Meta-analysis of Censored Adverse Events

Xinyue QI, Shouhao ZHOU*, Christine B. PETERSON, Yucai WANG, Xinying FANG, Michael L. WANG, and Chan SHEN

Abstract

Meta-analysis is a powerful tool for assessing drug safety by combining treatment-related toxicological findings across multiple studies, as clinical trials are typically underpowered for detecting adverse drug effects. However, incomplete reporting of adverse events (AEs) in published clinical studies is frequently encountered, especially if the observed number of AEs is below a pre-specified study-dependent threshold. Ignoring the censored AE information, often found in lower frequency, can significantly bias the estimated incidence rate of AEs. Despite its importance, this prevalent issue in meta-analysis has received little statistical or analytic attention in the literature. To address this challenge, we propose a Bayesian approach to accommodating the censored and possibly rare AEs for meta-analysis of safety data. Through simulation studies, we demonstrate that the proposed method can improve accuracy in point and interval estimation of incidence probabilities, particularly in the presence of censored data. Overall, the proposed method provides a practical solution that can facilitate better-informed decisions regarding drug safety.

KEYWORDS AND PHRASES: Adverse drug reaction, Bayesian inference, Drug safety, Incomplete reporting, MAGEC, Meta-analysis.

1. INTRODUCTION

This paper focuses on statistical modeling strategies for analyzing drug safety data across multiple studies. Our motivation stems from a systematic review of treatment-related adverse events (AEs) of programmed cell death 1 (PD-1) and PD-1 ligand 1 (PD-L1) inhibitors for cancer immunotherapy [54]. The PD-1 pathway is up-regulated in many tumors and in their microenvironment. Blockade of this pathway with antibodies to PD-1 or its ligands has led to remarkable clinical responses in various types of cancer [29], and is considered one of the most important breakthroughs in the treatment of cancer. These novel immune checkpoint inhibitors are clinically less toxic than traditional cancer treatments such as chemotherapy and radiation therapy, but can occasionally cause serious and sometimes life-threatening immune-related AEs. Because individual clinical trials are often underpowered for detecting adverse drug effect, combining evidence from multiple studies is an attractive approach to examine the toxicological profile of PD-1 and PD-L1 inhibitors.

Meta-analysis synthesizes findings from multiple independent clinical studies and provides a more powerful analysis than from a single study [51, 20]. However, one unique challenge in meta-analysis of drug safety is the incompleteness of AE reporting, which hinders the accuracy to quantify and understand the incidence of treatment-related AEs [13]. A significant portion of treatment-related AEs, despite having been counted in individual clinical trials, may not

*Corresponding author.

be publicly reported due to the rarity of such events. In the motivating anti-PD-1/PD-L1 example, many AEs were missing if their study-level observed frequencies were lower than a pre-determined reporting cutoff (e.g. 3% or 5% of the study sample size). Table 1 provides several sample studies on reporting of pneumonitis, a type of AE regarding lung inflammation, to illustrate the AE data structure in a typical meta-analysis. For instance, Powles *et al.* [44] conducted an open-label randomized controlled trial to assess the safety and efficacy of Atezolizumab. Among 459 patients treated with Atezolizumab, the AE reporting criteria was $\geq 5\%$ for all-grade and $\geq 2\%$ for grade 3 or higher, suggesting the missing pneumonitis count could be any number from 0 to 22 for all-grade and from 0 to 9 for grade 3-5. In another small multicenter study of Pembrolizumab, Nanda *et al.* [37] only listed the exact count of all grade AEs occurring in at least two patients (implying the reporting cutoff of 1 for censored AEs). Because pneumonitis wasn't reported in their results, it implies 0 or 1 case of all grade pneumonitis rather than the absence of pneumonitis. Likewise, Nanda *et al.* [37] only listed grade 3-5 AEs occurring in at least one patient, which indicates a reporting cutoff of 0 and implies that the actual grade 3-5 pneumonitis count, though unreported, was 0. Subsequently, if the analysis was conducted only based on the likelihood of the reported AE frequencies but ignoring all the left censored data, the inferences on incidence rates could be significantly overestimated.

Although censored AE data is a common issue in drug safety analysis, it has been largely overlooked in the de-

Table 1. A sample of AE observations for the motivating real-data application in Section 4. "–" indicates the missing data (i.e., unreported in publicly available data). The left-censored cutoffs for all grade AEs and grade 3-5 AEs could be different within a study. Given the cutoff of 0 for grade 3-5 AEs in 2016-Nanda-J Clin Onco study, the actual grade 3-5 pneumonitis count, though unreported, was exactly 0. In the proposed Bayesian modeling framework, identical inferences are generated by either treating it as a censored observation or observed.

