.221
	OR_{32}	0.500		0.148	0.147	0.146	0.169	0.194	0.194	0.193	0.219
S2	d_1	0.025	40%	**0.004**	0.009	0.008	0.015	**0.005**	0.011	0.010	0.016
	d_2	0.025	40%	**0.004**	0.009	0.008	0.015	**0.005**	0.011	0.010	0.016
	d_3	0.013	40%	**0.003**	0.007	0.006	0.013	**0.004**	0.008	0.007	0.014
	OR_{21}	1.000		0.248	0.220	0.218	**0.196**	0.329	0.289	0.287	**0.253**
	OR_{31}	0.500		**0.155**	0.156	0.156	0.174	**0.207**	0.210	0.210	0.228
	OR_{32}	0.500		**0.151**	0.154	0.154	0.174	**0.200**	0.206	0.206	0.226
S3	d_1	0.025	80%	**0.005**	0.021	0.019	0.026	**0.006**	0.023	0.021	0.028
	d_2	0.025	80%	**0.005**	0.021	0.019	0.026	**0.006**	0.023	0.021	0.028
	d_3	0.013	80%	**0.003**	0.016	0.015	0.021	**0.004**	0.017	0.016	0.022
	OR_{21}	1.000		0.290	0.237	0.239	**0.229**	0.391	0.316	0.319	**0.302**
	OR_{31}	0.500		**0.177**	0.197	0.196	0.206	**0.241**	0.266	0.266	0.273
	OR_{32}	0.500		**0.176**	0.199	0.197	0.208	**0.239**	0.266	0.266	0.274
S4	d_1	0.025	0%	**0.004**	0.004	0.004	0.009	**0.005**	0.005	0.005	0.010
	d_2	0.025	40%	**0.004**	0.009	0.008	0.015	**0.005**	0.011	0.010	0.016
	d_3	0.013	80%	**0.003**	0.016	0.015	0.021	**0.004**	0.017	0.016	0.022
	OR_{21}	1.000		**0.248**	0.465	0.448	0.288	**0.330**	0.601	0.582	0.377
	OR_{31}	0.500		**0.160**	0.772	0.691	0.530	**0.214**	0.880	0.801	0.608
	OR_{32}	0.500		**0.158**	0.425	0.381	0.365	**0.208**	0.511	0.468	0.438

Table 2.1: Mean absolute deviations (MADs) and root mean square errors (RMSEs) of drug effects (d) and odds ratios (ORs) for five methods, Bayesian method of censored data (BMCD), pooled estimation method after continuity correction (PEM), normal approximate method (NAM), logistic regression model (LRM), as well as normal approximate method with robust variance estimation (RVE) under four scenarios: (S1) 0% censoring; (S2) 40% censoring; (S3) 80% censoring; and (S4) mixed censoring.

the other four methods are all off-target in estimating CPs of ORs. When the censoring pattern is mixed, the bias in estimating incidence rates impacts both point and interval estimations of ORs for the other methods in Scenario 4.

The major advantage of the proposed method (BMCD) is that it can avoid bias in estimating the incidence rates and ORs under different percentages of data censoring. Across all scenarios considered above, the BMCD is more powerful and robust than the other four methods in dealing with rare and censored event data. The BMCD also outperforms the other four methods in estimating incidence rates as well as ORs when AEs have low incidence and when a high proportion of AEs are censored. Furthermore, the

quality of an estimator can be measured by its efficiency, which is defined as the asymptotic variance of an estimator [52]. The larger the variance, the lower the efficiency of an estimator. Here, the asymptotic relative efficiency (RF) is given to examine the amount of information loss in comparing two scenarios. Information loss in informative censoring may lead to an inefficient estimator. According to the variance of the point estimator from BMCD, regarding the drug effects, the RFs of two estimators by comparing high percentage of censoring (Scenario 3) to no censoring (Scenario 1) are 0.73, 0.76 and 0.78, respectively. In other words, 80% of censoring only results in 27%, 24%, and 22% loss of efficiency in estimating incidence rates, respectively, compared with no censoring. Meanwhile, the relative efficiency of Scenario 3, with respect to Scenario 1, is approximately 0.70 on average for estimators corresponding to ORs, suggesting that only 30% of information is lost under 80% of informative censoring.

2.6 Application

In this section, we apply the proposed Bayesian method of joint modeling to the real data meta-analysis of all-grade & Grade 3-5 adverse events (AEs) with censored information [2]. The goal is to evaluate the incidence rates of treatment-related AEs of two PD-1 and three PD-L1 inhibitors in a meta-analysis of 125 clinical studies.

The joint model is implemented in the statistical software R and JAGS [18], which uses a Markov Chain Monto Carlo (MCMC) algorithm to generate samples from the posterior distribution of the parameters of interest. Along with listing the data and setting the initial values of model parameters, we defined the likelihood functions and priors of a Bayesian model before compilation in JAGS. We run three parallel chains for the model. For each MCMC chain, after discarding the burn-in period of 30,000 iterations, the 3 chains showed good mixing and successful convergence to the target

distribution. We eventually obtain 10,000 posterior samples per chain by retaining one sample out of three. The 30,000 posterior samples of model parameters such as incidence rates of the 75 all-grade AEs and 20 drug-dose effects are saved for inference. By contrast, if those censored outcomes were treated as MCAR, by ignoring them in the analysis, the estimated all-grade incidence rate would be biased and overestimated by 40%.

2.6.1 Main Results

According to the subgroup analysis of incidence rates of AEs by 75 subtypes of AEs, Figure 2.4 shows the most common all-grade AE is abdominal pain (0.184; 95% credible interval [CrI], 0.168-0.198), followed by hypophysitis (0.107; 95% CrI, 0.096-0.116), pneumonia (0.095; 95% CrI, 0.086-0.104), type I diabetes (0.094; 95% CrI, 0.084-0.102), that had at least posterior median of incidence probability of 9%.

Figure 2.5 illustrates the corresponding proportions of Grade 3 or higher (G35) AEs among patients who experienced all-grade (G15) AEs. Among the 75 AEs under investigation, AST increased (64.46%; 95% CrI, 51.32%-76.47%) and platelet count decreased (61.92%; 95% CrI, 39.04%-81.82%) listed as top 2 highest proportions of grade 3 or higher AEs, suggesting more than 60% of patients developed severe platelet count reduction or rise in AST levels after cancer immunotherapy. Other AEs with higher proportions are erythema (52.73%; 95% CrI, 42.19%-63.31%), hypotension (49.19%; 95% CrI, 39.71%-58.98%), and colitis (46.10%; 95% CrI, 33.62%-59.21%). According to the forest plot, 16 out of 75 (21.3%) severe AEs were seldom ($\leq 2\%$) developed among patients.

Figure 2.3 in *Appendix 2.8* displays a forest plot for the incidence rates of all-grade AEs and their 95% CrIs by drug and dose, there are no significance differences in the incidence of all-grade AEs among different dosing schedules for anti-PD-1/PD-L1 drugs. Based on the posterior inference from the joint model of multinomial outcome (Y^0, Y^{12}, Y^{35}), we find that the overall incidence of all-grade AE is 1.41% (95% CrI, 1.27% -

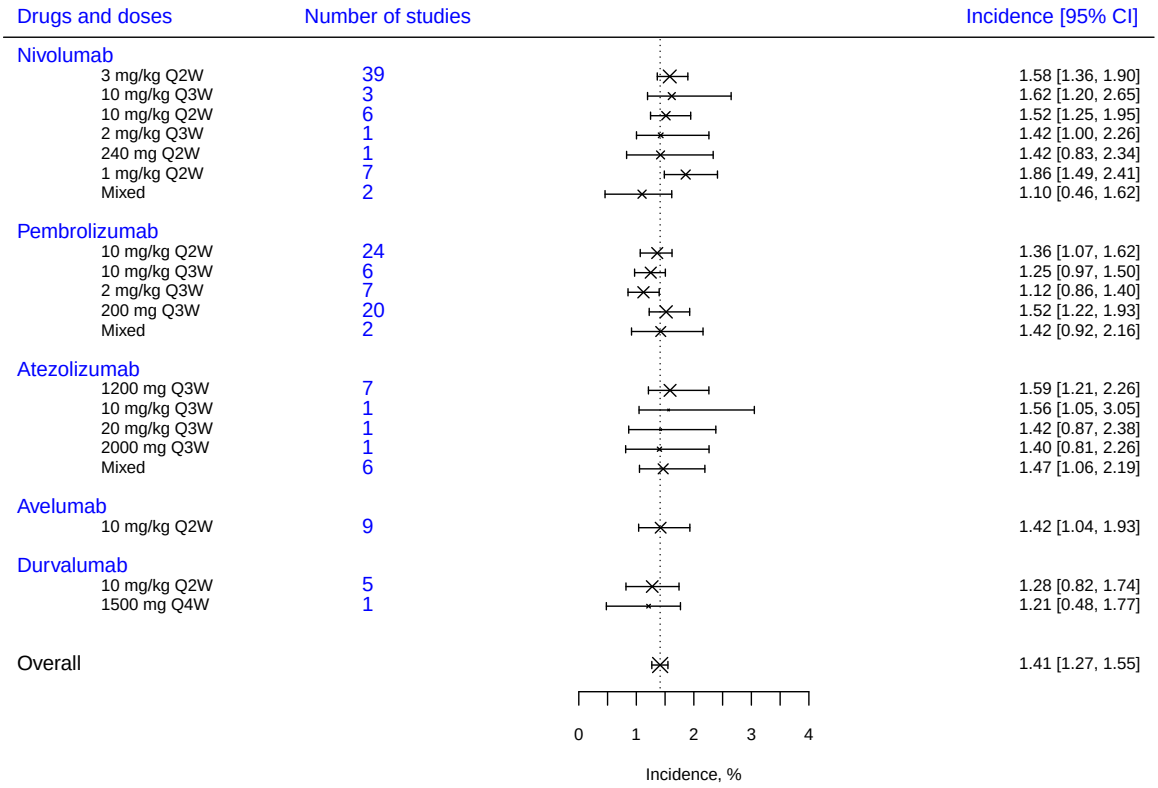

Figure 2.3: Incidence of all-grade AEs by drug and dose

1.55%), the overall incidence of grade 3 or higher AE is 0.11% (95% CrI, 0.09% - 0.13%).

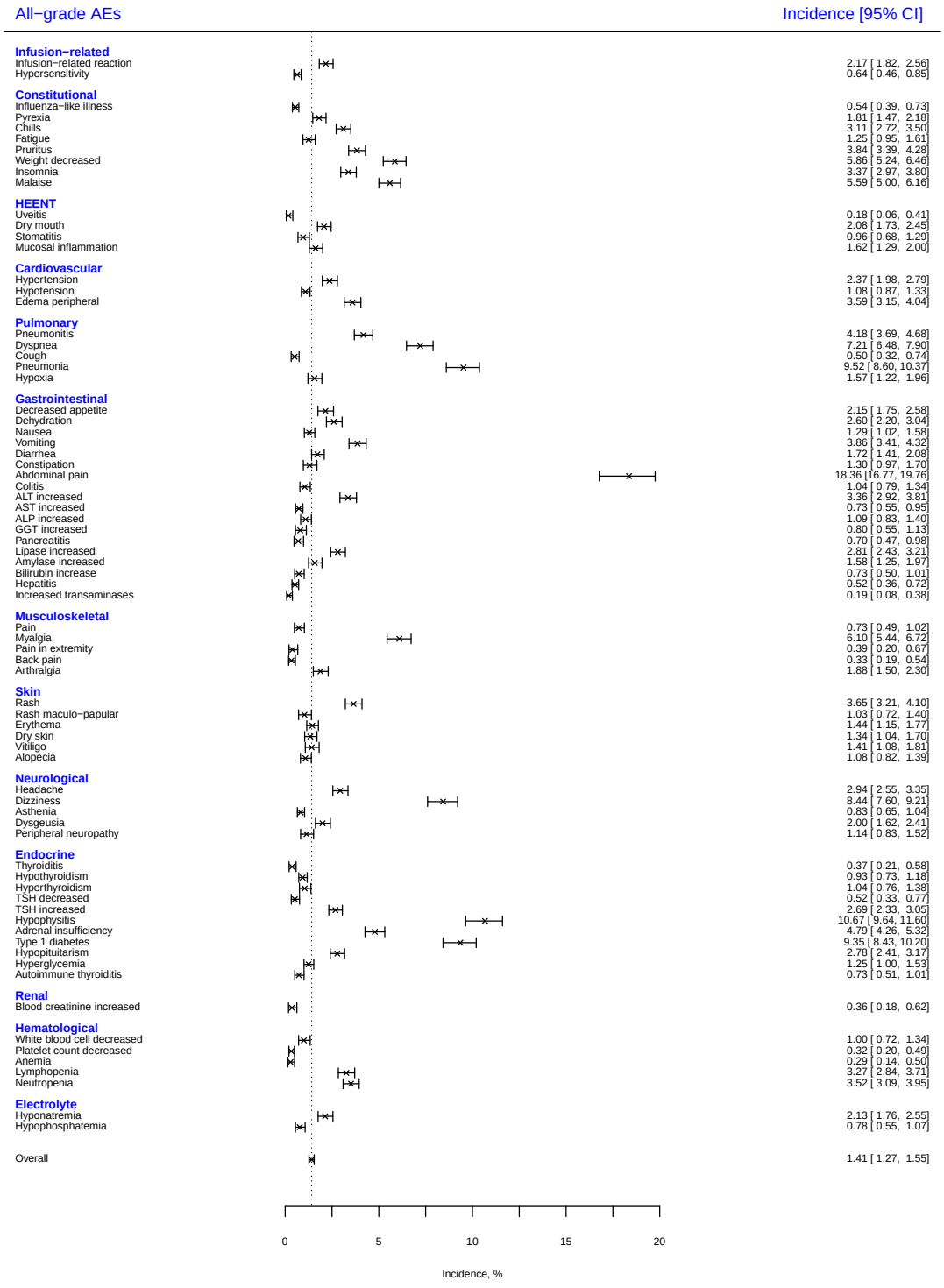

Figure 2.4: Incidence of all-grade AEs by AE subtype

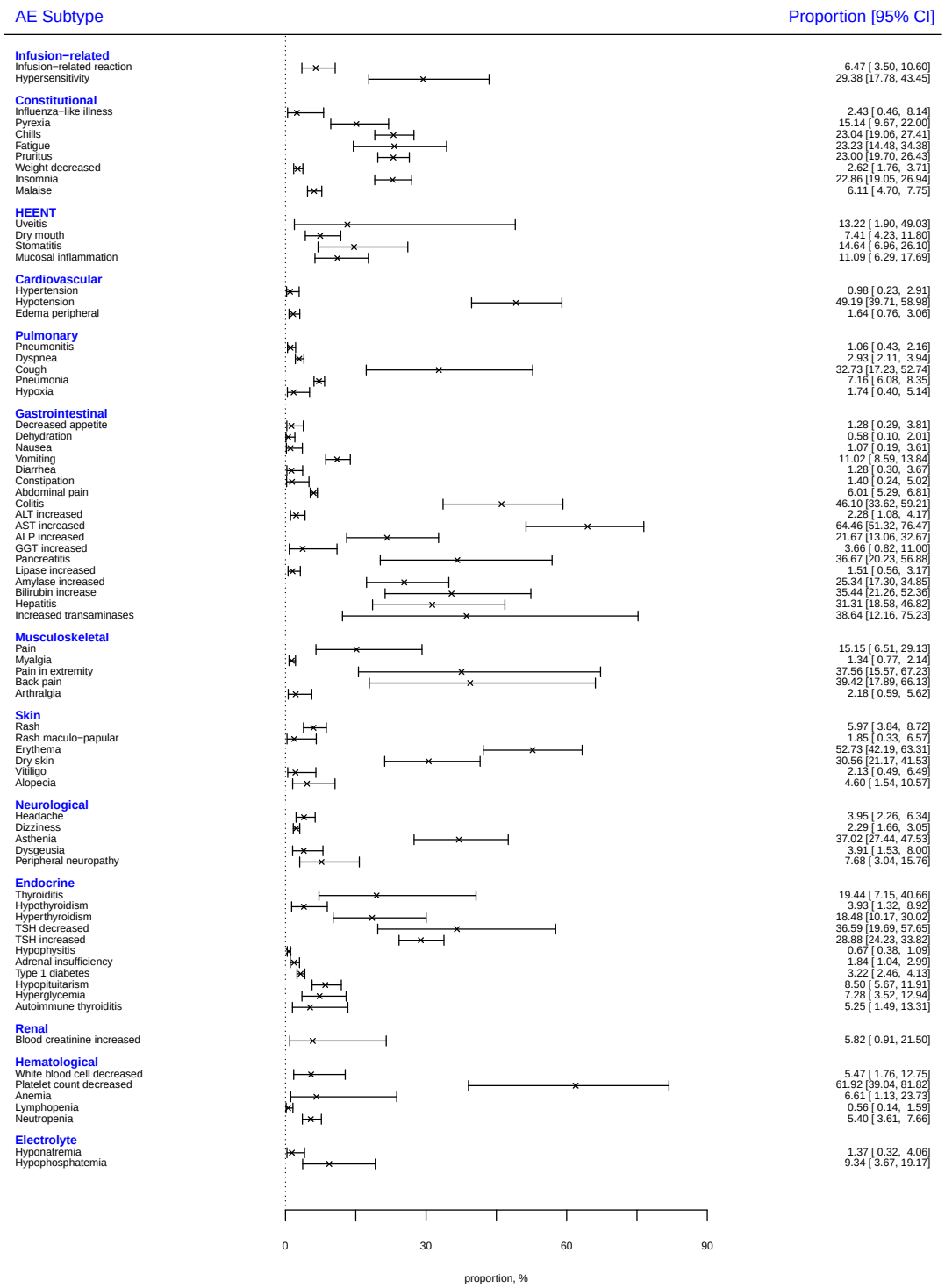

Figure 2.5: Proportion of Grade 3 or higher AEs by AE subtype

2.6.2 Sensitivity Analysis

This sensitivity analysis aims to test the robustness of posterior inference in the presence of uncertainty and explore the effects of input variables on the output. In our case, it helps to demonstrate whether our AE data are coarsened at random (CAR), suggesting the outcome (Y^0, Y^{35}) is conditionally independent of coarsening mechanism (M_1, M_2) given a set of study-level covariates (z).

