Received: 30 January 2022 | Revised: 16 June 2022 | Accepted: 13 July 2022

DOI: 10.1002/jrsm.1597

RESEARCH ARTICLE

Research Synthesis Methods WILEY

Bayesian sparse modeling to identify high-risk subgroups in meta-analysis of safety data

Xinyue Qi[1] | Shouhao Zhou[2] | Yucai Wang[3] | Christine Peterson[1]

[1]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[2]Department of Public Health Sciences, Penn State College of Medicine, Hershey, Pennsylvania, USA

[3]Division of Hematology, Mayo Clinic, Rochester, Minnesota, USA

Correspondence
Christine Peterson, Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.
Email: cbpeterson@mdanderson.org

Funding information
NIH/NCI CCSG grant, Grant/Award Number: P30CA016672

Abstract

Meta-analysis allows researchers to combine evidence from multiple studies, making it a powerful tool for synthesizing information on the safety profiles of new medical interventions. There is a critical need to identify subgroups at high risk of experiencing treatment-related toxicities. However, this remains quite challenging from a statistical perspective as there are a variety of clinical risk factors that may be relevant for different types of adverse events, and adverse events of interest may be rare or incompletely reported. We frame this challenge as a variable selection problem and propose a Bayesian hierarchical model which incorporates a horseshoe prior on the interaction terms to identify high-risk groups. Our proposed model is motivated by a meta-analysis of adverse events in cancer immunotherapy, and our results uncover key factors driving the risk of specific types of treatment-related adverse events.

KEYWORDS

censored data, horseshoe prior, interaction selection, rare events, subgroup analysis

Highlights

What is already known?

Meta-analysis is a widely-used and powerful tool to synthesize evidence on the safety of a medical treatment. Identification of subgroups at high risk of treatment-related toxicities is critical to informing treatment recommendations. However, existing meta-analysis methods may not be appropriate to address this question due to several challenges: the rarity of many adverse events, incomplete reporting of adverse event data, and the large number of potential risk factors for different types of adverse events.

What is new?

In this paper, we present a Bayesian hierarchical model that incorporates a horseshoe prior on the interaction terms in a multivariable regression to identify high-risk groups. Importantly, our model framework can handle both rare and censored events. Our simulations show that our proposed approach can accurately identify interactions corresponding to subgroups at elevated risk in the presence of both sparsity and censoring of the outcomes. We also apply our method to the meta-analysis of real clinical trial data, providing new insights on potential risk factors for toxicities of PD-1/PD-L1 inhibitors, a popular class of immunotherapy drugs used to treat cancer.

Potential impact for RSM readers

The method proposed in this study can be applied to meta-analyses of safety data in other disease settings in the future. To facilitate the dissemination of the method, we have provided the code to implement the model in JAGS, a free and easy-to-use platform for Bayesian modeling. The application of our method to accurately identify patient subgroups at increased risk of adverse events can potentially guide monitoring and prevention strategies to minimize the impact of treatment-related toxicities.

1 | INTRODUCTION

For both the approval of new treatments and postmarketing surveillance of existing medications, it is important to characterize drug safety profiles, but individual randomized controlled trials, which are typically powered to evaluate efficacy endpoints, are often too small to reliably estimate risks. In this setting, meta-analysis, which is an efficient and powerful way to pool evidence across studies, becomes a natural approach. Special attention has recently been given to the use of meta-analysis to assess drug safety[1]; in particular, the FDA has issued draft guidance on meta-analyses of adverse events related to the administration of a medication.[2]

Meta-analyses of clinical trials may focus on efficacy or safety; each pose unique methodological challenges.[3,4] However, combining evidence on adverse events is more difficult than combining evidence on efficacy.[5] This is partly because adverse events of interest may be rare or incompletely reported.[6,7] In particular, clinical trials may only report events with counts higher than a prespecified cutoff; in this case, the counts for infrequent events below the threshold can be considered as left-censored data.

To date, the relevant literature on statistical methods for meta-analysis of adverse events has mostly focused on the challenge of their rarity. Traditional meta-analysis methods for binary outcomes, in particular, the fixed effects model[8] and the random effects model,[9] often yield biased estimates for rare events.[10] Although both exact and asymptotic methods have been proposed to address this challenge,[10–12] these methods are only suitable for completely observed data. The Bayesian framework offers an attractive alternative for meta-analysis[13,14] as Bayesian methods can naturally handle heterogeneity across studies, rare events,[11] and censored outcomes.[15]

However, these existing methods fail to address a critical challenge in safety meta-analysis: unlike efficacy endpoints, which are limited in number, adverse events may be classified into a large number of categories, and risk may differ by disease status (e.g., cancer type) and intervention (e.g., type of drug, duration, and dosing) across these different categories. It is crucial and of overwhelming clinical interest to identify high-risk subgroups for toxicity monitoring. Critically, this setting poses unique challenges for statistical modeling because of the low frequency of the observed outcome variables and the large number of potential risk factors for different types of adverse events. We address this problem using the Bayesian meta-analysis framework, incorporating sparsity-inducing priors to identify pairwise interactions among study-level covariates. The statistical novelty of our approach lies in how we bring Bayesian variable selection methods to bear on the challenge of identifying groups at high risk for specific classes of toxicities, which requires careful consideration of the model specification. To enable the application of our method to future studies, we provide an easy-to-use JAGS implementation of our proposed Bayesian model.

The rest of this article is organized as follows. We begin with a discussion of our motivating data set on the safety of immunotherapy drugs in Section 2. In Section 3, we propose a novel Bayesian method for meta-analysis of safety data, which allows for the identification of high-risk subgroups through the selection of a sparse set of interaction terms. In Section 4, we compare the performance of the proposed approach to alternative methods using simulated data with missing and rare outcomes; we demonstrate that we can achieve good selection performance and accurate risk prediction in this context. In Section 5, we apply the proposed method to our motivating data set on meta-analysis of immunotherapy trials, identifying high-risk groups by selecting a set of key interaction terms. Some concluding remarks and discussion are provided in Section 6.

2 | MOTIVATING APPLICATION TO META-ANALYSIS OF IMMUNOTHERAPY SAFETY

Immunotherapy is a type of cancer treatment that boosts the body's natural immune defenses to kill tumor cells. In developing our proposed model, we are particularly motivated by the challenge of characterizing the safety of immunotherapy drugs; they are generally much less toxic

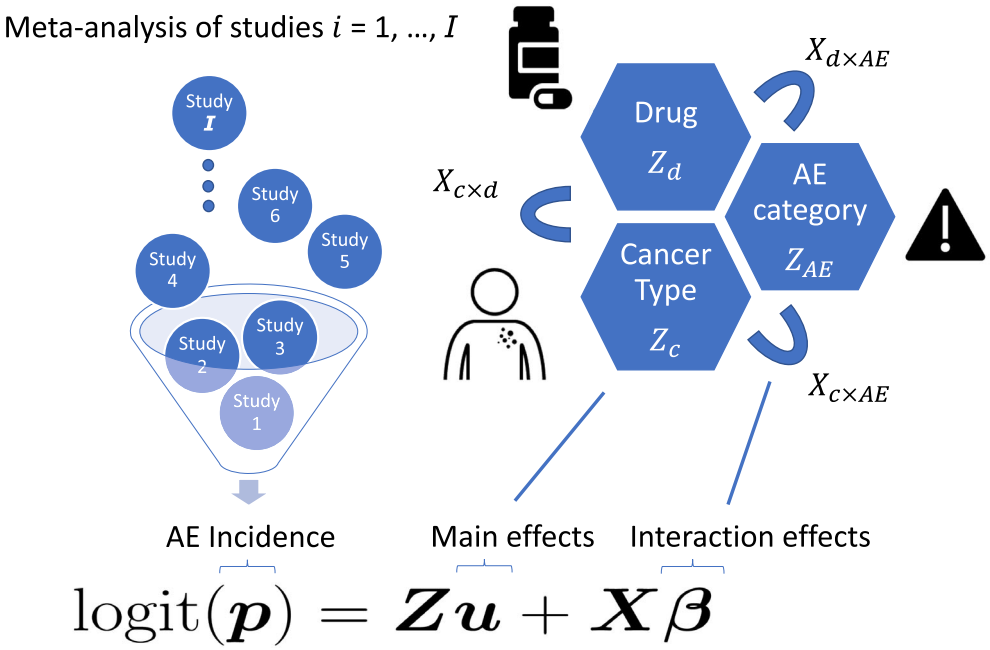

FIGURE 1 Illustrative schema of sparse meta-analysis of adverse events (AEs) with interaction selection [Colour figure can be viewed at wileyonlinelibrary.com]