Study Source	No. of Treated Patients	Drug	Cutoff (all grade)	Pneumonitis (all grade)	Cutoff (grade 3-5)	Pneumonitis (grade 3-5)
2015-Robert-N Engl J Med2 [46]	206	Nivolumab	4	–	4	–
2018-Lee-Lung Cancer [30]	100	Nivolumab	0	3	0	2
⋮		⋮				
2016-Nanda-J Clin Oncol [37]	32	Pembrolizumab	1	–	0	–
2017-Hsu-J Clin Oncol [23]	27	Pembrolizumab	1	3	1	2
⋮		⋮				
2018-Horn-Eur J Cancer [22]	89	Atezolizumab	4	–	4	–
2018-Powles-Lancet [44]	459	Atezolizumab	22	–	9	–
⋮		⋮				

velopment of meta-analysis methodology. In fact, the focus of attention in meta-analyses of treatment-related AEs has been mostly on the rarity of such events [2]. Standard methods to model binary patient outcomes such as AE incidences rely on either approximation methods based on the normal distribution or exact methods based on the binomial distribution [18]. When the observed events are rare, approximation approaches may provide poor estimates of the true incidences and lead to significantly biased results [33, 7, 3, 28]. Some recent efforts have been made to overcome this limitation, including the Poisson random-effect model to estimate relative risk between two treatment groups [6], and asymptotically unbiased estimation for the treatment effect and heterogeneity parameter in the random-effect model [2]. However, these methods were all developed for meta-analysis without missing data.

Most of the research on missing data in meta-analysis focuses on situations when the estimate from the whole study is missing [39], or on the analysis of treatment efficacy in different patterns [56, 57, 19, 36]. Due to the lack of appropriate analytic methods to address the problem of censored AE data, in current meta-analytic applications, most studies either totally ignored the AEs with low incidence, or completely discarded the studies with missing AE data [47], contributing to substantial publication selection bias or estimation error.

In this paper, we propose a general Bayesian approach to model censored AE data in a meta-analysis, with an aim to deliver exact inference in the estimation of AE incidence with proper uncertainty quantification. The rest of the article is organized as follows. In Section 2, we present the Bayesian modeling framework and implement the proposed approach in Just Another Gibbs Sampler (JAGS) with a tailored presentation for model assessment. In Section 3, we conduct numerical studies under different censoring scenarios by comparing the proposed Bayesian model of censored

data with other popular methods. Real data meta-analysis results demonstrating the advantage of the proposed approach are presented in Section 4. Lastly, some concluding remarks and discussions are provided in Section 5.

2. METHODS

We propose a new method, named Bayesian meta-analysis of adverse drug effects with censored data (MAGEC), to accommodate censored AE data in meta-analysis. It is a general framework that incorporates cumulative probabilities of the partial information contained in the censored data into the likelihood function, such that the meta-analysis can yield proper parameter estimation and statistical inference.

2.1 Modeling of Censored Adverse Events

Let Y_j, $j = 1, \ldots, J$, denote a safety response of interest in the jth study, which are collected in individual trials but may not be fully reported. For censored outcomes, the censoring mechanism can be defined by bounding variables (A, B), with semi-closed boundaries $(A_j, B_j]$ for response variable Y_j. Here, both bounding variables could be covariate-dependent. Often, we assume that the outcome model and the censoring mechanism are independent, so the censoring mechanism does not contribute to the inference and model estimation. It is a fundamental assumption for censored data behind most statistical methodologies [58]. In survival analysis, it is also known as noninformative censoring. Assume the random variable Y has the right continuous cumulative distribution function $F_Y(y) = P[Y \leq y]$. Denote $f_Y(y)$ the probability density/mass function of $F_Y(y)$.