In our case, based upon the domain knowledge, we assume that the M_1 is conditional independent of Y^{35} given Y^0 and M_2; and, the M_2 is conditional independent of Y^0 given Y^{35} and M_1. Meanwhile, we aim to fit a simplified model through estimating minimum number of parameters. Therefore, rather than modeling marginal distribution of M_1 and M_2 directly, we start with modeling conditional distribution of M_1 on M_2 and that of M_2 on M_1, denoted as q_{1j} and q_{2j}, respectively, on the basis of *eq* (*2.11*) in the *proposition 2.1* proved in [53]. It is straightforward to model the conditional distribution of no coarsening for Y^0 given the coarsening process for Y^{35} and that of no coarsening for Y^{35} given the coarsening process for Y^0 as follows,

$$q_{1j} = q_{1,M_2} = \Pr(M_1 = 0 \mid M_2) = \text{logit}^{-1}\left(\phi_1 + \sum_{i=1}^{4} \gamma_i z_i + \vartheta_1 Y^0 - \delta(1 - M_1)(1 - M_2)\right)$$

$$q_{2j} = q_{2,M_1} = \Pr(M_2 = 0 \mid M_1) = \text{logit}^{-1}\left(\phi_2 + \sum_{i=1}^{4} \gamma_i z_i + \vartheta_2 Y^{35} - \delta(1 - M_1)(1 - M_2)\right)$$

where ϑ_1, ϑ_2 serve as sensitivity parameters to identify CAR, and δ is a correlation coefficient representing the relationship between M_1 and M_2. Note that, in the model fitting, we shift Y^0 and Y^{35} by subtracting the corresponding mean values, without changing the spread of data. Furthermore, the conditional distribution of the coarsening mechanism depends on a set of study-level covariates $\mathbf{z}$. Here, we use an inverse logit to model the conditional probability of no coarsening given another coarsening pattern

and other study-level covariates $\{z_1, \cdots, z_4\}$: z_1 denotes the study-level sample size in the logarithm scale, $\log(n)$; z_2 is an indicator function of immune-related AEs, $\mathbb{I}$(irAEs), which only contributes to model the coarsening mechanism; z_3 and z_4 are cancer type and drug, respectively. Specifically, a study with a larger sample size is less likely to report a number of AEs because of a higher pre-determined reporting cutoff (e.g., due to publication space limit), but immune-related AEs (irAEs) are more likely to be reported even if the incidence is less than the cutoff as irAEs are of particular clinical interest.

The main objectives of the sensitivity analysis are (a) to test whether M_1, the coarsening mechanism for Y^0, is conditionally independent of Y^0 through ϑ_1; (b) to test whether M_2, the coarsening mechanism for Y^{35}, is conditionally independent of Y^{35} through ϑ_2; and (c) to check how M_1 and M_2 are correlated through δ. The coarsening patterns, M_1 and M_2, are not observed if the number of no-grade (G0) AE is $Y^0 \leq c^0$ and that of grade 3 or higher (G35) AE is $Y^{35} > c^{35}$, suggesting the coarsening mechanism is stochastic in nature. According to the complete coarsened-data likelihood, $\mathcal{L}_C(\theta \mid y_c)$ in (2.2), the likelihood function of the joint sensitivity model, which contains 9 components summarizing different characteristics of observed data, is presented in *Appendix* 2.8.

The regression coefficients $\hat{\gamma}_1 < 0$, $\hat{\gamma}_2 > 0$ in the conditional coarsening model confirm the domain knowledge based on the drug safety data. Moreover, the sensitivity parameters $\vartheta_1 = \vartheta_2 = 0$ would imply that the coarsening mechanisms for Y^0 and Y^{35} are ignorable. The 95% credible interval (CrI) on $\hat{\vartheta}_1$ and $\hat{\vartheta}_2$ cover the value of zero, further demonstrating the assumption that the AE data is conditionally independent of the coarsening mechanism given other study-level covariates.

2.7 Conclusions and Discussions

In this paper, we have shown the use of Bayesian hierarchical model in the meta-analysis setting when the study-level events are rare and left-censored. In the presence of data that are MNAR, it is essential to properly address the censoring which can lead to overestimation of model parameters if ignored. For CAR, inference can be based on the observed coarsened data, while the coarsening mechanism can be ignored. Under this condition, likelihood-based analyses on the observed data provide valid results. The proposed Bayesian approach is capable of limiting information loss and providing efficient estimation of drugs effects and odds ratios. Simulation results in Section 2.5 suggest the proposed approach outperforms four other methods when the AEs have low incidence and a high degree of censored information. And, our approach gives the relative efficiencies of high censoring with respect to no censoring for drug effects and odds ratios of 0.76 and 0.70 on average, respectively.

Other than assessing the toxicity profile, the proposed method can be extended to high-dimensional genomic data, in which large number of genes can be tested to estimate the mutation rate across studies. For such an extension, if some numbers of mutation, the primary outcome, are MNAR, they should be considered in the model using pre-specified cutoff value determined by gene selection criteria. Our approach could also be extended to analyze other data types/structures including time-to-event data with right-censoring, count data and ranking data [35], as well as apply to many other fields such as behavior science [36], environmental science [37] and food science [38].

In this work we only focused on the left censoring cases for all-grade AEs when lower than pre-specified cutoff values. However, in estimation of the all-grade AEs in meta-analysis, right censoring may also occur when some studies only report Grade 2 or higher AEs instead of all-grade AEs [54]. Furthermore, the proposed model can also be extended by identifying two-way interaction effects such as drug-by-cancer, drug-by-AE,

and cancer-by-AE interactions through Bayesian model selection for better personalized decision making.

Our work could have profound impact on public health policy. For example, our unbiased estimation on AE incidence by severity can provide a reliable probability of toxicity event in the decision tree model for calculating incremental cost-effectiveness ratio in the cost-effectiveness analysis. We believe our work on AEs may assist healthcare policymakers to reasonably budget for immunotherapy treatments for various types of cancer and the associated treatments for AEs.

2.8 Appendix

Proof of Lemma 1. For any deterministic rule $r(y_c \mid y, m)$, the conditional distribution of Y_c given $Y = y$ is

$$k(y_c \mid y, z, \theta, \gamma) = \int_{\Psi} r(y_c \mid y, m, \theta, \gamma) f_M(m \mid y, z, \gamma) dm = \int_{\Psi} r(y_c \mid y, m) f_M(m \mid y, z, \gamma) dm.$$

Given any $M = m$, z and γ, $r(y_c \mid y, m) f_M(m \mid y, z, \gamma)$ takes the same value for all $y \in y_c$ if $f_M(m \mid y, z, \gamma)$ takes the same value for all $y \in y_c$.

Proof of Proposition 1. To illustrate if the likelihood from JAGS model is identical to its exact likelihood, we start with deriving the formula of likelihood presented in the censored JAGS model, which consists of two major components, observed case and one-sided censored case. The full likelihood in the censored JAGS model can be written as:

$$\begin{aligned} \mathcal{L} &= \prod_{i \in O} f_Y(y_i) \prod_{j \in C} \left\{ [F_Y(c_j)]^{I(W_j=1)} [1 - F_Y(c_j)]^{I(W_j=0)} \right\} && (2.7) \\ &= \prod_{i \in O} f_Y(y_i) \prod_{\substack{j \in C \\ \{W_j=1\}}} F_Y(c_j) && (2.8) \end{aligned}$$

where O is the set of fully-observed event outcomes, C is the set of censored outcomes. W is a binary indicator for left-censoring, following a Bernoulli distribution with cumulative probability of left-censored case with a cutoff c, denoted as $F_Y(c)$. For missing data with left-censoring only, the subset $\{W_j = 0\}$ in C is empty, so that equation 2.8 holds. In general, equation 2.7 can also accommodate missing data with right-censoring only by specifying them with $W = 0$, or both types of left- and right-censoring.

The deterministic rule for $(Y^0, Y^{35}) \xmapsto{M_1, M_2} (Y_c^0, Y_c^{35})$ in Section 2.3 is given by

$$
r(y_c^0, y_c^{35} \mid y^0, y^{35}, m_1, m_2) =
\begin{cases}
1, \quad \text{if} \begin{cases}
y_c^0 = y^0, y_c^{35} = y^{35}. \\
y^0 > c^0, m_1 = 0, y_c^0 = y^0; y^{35} > c^{35}, y_c^{35} = y^{35}. \\
y^0 > c^0, m_1 = 1, y_c^0 = \{c^0 + 1, \cdots, n\}; y^{35} > c^{35}, y_c^{35} = y^{35}. \\
y^0 \leq c^0, y_c^0 = y^0; y^{35} \leq c^{35}, m_2 = 0, y_c^{35} = y^{35}. \\
y^0 \leq c^0, y_c^0 = y^0; y^{35} \leq c^{35}, m_2 = 1, y_c^{35} = \{0, \cdots, c^{35}\}. \\
y^0 > c^0, m_1 = 0, y_c^0 = y^0; y^{35} \leq c^{35}, m_2 = 0, y_c^{35} = y^{35}. \\
y^0 > c^0, m_1 = 0, y_c^0 = y^0; y^{35} \leq c^{35}, m_2 = 1, y_c^{35} = \{0, 1, \cdots, \min(c^{35}, n - y^0)\}. \\
y^0 > c^0, m_1 = 1, y_c^0 = \{c^0 + 1, \cdots, n\}; y^{35} \leq c^{35}, m_2 = 0, y_c^{35} = y^{35}. \\
y^0 > c^0, m_1 = 1, y_c^0 = \{c^0 + 1, \cdots, n\}; y^{35} \leq c^{35}, m_2 = 1, y_c^{35} = \{0, \cdots, c^{35}\}; y_c^0 + y_c^{35} \leq n.
\end{cases} \\
0, \quad \text{otherwise.}
\end{cases}
\tag{2.9}
$$

Lemma 2. *The data y^0, y^{35} are both coarsened at random (CAR) if the observed data y_c^0, y_c^{35} are coarsened by a deterministic rule $r(y_c^0, y_c^{35} \mid y^0, y^{35}, m_1, m_2, \theta, \gamma) = r(y_c^0, y_c^{35} \mid y^0, y^{35}, m_1, m_2)$, and given any z and γ, $f_{\mathbf{M}|\mathbf{Y}}(m_1, m_2 \mid y^0, y^{35}, z, \gamma)$ takes the same value for all $y^0 \in y_c^0, y^{35} \in y_c^{35}$.*

Proof of Lemma 2. For any deterministic rule $r(y_c^0, y_c^{35} \mid y^0, y^{35}, m_1, m_2)$, the conditional distribution of $\mathbf{Y}_c$ given $\mathbf{Y} = (y^0, y^{35})$ is

$$
\begin{aligned}
k(\mathbf{y}_c \mid \mathbf{y}, z, \theta, \gamma) &= \int_\Psi r(y_c^0, y_c^{35} \mid y^0, y^{35}, m_1, m_2, \theta, \gamma) f(m_1, m_2 \mid y^0, y^{35}, z, \gamma) dm_1 dm_2 \\
&= \int_\Psi r(\mathbf{y_c} \mid \mathbf{y}, \mathbf{m}) f_{\mathbf{M}|\mathbf{Y}}(\mathbf{m} \mid \mathbf{y}, z, \gamma) d\mathbf{m}
\end{aligned}
$$

Given any $M_1 = m_1, M_2 = m_2$, z and γ, $r(\mathbf{y_c} \mid \mathbf{y}, \mathbf{m}) f_{\mathbf{M}|\mathbf{Y}}(\mathbf{m} \mid \mathbf{y}, z, \gamma)$ takes the

same value for all $y^0 \in y_c^0, y^{35} \in y_c^{35}$ if $f_{\mathbf{M}|\mathbf{Y}}(\mathbf{m} \mid \mathbf{y}, z, \gamma)$ takes the same value for all $y^0 \in y_c^0, y^{35} \in y_c^{35}$.

The complete coarsened-data likelihood for a joint coarsening model of multinomial outcome in Section 2.3 can be written as

$$
\begin{aligned}
\mathcal{L}_C(\theta \mid y_c) \propto f(y_c \mid \theta) &= \int_{\Psi} \int_{\Xi} f_{Y,M}(y, m \mid n, x, z, \theta, \gamma) r(y_c \mid y, m, c) dy dm \\
&= \prod_{y_c^0 \le c^0, y_c^{35} > c^{35}} \text{Multi}(y^0 = y_c^0, y^{35} = y_c^{35} | n, x, \theta) \\
&\quad \prod_{y_c^0 \in \{c^0+1, \ldots, n-y_c^{35}\}, y_c^{35} > c^{35}} \text{Multi}(y_c^0, y_c^{35} \mid n, x, \theta) \Pr(m_1 = 0 \mid y_c^0, y_c^{35}, z, \gamma) \\
&\quad \prod_{y_c^0 = \{c^0+1, \ldots, n-y_c^{35}\}, y_c^{35} > c^{35}} \sum_{k=c^0+1}^{n-y^{35}} \text{Multi}(k, y_c^{35} \mid n, x, \theta) \Pr(m_1 = 1 \mid k, y_c^{35}, z, \gamma) \\
&\quad \prod_{y_c^0 \le c^0, y_c^{35} \in \{0, \ldots, c^{35}\}} \text{Multi}(y_c^0, y_c^{35} | n, x, \theta) \Pr(m_2 = 0 \mid y_c^0, y_c^{35}, z, \gamma) \\
&\quad \prod_{y_c^0 \le c^0, y_c^{35} = \{0, \ldots, c^{35}\}} \sum_{k=0}^{c^{35}} \text{Multi}(y_c^0, k | n, x, \theta) \Pr(m_2 = 1 \mid y_c^0, k, z, \gamma) \\
&\quad \prod_{y_c^0 \in \{c^0+1, \ldots, n-y_c^{35}\}, y_c^{35} \in \{0, \ldots, c^{35}\}} \text{Multi}(y_c^0, y_c^{35} \mid n, x, \theta) \Pr(m_1 = 0, m_2 = 0 \mid y_c^0, y_c^{35}, z, \gamma) \\
&\quad \prod_{y_c^0 \in \{c^0+1, \ldots, n\}, y_c^{35} = \{0, \ldots, min(c^{35}, n-y_c^0)\}} \sum_{k=0}^{c^{35}} \text{Multi}(y_c^0, k \mid n, x, \theta) \Pr(m_1 = 0, m_2 = 1 \mid y_c^0, k, z, \gamma) \\
&\quad \prod_{y_c^0 = \{c^0+1, \ldots, n-y_c^{35}\}, y_c^{35} \in \{0, \ldots, c^{35}\}} \sum_{k=c^0+1}^{n-y^{35}} \text{Multi}(k, y_c^{35} \mid n, x, \theta) \Pr(m_1 = 1, m_2 = 0 \mid k, y_c^{35}, z, \gamma) \\
&\quad \prod_{\substack{y_c^0 = \{c^0+1, \ldots, n\}, y_c^{35} = \{0, \ldots, c^{35}\} \\ y_c^0 + y_c^{35} \le n}} \sum_{k_1=c^0+1}^{n-k_2} \sum_{k_2=0}^{c^{35}} \text{Multi}(k_1, k_2 \mid n, x, \theta) \Pr(m_1 = 1, m_2 = 1 \mid k_1, k_2, z, \gamma)
\end{aligned}
\tag{2.10}
$$

[Chapter 2: manuscript is in preparation for publication.]