than traditional cancer treatments, such as chemotherapy, but they can occasionally trigger life-threatening inflammatory and immune-related adverse events.[16] Our motivating data set is from Reference 17, who examined the incidence of adverse events associated with PD-1 and PD-L1 inhibitors, two popular classes of immunotherapy drugs, across 125 studies reported in the literature. Some of the most frequent adverse events in this data set include fatigue, diarrhea, rash, and nausea. Life-threatening adverse events such as pneumonitis and cardiovascular failure were rare, but did lead to 82 treatment-related deaths across all studies. Importantly, this meta-analysis included data from studies on a variety of cancer types (including melanoma, lung cancer, and gastrointestinal cancer) that used a variety of drugs (including nivolumab, pembrolizumab, and atezolizumab).[17] estimated the incidence rates of adverse events for different classes of drugs, cancer types, and categories of adverse events; however, they were not able to identify risk factors for specific types of adverse events or the potential interactions of drug and cancer type due to the challenge of multiplicity in considering all such combinations.

This challenge represents a form of subgroup analysis. In the medical literature, such analyses are of interest to identify high-risk subgroups and reveal clinically relevant differences across patient groups, but spurious findings may occur without appropriate adjustment because of multiple subgroups being investigated and therefore multiple hypotheses being tested.[18–20] The importance of providing valid conclusions from subgroup analyses in meta-analysis has been well accepted in the literature,[21] but in practice, the multiple testing issue is seldom addressed.[22] Advanced methods to account for multiplicity in meta-analytic subgroup analysis are still underdeveloped, especially in the presence of high-dimensional interactions. This gap in the literature motivates our current work.

To illustrate our motivating application, we provide a schematic diagram in Figure 1 which depicts the problem of identifying the potential two-way interactions among drug, cancer type, and category of adverse event. While the primary goal of incorporating interaction effects is to identify subgroups of patients at high risk for specific classes of adverse events, estimating these interaction effects can also enable more accurate risk prediction. This knowledge is crucial for precision medicine to inform efforts for early detection and close monitoring of symptoms and enable proper management of adverse events.[23] In the current work, we incorporate selection of interaction terms using sparsity-inducing priors within the framework of Bayesian meta-analysis. We discuss our proposed model in more detail in the following section. To make the model setup more concrete, we provide the first few lines of the case study data in Table 1 as an illustration. We also relate the columns in this data set to our notation, which will be introduced as a part of the model specification in Section 3.1.

3 | METHODS

In this section, we describe our proposed modeling approach for meta-analysis of safety events. A key goal of our method is to identify subgroups of patients who are at high risk for certain types of adverse events. We achieve this through the selection of interaction terms

TABLE 1 An illustrative snippet of the case study data described in Section 2

Study	N	Type of adverse event	Drug	Cancer	Ypye of adverse event × cancer	Type of adverse event × drug	Cancer × drug	Cutoff	Y
1	131	1	1	1	1	1	1	6	14
1	131	2	1	1	2	2	1	6	-
1	131	3	1	1	3	3	1	6	12
1	131	4	1	1	4	4	1	6	10
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2	154	1	2	2	2	2	4	15	-
2	154	2	2	2	4	4	4	15	-
2	154	3	2	2	6	6	4	15	-
2	154	4	2	2	8	8	4	15	22
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	N_i	j	$Z_{ij,1}$	$Z_{ij,2}$	$X_{ij,1}$	$X_{ij,2}$	$X_{ij,3}$	c_i	Y_{ij}

Note: The index i represents the study, N_i denotes the sample size, j indexes the type of adverse event, $Z_{ij,k}$ denotes the column of the design matrix $\boldsymbol{Z}$ corresponding to the effect of covariate k on adverse event j in study i, $X_{ij,k}$ denotes the column of the design matrix $\boldsymbol{X}$ corresponding to the interaction effect of interaction k on adverse event j in study i, c_i denotes the study-specific censoring threshold, and Y_{ij} denotes the true number of adverse events of type j in study i. A hyphen is used to represent counts that were subject to censoring.

within a Bayesian modeling framework that can handle both rare and censored events, as described in the following subsections.

3.1 | Model specification

We consider a meta-analysis setting in which we would like to aggregate evidence regarding drug safety across a set of I studies. The toxicity outcomes for the ith study, where $i = 1, ..., I$, consist of a vector of counts $\boldsymbol{Y}_i = (Y_{i1}, ..., Y_{iJ})$, where Y_{ij} denotes the observed number of toxicities in the jth category. We assume that these counts arise from the binomial distribution

$$Y_{ij} \sim \text{Binomial}\Big(N_i, p_{ij}\Big), \tag{1}$$

where N_i represents the total number of subjects included in the safety analysis for the ith study, and p_{ij} represents the study-specific probability of the jth type of adverse event. By directly modeling the study-level counts, we are able to handle rare events and potential left-censoring, as discussed below.

3.1.1 | Incorporation of covariates

The key innovation of our proposed model lies in how we incorporate sparsity in linking the probability of an adverse event to study-level covariates. We are interested in determining the effects of factors such as drug, cancer type, and adverse event category on the probability of toxicity. To do this, we rely on a logistic regression model linking the probability p_{ij} of an adverse event of type j in study i to study-level factors. Importantly, these factors may interact with each other. For example, the effect of a drug on the incidence of an adverse event may also depend on the type of cancer being treated. Therefore, our proposed Bayesian meta-analysis model includes additive effects for not only study-level covariates, but also for the set of candidate two-way interactions, such as cancer × drug, cancer × type of adverse event, and drug × type of adverse event. These interaction terms are subject to selection, which enables us to identify combinations that are potentially more dangerous than expected on the basis of modeling only the main effects.

Our proposed Bayesian approach can be seen as a random-effects model,[24] as it allows us to decompose heterogeneity across studies into components attributed to the covariates included in the model, such as cancer type or study drug, and residual heterogeneity due to study-specific differences in patient population or clinical practice across institutions. Both the covariate and study effects are treated as Bayesian parameters, with an appropriate choice of prior distribution. In our hierarchical model, we additionally place a prior on the standard deviations of these distributions. We are then able to characterize heterogeneity across studies using the corresponding posterior distribution for these quantities.

More formally, consider the logistic regression model including pairwise interaction effects:

$$\text{logit}(\boldsymbol{p}) = \boldsymbol{Zu} + \boldsymbol{X\beta}, \tag{2}$$

where $\boldsymbol{p}$ is a vector of length $I \times J$ with entries corresponding to the probabilities of adverse event incidence p_{ij} in equation (1), $\boldsymbol{Z}$ 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. The coefficient vector $\boldsymbol{u}$ is a Q-vector of random marginal effects such as those for study, therapeutic regimen, cancer type, and category of adverse event. The coefficient vector $\boldsymbol{\beta}$ is a K-vector of interaction effects.