In a meta-analysis of safety data, the frequency of adverse events may not always be reported. For example, left censoring occurs when some severe (grade 3 or higher) events are not observed due to low incidence. In this case, the cutoff boundaries are not random but fixed and study-specific,

which spontaneously satisfies the noninformative censoring assumption. Denote the fixed cutoff by c_j for each study. The number of subjects having a specific event in the jth study Y_j follows a binomial distribution with study-level sample size n_j and AE incidence probability θ_j

$$Y_j \sim Bin(n_j, \theta_j).$$

Therefore, the likelihood function for both fully observed and censored events can be written by:

$$
\begin{aligned}
\mathcal{L} &= \prod_{o=1}^{O} f_Y(y_o) \prod_{l=1}^{L} [F_Y(c_l) - 0] \prod_{r=1}^{R} [F_Y(n_r) - F_Y(c_r)] \\
&= \prod_{o=1}^{O} f_Y(y_o) \prod_{l=1}^{L} F_Y(c_l) \prod_{r=1}^{R} [1 - F_Y(c_r)] \\
&= \prod_{o=1}^{O} f_Y(y_o) \prod_{l=1}^{L} \sum_{k_l=0}^{c_l} f_Y(k_l) \prod_{r=1}^{R} [1 - \sum_{k_r=0}^{c_r} f_Y(k_r)], \quad (2.1)
\end{aligned}
$$

where O is the number of fully-observed AE outcomes, L the number of left-censored outcomes, and R the number of right-censored outcomes. c_l and c_r are cutoff values for left-censored and right-censored data, respectively.

The likelihood (2.1) can simultaneously handle both cases of left- and right-censoring for meta-analysis of safety data. In meta-analysis of the grade 3-5 AEs, we often only observed left censoring due to reporting cutoffs. In meta-analysis of the all-grade AEs, right censoring may also occur when some studies report grade 2 or higher AEs only rather than all-grade AEs [4]. If Y_j is left-censored, then Y_j lies in the semi-closed interval $(A_l = 0^-, B_l = c_l]$, where c_l is the corresponding cutoff value for left-censored data. If Y_j is right-censored, then Y_j lies in the semi-closed interval $(A_r = c_r, B_r = n_r]$, where c_r is the corresponding cutoff value for right-censored data and n_r is the total number of subjects in the rth study. This is specified to be consistent with the JAGS model implementation in Section 2.2. Compared with multiple imputation approaches that often requires the assumption of normality and involves intensive computation for Bayesian modeling [26], the proposed approaches are computationally efficient to model incomplete data.

The incidence probability can be decomposed through the link function $g(\theta)$,

$$g(\theta_j) = \mathsf{logit}(\theta_j) = \mu + \alpha_j + \boldsymbol{X_j}\boldsymbol{\beta}, \quad (2.2)$$

where $\boldsymbol{X_j}$ is a design matrix for study-level covariates from the jth study, and we consider the logit link as the default link function. Given the overall mean μ and the study covariate effects $\boldsymbol{\beta}$ associated with $\boldsymbol{X_j}$, we assume the study-specific random effect α_j is conditionally independent with mean 0 and variance σ_α^2.

For nonhierarchical intercept μ or slope parameters in $\boldsymbol{\beta}$ in the linear term of logistic model, a symmetric Cauchy distribution with center 0 and scale 2.5 was suggested [16]. For random effects in the linear term, which include the effects of various study-level factors and study-specific effects α_j, we assume they follow mean-zero normal distribution with different variances, with the prior distributions of standard deviation parameters following half-Cauchy distribution and a finite scale parameter of 10 or 25 [14]. In practice, the normality assumption can also be replaced by using heavy-tailed distributions such as t-distribution. In the illustrative anti-PD-1/PD-L1 example, the meta-analysis included 125 studies with a total of 20,218 patients. To identify possible sources of heterogeneity between studies, the following pieces of study-level information were also extracted: trial name, number of treated patients, selected immunotherapy drug, dosing schedule, cancer type, AE frequency (which could be missing), and the study-specified censoring cutoff for AE reporting. In particular, the study-level factors could include the drug type or drug/dose combinations, and/or cancer type, such that it is possible to estimate the AE incidence of a new drug or a new cancer type in a future trial with Bayesian credible interval.

Lastly, the model specification of the proposed Bayesian modeling framework only involves discrete data in the likelihood function, while the prior distribution of parameters is proper. As a result, it is easy to verify the property of Bayesian posterior inference. Specifically, we can have the following remark:

Remark 1. The posterior distribution of Bayesian models in the MAGEC framework is always proper.

Overall, the proposed Bayesian framework can help to streamline the modeling process of drug safety data and enhance the reporting reliability of meta-analysis.

2.2 Model Implementation using JAGS

To carry out posterior inference and assess the performance of the proposed Bayesian model, we apply Just Another Gibbs Sampling (JAGS) to generate samples from the posterior distribution. JAGS makes Bayesian hierarchical models easy to implement using Markov Chain Monte Carlo (MCMC) simulation [40] in R and other computational software. In the presence of censored data in the response variable, an existing function, known as dinterval distribution, is commonly used to model censored data [27, 43]. However, such model specification for censored data in JAGS intrinsically yields a mis-specified deviance function [42]. More technical details are provided in Qi *et al.* (2022) [45].