Supplementary Materials: Sensitivity Analysis

Consider a bivariate Bernoulli random vector of coarsening mechanism, (M_1, M_2), we let the joint density of M_1 and M_2 be $p_{ij} = \Pr(M_1 = i, M_2 = j)$, where $i, j = \{0, 1\}$, suggesting $p_{ij} = (p_{00}, p_{01}, p_{10}, p_{11})$ with condition $\sum_{i\in\{0,1\}} \sum_{j\in\{0,1\}} p_{ij} = 1$. Moreover, the marginal distribution of a random variable in a bivariate Bernoulli vector is a univariate Bernoulli distribution in *proposition 2.1* proved by Dai, Ding and Wahba [53]. Therefore, the marginal distribution of M_1 and M_2 in a bivariate Bernoulli vector (M_1, M_2) can be specified as follows,

$$M_1 = 0 \mid Y^0, Y^{35} \sim \text{Bernoulli}(p_{01} + p_{00}), \qquad M_2 = 0 \mid Y^0, Y^{35} \sim \text{Bernoulli}(p_{10} + p_{00})$$

By defining the natural parameters f^1, f^2, f^{12}, we can write out the conditional distribution of M_1 given M_2 and that of M_2 given M_1, denoted as q_{1j} (q_{2j}), where $j = 0, 1$.

$$q_{11} = \Pr(M_1 = 0 \mid M_2 = 1) = \frac{p_{01}}{p_{01} + p_{11}} = \frac{1}{1 + \exp(f^1 + f^{12})}$$

$$q_{10} = \Pr(M_1 = 0 \mid M_2 = 0) = \frac{p_{00}}{p_{00} + p_{10}} = \frac{1}{1 + \exp(f^1)}$$

$$q_{21} = \Pr(M_2 = 0 \mid M_1 = 1) = \frac{p_{10}}{p_{10} + p_{11}} = \frac{1}{1 + \exp(f^2 + f^{12})}$$

$$q_{20} = \Pr(M_2 = 0 \mid M_1 = 0) = \frac{p_{00}}{p_{00} + p_{01}} = \frac{1}{1 + \exp(f^2)}$$

where

$$\exp(f^1) = \frac{1 - q_{10}}{q_{10}}, \quad \exp(f^2) = \frac{1 - q_{20}}{q_{20}},$$

$$\exp(f^{12}) = \frac{1 - q_{11}}{q_{11}} \cdot \frac{q_{10}}{1 - q_{10}} = \frac{1 - q_{21}}{q_{21}} \cdot \frac{q_{20}}{1 - q_{20}},$$

$$\exp(f^1 + f^2 + f^{12}) = \frac{1 - q_{20}}{q_{20}} \cdot \frac{1 - q_{11}}{q_{11}} = \frac{1 - q_{10}}{q_{10}} \cdot \frac{1 - q_{21}}{q_{21}}$$

$$\Rightarrow f^1 = \log(\frac{1-q_{10}}{q_{10}}) = -\text{logit}(q_{10}) = -(\phi_1 + \sum_{i=1}^{4}\gamma_i z_i + \vartheta_1 Y_1)$$

$$\Rightarrow f^2 = \log(\frac{1-q_{20}}{q_{20}}) = -\text{logit}(q_{20}) = -(\phi_2 + \sum_{i=1}^{4}\gamma_i z_i + \vartheta_2 Y_2)$$

$$\Rightarrow f^{12} = \log(\frac{1-q_{11}}{q_{11}}) + \log(\frac{q_{10}}{1-q_{10}}) = -\text{logit}(q_{11}) + \text{logit}(q_{10}) = -\delta$$

$$\Rightarrow f^{12} = \log(\frac{1-q_{21}}{q_{21}}) + \log(\frac{q_{20}}{1-q_{20}}) = -\text{logit}(q_{21}) + \text{logit}(q_{20}) = -\delta$$

Given the relationship between q and p, we can calculate the joint probability density functions as follow,

$$p_{00} = \frac{1}{1 + \frac{1-q_{10}}{q_{10}} + \frac{1-q_{20}}{q_{20}} + \frac{1-q_{20}}{q_{20}}\frac{1-q_{11}}{q_{11}}}, \quad p_{01} = \frac{\frac{1-q_{20}}{q_{20}}}{1 + \frac{1-q_{10}}{q_{10}} + \frac{1-q_{20}}{q_{20}} + \frac{1-q_{20}}{q_{20}}\frac{1-q_{11}}{q_{11}}}$$

$$p_{10} = \frac{\frac{1-q_{10}}{q_{10}}}{1 + \frac{1-q_{10}}{q_{10}} + \frac{1-q_{20}}{q_{20}} + \frac{1-q_{20}}{q_{20}}\frac{1-q_{11}}{q_{11}}}, \quad p_{11} = \frac{\frac{1-q_{20}}{q_{20}}\frac{1-q_{11}}{q_{11}}}{1 + \frac{1-q_{10}}{q_{10}} + \frac{1-q_{20}}{q_{20}} + \frac{1-q_{20}}{q_{20}}\frac{1-q_{11}}{q_{11}}}$$

In the sensitivity analysis, we assume the censored AE outcome data, following a binomial distribution but being truncated below at value 0 and above at cutoff value c^{35} for Y^{35}, or being truncated below at value c^0 and above at study size at n for Y^0 . In the presence of *multinomial* outcomes that are partially observed, such as left-censored Y^{35}, we can evaluate if the coarsening mechanism for Y^{35}, M_2, and G35 AE counts, Y^{35}, are conditionally independent given the coarsening mechanism of G0 AE, by generating random samples of Y^{35} from a *univariate* conditional distribution of truncated binomial distribution, denoted as TBinomial$(n^*, p^*; a, b)$. For example,

1. when G0 AE is observed, but G35 AE is censored,

$$Y^0 \sim \text{Bin}(n, \theta_1), \qquad Y^{35} \mid Y^0 \sim \text{TBinomial}(n - y^0, \frac{\theta_3}{\theta_2 + \theta_3}; 0, c^{35})$$

2. when G0 AE is censored, but G35 AE is observed,

$$Y^0 \mid Y^{35} \sim \text{TBinomial}(n - y^{35}, \frac{\theta_1}{\theta_1 + \theta_2}; c^0 + 1, n), \qquad Y^{35} \sim \text{Bin}(n, \theta_3)$$

3. when both G0 AE and G35 AE are censored,

$$Y^0 \sim \text{TBinomial}(n, \theta_1; c^0 + 1, n), \qquad Y^{35} \mid Y^0 \sim \text{TBinomial}(n - y^0, \frac{\theta_3}{\theta_2 + \theta_3}; 0, c^{35})$$

Chapter 3

JOURNAL ARTICLE 2

Bayesian Analysis of Censored data in JAGS

3.1 Introduction

Censored data are commonly observed in different disciplines such as economics, engineering and life sciences [55, 56, 57]. Given the uncertainty in censored data, the modeling and analysis fit naturally in the Bayesian framework by using expectation–maximization (EM), data-augmentation (DA) and Markov chain Monte Carlo (MCMC) algorithms [58, 59]. For example, in highly fractionated experiments, frequentist likelihood-based estimates may not even exist for simple models consisting of only main effects, while Bayesian approach offers a straightforward implementation strategy [60]. When the outcome cannot be fully observed, censored data can be treated as additional parameters from a fully Bayesian perspective, with a likelihood function specifying joint modeling for both observed and censored data. The Bayesian approach has multiple advantages in the

presence of censored data or inadequate sample size, and for nested/non-nested model comparisons [61]. Compared with multiple imputation, Bayesian modeling is robust in statistical inference even when a large proportion of missing data is present [62].

Just Another Gibbs Sampling (JAGS) is an object-oriented software to generate posterior samples using MCMC simulations [18]. It simplifies the implementation of Bayesian hierarchical models by only requiring the specification of likelihood functions and prior distributions, making it unnecessary to specify the conditional distributions for model parameters, especially when the closed form expressions are not available. JAGS also clarifies certain confusing aspects for missing data in WinBUGS and OpenBUGS [63, 64]. To distinguish the concepts of censoring and truncation, it introduces a degenerate dinterval distribution function for general interval-censored data [18].

For Bayesian inference especially with complicated model features, model selection is a critical component to identify an approximate model best describing the information in the data. Among many popular approaches, the seminal work of deviance information criterion (DIC; [47]) was proposed based on Kullback-Leibler (K-L) divergence [65] and embedded in JAGS as part of the dic module based on the posterior samples obtained from MCMC simulations. However, when the outcome variables are censored, the built-in function dinterval returns a constant value of 1 for the likelihood calculation [19, 20], which is equivalent to ignoring all of the censored observations in the deviance monitor of the dic module. As a result, it fails to calculate DIC for model comparison, which may limit the broader usage of JAGS for Bayesian modeling of censored data [21].

Therefore, we propose an alternative modeling strategy for analysis of censored outcomes in JAGS. It is a universal approach that automatically returns the correct deviances for both observed and censored data, such that DIC and penalized expected deviance [66] can be properly and simultaneously calculated using posterior samples from MCMC simulations; thus Bayesian model selection for censored data modeling can be

conducted using JAGS without analytical customization of the deviance of the model. The proposed approach is applicable to many different Bayesian model structures, such as Bayesian tobit regression model [67], semiparametric accelerated failure time (AFT) models for censored survival data [33], illness-death model using Bayesian approach for semicompeting risks data [34], Bayesian hierarchical model for censored normal outcome [32], and Bayesian Thurstonian models for ranking data [35], among many.

The rest of the paper is organized as follows. The default approach for censored data modeling using built-in function in JAGS is introduced in Section 3.2. The alternative strategy for correct deviance computation is proposed in Section 3.3. In Section 3.4, we use a right-censored survival example to illustrate the discrepancy in deviance functions using both approaches, and applied Bayesian model selection using the correctly specified likelihood in an application to drug safety for cancer immunotherapy. Concluding remarks and discussions are given in Section 3.5.

3.2 Default procedure for censored data modeling in JAGS

Censoring occurs when the value of an observation is only partially observed, which is common in medical research. For analysis of censored observations in JAGS, a default approach is to use the built-in dinterval distribution function for model specification and posterior sampling. The Model 1 below illustrates a general form of model specification for censored data analysis in JAGS. It helps modeling three types of censoring: right-censoring, left-censoring and interval-censoring [20].

```
model{ # Model 1
      for (o in 1:O){ # O is the number of observed cases;
            Y[o] ~ f(theta[o])
```

```
    }
    for (j in 1:J){ # J is the number of censored observations;
        # Left censoring (R=0): lim[j,] = c(cut[j], inf);
        # Right censoring (R=2): lim[j,] = c(-inf, cut[j]);
        # Interval censoring (R=1): lim[j,] = c(cut1[j], cut2[j]);
        R[j] ~ dinterval(Y[j], lim[j,])
        Y[j] ~ f(theta[j])
    }
    # prior for theta's
}
```

where the outcome of interest, Y, which can be either observed or censored (coded as NA in the data table), follows density distribution f with parameter θ. R is a censoring variable following an interval distribution. If $R = 1$, then the outcome is interval-censored; If $R = 0$, the data is left-censored while outcome contains partial information which is less than a lower limit; If $R = 2$, the data is right-censored, which is above a certain cutoff value. lim[,] is a vector of length 2, which contains a pair of cutoff values for each unobserved outcome data, as illustrated in the comment lines above.

However, dinterval() function has a limitation in deviance calculation when we assess model fit based upon deviance-based statistics. When an existing function, dic.samples, in the rjags package [68] is applied to call the dic module and to generate penalized deviance samples within R [69], the following warning message appears.

```
Warning message:
In dic.samples(model = model, n.iter = n.iter, type = "pD") :
Failed to set mean monitor for pD
Support of observed nodes is not fixed
```

By default, the dic module was created to monitor and record the likelihood/deviance

of a JAGS model at each iteration. In the presence of censored outcomes, even if the dinterval() function can generate the proper posterior distribution of the parameters in JAGS, the likelihood function is misspecified with *the wrong focus* of inference on the censored outcome variable [21]. Instead, a constant value of 1 for the likelihood function, or equivalently, a constant value of 0 for the deviance function, is counted for the censored outcomes in the deviance monitor. The posterior mean deviance computed from the dic module using the default procedure dinterval() is, in fact, the posterior mean deviance of observed data only. It suggests that the posterior mean deviance extracted from the dic module in JAGS should not be used in model assessment [19].

It is always possible to manually calculate the deviance by definition for model selection using posterior samples [47], which actually contradicts with the design of JAGS to develop a convenient tool for the Bayesian analysis of complex statistical models using MCMC methods. It also adds additional technical obstacles, especially for non-statistician practitioners. For example, the likelihood or deviance function for each candidate model need be specified individually. In the calculation of DIC, the value of deviance function at the posterior mean or mode has to be evaluated externally. To avoid those difficulties, in the next section we will explore the alternative modeling strategy in JAGS which can not only produce correct inference for posterior distributions but also automatically specify correct deviances in the dic module for censored observations.

3.3 Alternative modeling strategy in JAGS

Rather than handling censored data with the dinterval function in the JAGS Model 1, we develop an alternative modeling strategy to specify the proper deviance. Based on the type of censoring for each observation, we divide the data into 3 subgroups: observed, left- or right-censored, and interval-censored. For incomplete observations, we introduce

ancillary indicator variables Z_1 for left- and right- censored data and Z_2 for interval-censored data. Hence, we can present the alternative JAGS model specification (Model 2) in a general form:

```
model{ # Model 2
      # block 1: fully-observed
      for (o in 1:O){
            Y[o] ~ f(theta[o])
      }
      # block 2: left/right censoring
      for (c in 1:C){
            Z1[c] ~ dbern(p[c])
            p[c] <- F(cut[c], theta)
      }
      # block 3: interval censoring
      for (i in 1:I){
            Z2[i] ~ dbern(p[i])
            p[i] <- F(cut2[i], theta) - F(cut1[i], theta)
      }
      # prior for theta's
}
```

Every subgroup is self-blocked with a separate section of the likelihood in JAGS, where O is the set of observed data, C is the set of left/right-censored observations, and I is the set of interval-censored observations. Z_1 is a binary random variable, where $Z_1 = 1$ if it is left-censored, or $Z_1 = 0$ if right-censored. The probability of success p in Bernoulli distribution of Z_1 is defined by the cumulative distribution F for the censored outcomes, which neatly identifies the probabilities for both left-censored and right-censored data

with properly specified cutoffs. For interval censored observations, we set $Z_2 = 1$ and the probability of success in Bernoulli distribution is the incremental change of the values in F function between the cutoffs, corresponding to the unobserved outcome which lies in a semi-closed interval.