Our goal is to both estimate $\boldsymbol{u}$ and identify a subset of interaction terms to include in the model. In the next section, we describe our choice of prior on the parameters $\boldsymbol{u}$ and $\boldsymbol{\beta}$ that allows us to achieve this goal within a Bayesian modeling framework.

3.1.2 | Prior on main effects

We choose to keep all Q random marginal (main) effects in the model, as the study-level covariates of interest (cancer type, drug, and class of adverse event) are all clinically meaningful. To specify the prior on these marginal effects $\boldsymbol{u} = (u_1, ..., u_Q)$, we rely on the normal prior:

$$\boldsymbol{u} \mid \Sigma_u \sim \mathcal{N}(\boldsymbol{\mu}_0, \Sigma_u). \tag{3}$$

We follow the common choice of setting the prior mean $\boldsymbol{\mu}_0$ to be $\mathbf{0}$. The matrix Σ_u is a block diagonal matrix with four main-diagonal blocks such that each block (corresponding to a categorical variable) shares the same variance, $\left\{\sigma_u^{(b)}\right\}^2$, where $b = 1, ..., 4$. The superscript b stands for a categorical variable corresponding to that block. For the choice of prior on the standard deviations $\sigma_u^{(b)}$, we follow the recommendation in Reference 25, and use a weakly-informative prior: $\sigma_u^{(b)} \sim C^+(0, A)$, where C^+ denotes the half-Cauchy distribution, and A is a scale parameter set to 25.

3.1.3 | Prior on interaction effects

As discussed above, we would like to achieve sparsity in estimating the interaction effects. For this purpose, we adopt conditionally independent horseshoe priors[26,27] on the interaction terms. We rely on the horseshoe prior as it has been shown to have nice theoretical properties,[28,29] and has been widely adopted in applications such as machine learning.[30]

We now discuss the formulation of this approach in more detail, as this represents novel aspect of our meta-analysis model. The horseshoe prior on each interaction coefficient β_k can be represented as the following scale mixture of normals:

$$\beta_k \mid \lambda_k, \tau \sim \mathcal{N}\left(0, \lambda_k^2 \tau^2\right) \tag{4}$$

$$\lambda_k \sim C^+(0, 1),$$

where λ_k is a *local* shrinkage parameter, and τ is an overall *global* shrinkage parameter. The global parameter τ is responsible for shrinking all of the coefficients toward zero, while the local parameter λ_k, through its heavy tails, allows true signals to escape this shrinkage.

We now briefly review the intuition behind the horseshoe prior. Carvalho et al[26] introduced this prior in the simple setting where $\boldsymbol{\mu}$ represents a vector of normal means, and the observed data follow the distribution $(\mathbf{y}|\boldsymbol{\mu}) \sim (\boldsymbol{\mu}, \sigma^2 I)$, where I is the identity matrix. They then assume $\sigma^2 = \tau^2 = 1$, define

$$\kappa_k = \frac{1}{1 + \lambda_k^2} \tag{5}$$

as the "shrinkage factor" corresponding to μ_k, and show that $E(\mu_k|y) = \{1 - E(\kappa_k|y_k)\}y$. The name "horseshoe prior" is based on the fact that the prior induced on κ_k resembles a U-shape, with very little shrinkage on strong signals (corresponding to the peak when κ_k is close to 0) and almost total shrinkage on true zeros (corresponding to the peak when κ_k is close to 1).

The global shrinkage parameter τ enables adaptivity to the overall sparsity level. As $\tau \to \infty$, the priors for all variables become diffuse, with little shrinkage, and as $\tau \to 0$, all priors strongly favor values of β_k close to 0. As recommended in Reference 26, we assume a half-Cauchy prior

$$\tau \sim C^+\left(0, \tau_0^2\right), \tag{6}$$

with the fixed hyperparameter τ_0^2 set to 1. While subsequent work has further explored the choice of hyperprior on the global shrinkage parameter,[31] we found that the standard half-Cauchy density proposed in Reference 26 worked well in both our simulations and our real data analysis. We performed a sensitivity analysis to assess the impact of varying τ_0 on posterior inference, as discussed in the Supplementary Material, and found that our results were not overly sensitive to the choice of τ_0.

3.1.4 | Handling of censored observations

A source of reporting bias in meta-analyses of toxicity outcomes is that events with very low frequency (below a pre-specified study-specific threshold) may be omitted from the final trial report. Since these counts are missing not at random, ignoring the missing values will result in bias in the estimation of the incidence probabilities. Therefore, we take censored observations into account through the likelihood function, following the approach proposed in Reference 15. We let $\boldsymbol{\delta}_i = (\delta_{i1}, ..., \delta_{iJ})$ serve as a vector of censoring indicators, where $\delta_{ij} = 1$ if the count for the jth toxicity outcome in the ith study is fully observed and $\delta_{ij} = 0$ if the observation is censored. Taking advantage of the study-specific cutoff values in handling the missing events, we consider the missing observations as left-censored data, with the unobserved number of toxicities for the ith study taking a value in the range from 0 to the censoring threshold c_i. We therefore rely on the cumulative binomial distribution, denoted as $F\left(c_i; N_i, p_{ij}\right)$, to model the incidence probabilities for the left-censored adverse events. The full likelihood for both the fully observed data and left-censored data can then be written:

$$\begin{aligned}\mathscr{L} &= \prod_{i=1}^{I}\prod_{j=1}^{J}\left[f_Y\left(y_{ij}\right)\right]^{\delta_{ij}=1}\left[F_Y(c_i)\right]^{\delta_{ij}=0} \\ &= \prod_{i=1}^{I}\prod_{j=1}^{J}\left[f_Y\left(y_{ij}\right)\right]^{\delta_{ij}=1}\left[\sum_{k_i=0}^{c_i} f_Y(k_i)\right]^{\delta_{ij}=0}\end{aligned}$$

where $f_Y(y)$ represents the probability density function of Y, and $F_Y(y) = P[Y \le y]$ represents the cumulative distribution function of Y. If the censoring indicator $\delta_{ij} = 1$, then the jth toxicity outcome in the ith study was observed. If $\delta_{ij} = 0$, then the toxicity outcome was censored; the corresponding term in the likelihood reflects the information contained in the censoring that its value was a count less than or equal to the reporting threshold c_i.

3.2 | Model implementation in JAGS

To model the incidence probabilities for left-censored adverse events, we construct the exact likelihood function for censored observations through ancillary variables $W_{ij} = 1$ following the Bernoulli distribution

$$W_{ij} \sim \text{Bernoulli}\left(q_{ij}\right). \tag{7}$$

The Bernoulli probability $q_{ij} = F\left(c_i; N_i, p_{ij}\right) = \sum_{k_{ij}=0}^{c_{ij}} \binom{N_{ij}}{k_{ij}} p_{ij}^{k_{ij}} \left(1 - p_{ij}\right)^{N_{ij}-k_{ij}}$ represents the study-specific probability of left censoring for the jth type of adverse event in the ith study. As noted in Reference 32, this formulation results in the correct likelihood and deviance functions in JAGS for Bayesian modeling with left-censored observations.