Alternatively, we apply a simple but effective approach to censored data specification. To facilitate model implementation for censored observations (when $\delta_j = 0$) and avoid the miscalculation of deviance via the dinterval() function in JAGS. Here, we utilize the idea of data augmentation by introducing ancillary indicator variables W_j. Each W_j separates the left-censored from right-censored observations ($W_j = 1$ if left-censored, 0 if right-censored). By

assuming W_j follows a Bernoulli distribution, we have the density function

$$f_W(w_j) = F_Y(c_l)^{w_j}[1 - F_Y(c_r)]^{1-w_j}$$
$$= \left[\sum_{k_l=0}^{c_l} f_Y(k_l)\right]^{w_j} \left[1 - \sum_{k_r=0}^{c_r} f_Y(k_r)\right]^{1-w_j}, \quad (2.3)$$

which exactly matches the second and third terms for censored observations in (2.1), with the cumulative binomial distribution of incidence probability of AE in the jth study, restricted by a pre-determined study-level cutoff value c_j. This approach can also be extended to interval-censored data [45].

A JAGS model specification for the application is provided in the *Supplementary Material* for illustrative purposes. Together with fully observed data that follow a binomial distribution, the full likelihood implemented in a JAGS model is, in fact, identical to the exact likelihood of observed and censored cases in (2.1). This creates the correct focus of model parameters and produces the proper posterior samples, as well as simultaneously computes the correct deviance for model selection. For example, by calling the deviance module in JAGS, correct deviance information criterion (DIC) [49], penalized deviance [41], posterior averaging information criterion (PAIC) [60], or widely applicable information criterion (WAIC) [55, 53] can be conveniently derived to assess candidate models for Bayesian model selection. It is important and beneficial to select the best model, especially in the presence of complicated model features. Also, from the point of view of convenience, the sampling procedure can smoothly start with default initial values for the censored data. The standard diagnostic tools can be applied to assess the posterior convergences of the model [15].

3. SIMULATION

In this section, we conduct a simulation study to assess the performance of the proposed Bayesian model for estimating the incidence rates in the meta-analysis of adverse events (AEs) with censored information. We compare it with that of other popular methods applied under a standard setting [47]. Other than the proposed Bayesian model, MAGEC, all other methods do not account for censoring.

3.1 Settings

The total number of studies for each probability level of AE incidence is fixed at 10 ($J = 10$) to reflect the typical number of studies in a meta-analysis for subgroups. We also consider $J = 25$ and 50 as comparisons. The outcome of interest, the number of AEs for each study, is generated from a binomial distribution with the number of patients ($n = 100$) and probability of events ($p = 0.2, 0.05$, and 0.01, respectively). The probability of incidence is designed to cover the typical range of the probability levels of AE

incidence. The data generation incorporates heterogeneity between studies into the incidence probability of AE. To be specific, the coefficient of study effect was generated from a normal distribution at mean 0 and standard deviation of 0.2.

To assess the performance of the proposed model that incorporates both observed and censored data, we consider six scenarios: (1) no censoring; (2) a low percentage (20%) of left-censoring for all three incidence levels; (3) a moderate percentage (40%) of left-censoring; (4) a high percentage (60%) of left-censoring; (5) a low/moderate percentage (20%/40%) of right-censoring; and (6) a moderate percentage (40%) of left-censoring with a low percentage (20%) of right censoring. In Scenario 1, the number of events for all studies is fully observed. In Scenarios 2, 3 and 4, which include observations that could only be informatively left-censored for zero and low incidence counts to mimic real-world cases. Therefore, in Scenario 2, we treat the 20% of AE data with the lowest observed incidence as censored data and the 80% of AE data that have a 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, 40% of AE data with low incidence are treated as censored and the remaining 60% are treated as observed. Lastly, in Scenario 4, the top 40% of studies are treated as observed data, and the remaining 60% are treated as censored data. Accordingly, the study-specific left-censored cutoff value is selected from the sorted outcomes in a descending order based on the percentage of left-censoring after data generation. For example, in the scenario of 60% left-censoring, the cutoff value corresponds to the $((1-60\%)J+1)^{\text{th}}$ outcome among sorted values. However, the right-censoring in AE reporting is a different censoring mechanism (e.g., the reporting of grade 2 or higher AEs rather than all-grade AEs) and likely from the random selection of studies. Therefore, in Scenario 5, 20% of studies are randomly treated as right-censored; in Scenario 6, 40% of AE data with low incidence are treated as left-censored data first and then randomly selecting one third of the remaining 60% (e.g., 20% among all) of data as right-censored. The study-specific right-censored cutoff value is generated from a binomial distribution by assuming the probability of grade 2 or higher AE is 50% of all-grade AE incidence.