The JAGS Model 2 encompasses a broad range of model structures. The censored regression models, which are also called tobit models, usually have data both in blocks 1 and 2 with normally distributed or t-distributed errors [67, 70]. Some extensions include time-series analysis [71], longitudinal data analysis [72] and spatial analysis [73]. In the context of survival data analysis, some commonly assumed parametric distributions F include exponential, Weibull, generalized gamma, log-normal, and log-logistic [74, 75],since the event times are positively valued with a skewed distribution, making the symmetric normal distribution a poor choice for fitting the data closely. Additionally, it is unnecessary to assume a known censoring time. Because the cutoff can be either pre-specified with a fixed value or modeled as a random variable, the proposed approach naturally accommodates models with unobserved, stochastic censoring thresholds [76].

Even for non-censored data, the proposed modeling strategy can still be useful in some situations for computational advantages. After converting the standard model to a latent-variable formulation, we can adapt logit, probit or complementary log-log models as a type of block 2 data with Z_1 defined as the binary outcome and cut (cutoff) treated as fixed at 0 [77]. It is also possible to extend the proposed approach for ordered probit analysis [78], which accommodates many applications in economics and marketing [79].

Next, we justify that the proposed alternative procedure constructs the correct likelihood function for censored outcomes. In likelihood-based inference, the full likelihood for observed and censored data comprises four key components: observed case, left-censored case, right-censored case and interval-censored case. For observed data, the likelihood is simply a product of individual probability density/mass function of ob-

served outcome. For any type of censored cases, the likelihood can be presented in a form of $F_Y(b) - F_Y(a)$, defining the probability of a censored outcome Y observed in the semi-closed interval, $(a, b]$. Here, $F_Y(y) = \mathrm{P}(Y \leq y)$ denotes the cumulative distribution function of the random outcome variable if it is fully observed. If the outcome variable is left-censored at a cutoff, y_l, then $F_Y(b) = F_Y(y_l)$ and $F_Y(a) = F_Y(-\infty) = 0$. If data is right-censored with a lower bound, y_r, then $F_Y(a) = F_Y(y_r^-)$ and $F_Y(b) = F_Y(+\infty) = 1$. For interval-censored data, the likelihood function is the product of $\Pr(u_i \leq Y \leq v_i) = F_Y(v_i) - F_Y(u_i^-)$, where u_i and v_i are a pair of interval thresholds, which could vary for every observation. Therefore, the exact likelihood function is given by:

$$\mathcal{L}_{exact}(\theta; y) = \prod_{o \in O} f_Y(y_o) \prod_{l \in L} F_Y(y_l) \prod_{r \in R} \left[1 - F_Y\left(y_r^-\right)\right] \prod_{i \in I} \left[F_Y(v_i) - F_Y\left(u_i^-\right)\right], \quad (3.1)$$

where O is the set of observed outcome, L (or R) is the set of left (or right) censored observations, and I is the set of interval-censored data with u_i and v_i denoting the lower and upper bound of the ith interval-censored observation.

In the JAGS Model 2, we can specify the cutoff value $\mathsf{cut} = y_l$ if data are left-censored, $\mathsf{cut} = y_r^-$ if data are right-censored, and $(\mathsf{cut1}, \mathsf{cut2}) = (u_i^-, v_i)$ if data are interval censored. Defining $\mathsf{F} = F_Y$, we have the following property for the likelihood from the proposed JAGS model.

Proposition 2. *The likelihood generated from the JAGS Model 2 using Bernoulli distribution with the cumulative probabilities for censored data is identical to the exact likelihood (3.1).*

Proof. To illustrate that the likelihood from the JAGS Model 2, $\mathcal{L}_{jags}$, is identical to its exact likelihood, $\mathcal{L}_{exact}$, we start with deriving the formula for the likelihood presented in the censored JAGS model, which has three major components: observed case, one-sided

censored case, and interval-censored case. The full likelihood, $\mathcal{L}_{jags}$, can be written as:

$$\begin{aligned}
\mathcal{L}_{jags}(\theta; y) &= \prod_{o \in O} f_Y(y_o) \prod_{c \in C} \left\{ [F(\text{cut}_c)]^{I(Z_{1,c}=1)} [1 - F(\text{cut}_c)]^{I(Z_{1,c}=0)} \right\} \\
&\quad \prod_{i \in I} [F(\text{cut2}_i) - F(\text{cut1}_i)]^{I(Z_{2,i}=1)} \\
&= \prod_{o \in O} f_Y(y_o) \prod_{\substack{c \in C \\ \{Z_{1,c}=1\}}} F(\text{cut}_c) \prod_{\substack{c \in C \\ \{Z_{1,c}=0\}}} [1 - F(\text{cut}_c)] \\
&\quad \prod_{\substack{i \in I \\ \{Z_{2,i}=1\}}} [F(\text{cut2}_i) - F(\text{cut1}_i)] \\
&= \prod_{o \in O} f_Y(y_o) \prod_{l \in L} F_Y(y_l) \prod_{r \in R} \left[1 - F_Y\left(y_r^-\right)\right] \prod_{i \in I} \left[F_Y(v_i) - F_Y\left(u_i^-\right)\right].
\end{aligned} \tag{3.2}$$

□

Proposition 2 demonstrates that the proposed alternative modeling strategy in the JAGS Model 2 has a correctly specified likelihood function for censored data, which warrants the JAGS Model 2 to generate proper posterior samples and deliver valid statistical inference. For K-L based model comparison, especially when there are complicated model features, it is convenient to employ the automatic computation of deviance function and model selection criteria. Because the computation is implemented *via* the built-in dic module in JAGS, we empirically compare the deviance reported from the JAGS Model 2 to the deviance manually calculated using posterior samples in the next Section.

3.4 Illustrative Examples

In this section, two real data applications are examined with the proposed approach. The first example in Section 3.4.1 applies both the default approach and the alternative strategy to model time-to-event outcome with right censoring. The reported deviance of the model is assessed with the true value calculated manually based on the full likeli-

hood function. It demonstrates that the alternative strategy can not only properly draw posterior samples in JAGS, but also automatically deliver the correct deviance for model assessment. The second example in Section 3.4.2 shows that the proposed approach is capable of comparing censored data models by DIC [47] and penalized expected deviance (PED, [66]) simultaneously, using a drug safety subset [2] in which some of the outcome data are missing not at random (MNAR).

3.4.1 survival data

Right censoring is common in the time-to-event data of survival analysis.The first example is from a classical right-censored survival dataset on acute myeloid leukemia [80]. Individual patient-level data were collected along with survival or censoring time to test whether the standard course of chemotherapy should be maintained for additional cycles or not. The Bayesian survival analysis is conducted using MCMC simulation and implemented in JAGS 4.3.0 software [20] and R version 3.4.1. The JAGS codes for both models are attached in *Appendix* 3.6. We run three parallel chains for the model and discard 30,000 iterations of burn-in, followed by 10,000 posterior samples of hazard rates per MCMC chain with thinning in the exponential survival regression model. Once the posterior samples are obtained, the deviance function of the model based on the exact likelihood function is manually calculated, and compared with the calculated deviance using dic.samples() function in the rjags package with additional 10,000 iterations.

Figure 4.1(a) and 4.1(b) compare the kernel density plots of posterior samples for coefficients in the exponential survival regression model between the default approach using dinterval() and the alternative strategy. The proposed approach has almost identical distribution to the default approach using dinterval() in estimation of the coefficient parameters. The output of dic.samples() function for mean deviance estimation is plotted in Figure 4.1(c), where the solid vertical line shows the mean deviance using dinterval()

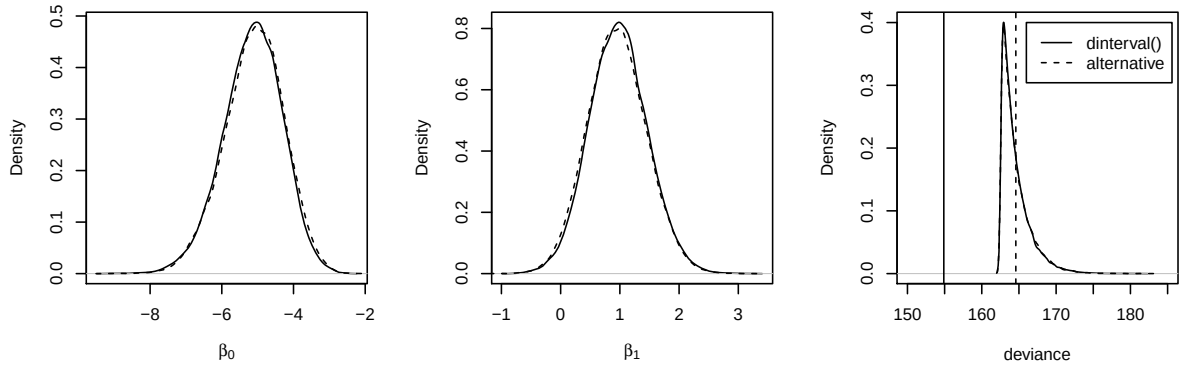

Figure 3.1: (a): A kernel density plot of regression coefficient β_0 in the exponential survival regression model comparing two methods; (b): A kernel density plot of regression coefficient β_1 comparing two methods; (c): A kernel density plot of deviance functions comparing two methods by manual computation of deviance from posterior samples (based upon the exact likelihood). The two vertical lines show the mean deviances generated via the dic.samples() function by the two methods.

function and the dashed vertical line using the proposed alternative strategy. Based on 30,000 posterior samples of each method, we also manually calculate the deviance based on the exact likelihood (3.1) and plot their kernel density curves displayed in the last panel. The result demonstrates that the proposed JAGS Model 2 provides the correct value of mean deviance, while the estimate using dinterval() function is significantly biased due to the deviances ignored for censored outcomes.

3.4.2 binomial data

The second example is from an application to assess drug safety for cancer immunotherapy, known as programmed cell death protein (PD-1) and programmed death-ligand 1 (PD-L1) inhibitors. In clinical practice, it is important to investigate the incidences of treatment-related adverse events (AEs) and to better understand the safety profiles of these immuno-oncology drugs. In this illustrative example, we apply the alternative strategy after extracting all-grade pneumonitis (a specific type of AE for inflammation

of lung tissue) data from a recent meta-analysis [2]. The primary response is a binomial outcome for the number of pneumonitis cases that could be censored; some rare pneumonitis data may be missing due to low incidence. Usually, the less frequently observed AEs are less likely to be disclosed, given the prevalent manuscript word count limitations for clinical trial publications in medical journals. For each censored AE, a study-specific cutoff value can be identified; only the AEs either of special interest or with observed incidence exceeding the cutoff were reported. To take those non-ignorable censored data into account, we considered study-level rare binomial AE outcome data within the data coarsening framework [13] to examine the impact of stochastic censoring mechanism. If the data are coarsened at random, then we can construct the resultant likelihood ignoring the coarsening mechanism and model the outcome data only, as is presented below. The complete likelihood can be represented and modeled using *selection model* factorization including sensitivity analysis [81].

In the Bayesian context, we compare seven distinct censored binomial models for all-grade pneumonitis data to examine the model performance using the proposed strategy. To apply the JAGS Model 2, an outcome variable Z_1 is incorporated for censoring status in block 2. In Model A, a baseline beta-binomial model by complete pooling is to estimate the overall incidence of AE, in which no additional effect is included. In Model B, two-group drug effect is incorporated into the baseline model, and then we can estimate the AE incidences for two drug groups (PD-1 vs. PD-L1 inhibitors). To allow for five drug-specific (Nivolumab vs. Pembrolizumab vs. Atezolizumab vs. Avelumab vs. Durvalumab) effect on the incidence of AE, we begin with modeling drug effects without any link function as Model C, and then extend to specify half-Cauchy prior [46] to the standard deviation of drug effect with logit, cloglog, and probit link functions in Model D-F, respectively. Lastly, we include a saturated model G to estimate the incidence rate corresponding to each study without pooling. Mean deviance ($\bar{D}$), effective number of

parameters (p_D), DIC, optimism (p_{opt}), and PED are all calculated and compared based on the seven candidate models described above. The model assessment results obtained from the proposed JAGS models are summarized in Table 3.1.

Model	$\bar{D}$	p_D	DIC	p_{opt}	PED
A	380.85	0.99	381.84	2.05	382.90
B	371.11	1.99	373.10	4.26	375.37
C	343.14	4.61	347.75	10.65	353.79
D	343.35	4.56	347.91	11.02	354.37
E	343.39	4.54	347.93	13.19	356.58
F	343.38	4.61	347.99	10.28	353.66
G	269.30	94.60	363.90	865.69	1134.99

Table 3.1: Model Comparison: posterior mean deviance ($\bar{D}$), effective number of parameters (p_D), deviance information criterion (DIC), optimism (p_{opt}), and penalized expected deviance (PED) from modeling observed and censored all-grade AE (pneumonitis) data. $DIC = \bar{D} + p_D$, $PED = \bar{D} + p_{opt}$.

Per the results summarized in Table 3.1, there is no significant discrepancy on either DICs or PEDs among Model C-F, indicating that the data are not sensitive to the choice of link functions. In general, models with drug-specific effects (Model C-F) outperform the baseline model with complete pooling (Model A) and the model with PD-1/PD-L1 effect (Model B); the beta-binomial model without pooling (Model G) overfits the data. All results are simultaneously computed from dic.samples() function in the rjags R package.

3.5 Discussion

In this paper we propose an alternative strategy to apply Bayesian modeling for censored data in JAGS. It specifies the correct deviances for censored observations such that the model selection methods DIC and PED can be easily calculated from the built-in dic module. The proposed approach can also simplify the calculation of other popular Bayesian K-L based measures such as the Bayesian predictive information criterion

(BPIC, [82]) and the widely applicable information criterion (WAIC, [83]). Though not explicitly specified, the proposed approach can be easily extended to model truncated data, for example, left-truncated right-censored observations in survival analysis. Even for non-censored data such as binary outcomes, the proposed approach can still be useful for computational advantages.

The proposed method may have a similar model presentation to the EM algorithm [58] to handle censored data, for example, in tobit or probit regression modeling [84, 85]. In Bayesian contexts, the EM-type algorithms are designed to apply parameter optimization in the posterior mode estimation, while the goal is to achieve the automatic calculation of deviance with the posterior distribution estimation. DA is another relevant approach to estimate the posterior distribution, which constructs computationally convenient iterative sampling via the introduction of unobserved data or latent variables [59, 67, 78]. Different from our approach, it requires the sampling of the unobserved data, which may alter the deviance in application of K-L based model selection [47].

Censoring is frequently observed in real-world data analysis. In addition to normally distributed data in censored regression models, various types of outcome, including survival data [61], binomial data [2], count data [86] and ranking data [35], can all be modeled by the proposed alternative strategy when censoring occurs. Not only to the medical sciences, the proposed strategy can also be applied to many other fields, such as, in measuring the performance of timing asynchronies using censored normal sensorimotor synchronization data in behavioral science [36], comparing industrial starch grain properties with ordered categorized data in agriculture [87], exploring forest genetics by modeling censored growth strain data for narrow-sense heritability estimation in environmental science [37], determining the importance of influential factors to lower the risk of food contamination for censored microbiological contamination data in food science [38], and modeling the interval-censored as well as right-censored time to dental health

event in primary school children for public health science [39]. In summary, the proposed JAGS Model 2 can encompass a broad range of popular model structures and be utilized in a wide spectrum of applications.

3.6 Appendix

The following is the JAGS textsf for survival regression model in Section 3.4.1.

```
# The default approach implemented in JAGS
model{
      for (j in 1:J){
            R[j] ~ dinterval(Y[j],lim[j]) # right-censored
            Y[j] ~ dexp(lambda[j])
            lambda[j] <- exp(b0+b1*group[j])
      }
      b0 ~ dnorm(0, tau0) # tau0 fixed at 0.01
      b1 ~ dnorm(0, tau1) # tau1 fixed at 0.01
}

# The proposed approach implemented in JAGS
model{
      for (o in 1:O){
            Y[o] ~ dexp(lambda.adj[o]) # observed
            lambda[o] <- exp(b0 + b1*group[o])
      }
      for (c in 1:C){
            Z[c] ~ dbern(p.adj[c])              # censoring status
            p[c] <- pexp(cut[c],lambda[c+O])    # cumulative exp. dist.
            lambda[c+O] <- exp(b0 + b1*group[c+O])
      }
      b0 ~ dnorm(0, tau0) # tau0 fixed at 0.01
      b1 ~ dnorm(0, tau1) # tau1 fixed at 0.01
}
```

[Chapter 3: manuscript has been submitted for publication.]