Given the likelihood and priors defined above, we consider the joint posterior

$$\begin{aligned}p(\boldsymbol{u}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma_u, \tau | \boldsymbol{Y}_i, \boldsymbol{W}_i) \propto \prod_{i=1}^{I} p(\boldsymbol{Y}_i, \boldsymbol{W}_i | \boldsymbol{u}, \boldsymbol{\beta}) \\ \prod_{q=1}^{Q} p(u_q | \sigma_u) \prod_{k=1}^{K} [p(\beta_k | \lambda_k, \tau) p(\lambda_k)] p(\sigma_u) p(\tau),\end{aligned}$$

where $\boldsymbol{W}_i = (W_{i1}, ..., W_{iJ})$ and $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_K)$. Since this posterior is not tractable, meaning that there is no closed form for the full conditional distributions, we rely on Markov chain Monte Carlo (MCMC) to obtain a sample from the posterior. Here we utilize Just another Gibbs Sampling (JAGS), which enables MCMC sampling for Bayesian hierarchical models on the basis of the model specification.[33,34] The full specification for the proposed model in JAGS is provided in the Supplementary Material. The simulation and case study results provided in Sections 4 and 5 were obtained using this JAGS implementation.

4 | SIMULATION STUDIES

In this section, we compare the performance of our proposed approach to that of existing meta-analysis methods on simulated data and discuss sensitivity to the choice of prior hyperparameters.

4.1 | Simulation set-up

4.1.1 | Data generation

For each simulation scenario, we generate 200 simulated data sets, with 100 used for training and 100 used for testing. Each data set includes $n = 1000$ observations, corresponding to $I \times J = 1000$ potentially observed counts y_{ij} across $I = 100$ studies and $J = 10$ types of adverse events. To generate the simulation truth for $\boldsymbol{Y}_{train}$ and $\boldsymbol{Y}_{test}$, we set the total number of main effects Q to be 125, corresponding to 100 studies, 10 cancer types, 10 categories of adverse events (AEs), and 5 drugs. The total number of

candidate pairwise interactions K is then 200, including 50 AE × drug interactions, 100 AE × cancer interactions, and 50 cancer × drug interactions. We generate a design matrix ($\boldsymbol{Z}$) for the marginal effects, and draw the corresponding coefficients (a vector of length $Q = 125$) from the normal distribution $u_q \sim \mathcal{N}(-1, 0.4^2)$. For the interaction effects, we generate another design matrix ($\boldsymbol{X}$), and construct the corresponding coefficient vector $\boldsymbol{\beta}$ by assuming that 20 (of 200) are true (non-zero) interactions with a value of 1 ($\beta = 1$) and the rest are zeros, so that the sparsity level is 10%. Given these inputs, the true outcome $\boldsymbol{Y}_{train}$ for each simulated data set is generated from a binomial distribution with the number of patients within each study as $n_i = 100$ and the toxicity probability set to $p = \text{logit}^{-1}(\boldsymbol{Zu} + \boldsymbol{X\beta})$. The test set $\boldsymbol{Y}_{test}$ is generated from a binomial distribution using the same parameters. We generate a separate test set to enable comparison of predictive performance across methods.

We consider three simulation scenarios with increasing percentages of censored observations. In Scenario 1, all adverse events are fully observed. In Scenario 2, 40% of the observations are censored, while in Scenario 3, 80% of the observations are censored. In the second and third scenarios, where there is missing data, the missingness is designed to follow the real-world pattern, where adverse events with zero or low incidence are censored. This represents informative censoring, since the missing data correspond to counts that fall beneath a specific threshold.

4.1.2 | Methods compared

We include the following methods in our performance comparison:

- *BMCD*: Bayesian model for censored data, which is a marginal model without interaction terms, as described in Reference 17.
- *sBMI*: sparse Bayesian Model with Interaction selection, that is, our proposed Bayesian model for censored data with selection of interaction terms using a horseshoe prior.
- *glmIA*: a classical generalized linear regression model with a logit link that includes all main effects and true interactions.
- *glm*: a logistic regression model with all main effects included, but no interaction terms
- *glmnet*: a LASSO logistic regression model with all main effects included, and interaction selection using an l_1 penalty.[35]

Importantly, the only existing method, which allows for the selection of interactions is glmnet: the remaining alternatives either include no interaction terms or rely on knowledge of the ground truth in including interactions. We also note that due to the large number of candidate interactions, it is not possible to fit a dense model including all potential interaction terms.

Both Bayesian models are implemented in JAGS. In BMCD (the Bayesian model with main effects only), we estimate the 200 main study-level factors in the marginal model in JAGS, and obtain the estimated marginal effects ($\widehat{\boldsymbol{u}}$) using the MCMC results. In sBMI, we fit a model that includes both the main study-level factors and candidate two-way interactions and obtain the posterior samples of both the marginal effects ($\widehat{\boldsymbol{u}}$) and the interaction terms ($\widehat{\boldsymbol{\beta}}$). For model fitting in JAGS, we discard the first 30,000 iterations as burn-in and retain 30,000 posterior samples after the burn-in to use as the basis for posterior inference, for example, estimation of the posterior median and 95% credible intervals. The glmIA model represents the best possible performance for a classical generalized linear model, as knowledge of the ground truth was used to include only the true interactions. The final two methods included in the comparison, glm and glmnet, represent realistic approaches in the frequentist framework.

On the basis of the fitted models, we obtain predictions $\widehat{\boldsymbol{Y}}$ for the test sets as follows. For models with main effects only, the test predictions are generated from a binomial distribution with sample size n_i and the estimated incidence probability $\widehat{p} = \text{logit}^{-1}(\boldsymbol{Z}\widehat{\boldsymbol{u}})$, while for models with both main effects and interactions, we take $\widehat{p} = \text{logit}^{-1}\left(\boldsymbol{Z}\widehat{\boldsymbol{u}} + \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)$. For the Bayesian methods, $\widehat{\boldsymbol{u}}$ is taken to be the medians of the posterior samples for the marginal effects. For sBMI, $\widehat{\boldsymbol{\beta}}$ is similarly taken as the posterior medians of the sampled values. For model selection, we adopt the approach of Reference 36 and consider an interaction term β_k as "selected" if its marginal 95% credible interval does not contain 0.

4.1.3 | Performance metrics

We assess model performance in terms of prediction error on the test data and accuracy in the selection of the model coefficients. More specifically, we compare the five methods using the following metrics:

- *MSPE*: mean squared prediction error, the mean squared difference between the true value of the outcome (Y) from the test dataset and the fitted value ($\widehat{Y}$) from the model.

$$\text{MSPE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2$$

- *TPR*: true positive rate, sensitivity in identifying the true nonzero interaction terms

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

- *FPR*: false positive rate, the proportion of true zero interactions selected

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

- *MCC*: the Matthews correlation coefficient, a measure of the overall interaction selection accuracy

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

In the formulas above, *TP* (true positive) represents the number of actual non-zero interactions that are selected, *TN* (true negative) is the number of actual zeros that are not selected, *FP* (false positive) is the number of actual zeros that are selected, and *FN* (false negative) is the number of non-zeros that are not selected. We take the MCC to be zero (rather than undefined) if any term in the denominator is zero.

4.2 | Simulation results

We illustrate the results of our proposed method and provide an in-depth comparison of its performance with alternative methods. To provide insight into the estimated interaction effects obtained using the horseshoe prior, we visualize the sBMI modeling results for a single simulated data set by depicting the 95% posterior credible intervals for the full set of candidate interaction terms in the simpler setting with no censoring (top of Figure 2) and the most challenging setting with 80% censoring (bottom of Figure 2). In these figures, we show the set of 200 interaction terms along the x-axis. Under the simulation truth, there are 20 true nonzero interaction effects (plotted in green) that are equal to 1, while the remaining 180 interactions (plotted in red) are truly 0. The vertical orange lines correspond to the 95% posterior credible intervals. True positive selections correspond to true signals with 95% credible intervals for the corresponding interaction coefficient (β) that do not overlap zero; false negatives arise when the 95% credible interval covers zero. Similarly, for actual zero interaction terms, if the 95% posterior credible interval covers 0, this represents a true negative, while if the 95% credible does not cover 0, that term is a false positive. 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). When the rate of missingness increases, the 95% credible intervals for the interaction coefficients become wider, but the proposed sBMI approach can still recover the underlying truth at TPR = 0.65, with FPR = 0 in Scenario 3 on the basis of a single simulated dataset.