We compare the proposed model, MAGEC, with four other methods: the pooled estimation method after continuity correction (PEM) [52, 24], the normal approximation method (NAM) [38], the logistic regression method (LRM), and the normal approximation method with robust variance estimator (RVE) [34]. In MAGEC, following the recommendation of the weakly-informative prior distribution for logistic regression models [16], we assign a half-Cauchy prior distribution on the standard deviation of study effects. In PEM [52], we pool observations by incidence level and the 95% nominal confidence intervals (CIs) for AE incidence levels are calculated by the exact binomial test. In NAM, a
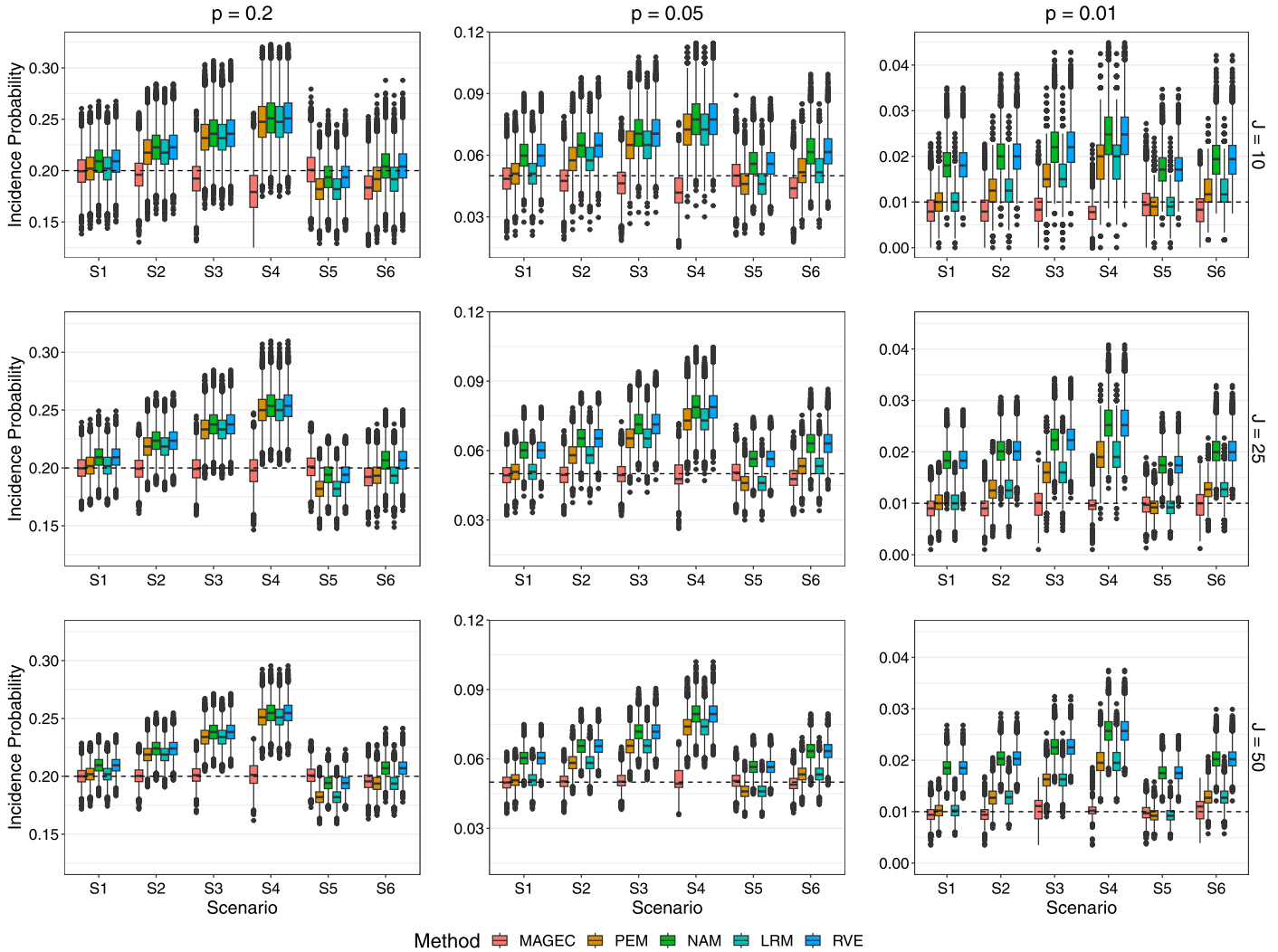
Figure 1: Boxplot of point estimators for AE incidence using five methods. We compared the proposed Bayesian model (MAGEC), pooled estimation method after continuity correction (PEM), normal approximate method (NAM), logistic regression model (LRM), and normal approximate method with robust variance estimation (RVE) under different total number of studies ($J = 10, 25, 50$) and six scenarios: (S1) 0% left censoring; (S2) 20% left censoring; (S3) 40% left censoring; (S4) 60% left censoring; (S5) 20% right censoring; and (S6) 40% left censoring & 20% right censoring.