Chapter 4

JOURNAL ARTICLE 3

Bayesian Interaction Selection for Meta-analysis with Censored Rare Events

4.1 Introduction

In this article, we are interested to identify the potential two-way interactions between various study-level factors in a meta-analysis of rare and censored adverse events (AEs). As an extension of a recent meta-analysis of clinical trials reporting on AEs associated with PD-1 (programmed cell death protein 1; a protein on the surface of T cells) and PD-L1 (programmed cell death ligand 1; a protein on the cancer cells) inhibitors [2], the primary motivations for incorporating interactions into a Bayesian multilevel regression model include (1) obtaining more accurate toxicity profiles of immunotherapy drugs, (2) helping early detection and close monitoring of symptoms for proper management of

adverse events, and (3) identifying the high-risk subgroups of patients, in the presence of missing outcome data. The major statistical question we addressed in this work is how to select the promising subsets of interaction terms when the search space is large. In general, the majority of features could be irrelevant. In our context, the idea of sparsity comes from the true interactions should be reasonably sparse, suggesting the AE incidence is not impacted by too many interactions. By fitting a sparse linear model, we only select a subset of interaction terms which are non-zeros. It is beneficial both in terms of enabling scientific interpretation by identifying the truly important interactions, and in terms of statistical modeling, as the assumption of sparsity reduces the number of parameters to be estimated.

A number of papers in both the Bayesian and frequentist frameworks proposed approaches on how to achieve sparsity, with the most popular approach being the LASSO [22]. The major existing approaches to the Bayesian variable selection in the context of regression modeling include indicator model selection, stochastic search variable selection (SSVS), Laplacian shrinkage and shrinkage by horseshoe. Two methods of indicator model selection are Kuo & Mallick (K&M; [24]) and Gibbs Variable Selection (GVS; [25]). The K&M method assumes the prior distributions of the auxiliary indicator variables and regression coefficients are independent, whereas GVS assumes that two prior distributions are conditionally independent of each other. As a Bayesian alternative to LASSO, Bayesian LASSO (Laplacian shrinkage) uses a double-exponential prior distribution to each model parameter [26]. SSVS is a procedure to identify the promising subsets of predictors which have higher posterior probability using Gibbs sampling [28]. To mitigate the computational challenge for high-dimensional data, [31] proposed an alternative approach to stochastic search, known as Expectation-Maximization (EM) variable selection, which relies on the basis of the EM algorithm to quickly identify the promising subsets in the high-dimensional setting. To achieve sparsity in the Bayesian variable se-

lection, recently, a horseshoe prior has become a widely used choice for shrinkage [30, 29], which is also free of tuning parameters.

However, interaction selection problem in the presence of informative censoring has been barely addressed. In this article, from the Bayesian perspective, we develop a sparse interaction selection method along with the prior specification on shrinkage parameters for the interaction coefficients to handle censored and rare event in a meta-analysis setting. The major challenge of Bayesian interaction selection is that the possible sets of two-way interactions could be high dimensional. How to specify proper priors and control the shrinkage properties for coefficients of interactions in the sparse Bayesian hierarchical model framework are crucial.

The rest of this article is organized as follows. In Section 4.2, we present the proposed model and prior specifications with aim to select sparse interactions in the Bayesian framework, along with discussing the criteria for the performance of interaction selection. In Section 4.3, we compare the simulation results of the proposed approach with others in the presence of missing outcome data. In Section 4.4, we apply the proposed method to a meta-analysis of immunotherapy trials reporting adverse events for identifying high risk groups by selecting a set of significant interactions that highly impact on AE incidence. Some concluding remarks and discussion are given in Section 4.5.

4.2 Methods

The goal of interaction selection is to recognize the high risk subgroups that have higher probability of AE incidence in a sparse model. In the context of univariate binomial safety data, other than including marginal effects in the model, we are also interested in identifying pairwise interactions by specifying priors to control shrinkage properties for

those that are believed to be highly relevant to the AE incidence.

4.2.1 Bayesian Interaction Selection

The proposed Bayesian multilevel logistic regression model starts with the following additive study-level factors such as study, therapeutic regimen (drug), cancer type and AE category. However, these independent categorical factors may interact with each other. For example, the effect of drug on the incidence of an adverse event may also depend upon AE subgroup. The drug $\times$ AE interaction indicates that AE category influences the relationship between drug and AE incidence. Therefore, we first extend the main effect model by adding two-way interactions to help us to target interesting candidate features. We also consider to select from other candidate interactions, such as cancer $\times$ AE and cancer $\times$ drug.

4.2.2 Model Specification

Let Y_{ij} be the binomial outcome for jth AE in the ith study. Consider the Bayesian multilevel logistic regression model involving interactions effects between two factors:

$$Y_{ij} \sim \text{Binomial}(N_i, p_{ij}), \qquad \text{logit}(p_{ij}) = \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{X}\boldsymbol{\beta},$$

where $i = 1, \cdots, I,$ and $j = 1, \cdots, J.$ p_{ij} is a length-$(I \times J)$ vector of AE incidence, Y_{ij} is a length-$(I \times J)$ vector of observed number of AE within each study, $N_{ij} = N_i$ is the study-level sample size of ith study, and $\boldsymbol{Z}$ is a known $(I \times J) \times q$ design matrix for random effects, and $\boldsymbol{X}$ is a known $(I \times J) \times k$ design matrix for interactions. $\boldsymbol{u}$ is a q-dimensional of random marginal effects such as study, therapeutic regimen, cancer types and AE categories. $\boldsymbol{\beta}$ is a k-dimensional of interaction effect coefficients. Our goal is to find a subset of interactions terms to include in the model. With k potential

two-way interactions, we introduce k binary indicator (latent) variables corresponding to candidate interactions, $\delta_1, \cdots, \delta_k$. If $\delta_k = 1$ $(k = 1, \cdots, K)$, then X_k is included in the model. If $\delta_k = 0$, then X_k is excluded. The posterior probabilities of model parameters can be generated using MCMC and the model including a smaller subset of strongest interactions terms with marginal inclusion probability greater or equal to 0.5 will be selected.

4.2.3 Prior Specifications

We consider all q random marginal/main effects to be included in the model and kth interaction to be included if the corresponding coefficient $\beta_k \neq 0$. To specify the prior for each coefficient of marginal effect, we use a normal prior, which is of the form : $u_q \sim N(\mu, \sigma_u^2)$, where the most common choice for μ is zero and the standard deviation parameter, $\sigma_u > 0$. Moreover, for the choice of prior on σ_u, following the recommendation in [46], we use a weakly-informative prior: $\sigma_u \sim C^+(0, A)$, where C^+ denotes half-Cauchy distribution and A is a scale parameter, which is chosen to be 25.

To perform sparse Bayesian interaction selection, we place priors on the individual interactions, as well as independent priors on the hyperparameters. Specifically, we use spike and slab priors to discriminate between the truly important interaction effects and the negligible ones. The stochastic search variable selection (SSVS) technique was developed to identify regression coefficients via Gibbs sampling [28]. In this project, we use SSVS to recognize promising subsets of interaction terms rather than marginal effects in the model. We define a normal mixture prior for each of interaction coefficients: $\beta_k|\delta_k \sim (1 - \delta_k) \times N(0, \sigma_k^2) + \delta_k \times N(0, c_k^2\sigma_k^2)$. Given the relationship between precision and variance: $\tau_k = 1/\sigma_k^2$, the choice of c_k, τ_k is started with a fixed value by setting c_k large enough (e.g. 100) and precision τ_k at 4. Alternatively, we can also assign a uniform prior on the standard deviation of the spike component, e.g., $\sigma_k \sim \text{Unif}(0, 1)$. The prior

probability of inclusion (suggesting interaction k has a nonzero effect), $\Pr(\delta_k = 1) = 1 - \Pr(\delta_k = 0) = p_k$, is fixed at 0.5.

If β is believed to be sparse, we may also place a horseshoe prior [29] on each of interaction coefficient: $\beta_k|\lambda_k, \tau \sim N(0, \lambda_k^2\tau^2)$, in which λ_k is a local shrinkage parameter and τ is an overall (global) shrinkage parameter. Both hyperparameters follow standard half-cauchy distributions [30]: $\lambda_k, \tau \sim C^+(0, 1)$. The advantage of adopting horseshoe prior is no tuning of any hyper-parameter is required in the model specification.

4.2.4 Model Implementation in JAGS

For Bayesian inference in logistic regression, after determining the full likelihood function of the data and forming a prior distribution over regression parameters, we can find the posterior distribution over all parameters via Bayes theorem. Since the closed form expression of full conditional is not available, we will implement in Just another Gibbs Sampling (JAGS) using Markov Chain Monto Carlo (MCMC) algorithm to approximate/generate samples from the marginal posterior distribution for each regression coefficient.

To overcome the difficulty of modeling censored outcome data, we use an alternative modeling strategy to define a binary censoring status variable, W ($W = 1$ if left-censored, 0 if right-censored), in JAGS. The details can be found in the second aim. In the presence of left-censored AE data with a pre-specified study-level cutoff value, c, instead of modeling binomial outcome with density $f_Y(y|n, p)$ directly, we may construct the censored data likelihood by cumulative distribution, $F_Y(c|n, p) = \sum_{k=0}^{c} \binom{n}{k} p^k (1-p)^{n-k}$, as the probability of left censoring, to help estimate AE incidence. The full model specification of censored data with horseshoe prior for interaction selection in JAGS is in the *Appendix* 4.6. In the next two sections, we will program the simulation study as well as case study in JAGS.

4.3 Simulation Study

4.3.1 Settings

In this section, to show the importance of including two-way interaction terms in the model and to find an optimal interaction selection method, we conduct a simulation study to evaluate the performance of our two proposed sparse Bayesian models with interaction (sBMI), an sBMI under horseshoe prior (sBMIhs) and an sBMI under spike-and-slab prior (sBMIss), with that of other competitors. These include a Bayesian model of censored data (BMCD), which is a marginal model without interactions [2]; glmIA, a classical generalized linear regression model with logit link including true interactions as a reference model; and glm, a logistic regression model without interaction. We generate 100 datasets, each of which contains the training and test set of $n = 1,000$.

To assess the performance of the proposed method, we consider three scenarios. The first scenario (Scenario 1) is when AEs are fully observed. The second and third scenarios (Scenario 2 & 3) are when some of AEs are missing (censored). In scenarios with censored observations, the AE outcome data with low incidence are informatively censored to mimic real-world cases, where low and zero events are often censored. Therefore, the second scenario (Scenario 2) treats 40% and the third scenario (Scenario 3) treats 80% of AE outcome data with low incidences as censored data and keep the remaining observations with higher incidences as observed data.

In the training dataset, we use the same procedure to generate the true value of AE outcome (Y_{train}) for all models. Let the total number of main effects, n_{ME}, be 125, including 100 studies, 10 cancer types, 5 drugs and 10 AE subtypes; let the total number of candidate pairwise interactions, n_{IA}, be 200, including 100 $\mathsf{AE} \times \mathsf{drug}$ interactions, 100 $\mathsf{AE} \times \mathsf{cancer}$ interactions, and 50 $\mathsf{cancer} \times \mathsf{drug}$ interactions; and set the total number of observations, n_{obs}, at 1,000. Then, we generate a design matrix (Z) for marginal

effects and specify the marginal effects (a vector of length $n_{ME} = 125$) by a normal distribution, $u \sim \mathrm{N}(-1, 0.4^2)$. Meanwhile, we generate another design matrix (X) for interaction effects by assuming 20 (out of 200) as selected (nonzero) interactions with value of 1 ($\beta = 1$) and the rest of them as zeros, in the case of sparse interaction signals (assuming the level of sparsity at 10%). Thereafter, the true AE outcome (Y_{train}) of each observation is generated by a binomial distribution with number of patients within study ($n = 100$) and toxicity probability at $p = \mathrm{logit}^{-1}(\boldsymbol{Zu} + \boldsymbol{X\beta})$.

All three Bayesian censored models are implemented in JAGS. In BMCD (model with main-effect only), we fit 200 main study-level factors into the marginal model in JAGS and eventually obtain the estimated marginal effects ($\hat{u}$) using MCMC techniques. In sBMIhs, we fit main study-level factors as well as candidate two-way interactions into the proposed interaction selection model and eventually obtain the posterior samples of marginal effects ($\hat{u}$) and interaction terms ($\hat{\beta}$). In sBMIss, we fit the same training data sBMIhs, but different prior specification for interaction coefficients. For model fitting in JAGS, we discard the 30,000 iterations of burn-in and obtain 30,000 posterior samples for each parameter of interest as well as the corresponding posterior median and 95% credible intervals. In contrast, glmIA aims to be treated as a gold standard by including all nonzero interactions in a generalized linear model with logit link. Moreover, glm is a multiple logistic regression model without interactions and it treats the marginal effects as fixed categorical covariates.

We use the same generation procedure as training set for test dataset with all 125 marginal effects and 200 candidate pairwise interaction effects. The true AE outcome in the test dataset is generated by $Y_{test} = \tilde{Y} \sim \mathrm{Bin}(n, p)$, where the true probability of incidence, $p = \mathrm{logit}^{-1}(\boldsymbol{Zu} + \boldsymbol{X\beta})$. Thereafter, in BMCD (main-effect only), the predicted AE outcome, $\hat{Y}$, is generated from a binomial distribution with sample size n and the estimated incidence probability, $\hat{p} = \mathrm{logit}^{-1}(\boldsymbol{Z\hat{u}})$, where $\hat{u}$ are the medians of

posterior samples for marginal effects. In contrast, in sBMIhs and sBMIss, the predicted AE outcome, $\hat{Y}$, is generated from a binomial distribution with sample size n and the estimated probability of incidence, $\hat{p} = \text{logit}^{-1}(\boldsymbol{Z}\hat{\boldsymbol{u}} + \boldsymbol{X}\hat{\boldsymbol{\beta}})$, in which, the estimated marginal effects and interactions are extracted from the posterior medians of samples using MCMC. After generating both Y_{test} and $\hat{Y}$ for each model, we calculate the average value of MSPE across all replications using formula below.

All five models are compared using the mean squared prediction error (MSPE), which is the mean squared difference between the true value of the outcome ($\tilde{Y}$) from the test dataset and the fitted value ($\hat{Y}$) from the model. To evaluate the performance of the proposed model with selected interactions, we calculate the mean squared error (MSE) of regression coefficients, sensitivity (true positive rate; TPR) and specificity (true negative rate; TNR). In the context of interaction selection, true positive rate (TPR; the proportion of actual positives correctly identified) is defined as the probability that actual nonzero interactions are correctly selected, and false positive rate (FPR = 1 - TNR; the proportion of actual negatives that are identified as positives) is defined as the probability that actual zero interactions are finally selected. We can define the sensitivity (TPR), specificity (1-FPR), mean MSPE, and MSE as follows:

$$\text{MSPE} = \frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \hat{Y}_i)^2, \qquad \text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(u_i - \hat{u}_i)^2$$

$$\text{TPR} = \frac{TP}{TP + FN} = \frac{\text{number of actual non-zeros are selected}}{\text{total number of non-zeros}}$$

$$\text{FPR} = \frac{FP}{TN + FP} = \frac{\text{number of actual zeros are selected}}{\text{total number of zeros}}$$

Furthermore, we include Matthews correlation coefficient (MCC), a measure of the overall interaction selection accuracy, to evaluate the identification performance of each selection

method.

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

where TP represents the number of actual non-zero interactions that are selected, TN is the number of actual zeros that are not selected, FP is the number of actual zeros that are selected, and FN is the number of non-zeros that are not selected.

In the Bayesian selection model using spike-and-slab prior, the posterior probability of inclusion (PPI) for kth interaction is defined as

$$\text{PPI}_k = \frac{\sum_{i=1}^{I} \hat{\delta}_k^{(i)}}{I}$$

where I = total number of MCMC iterations after burn-in.