We provide results comparing the performance of the five methods across 100 simulated data sets, summarized in Table 2. In this table, we provide the mean MSPE obtained by averaging over all iterations, along with the standard error (SE). We also report the TPR, FPR, and MCC for sBMI and glmnet, which are the only methods that perform selection. glmnet * results reflect averages over the simulated data sets where glmnet did not suffer convergence issues. The median of the true values of Y over 100 simulated datasets is 2.290. As shown in Table 2, the proposed method sBMI has the best predictive performance across the range of censoring scenarios considered. When the data are fully observed, sBMI, glmIA, and glmnet (the methods that estimate interaction terms) have the best predictive accuracy, while BMCD and glm (the Bayesian and frequentist methods which only include the main effects) perform poorly. This demonstrates the importance of including interaction terms in the model. The performance of BMCD (main-effect only) and sBMI is relatively stable when the percentage of missingness increases, demonstrating the benefits of handling censored rare event outcomes under the Bayesian framework. Such robustness results from the ability to accommodate the censored outcomes in the likelihood function. For interaction identification, sBMI has an FPR that is very close to zero in all scenarios, which indicates that very few zero interactions are incorrectly selected. Also, sBMI has the highest TPR (over 0.7) in Scenario 1. In the case when 80% of outcomes are censored (Scenario 3), TPRs over 0.55 are still achieved without suffering high FPRs in both Bayesian interaction

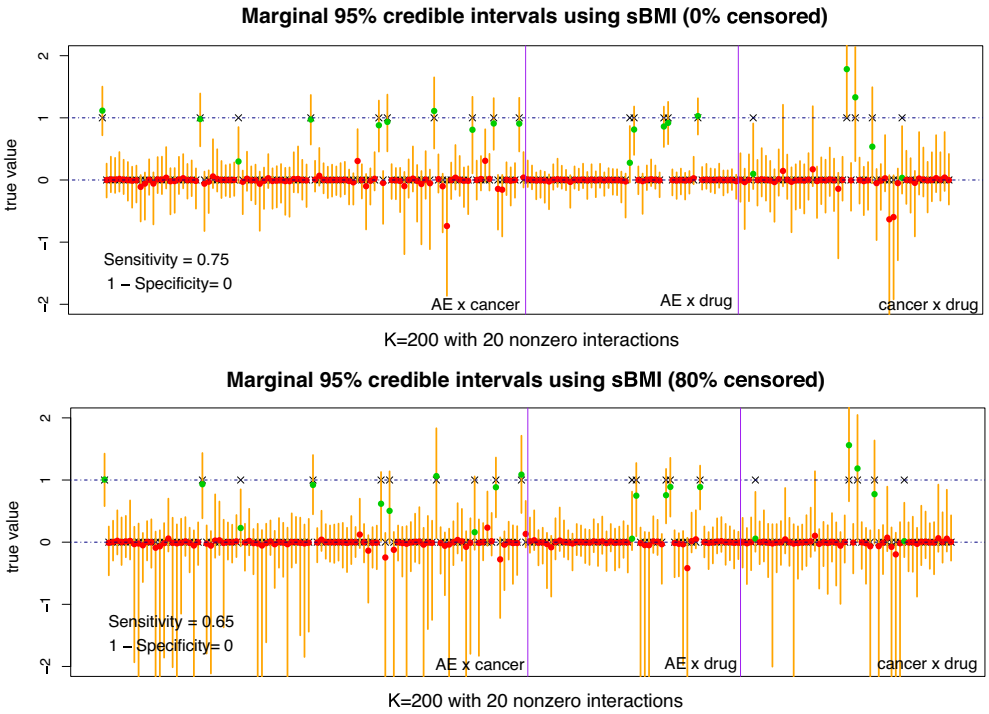

FIGURE 2 95% marginal credible intervals for the interaction coefficients obtained using sBMI on a single simulated fully-observed data set (top) and a data set with 80% censoring (bottom), in a setting with $n_{IA} = 200$ and 20 true non-zero interactions [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Performance comparison described in Section 4.2

Scenario	Method	Median ($\widehat{Y}$)	MSPE (SE)	TPR	FPR	MCC
0% censored	BMCD	2.595	11.473 (2.954)	-	-	-
	sBMI	2.280	7.314 (1.428)	0.708	6e-5	0.827
	glmIA	2.210	7.849 (1.817)	-	-	-
	glm	2.410	11.717 (3.244)	-	-	-
	glmnet*	2.319	7.502 (1.529)	0.736	0.251	0.327
40% censored	BMCD	2.750	11.618 (2.939)	-	-	-
	sBMI	2.495	7.418 (1.463)	0.687	6e-5	0.813
	glmIA	4.175	41.799 (14.087)	-	-	-
	glm	4.485	37.423 (12.793)	-	-	-
	glmnet*	4.380	11.633 (2.336)	0.721	0.270	0.296
80% censored	BMCD	3.095	12.318 (3.243)	-	-	-
	sBMI	2.365	7.766 (1.607)	0.553	6e-5	0.723
	glmIA	8.180	73.471 (24.081)	-	-	-
	glm	8.545	68.971 (22.119)	-	-	-
	glmnet*	9.169	24.880 (4.916)	0.444	0.205	0.196

Note: Results are summarized by median estimated Y, average mean squared prediction error (MSPE) on test data with standard error (SE), true positive rate (TPR), false positive rate (FPR), and Matthews correlation coefficient (MCC) for five methods: Bayesian marginal-only model (BMCD), sparse Bayesian model with interaction selection (sBMI), logistic regression with all true interactions (glmIA), logistic regression without interactions (glm), and generalized linear model with lasso regularization (glmnet), under 0%, 40%, and 80% censoring. glmnet* results reflect averages over the simulated data sets where glmnet did not suffer convergence issues. The median of the true values of Y over 100 simulated datasets is 2.290.

selection models. In all three scenarios, sBMI achieves a much higher MCC, with values of 0.83, 0.81 and 0.72, respectively.

The non-Bayesian logistic regression approaches have some key limitations that are evident in our simulation results. Compared to the proposed model, glmIA, which

includes all true nonzero interactions, has similar prediction accuracy on the basis of the value of MSPE in Scenario 1; however, in the presence of censoring (Scenarios 2 and 3), its prediction accuracy worsens, since only observed cases are included in the model. Furthermore, glmIA relies on knowledge of the truth, and therefore does not represent a practical modeling approach. The penalized logistic regression approach, glmnet, faced performance challenges in the interaction selection because of the presence of multiple high-dimensional categorical variables in the model. In particular, some simulated datasets resulted in convergence issues when fitting the regression model with glmnet in R. Therefore, the results from glmnet* presented in Table 2 are summarized over 83 of the 100 simulated data sets that did not result in a convergence issue. Moreover, the MCC for interaction selection was low in all three scenarios.

Each simulation study was performed in JAGS using R version 4.1.0[37] on a single core of a computing cluster. For a simulated dataset with 125 main effects and 200 candidate interactions (including 20 true interaction effects) and a sample size of 100 trials, the proposed model runs in 35 min in the 40% censoring scenario and 40 min in the 80% censoring scenario, where only 200 out of $n = I \times J = 1000$ AE counts were observed. We also considered how the number of candidate interactions relative to the sample size influences the accuracy of the results. In the simulation scenario with 80% censoring, the ratio of the number of interactions versus the total number of observations is $\frac{K}{I \times J} = 0.2$, and the proposed method achieves a TPR of 0.553. We examined the impact on performance of varying this ratio. We found that accuracy increases dramatically when the ratio is lower than 0.2, while higher ratios are more challenging. In particular, we observed that the TPR decreases to 0.4 if the ratio is increased to 0.3.