standard method in practice [38], we use a normal likelihood procedure to estimate the incidence rate by taking the inverse logit of the observed logit incidence [18] of each incidence level weighted by its within-level variance. In LRM, we estimate the AE incidences by an exact method through fitting a generalized linear model with a logit link. In addition, we compare the performance of NAM with and without robust variance estimators [34]. 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.

We assess model performance in terms of accuracy in the estimation of the AE incidences. More specifically, we compare the five methods using the following metrics: mean ab-

solute deviation (MAD), root mean squared error (RMSE) of the point estimates (for MAGEC, we use the posterior median), and coverage probability (CP) of the 95% nominal confidence interval (for MAGEC, we use the 95% posterior credible intervals), of the incidence rate parameters of interest in each scenario.

3.2 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 1 gives boxplots for the point estimates of incidence probabilities by scenario, method and the total number of studies. CPs of three parameters of interest by scenario,
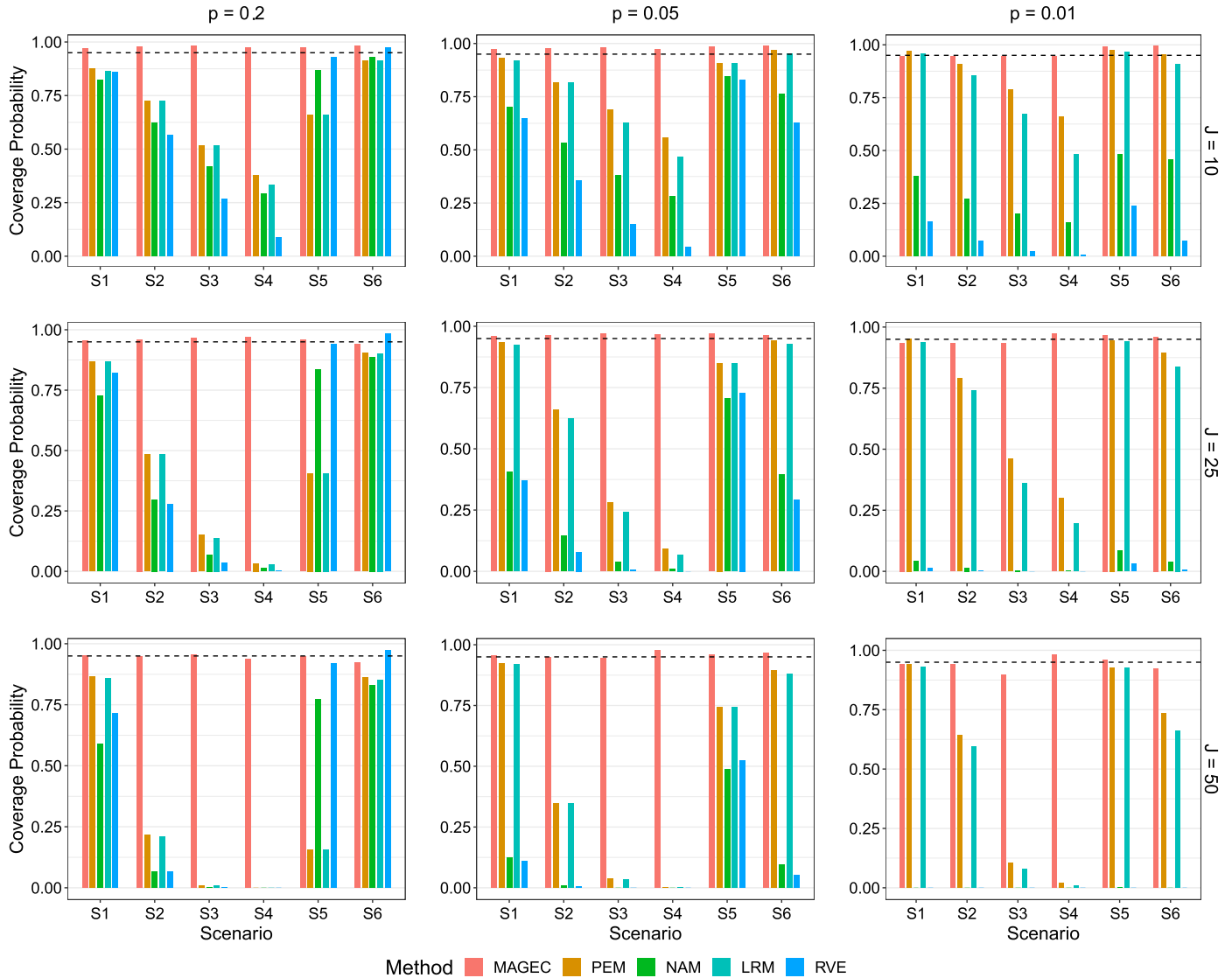
Figure 2: Coverage probabilities (CPs) of AE incidences using five methods. We compared the proposed Bayesian model (MAGEC), pooled estimation method after continuity correction (PEM), normal approximate method (NAM), logistic regression model (LRM), as well as normal approximate method with robust variance estimation (RVE) under different total number of studies ($J = 10, 25, 50$) and six scenarios: (S1) 0% left censoring; (S2) 20% left censoring; (S3) 40% left censoring; (S4) 60% left censoring; (S5) 20% right censoring; and (S6) 40% left censoring & 20% right censoring.