4.3.2 Simulation Results

The purpose of simulation study is to show that the proposed Bayesian selection model gives the unbiased estimation on probability of incidence across subgroup to identify high-risk groups, as well as identifies the candidate interactions in the presence of missing data.

In the case of the sparse models, uncertainty qualification by marginal credible intervals has been demonstrated to be an effective tool for detecting true and false discoveries of signals [88]. Therefore, we visualize the 95% credible interval (CrI) for each interaction based on two prior specifications under different percentages of missingness by plots. Figures 4.1 and 4.2 show the posterior medians of 200 interaction coefficients with 95% CrIs at sparsity level of 10% (number of signals = 200 × 10% = 20) under 0% and 80% of missingness, respectively. Each figure is based on a single simulated dataset. The true values of 20 nonzero interaction effect (plotted in green) equal to 1 and the remaining 180 interactions at 0 (plotted in red). For actual nonzero interactions, if 95%

CrI of each interaction coefficient (β) covers zero, it is not successfully selected. However, for those actual zero interactions, if 95% CrI of each β does not cover zero, it is mistakenly selected. As shown, the proposed selection models detect the true interactions well (i.e., sensitivity at 0.75 and specificity at 1) in the presence of no missingness (Scenario 1). In contrast, when the missingness gets higher, the 95% CrIs of interaction coefficients become wider, but the proposed selection models can still recover the underlying truth at TPR = 0.65 and TPR = 0.60, with FPR = 0 in Scenarios 2 and 3 based on a single simulated dataset.

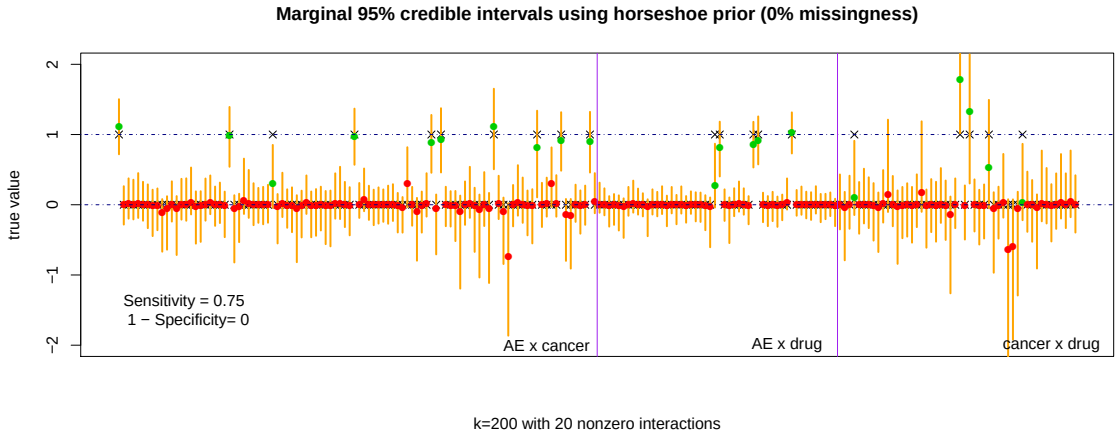

Figure 4.1: 95% marginal credible intervals of interaction coefficients based on horseshoe prior for just a single simulated fully-observed data set, in a setting with $n_{IA} = 200$ and 20 non-zero interactions.

The results of posteriors are based on 100 simulated training data sets. For each model, we repeat the same data generation procedure in order to enable the comparison of simulation results. After we extract the posterior medians of main effects ($\hat{u}$) from all three Bayesian models and interaction effects ($\hat{\beta}$) from sBMIhs and sBMIss, we plug estimates from these models to obtain the predicted AE incidence ($\hat{p}$), and then generate the predicted AE outcome ($\hat{Y}$). For each iteration, we are capable of calculating MSPE, the expected value of squared difference between the estimated AE outcome and true

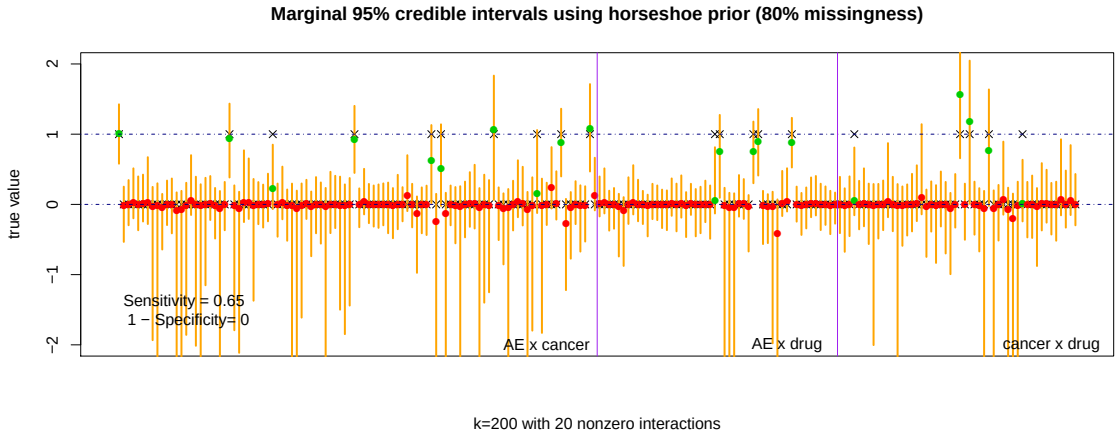

Figure 4.2: 95% marginal credible intervals of interaction coefficients based on horseshoe prior for just a single simulated data set under 80% missingness, in a setting with $n_{IA} = 200$ and 20 non-zero interactions.

value. Table 4.1 summarizes the MSPE, MSE, TPR, FPR, and MCC by averaging over all iterations. Note that the true value of regression coefficients for marginal effects are the same for both training and test datasets. The results are based on averaging over all 100 simulations.

The simulation results of each model are summarized in Table 4.1. As shown, when all the data are fully observed in the analysis, the prediction accuracy of BMCD and glm - marginal models without considering any interactions, but under different frameworks - is the worst among all models in Scenario 1, demonstrating the potential interactions between main effects. The prediction accuracy is then improved by the proposed sparse interaction selection models. sBMIhs and sBMIss have superior prediction performance compared with others in the presence of no/medium/high percentage of missingness. Moreover, it is remarkable that the performance of BMCD (main-effect only), sBMIhs and sBMIss is relatively stable when the percentage of missingness increases, further demonstrating the benefits of handling censored rare event outcomes under the Bayesian framework. Such robustness results from the ability to accommodate the informatively

Scenario	Method	median($\tilde{Y}$;$\hat{Y}$)	MSPE(SE)	MSE(SE)	TPR	FPR	MCC
0%	BMCD	2.290;2.595	11.473(2.954)	1.300(0.099)	-	-	-
censored	sBMIhs	2.290;2.280	**7.314(1.428)**	**0.531(0.042)**	0.708	6e-5	0.827
	sBMIss	2.290;2.265	**7.360(1.479)**	**0.529(0.041)**	0.724	2.2e-4	0.836
	glmIA	2.290;2.210	7.849(1.817)	-(-)	-	-	-
	glm	2.290;2.410	11.717(3.244)	-(-)	-	-	-
40%	BMCD	2.290;2.750	11.618(2.939)	1.291(0.099)	-	-	-
censored	sBMIhs	2.290;2.495	**7.418(1.463)**	**0.529(0.041)**	0.687	6e-5	0.813
	sBMIss	2.290;2.460	**7.486(1.515)**	**0.525(0.042)**	0.702	1.7e-4	0.822
	glmIA	2.290;4.175	41.799(14.087)	-(-)	-	-	-
	glm	2.290;4.485	37.423(12.793)	-(-)	-	-	-
80%	BMCD	2.290;3.095	12.318(3.243)	1.294(0.102)	-	-	-
censored	sBMIhs	2.290;2.365	**7.766(1.607)**	**0.541(0.041)**	0.553	6e-5	0.723
	sBMIss	2.290;2.320	**7.731(1.546)**	**0.538(0.042)**	0.582	1.7e-4	0.742
	glmIA	2.290;8.180	73.471(24.081)	-(-)	-	-	-
	glm	2.290;8.545	68.971(22.119)	-(-)	-	-	-

Table 4.1: The mean squared prediction error (MSPE) with sample standard deviation (SSD) for the test data, the mean squared error (MSE) for regression coefficients, the average true positive rate (TPR), the average false positive rate (FPR), and Matthews correlation coefficient (MCC) for five models, marginal-only model (BMCD), sparse Bayesian model of interaction selection using horseshoe prior (sBMIhs), sparse Bayesian model of interaction selection using spike-and-slab prior (sBMIss), logistic regression model with all true interactions (glmIA), and logistic regression model without interactions (glm) under 0%, 40%, and 80% censoring (missing not at random).

censored outcomes in the likelihood function. Compared to the proposed models, glmIA, which includes all nonzero true interactions, has similar prediction accuracy based on the value of MSPE in Scenario 1; however, in the presence of missingness (Scenarios 2 and 3), its prediction accuracy worsens due to only observed cases being included in the model. Furthermore, glmIA has a limitation to handle multiple categorical covariates with high-dimensional categories. For interaction identification, sBMIhs and sBMIss have lower FP rates which are very close to zero in all scenarios, which suggests very few zero interactions are incorrectly selected. Also, sBMIhs and sBMIss have highest TP rates (over 0.7) in Scenario 1. In the case when 80% of AE outcomes are censored (Scenario 3), TP rates over 0.55 are still achieved without suffering high FP rates in both Bayesian

interaction selection models. In all three scenarios, the sBMIhs gives the value of MCC at 0.83, 0.81 and 0.72 on average, respectively.

4.4 Case Study

In this section, we utilize the proposed Bayesian interaction selection method to identify pairwise interaction effects in the real data meta-analysis of treatment-related all-grade (Grade 1-5) and Grade 3 or higher (Grade 3-5) adverse events (AEs) with informative censoring [2]. We obtained the study-level safety data from systematic review of 125 clinical studies reporting AEs for two anti-PD-1 drugs (Nivolumab, Pembrolizumab) and three anti-PD-L1 drugs (Atezolizumab, Avelumab and Durvalumab) published from 2011-2018. Other than the study-level AE outcome data, we have the following study-level information including drug, AE subtype, cancer type, criteria of AE reporting. For the severity level of AE outcome, the number of Grade 1-5/Grade 3-5 (G15/G35) AE was recorded, however, more than 60% of AEs on average for each study was left-censored due to lower than a pre-specified cutoff value. Therefore, censored AE outcomes were treated as missing not at random (MNAR) in the analysis. The objective is to evaluate the incidence probability of AEs by subgroups and identify the high risk groups by selecting those nonzero two-way interactions in the presence of left-censored study-level data.

The proposed Bayesian interaction selection model under horseshoe prior is implemented in the statistical software R and JAGS [18], which uses a Markov Chain Monto Carlo (MCMC) algorithm to generate samples from the posterior distribution of the parameters of interest. Along with listing the data and setting the initial values of model parameters, we specified the likelihood functions and prior distributions for the proposed model in JAGS. We run three parallel chains for the model. For each MCMC chain, after discarding the burn-in period of 30,000 iterations, the 3 chains showed good mixing and

successful convergence to the target distribution. We eventually obtain 10,000 posterior samples per chain by retaining one sample out of three. The 30,000 posterior samples of model parameters such as incidence probabilities of AEs by subgroups and candidate interactions are saved for the Bayesian inference.

Figure 4.3 displays the posterior medians and the corresponding 95% CrIs of those selected interaction coefficients by horseshoe prior. Of the 927 candidate interactions, 30 nonzero ones were selected under both horseshoe and spike-and-slab prior, suggesting a relatively consistent result on sparse interaction identification in the presence of censored outcome data.

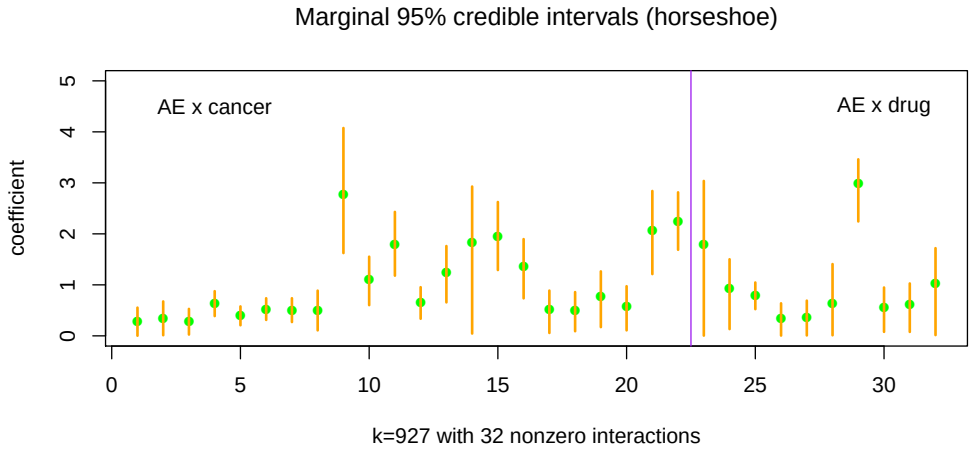

Figure 4.3: Significant interactions for all-grade AEs selected by horseshoe prior

Based on the posterior medians and the corresponding 95% credible intervals (95% CrI) of 927 interaction coefficients obtained from the proposed interaction selection model for all-grade AEs using horseshoe prior, 32 interactions having 95% CrIs not covering 0 were selected. Among all those 22 selected AE $\times$ cancer interactions, vitiligo $\times$ melanoma interaction has the largest effect ($\hat{\beta} = 2.766$; 95% CrI, 1.623-4.075) on all-grade AE incidence, and the estimated all-grade (G15) incidence probability of AE is $\hat{p}^{G15} = 0.167$ with 95% CrI (0.060,0.421), followed by neutropenia $\times$ hematologic malignancy interaction

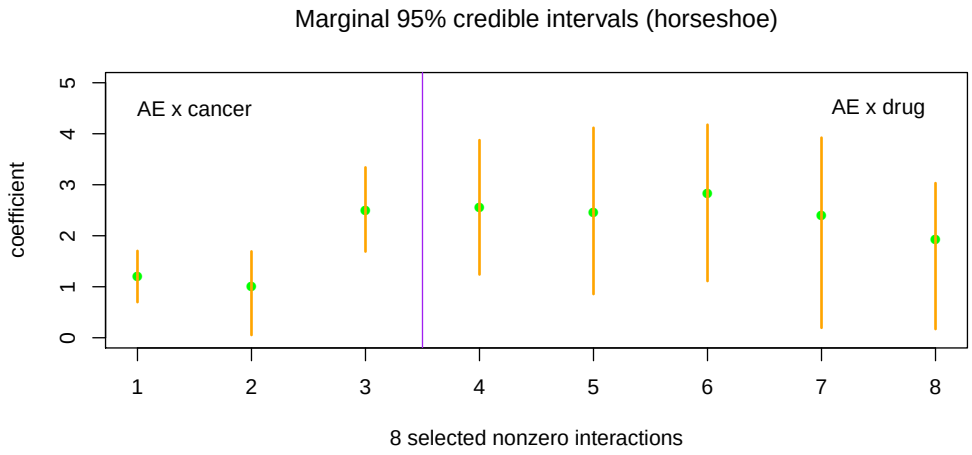

Figure 4.4: Significant interactions for Grade 3 or higher AEs selected by horseshoe prior

($\hat{\beta}$ = 2.252 (1.6881-2.814); $\hat{p}^{G15}$ = 0.106 (0.062-0.175)), and platelet count decreased × hematologic malignancy interaction ($\hat{\beta}$ = 2.072; 95% CrI, 1.211-2.841. $\hat{p}^{G15}$ = 0.091; 95% CrI, 0.040-0.178). Among those 10 selected AE × drug interactions, infusion-related reaction × avelumab interaction has the largest effect ($\hat{\beta}$ = 2.986; 95% CrI, 2.243-3.461) on AE incidence, and the estimated all-grade incidence probability of AE is $\hat{p}^{G15} = 0.178$ with 95% CrI (0.076,0.275), followed by Amylase increased × Nivolumab (0.061; 95% CrI, 0.011-0.191). The overall mean incidence probability of all-grade AE is 0.011 (95% CrI, 0.005-0.013).