5 | RESULTS FROM THE CASE STUDY

In this section, we utilize the proposed sparse Bayesian interaction selection method to identify key interaction effects in our motivating data set, the immunotherapy meta-analysis data considered in Reference 17. We introduced this data set at a high level in Section 1, and provide more detail here. The study-level safety data were collected from a systematic review of 125 clinical studies that reported 75 categories of adverse events for two anti-PD-1 drugs (nivolumab and pembrolizumab) and three anti-PD-L1 drugs (atezolizumab, avelumab, and durvalumab) published from 2011 to 2018. The majority of these studies were early-phase single-arm studies. Other than the study-level adverse event incidence data, we have the following study-level information: drug, type of adverse event, cancer type, and adverse event reporting criteria. In each study, the numbers of all-grade (G15, grades 1–5) and grade 3 or higher (G35, grades 3–5) adverse events were recorded; however, a mean of more than 60% of the adverse events for each study were left-censored, as the count for that adverse event type was lower than a prespecified cutoff value reported from each study. This threshold is captured as a part of the standard data elements in clinical trial reporting and publication, and varies from study to study. At ClinicalTrials.gov, the frequency threshold has an allowed maximum of 5%.[38] As an example, Study 1 in Table 1 corresponds to the CheckMate 017 trial,[39] which published the observed counts for all treatment-related adverse events that were reported in at least 5% of subjects. Given their sample size of 131, this means that adverse event types with counts ≤ 6 were not reported. Similar thresholds were applied in other studies, resulting in a substantial amount of censored data. Therefore, we treat the censored outcomes as censored data, following the approach described in Section 3. The objective is to evaluate the adverse event incidence probability by subgroup and identify the high-risk groups by selecting the nonzero interactions.

To perform inference, we applied the sBMI approach, using the JAGS implementation described in Section 3.2. We ran three parallel chains. For each MCMC chain, after discarding the burn-in period of 30,000 iterations, the three chains showed good mixing and successful convergence to the target distribution. We eventually obtained 10,000 posterior samples per chain by retaining one sample of three. We pooled these across the three chains to obtain 30,000 posterior samples as the basis for inference. To assess the convergence of the MCMC chains, we computed the potential scale reduction factor $\widehat{R}$. The mean $\widehat{R}$ value across all parameters in the G35 model was 1.04, and the mean for the G15 model was 1.02. These values are close to 1, indicating good convergence of the chains.[40]

The top panel of Figure 3 displays the posterior medians and the corresponding 95% credible intervals of the interaction coefficients selected by sBMI, which correspond to 32 interaction terms from a total of 927 considered. This highlights a key advantage of the sparsity-inducing prior formulation, which allows us to focus on a small set of interaction terms with strongest support from the observed data. As in the simulation study, we consider an interaction term to be selected if its 95% posterior credible interval (CrI) does not overlap 0. Among the 22 selected type of adverse event × cancer type interactions, the vitiligo × melanoma interaction has the largest effect ($\widehat{\beta} = 2.766$; 95% CrI: 1.623–4.075) on G15 adverse

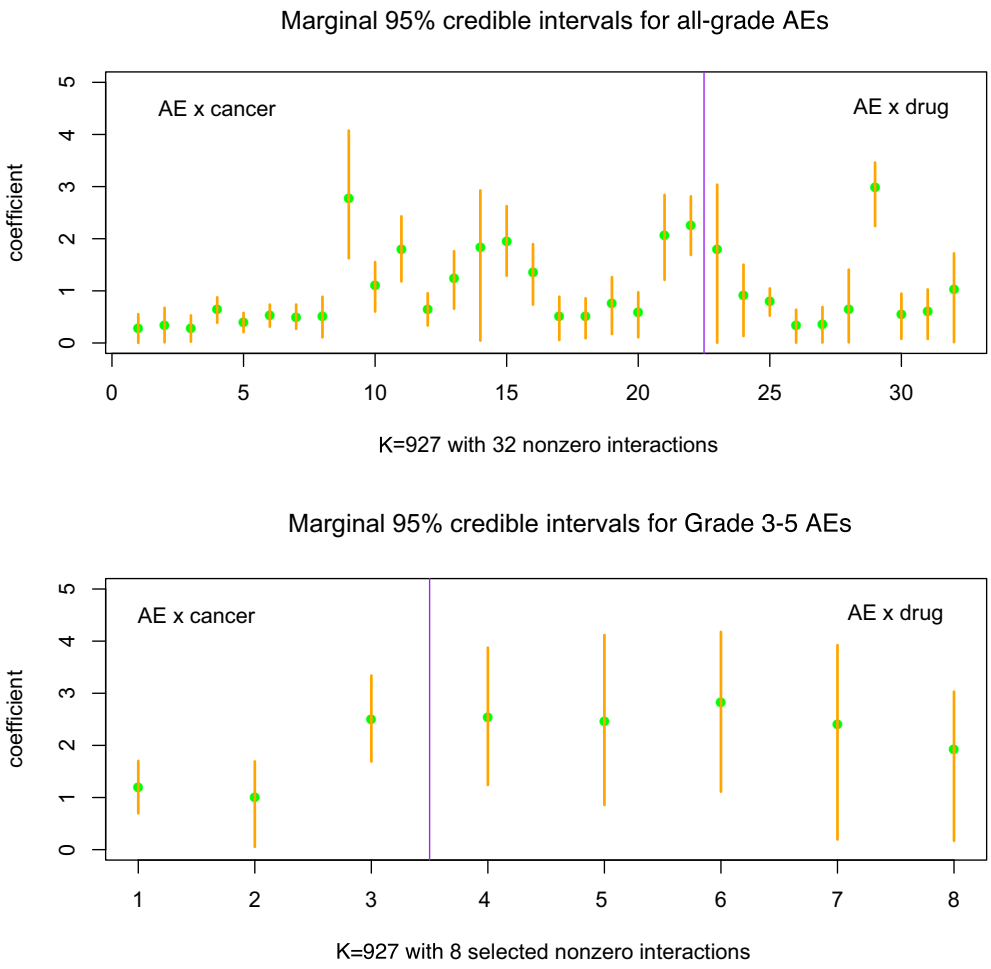

FIGURE 3 Selected interactions from sBMI for G15 adverse events (top) and for G35 adverse events (bottom) [Colour figure can be viewed at wileyonlinelibrary.com]

event incidence, with an estimated incidence probability of $\widehat{p}^{G15} = 0.167$ (95% CrI: 0.060–0.421). It is followed by the neutropenia × hematologic malignancy interaction ($\widehat{\beta} = 2.252$; 95% CrI: 1.688–2.814), with $\widehat{p}^{G15} = 0.106$ (95% CrI: 0.062–0.175), and the platelet count decreased × hematologic malignancy interaction ($\widehat{\beta} = 2.072$; 95% CrI: 1.211–2.841), with $\widehat{p}^{G15} = 0.091$ (95% CrI: 0.040–0.178). Among the 10 selected type of adverse event × drug interactions, the infusion-related reaction × avelumab interaction has the largest effect ($\widehat{\beta} = 2.986$; 95% CrI: 2.243–3.461) on adverse event incidence, with an estimated G15 incidence probability $\widehat{p}^{G15} = 0.178$ (95% CrI: 0.076–0.275), followed by the amylase increased × nivolumab interaction ($\widehat{\beta} = 0.061$; 95% CrI: 0.011–0.191). The overall mean incidence probability of G15 adverse events is 0.011 (95% CrI: 0.005–0.013).