method, and the total number of studies are displayed in bar charts in Figure 2. In Table 2, performance in terms of both MAD and RMSE of incidence rates based on the five methods are shown for the four data censoring scenarios. The table that presents two additional data censoring scenarios is available in the *Supplementary Material*.

When there is no censoring (Scenario 1), the proposed method (MAGEC) has CPs (Figure 2), MADs, and RMSEs (Table 2) for the incidence probabilities that are almost identical to those of PEM and LRM. Of the five methods compared, PEM can be considered the gold standard/benchmark for both interval and point estimation. Our results

indicate that MAGEC is not inferior to PEM. They also indicate that the CP for each probability level of AE incidence obtained from NAM and RVE appeared to be unstable in estimating the incidence probabilities of less frequent events compared with the other methods. The point estimates of the incidence probabilities 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 [33].

When 20% of data are censored (Scenario 2), the proposed method (MAGEC) performs better than the others in

Table 2. Mean absolute deviations (MADs) and root mean square errors (RMSEs) of drug incidence probabilities for five methods. We compared the proposed Bayesian model (MAGEC), 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 different study sample sizes ($J = 10, 25, 50$) and four scenarios: (S1) 0% left censoring; (S2) 20% left censoring; (S3) 40% left censoring; and (S4) 60% left censoring. LC = left censoring; RC = right censoring.