On the other hand, 8 pairwise interactions were selected based upon Grade 3 or higher (Grade 3-5) AEs using horseshoe prior. Among three selected AE × cancer interactions, neutropenia × hematologic malignancy interaction has the largest effect on Grade 3-5 (G35) AE incidence, and the estimated incidence probability of G35 AE is $\hat{p}^{G35} = 0.009$ (95% CrI; 0.004,0.021), followed by pneumonitis × lung cancer interaction and colitis × melanoma interaction. Among those 5 selected AE × drug interactions, infusion-related reaction × Avelumab interaction has the largest effect on G35 AE incidence, and the estimated all-grade incidence probability of G35 AE is $\hat{p}^{G35} = 0.012$ (95% CrI; 0.002,0.044),

followed by Lipase increased × Nivolumab interaction, Amylase increased × Nivolumab interaction, Lipase increased × Avelumab interaction, and γ-Glutamyl transferase increased (GGT) × Durvalumab interaction. The overall mean incidence probability of G35 AE is 0.0007 (95% CrI, 0.0005-0.0009).

4.5 Discussion

In this work, we have developed a sparse full Bayesian interaction selection model to simultaneously identify nonzero interactions and those high risk groups with higher incidence probability of AE, in the presence of informative censoring. Through simulations, we have demonstrated that the proposed interaction selection approach can improve the prediction accuracy and select those non-zero interactions when the underlying truth is sparse and the overall AE incidence is rare. We have also illustrated the proposed approach with a real-data meta-analysis to identify significant ones among high-dimensional candidate pairwise interactions. The results from this application can make a tremendous impact on cancer patients treated with immunotherapy drugs.

L1 regularization promotes sparsity, therefore, LASSO is preferred over ridge regression to achieve sparsity in the model features. However, LASSO may not be sufficiently sparse, especially when determining the optimal value for regularization/penalty parameter using cross-validation, the model becomes denser and the significant features are over-selected.

The interaction we identified between vitiligo, an autoimmune skin disorder, and melanoma is consistent with reports in the literature; for example, a prospective study of patients with metastatic melanoma treated with pembrolizumab observed a cumulative incidence of vitiligo of 25% [89]. Low neutrophil counts (neutropenia) and low platelet counts (thrombocytopenia) have both been reported as serious, but rare, hematological

irAEs [90], so the link to hematological malignancies is logical. Avelumab has been linked to infusion-related reactions, with a rate of 20% reported among subjects with urothelial cancer treated in a Phase II study [91]. Thus, our real-data meta-analysis demonstrates that the proposed interaction selection method yields clinically meaningful results. Understanding the risk of various irAEs based on a subject's cancer type and potential treatment options is a key question which can guide improved prevention and management of these events [92], and the proposed statistical framework can provide robust estimates to inform these decisions.

The proposed method is appropriate for other applications if outcomes are censored. Extending the current sparse interaction selection approach to handle three-way interactions, such as $\mathsf{AE} \times \mathsf{drug} \times \mathsf{cancer}$ should be possible, although the complexity of model fitting may increase. Another potential direction is to extend the approach to other types of priors along with appropriate calibrations on shrinkage parameters to capture much smaller effects and favor more sparsity. We hope to explore more clinical-type models to offer practical guidance on drug safety as future research of interest.

4.6 Appendix

JAGS for sparse Bayesian model of censored data with interaction selection based on horseshoe prior (sBMIhs).

```
model{
      for (j in 1:J1){ # observed
            Y[j] ~ dbin(theta[j], N[j]) # the likelihood
            logit(theta[j]) <- theta.v1[v1[j]] + theta.v2[v2[j]] +
            theta.v3[v3[j]] + theta.v4[v4[j]] + beta1[v5[j]] +
            beta2[v6[j]] + beta3[v7[j]]
      }
      for (j in 1:J2){ # censored
            Z[j] ~ dbern(p[j]) # the likelihoo
            p[j] <- pbin(cut[j], theta[j+J1], N[j+J1])  #Y<=cut
            logit(theta[j+J1]) <- theta.v1[v1[j+J1]] + theta.v2[v2[j+J1]] +
            theta.v3[v3[j+J1]] + theta.v4[v4[j+J1]] +
            beta1[v5[j+J1]] + beta2[v6[j+J1]] + beta3[v7[j+J1]]
      }
      for (i1 in 1:n.v1){
            theta.v1[i1] ~ dnorm(0, prec.v1)
            } # a half-cauchy prior on standard deviation
      prec.v1 <- pow(sigma.v1, -2)
      sigma.v1 ~ dt(0, a, 1)T(0,) # a=1/A^2, where scale parameter A=25

      ... # same priors on other main coefficients: v2, v3, v4
```

```
      for (k1 in 1:n.v5){
            beta1[k1] ~ dnorm(0, prec1[k1])
            prec1[k1] <- pow(sigma1[k1], -2)# precision = 1/variance
            sigma1[k1] <- lambda1[k1]*tau1
            lambda1[k1] ~ dt(0, 1, 1)T(0,)  # local shrinkage parameters
      } # horseshoe prior
      tau1 ~ dt(0, 1, 1)T(0,)  # global shrinkage parameter

      ... # same priors on other interaction coefficients: v6, v7
}
```

In contrast, interaction selection based on spike-and-slab prior (sBMIss) as follows:

```
model{
      # same model specification for likelihood and main coefficients as above.
      ...
      for (k1 in 1:n.v5){
            IndIA1[k1] ~ dcat(PInd1[])   # returns 1 or 2
            Ind1[k1] <- IndIA1[k1]-1     # returns 0 or 1
            beta1[k1] ~ dnorm(0, tauIA1[IndIA1[k1]])
      } # SSVS
      PInd1[1] <- 1/2
      PInd1[2] <- 1-PInd1[1]
      tauIA1[1] <- taub1       # spike: ind1 = 0
      tauIA1[2] <- taub1/100   # slab:  ind1 = 1
      taub1 <- pow(sdbeta1, -2)
      sdbeta1 ~ dunif(0, 1)
```

```
      ... # same priors on other interaction coefficients: v6, v7
}
```

[Chapter 4: manuscript is in preparation for publication.]

Some figures on pairwise interaction selections from the real-data application.

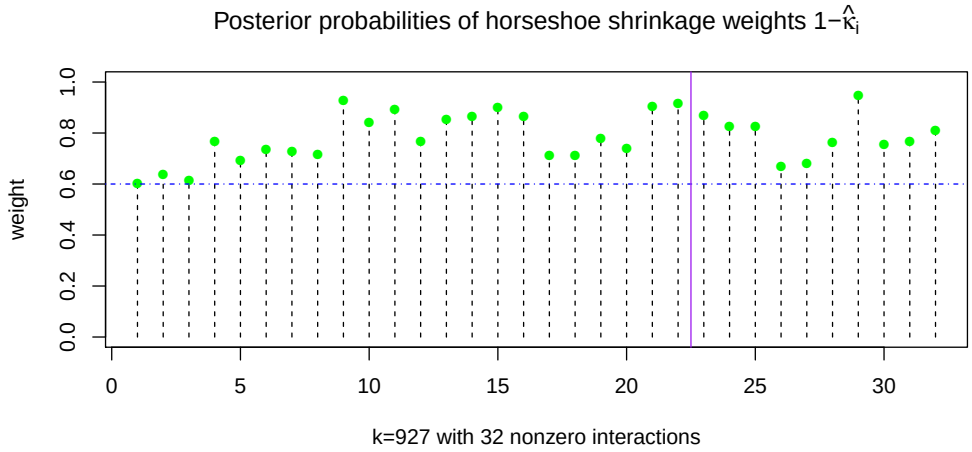

Figure 4.5: Shrinkage weights corresponding to 32 nonzero interactions for all-grade AEs selected by horseshoe prior

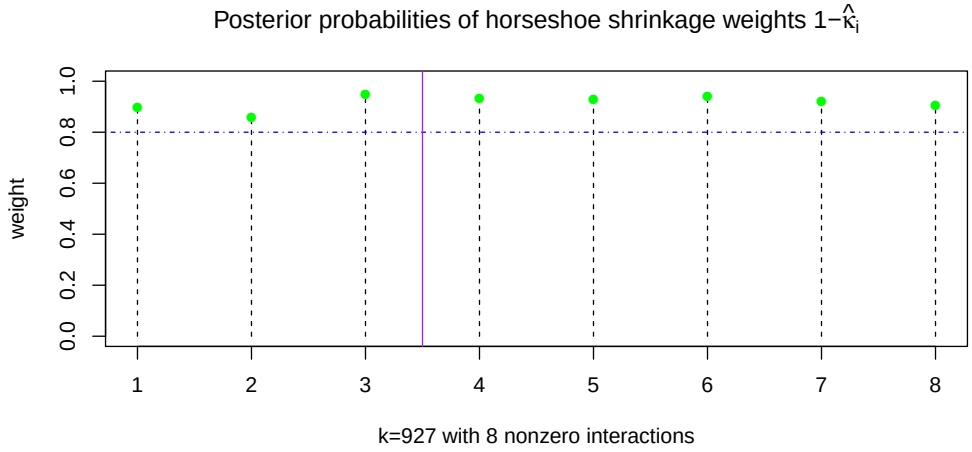

Figure 4.6: Shrinkage weights corresponding to 8 nonzero interactions for grade 3 or higher AEs selected by horseshoe prior.

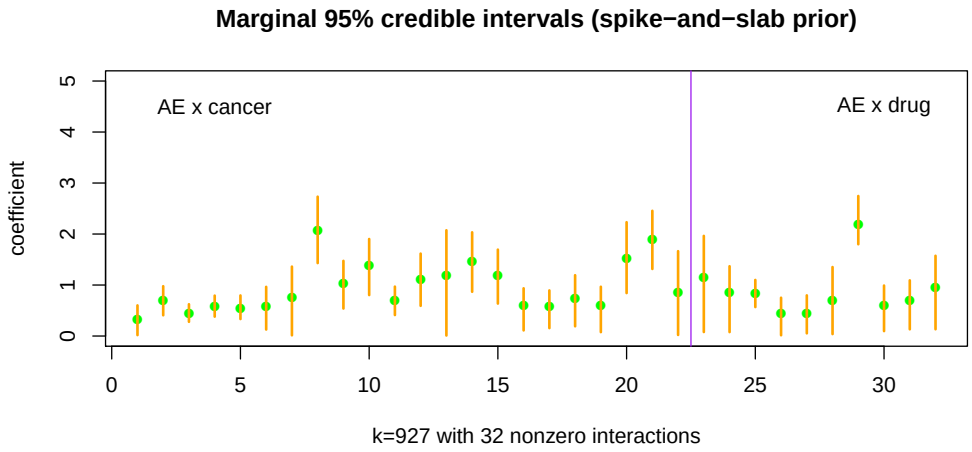

Figure 4.7: Significant interactions for all-grade AEs selected by spike-and-slab prior

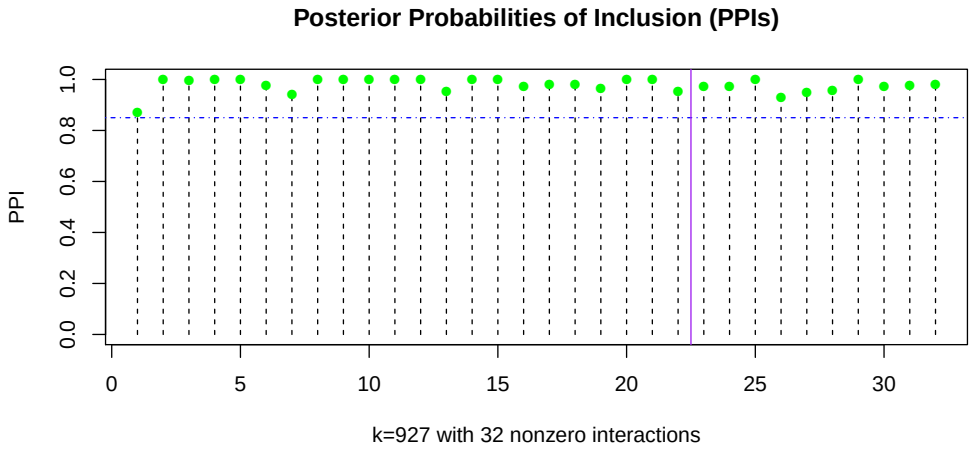

Figure 4.8: PPIs corresponding to 32 nonzero interactions for all-grade AEs selected by spike-and-slab prior.

Chapter 5

Summary and Future Work

5.1 Conclusion

A Bayesian hierarchical model is a natural choice to deal with sparse event data along with accounting for study-specific heterogeneity in a meta-analysis setting when the study-level outcome data are informatively censored. To better understand the toxicity profiles of immunotherapy drugs, in Chapter 2, we present our model to handle censored rare event outcome with missing not at random under the coarsened data framework. The data are coarsened at random, and model parameters and coarsening process are *a priori* independent, therefore, the resulting likelihood and Bayesian inferences are capable of ignoring the stochastic nature of the coarsening mechanism.

An unsolved statistical computing issue has been identified during implementing Bayesian model selection for censored data analysis in JAGS. To select the best one among a set of candidate Bayesian censored data models in the presence of complicated or high-dimensional features, in Chapter 3, we demonstrate that the alternative modeling strategy draws correct posterior samples, as well as helps simultaneously computing the true deviance on the basis of likelihood functions for censored model selection in JAGS.

The proposed algorithm can also be generalized to different types of data, models as well as many other disciplines.

On top of main effects in the model, we are also interested in exploring significant pairwise interactions by specifying priors to control shrinkage properties for selection of interactions. In Chapter 4, we aim to find the optimal method to identify two-way nonzero ones that are truly relevant to outcome in a large search space of possible sets of interactions. In addition to unbiasedly estimating the incidence probability across subgroups, the proposed Bayesian interaction selection method is able to select significant interactions among candidates and identify high-risk groups for practical guidance.

5.2 Discussion

The additional future directions will include the extension of the proposed Bayesian model of censored data in aim 1 to (1) model high-dimensional genomic data at the study-level or patient level and estimate the probability of genetic mutation by subgroup analysis; (2) handle right-censored AE data when only grade 2 or higher AEs were reported in a meta-analysis of trials reporting drug safety. The proposed algorithm in aim 2 can also facilitate the comparison among eight possible variations of deviance information criterion (DIC) for missing data models [93] , such as observed DICs, complete DICs, conditional DICs. It would be of interest in future work. Extending the current sparse interaction selection approach in aim 3 to handle three-way interactions, such as $\mathsf{AE} \times \mathsf{drug} \times \mathsf{cancer}$ should be possible, although the complexity of model fitting may increase. Another potential direction is to extend the approach to other types of priors along with appropriate calibrations on shrinkage parameters to capture much smaller effects and favor more sparsity. We hope to explore more clinical-type models to offer practical guidance on safety as future research of interest.

Bibliography

[1] J. J. Barendregt, S. A. Doi, Y. Y. Lee, R. E. Norman, and T. Vos, "Meta-analysis of prevalence," *J Epidemiol Community Health*, vol. 67, no. 11, pp. 974–978, 2013.

[2] Y. Wang, S. Zhou, F. Yang, X. Qi, X. Wang, X. Guan, C. Shen, N. Duma, J. V. Aguilera, A. Chintakuntlawar, *et al.*, "Treatment-related adverse events of pd-1 and pd-l1 inhibitors in clinical trials: a systematic review and meta-analysis," *JAMA oncology*, vol. 5, no. 7, pp. 1008–1019, 2019.

[3] A. J. Sutton, N. J. Cooper, P. C. Lambert, D. R. Jones, K. R. Abrams, and M. J. Sweeting, "Meta-analysis of rare and adverse event data," *Expert review of pharmacoeconomics & outcomes research*, vol. 2, no. 4, pp. 367–379, 2002.