For G35 adverse events, eight pairwise interactions were selected using the sBMI approach. Among the three selected type of adverse event × cancer type interactions, the neutropenia × hematologic malignancy interaction has the largest effect on adverse event incidence, with an estimated G35 adverse event incidence probability of $\widehat{p}^{G35} = 0.009$ (95% CrI: 0.004–0.021), followed by the pneumonitis × lung cancer interaction and the colitis × melanoma interaction. Among the five selected type of adverse event × drug interactions, the infusion-related reaction × avelumab interaction has the largest effect on G35 adverse event incidence, and the estimated G35 adverse event incidence probability is $\widehat{p}^{G35} = 0.012$ (95% CrI: 0.002–0.044), followed by the increased lipase × Nivolumab interaction, the increased amylase × Nivolumab interaction, the increased lipase × Avelumab interaction, and the increased γ-glutamyl transferase × Durvalumab interaction. The overall mean incidence probability of G35 adverse event is 0.0007 (95% CrI: 0.0005–0.0009), which is consistent with the rarity of these severe adverse events.

Our approach allows us to characterize heterogeneity across studies by breaking it down into components attributable to adverse event type, cancer type, drug, and residual heterogeneity due to variation across studies. We report the posterior medians and credible intervals for the corresponding standard deviation terms for the G15 and G35 models in Table 3. These results suggest that the study drug was the leading source of heterogeneity in the risk of adverse events across studies.

TABLE 3 A summary of heterogeneity for the G15 and G35 models. The reported values are the posterior median for the standard deviation terms and the corresponding 95% credible intervals

Model	σ_{AE}	σ_{cancer}	σ_{drug}	σ_{study}
G15	1.12 (0.79–1.64)	0.04 (0.001–11.0)	19.6 (4.42–77.8)	0.41 (0.31–0.56)
G35	1.22 (0.98–1.54)	0.23 (0.01–0.74)	8.43 (4.75–20.1)	0.74 (0.62–0.90)

The strongest main effects obtained from the sBMI analysis are consistent with those reported in Reference 17, a meta-analysis with no selection of interactions: specifically, the top five G15 adverse events found using sBMI are fatigue, diarrhea, nausea, pruritus, and rash, which were reported as well in Reference 17 as the most frequent G15 adverse events types. Another common finding is that the top three G35 adverse events are fatigue, increased aspartate aminotransferase, and anemia. However, as we discuss below, our proposed sBMI model enables additional insight into key interactions that elevate risk.

The interaction that we identified between vitiligo, an autoimmune skin disorder, and melanoma is consistent with results reported in the literature; for example, a prospective study of patients with metastatic melanoma treated with pembrolizumab observed a cumulative incidence of vitiligo of 25%.[41] Low neutrophil counts (neutropenia) and low platelet counts (thrombocytopenia) have both been reported as serious, but rare, hematological immune-related adverse events[42]; thus, 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.[43] Patients with hematological malignancies are more likely to develop cytopenia with cytotoxic chemotherapy because of disease related reduction of bone marrow reserve. The link of neutropenia and thrombocytopenia with hematological malignancies in our analysis is of clinical interest and suggests that these patients are also vulnerable to immunotherapy-induced cytopenia and calls for close monitoring for cytopenia. Thus, our real-data meta-analysis demonstrates that the proposed interaction selection method yields clinically meaningful results. Understanding the risk of various immune-related adverse events on the basis of a subject's cancer type and potential treatment options is key to guiding improved monitoring, prevention and management of these events,[23,44] and the proposed statistical framework can provide robust estimates to inform these decisions.

6 | DISCUSSION

In this work, we developed a sparse Bayesian interaction selection model to simultaneously identify nonzero interactions and high-risk groups with an elevated probability of adverse events, addressing a key challenge in the meta-analysis of safety data. Since our focus is on meta-analysis of safety data, our likelihood is designed to enable the aggregation of single-arm data in clinical studies. Through simulations, we demonstrated that the proposed interaction selection approach can improve prediction accuracy and accurately select non-zero interactions when the underlying truth is sparse and the overall adverse event incidence is rare. We illustrated the proposed approach with a real-data meta-analysis to identify key interaction terms among a high-dimensional set of candidate pairwise interactions; the approach demonstrated good performance, even in settings with a high percentage of censored data. Our simulation results show that the proposed model selection rule has a very low rate of false positive selections; this is consistent with prior theoretical work on the horseshoe prior that showed that selection of effects using marginal credible intervals is conservative, with selected effects tending to be true discoveries.[36]

Our simulation results indicate that the horseshoe prior is an effective approach to achieve sparsity in our setting of interest. In fact, as an alternative to frequentist hypothesis testing for subgroup analyses, a Bayesian hierarchical model is more flexible and straightforward.[45] Penalized regression approaches, such as the lasso, may not result in sufficiently sparse model selection when using standard approaches for penalty parameter selection, such as cross validation.[46] Moreover, we found that the frequentist models suffered from convergence issues in our simulation studies, as 17 of 100 simulated data sets reported a convergence issue when fitting the regression model with glmnet in R.

In the Bayesian literature, the development of sparsity inducing priors remains an active area of theoretical research - in particular, extensions of the horseshoe, including the regularized horseshoe[47] and the horseshoe+,[48] have been proposed in recent years. While the standard horseshoe allows coefficients with true values that are far from zeros to escape from shrinkage, the regularized horseshoe imposes shrinkage on these parameters. This may be useful in settings where the parameters are only weakly identified by the likelihood. In addition, Piironen and Vehtari[47]

demonstrate a link between the shrinkage parameter τ and the effective number of nonzero coefficients in the model. In settings where there is prior knowledge on the number of relevant variables, this can then be used to inform the choice of τ. This approach may be of interest for future investigation when such information is available.

Another avenue of future interest is the inclusion of a selection prior on the main effects in the model. In the context of Bayesian variable selection for regression models, Chipman[49] discussed the challenge of selecting interaction terms, and proposed the strong inheritance principle for interaction selection, which means that the interaction is active only if the corresponding main effects are active. Here, we automatically satisfy this principle by keeping all main effects in the model. An alternative is a weak hierarchy, where an interaction can be selected if at least one of the corresponding main effects are included. Model formulations allowing selection of main effects interactions would be a natural extension of our work. Selecting higher-order interactions (e.g., three-way interactions) is also potentially of interest, but would be more challenging in the current framework due to parameter identifiability issues.

Importantly, the proposed Bayesian method can handle major challenges encountered in meta-analysis of safety data (rare events, incomplete data, and clinical heterogeneity across studies) and enables control of the false positive rate in identification of the true interaction effects. In practice, sparse modeling for interaction effects can be applied alone for meta-analytic subgroup analyses without rare or censored events. Since our method is implemented in JAGS and the code is provided, our proposed model will be a relatively easy tool for researchers to use in future meta-analyses. We hope that future applications of this method will provide clinically useful insights into meta-analysis of adverse events and safety data in other disease settings, and ultimately guide recommendations for treatment and toxicity monitoring.

AUTHOR CONTRIBUTION

Xinyue Qi, Shouhao Zhou, and Christine Peterson conceptualized this research. Xinyue Qi wrote the R and JAGS code, implemented the simulations, carried out the case study, and produced the tables and figures summarizing the results. Yucai Wang contributed to the conceptualization and interpretation of the case study. All authors contributed to writing the manuscript, participated in revisions, and read and reviewed the final manuscript.

ACKNOWLEDGMENT

The authors are grateful for the editing service provided by the Research Medical Library, The University of Texas MD Anderson Cancer Center prior to the initial submission of the manuscript.

FUNDING INFORMATION

This work was supported by NIH/NCI CCSG grant P30CA016672.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The case study data, as well as the R code implementing the method and simulation studies, are available at online at https://github.com/xinyue-qi/Meta-analysis/.

ORCID

Xinyue Qi https://orcid.org/0000-0002-3954-0974
Shouhao Zhou https://orcid.org/0000-0002-8124-5047
Yucai Wang https://orcid.org/0000-0002-1576-8341
Christine Peterson https://orcid.org/0000-0003-3316-0468

REFERENCES

1. Singh S, Loke Y. Drug safety assessment in clinical trials: methodological challenges and opportunities. *Trials*. 2012;13(1):1-8.
2. United States Food and Drug Administration. Meta-analyses of randomized controlled clinical trials to evaluate the safety of human drugs or biological products guidance for industry. Draft Guidance Document; 2018.
3. Sutton A, Cooper N, Lambert P, Jones D, Abrams K, Sweeting M. Meta-analysis of rare and adverse event data. *Expert Rev Pharmacoecon Outcomes Res*. 2002;2(4):367-379.
4. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, et al. *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley; 2019.
5. Council for International Organizations of Medical Sciences. *Evidence Synthesis and Meta-Analysis for Drug Safety: Report of CIOMS Working Group X*. CIOMS. 2016.
6. Berlin J, Crowe B, Whalen E, Xia H, Koro C, Kuebler J. Meta-analysis of clinical trial safety data in a drug development program: answers to frequently asked questions. *Clin Trials*. 2013;10(1):20-31.
7. Stoto M. Drug safety meta-analysis: promises and pitfalls. *Drug Saf*. 2015;38(3):233-243.
8. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst*. 1959;22(4):719-748.
9. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188.
10. Bhaumik D, Amatya A, Normand S, et al. Meta-analysis of rare binary adverse event data. *J Am Stat Assoc*. 2012;107(498):555-567.
11. Cai T, Parast L, Ryan L. Meta-analysis for rare events. *Stat Med*. 2010;29(20):2078-2089.
12. Liu D, Liu R, Xie M. Exact meta-analysis approach for discrete data and its application to 2 × 2 tables with rare events. *J Am Stat Assoc*. 2014;109(508):1450-1465.

13. Smith T, Spiegelhalter D, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat Med*. 1995;14(24):2685-2699.
14. Warn D, Thompson S, Spiegelhalter D. Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales. *Stat Med*. 2002;21(11):1601-1623.
15. Qi X, Zhou S, Peterson C, et al. Meta-analysis of censored adverse events. *arXiv preprint arXiv:2101.07934*. 2021:1-6.
16. Abdel-Wahab N, Alshawa A, Suarez-Almazor M. *Adverse Events in Cancer Immunotherapy*. Springer; 2017:155-174.
17. Wang Y, Zhou S, Yang F, Qi X, Wang X, et al. Treatment-related adverse events of PD-1 and PD-L1 inhibitors in clinical trials: a systematic review and meta-analysis. *JAMA Oncology*. 2019;5(7):1008-1019.
18. Berry D. Subgroup analyses. *Biometrics*. 1990;46(4):1227-1230.
19. Bender R, Bunce C, Clarke M, et al. Attention should be given to multiplicity issues in systematic reviews. *J Clin Epidemiol*. 2008;61(9):857-865.
20. Sun X, Ioannidis J, Agoritsas T, Alba A, Guyatt G. How to use a subgroup analysis: users' guide to the medical literature. *JAMA*. 2014;311(4):405-411.
21. Trøstheim M, Eikemo M, Meir R, Hansen I, Paul E, et al. Assessment of anhedonia in adults with and without mental illness: a systematic review and meta-analysis. *JAMA Netw Open*. 2020;3(8):e2013233.
22. Polanin J, Pigott T. The use of meta-analytic statistical significance testing. *Res Synth Methods*. 2015;6(1):63-73.
23. Naing A, Hajjar J, Gulley J, Atkins M, Ciliberto G, et al. Strategies for improving the management of immune-related adverse events. *J Immunotherapy Cancer*. 2020;8(2):e001754.
24. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods*. 2010;1(2):97-111.
25. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Anal*. 2006;1(3):515-534.
26. Carvalho C, Polson N, Scott J. Handling sparsity via the horseshoe. In: *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS)*; 2009:73–80.
27. Carvalho C, Polson N, Scott J. The horseshoe estimator for sparse signals. *Biometrika*. 2010;97(2):465-480.
28. Datta J, Ghosh J. Asymptotic properties of Bayes risk for the horseshoe prior. *Bayesian Anal*. 2013;8(1):111-132.
29. Pas V, Kleijn B, Van Der Vaart A, et al. The horseshoe estimator: posterior concentration around nearly black vectors. *Electron J Stat*. 2014;8(2):2585-2618.
30. Bhadra A, Datta J, Li Y, Polson N. Horseshoe regularisation for machine learning in complex and deep models. *Int Stat Rev*. 2020;88(2):302-320.
31. Piironen J, Vehtari A. On the hyperprior choice for the global shrinkage parameter in the horseshoe prior. In: *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*; 2017:905–913.
32. Qi X, Zhou S, Plummer M. On Bayesian modeling of censored data in JAGS. *BMC Bioinformatics*. 2022;23(1):1-13.
33. Plummer M. *JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling*. Austria; 2003:1-10.
34. Plummer M. JAGS (Just Another Gibbs Sampler). 2010
35. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1-22.
36. DS PV, Szabó B, DA VV. Uncertainty quantification for the horseshoe (with discussion). *Bayesian Anal*. 2017;12(4):1221-1274.
37. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2021.
38. ClinicalTrials.gov results data element definitions for interventional and observational studies. 2021. https://prsinfo.clinicaltrials.gov/results_definitions.html (Accessed May 26, 2022)
39. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus docetaxel in advanced squamous-cell non–small-cell lung cancer. *N Engl J Med*. 2015;373(2):123-135.
40. Brooks S, Gelman A. General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat*. 1998;7(4):434-455.
41. Hua C, Boussemart L, Mateus C, Routier E, Boutros C, et al. Association of vitiligo with tumor response in patients with metastatic melanoma treated with pembrolizumab. *JAMA Dermatol*. 2016;152(1):45-51.
42. Delanoy N, Michot J, Comont T, Kramkimel N, Lazarovici J, et al. Haematological immune-related adverse events induced by anti-PD-1 or anti-PD-L1 immunotherapy: a descriptive observational study. *Lancet Haematol*. 2019;6(1):e48-e57.
43. Apolo A, Infante J, Balmanoukian A, Patel M, Wang D, et al. Avelumab, an anti–programmed death-ligand 1 antibody, in patients with refractory metastatic urothelial carcinoma: results from a multicenter, phase Ib study. *J Clin Oncol*. 2017;35(19):2117-2124.
44. Connolly C, Bambhania K, Naidoo J. Immune-related adverse events: a case-based approach. *Front Oncol*. 2019;9.
45. Jones H, Ohlssen D, Neuenschwander B, Racine A, Branson M. Bayesian models for subgroup analysis in clinical trials. *Clin Trials*. 2011;8(2):129-143.
46. Chetverikov D, Liao Z, Chernozhukov V. On cross-validated lasso in high dimensions. *arXiv preprint arXiv:1605.02214*. 2016.
47. Piironen J, Vehtari A. Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electron J Stat*. 2017;11(2):5018-5051.
48. Bhadra A, Datta J, Polson N, Willard B, et al. The horseshoe+ estimator of ultra-sparse signals. *Bayesian Anal*. 2017;12(4):1105-1131.
49. Chipman H. Bayesian variable selection with related predictors. *Canadian J Stat*. 1996;24(1):17-36.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Qi X, Zhou S, Wang Y, Peterson C. Bayesian sparse modeling to identify high-risk subgroups in meta-analysis of safety data. *Res Syn Meth*. 2022;1-14. doi:10.1002/jrsm.1597