[4] M. J. Bradburn, J. J. Deeks, J. A. Berlin, and A. Russell Localio, "Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events," *Statistics in medicine*, vol. 26, no. 1, pp. 53–77, 2007.

[5] D. K. Bhaumik, A. Amatya, S.-L. T. Normand, J. Greenhouse, E. Kaizar, B. Neelon, and R. D. Gibbons, "Meta-analysis of rare binary adverse event data," *Journal of the American Statistical Association*, vol. 107, no. 498, pp. 555–567, 2012.

[6] T. H. Hamza, H. C. van Houwelingen, and T. Stijnen, "The binomial distribution of meta-analysis was preferred to model within-study variability," *Journal of clinical epidemiology*, vol. 61, no. 1, pp. 41–51, 2008.

[7] K. Carriere, "How good is a normal approximation for rates and proportions of low incidence events?," *Communications in Statistics-Simulation and Computation*, vol. 30, no. 2, pp. 327–337, 2001.

[8] H. S. Luft and B. Brown Jr, "Calculating the probability of rare events: why settle for an approximation?," *Health services research*, vol. 28, no. 4, p. 419, 1993.

[9] P. W. Lane, "Meta-analysis of incidence of rare events," *Statistical methods in medical research*, vol. 22, no. 2, pp. 117–132, 2013.

[10] T. Cai, L. Parast, and L. Ryan, "Meta-analysis for rare events," *Statistics in medicine*, vol. 29, no. 20, pp. 2078–2089, 2010.

[11] D. Liu, R. Y. Liu, and M.-g. Xie, "Exact meta-analysis approach for discrete data and its application to 2× 2 tables with rare events," *Journal of the American Statistical Association*, vol. 109, no. 508, pp. 1450–1465, 2014.

[12] J. P. Higgins, I. R. White, and A. M. Wood, "Imputation methods for missing outcome data in meta-analysis of clinical trials," *Clinical Trials*, vol. 5, no. 3, pp. 225–239, 2008.

[13] D. F. Heitjan and D. B. Rubin, "Ignorability and coarse data," *The annals of statistics*, pp. 2244–2253, 1991.

[14] D. F. Heitjan, "Ignorability and coarse data: Some biomedical examples," *Biometrics*, pp. 1099–1109, 1993.

[15] D. F. Heitjan, "Ignorability in general incomplete-data models," *Biometrika*, vol. 81, no. 4, pp. 701–708, 1994.

[16] D. Mavridis, A. Chaimani, O. Efthimiou, S. Leucht, and G. Salanti, "Addressing

missing outcome data in meta-analysis," *Evidence-based mental health*, vol. 17, no. 3, pp. 85–89, 2014.

[17] Y. Yuan and R. J. Little, "Meta-analysis of studies with missing data," *Biometrics*, vol. 65, no. 2, pp. 487–496, 2009.

[18] M. Plummer, "Jags: A program for analysis of bayesian graphical models using gibbs sampling," in *Proceedings of the 3rd international workshop on distributed statistical computing*, vol. 124, Vienna, Austria., 2003.

[19] J. Kruschke, *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press, 2014.

[20] M. Plummer, "Jags version 4.3. 0 user manual. 28 june 2017," 2017.

[21] M. Plummer, "*SOURCEFORGE JAGS: Just Another Gibbs Sampler* help forum: dinterval()," 2012. https://sourceforge.net/p/mcmc-jags/discussion/610037/thread/fd7f3f7e/.

[22] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[23] R. B. O'Hara and M. J. Sillanpää, "A review of bayesian variable selection methods: what, how and which," *Bayesian analysis*, vol. 4, no. 1, pp. 85–117, 2009.

[24] L. Kuo and B. Mallick, "Variable selection for regression models," *Sankhyā: The Indian Journal of Statistics, Series B*, pp. 65–81, 1998.

[25] P. Dellaportas, J. J. Forster, and I. Ntzoufras, "On bayesian model and variable selection using mcmc," *Statistics and Computing*, vol. 12, no. 1, pp. 27–36, 2002.

[26] T. Park and G. Casella, "The bayesian lasso," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 681–686, 2008.

[27] P. J. Green, "Reversible jump markov chain monte carlo computation and bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.

[28] E. I. George and R. E. McCulloch, "Variable selection via gibbs sampling," *Journal of the American Statistical Association*, vol. 88, no. 423, pp. 881–889, 1993.

[29] C. M. Carvalho, N. G. Polson, and J. G. Scott, "The horseshoe estimator for sparse signals," *Biometrika*, vol. 97, no. 2, pp. 465–480, 2010.

[30] C. M. Carvalho, N. G. Polson, and J. G. Scott, "Handling sparsity via the horseshoe," in *Artificial Intelligence and Statistics*, pp. 73–80, 2009.

[31] V. Ročková and E. I. George, "Emvs: The em approach to bayesian variable selection," *Journal of the American Statistical Association*, vol. 109, no. 506, pp. 828–846, 2014.

[32] G. Carvajal, A. Branch, S. A. Sisson, D. J. Roser, B. van den Akker, P. Monis, P. Reeve, A. Keegan, R. Regel, and S. J. Khan, "Virus removal by ultrafiltration: understanding long-term performance change by application of bayesian analysis," *Water research*, vol. 122, pp. 269–279, 2017.

[33] S. K. Ghosh and S. Ghosal, "Semiparametric accelerated failure time models for censored data," *Bayesian Statistics and its Applications*, vol. 15, pp. 213–229, 2006.

[34] B. Han, M. Yu, J. J. Dignam, and P. J. Rathouz, "Bayesian approach for flexible modeling of semicompeting risks data," *Statistics in medicine*, vol. 33, no. 29, pp. 5111–5125, 2014.

[35] T. R. Johnson and K. M. Kuhn, "Bayesian thurstonian models for ranking data using jags," *Behavior research methods*, vol. 45, no. 3, pp. 857–872, 2013.

[36] R. Bååth, "Estimating the distribution of sensorimotor synchronization data: A bayesian hierarchical modeling approach," *Behavior research methods*, vol. 48, no. 2, pp. 463–474, 2016.

[37] N. T. Davies, L. A. Apiolaza, and M. Sharma, "Heritability of growth strain in eucalyptus bosistoana: a bayesian approach with left-censored data," *New Zealand Journal of Forestry Science*, vol. 47, no. 1, p. 5, 2017.

[38] P. Busschaert, A. Geeraerd, M. Uyttendaele, and J. Van Impe, "Hierarchical bayesian analysis of censored microbiological contamination data for use in risk assessment and mitigation," *Food microbiology*, vol. 28, no. 4, pp. 712–719, 2011.

[39] X. Wang, M.-H. Chen, and J. Yan, "Bayesian dynamic regression models for interval censored survival data with application to children dental health," *Lifetime data analysis*, vol. 19, no. 3, pp. 297–316, 2013.

[40] R. Little and D. Rubin, "Statistical analysis with missing data. wiley," *New York*, 2002.

[41] M. A. Silva, A. C. Swanson, P. J. Gandhi, and G. R. Tataronis, "Statin-related adverse events: a meta-analysis," *Clinical therapeutics*, vol. 28, no. 1, pp. 26–35, 2006.

[42] M. Plummer, "Jags (just another gibbs sampler). 2010," *Available from: www. mcmc-jags. sourceforge. net/)[Accessed on 8 December 2013]*, 2007.

[43] D. B. Rubin, "Inference and missing data," *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976.

[44] M. Jacobsen and N. Keiding, "Coarsening at random in general sample spaces and random censoring in continuous time," *The Annals of Statistics*, vol. 23, no. 3, pp. 774–786, 1995.

[45] R. J. Little, D. B. Rubin, and S. Z. Zangeneh, "Conditions for ignoring the missing-data mechanism in likelihood inferences for parameter subsets," *Journal of the American Statistical Association*, vol. 112, no. 517, pp. 314–320, 2017.

[46] A. Gelman, "Prior distributions for variance parameters in hierarchical models," *Bayesian analysis*, vol. 1, no. 3, pp. 515–534, 2006.

[47] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde, "Bayesian measures of model complexity and fit," *Journal of the royal statistical society: Series b (statistical methodology)*, vol. 64, no. 4, pp. 583–639, 2002.

[48] M. J. Sweeting, A. J. Sutton, and P. C. Lambert, "What to add to nothing? use and avoidance of continuity corrections in meta-analysis of sparse data," *Statistics in medicine*, vol. 23, no. 9, pp. 1351–1375, 2004.

[49] N. P. Jewell, *Statistics for epidemiology*. Chapman and Hall/CRC, 2003.

[50] D. B. Peizer and J. W. Pratt, "A normal approximation for binomial, f, beta, and other common, related tail probabilities, i," *Journal of the American Statistical Association*, vol. 63, no. 324, pp. 1416–1456, 1968.

[51] Y. Ma, H. Chu, and M. Mazumdar, "Meta-analysis of proportions of rare events–a comparison of exact likelihood methods with robust variance estimation," *Communications in Statistics-Simulation and Computation*, vol. 45, no. 8, pp. 3036–3052, 2016.

[52] G. Casella and R. L. Berger, *Statistical inference*, vol. 2. Duxbury Pacific Grove, CA, 2002.

[53] B. Dai, S. Ding, and G. Wahba, "Multivariate bernoulli distribution," *Bernoulli*, vol. 19, no. 4, pp. 1465–1483, 2013.

[54] J. R. Brahmer, C. G. Drake, *et al.*, "Phase i study of single-agent anti–programmed death-1 (mdx-1106) in refractory solid tumors: safety, clinical activity, pharmacodynamics, and immunologic correlates," *Journal of clinical oncology*, vol. 28, no. 19, p. 3167, 2010.

[55] A. Lewbel and O. Linton, "Nonparametric censored and truncated regression," *Econometrica*, vol. 70, no. 2, pp. 765–779, 2002.

[56] M. Hamada and C. Wu, "Analysis of censored data from highly fractionated experiments," *Technometrics*, vol. 33, no. 1, pp. 25–38, 1991.

[57] D.-G. D. Chen, J. Sun, and K. E. Peace, *Interval-Censored Time-to-Event Data: Methods and Applications*. CRC Press, 2012.

[58] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[59] M. A. Tanner and W. H. Wong, "The calculation of posterior distributions by data augmentation," *Journal of the American statistical Association*, vol. 82, no. 398, pp. 528–540, 1987.

[60] M. Hamada and C. Wu, "Analysis of censored data from fractionated experiments: a bayesian approach," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 467–477, 1995.

[61] J. G. Ibrahim, M.-H. Chen, and D. Sinha, *Bayesian survival analysis*. Springer Science & Business Media, 2013.

[62] J. C. Jakobsen, C. Gluud, J. Wetterslev, and P. Winkel, "When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts," *BMC medical research methodology*, vol. 17, no. 1, p. 162, 2017.

[63] D. Spiegelhalter, A. Thomas, N. Best, and D. Lunn, "Winbugs user manual," 2003.

[64] D. Lunn, C. Jackson, N. Best, A. Thomas, and D. Spiegelhalter, *The BUGS book: A practical introduction to Bayesian analysis*. CRC press, 2012.

[65] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[66] M. Plummer, "Penalized loss functions for bayesian model comparison," *Biostatistics*, vol. 9, no. 3, pp. 523–539, 2008.

[67] S. Chib, "Bayes inference in the tobit censored regression model," *Journal of Econometrics*, vol. 51, no. 1-2, pp. 79–99, 1992.

[68] M. Plummer, *rjags: Bayesian Graphical Models using MCMC*, 2019. R package version 4-10.

[69] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.

[70] J. S. Long, "Regression models for categorical and limited dependent variables (vol. 7)," *Advanced quantitative techniques in the social sciences*, p. 219, 1997.

[71] L.-F. Lee, "Estimation of dynamic and arch tobit models," *Journal of Econometrics*, vol. 92, no. 2, pp. 355–390, 1999.

[72] J. W. Twisk, *Applied longitudinal data analysis for epidemiology: a practical guide*. cambridge university press, 2013.

[73] X. Xu and L.-f. Lee, "Maximum likelihood estimation of a spatial autoregressive tobit model," *Journal of Econometrics*, vol. 188, no. 1, pp. 264–280, 2015.

[74] J. P. Klein and M. L. Moeschberger, *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media, 2006.

[75] J. D. Kalbfleisch and R. L. Prentice, *The statistical analysis of failure time data*, vol. 360. John Wiley & Sons, 2011.

[76] F. D. Nelson, "Censored regression models with unobserved, stochastic censoring thresholds," *Journal of econometrics*, vol. 6, no. 3, pp. 309–327, 1977.

[77] D. A. Freedman, *Statistical models: theory and practice*. cambridge university press, 2009.

[78] J. H. Albert and S. Chib, "Bayesian analysis of binary and polychotomous response data," *Journal of the American statistical Association*, vol. 88, no. 422, pp. 669–679, 1993.

[79] G. M. Koop, *Bayesian econometrics*. John Wiley & Sons Inc., 2003.

[80] R. G. Miller Jr, *Survival analysis*, vol. 66. John Wiley & Sons, 2011.

[81] R. J. Little and D. B. Rubin, *Statistical analysis with missing data*, vol. 793. John Wiley & Sons, 2019.

[82] T. Ando, "Bayesian predictive information criterion for the evaluation of hierarchical bayesian and empirical bayes models," *Biometrika*, vol. 94, no. 2, pp. 443–458, 2007.

[83] S. Watanabe, "Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory," *Journal of Machine Learning Research*, vol. 11, no. Dec, pp. 3571–3594, 2010.

[84] R. D. Bock and M. Aitkin, "Marginal maximum likelihood estimation of item parameters: Application of an em algorithm," *Psychometrika*, vol. 46, no. 4, pp. 443–459, 1981.

[85] C. Liu, D. B. Rubin, and Y. N. Wu, "Parameter expansion to accelerate em: the px-em algorithm," *Biometrika*, vol. 85, no. 4, pp. 755–770, 1998.

[86] G. L. de Oliveira, R. H. Loschi, and R. M. Assunção, "A random-censoring poisson model for underreported data," *Statistics in medicine*, vol. 36, no. 30, pp. 4873–4892, 2017.

[87] A. Onofri, H.-P. Piepho, and M. Kozak, "Analysing censored data in agricultural research: A review with examples and software tips," *Annals of applied biology*, vol. 174, no. 1, pp. 3–13, 2019.

[88] S. van der Pas, B. Szabó, and A. van der Vaart, "Uncertainty quantification for the horseshoe (with discussion)," *Bayesian Analysis*, vol. 12, no. 4, pp. 1221–1274, 2017.

[89] C. Hua, L. Boussemart, C. Mateus, E. Routier, C. Boutros, H. Cazenave, R. Viollet, M. Thomas, S. Roy, N. Benannoune, *et al.*, "Association of vitiligo with tumor response in patients with metastatic melanoma treated with pembrolizumab," *JAMA dermatology*, vol. 152, no. 1, pp. 45–51, 2016.

[90] N. Delanoy, J.-M. Michot, T. Comont, N. Kramkimel, J. Lazarovici, R. Dupont, S. Champiat, C. Chahine, C. Robert, C. Herbaux, *et al.*, "Haematological immune-related adverse events induced by anti-pd-1 or anti-pd-l1 immunotherapy: a de-

scriptive observational study," *The Lancet Haematology*, vol. 6, no. 1, pp. e48–e57, 2019.

[91] A. B. Apolo, J. R. Infante, A. Balmanoukian, M. R. Patel, D. Wang, K. Kelly, A. E. Mega, C. D. Britten, A. Ravaud, A. C. Mita, *et al.*, "Avelumab, an anti–programmed death-ligand 1 antibody, in patients with refractory metastatic urothelial carcinoma: results from a multicenter, phase ib study," *Journal of Clinical Oncology*, vol. 35, no. 19, p. 2117, 2017.

[92] C. Connolly, K. Bambhania, and J. Naidoo, "Immune-related adverse events: A case-based approach," *Frontiers in Oncology*, vol. 9, 2019.

[93] G. Celeux, F. Forbes, C. Robert, and M. Titterington, "Deviance information criteria for missing data models," *Bayesian Analysis*, vol. 1, no. 4, pp. 651–674, 2006.