Study Sample Size	Scenario	True Value	% of LC	% of RC	Mean Absolute Deviation				Root-mean-squared Error			
					MAGEC	PEM	LRM	NAM/RVE	MAGEC	PEM	LRM	NAM/RVE
10	S1	0.2	0%	0%	0.013	0.013	0.013	0.015	0.014	0.016	0.016	0.019
		0.05	0%	0%	0.006	0.006	0.006	0.011	0.008	0.008	0.008	0.013
		0.01	0%	0%	0.003	0.003	0.003	0.008	0.004	0.003	0.003	0.009
	S2	0.2	20%	0%	0.014	0.020	0.020	0.024	0.014	0.020	0.020	0.024
		0.05	20%	0%	0.006	0.009	0.009	0.015	0.008	0.012	0.012	0.017
		0.01	20%	0%	0.003	0.004	0.004	0.010	0.004	0.005	0.005	0.011
	S3	0.2	40%	0%	0.015	0.033	0.033	0.037	0.019	0.037	0.037	0.041
		0.05	40%	0%	0.007	0.015	0.015	0.021	0.009	0.017	0.017	0.023
		0.01	40%	0%	0.003	0.006	0.006	0.012	0.004	0.007	0.007	0.013
	S4	0.2	60%	0%	0.026	0.049	0.049	0.052	0.032	0.053	0.053	0.056
		0.05	60%	0%	0.009	0.023	0.023	0.028	0.011	0.025	0.025	0.030
		0.01	60%	0%	0.003	0.009	0.009	0.015	0.004	0.011	0.011	0.016
25	S1	0.2	0%	0%	0.008	0.008	0.008	0.013	0.010	0.010	0.010	0.014
		0.05	0%	0%	0.004	0.004	0.004	0.010	0.005	0.005	0.005	0.011
		0.01	0%	0%	0.002	0.002	0.002	0.009	0.002	0.002	0.002	0.009
	S2	0.2	20%	0%	0.014	0.019	0.019	0.024	0.017	0.022	0.022	0.026
		0.05	20%	0%	0.004	0.008	0.008	0.015	0.005	0.010	0.010	0.016
		0.01	20%	0%	0.002	0.003	0.003	0.010	0.002	0.004	0.004	0.011
	S3	0.2	40%	0%	0.009	0.034	0.034	0.038	0.011	0.036	0.036	0.040
		0.05	40%	0%	0.004	0.015	0.015	0.022	0.005	0.016	0.016	0.022
		0.01	40%	0%	0.002	0.006	0.006	0.012	0.003	0.007	0.007	0.013
	S4	0.2	60%	0%	0.011	0.050	0.050	0.054	0.014	0.052	0.052	0.056
		0.05	60%	0%	0.005	0.023	0.023	0.029	0.006	0.025	0.025	0.030
		0.01	60%	0%	0.001	0.010	0.010	0.016	0.002	0.010	0.010	0.016
50	S1	0.2	0%	0%	0.006	0.006	0.006	0.010	0.007	0.008	0.008	0.012
		0.05	0%	0%	0.003	0.003	0.003	0.010	0.003	0.004	0.004	0.011
		0.01	0%	0%	0.001	0.001	0.001	0.009	0.002	0.001	0.001	0.009
	S2	0.2	20%	0%	0.006	0.019	0.019	0.024	0.008	0.020	0.020	0.025
		0.05	20%	0%	0.003	0.008	0.008	0.015	0.003	0.009	0.009	0.016
		0.01	20%	0%	0.001	0.003	0.003	0.010	0.002	0.003	0.003	0.011
	S3	0.2	40%	0%	0.007	0.034	0.034	0.038	0.008	0.035	0.035	0.039
		0.05	40%	0%	0.003	0.016	0.016	0.022	0.004	0.016	0.016	0.022
		0.01	40%	0%	0.002	0.006	0.006	0.013	0.002	0.007	0.007	0.013
	S4	0.2	60%	0%	0.008	0.051	0.051	0.055	0.010	0.052	0.052	0.056
		0.05	60%	0%	0.004	0.024	0.024	0.030	0.004	0.024	0.024	0.030
		0.01	60%	0%	0.001	0.010	0.010	0.016	0.001	0.010	0.010	0.016

estimating incidence probabilities; with an increasing total number of studies, its performance in Scenario 2 is comparable to that in Scenario 1. Because censored observations are ignored under the other four methods (PEM, NAM, LRM, and RVE), it is unsurprising that their point estimates for the incidence probabilities are overestimated and that their CPs in Scenario 2 are much lower than those in Scenario 1. In contrast, the performance of MAGEC in Scenario 2 is almost identical to its performance in Scenario 1 for both interval and point estimates, demonstrating its robustness of censored data.

In a more extreme scenario where 40% of data are cen- sored (Scenario 3), the proposed method (MAGEC) per- forms well, with little information loss compared with Sce- narios 1 and 2. However, all other estimators of AE incidence lead to inferior CP due to an increased percentage of cen- soring. The point estimations obtained from PEM, LRM, 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 are larger deviations from the true values of the incidence probabilities compared with those in Scenario 2. With an increasing total number of studies, their CPs in Scenario 3 are much lower that those in Scenario 1 and 2. When the percentage of censoring reaches 60%, the

performance of MAGEC is still the best among all methods. When we incorporate right-censored AE data, MAGEC consistently provides the lowest errors across all settings. In Scenario 6 with censoring possibly on both sides, the results from four other methods are unreliable. For example, the observed errors sometimes are smaller in Scenario 6 (e.g., MAD = 0.015 in PEM) than in Scenario 3 or 5 for one-sided censored data alone (e.g., MAD = 0.033 or 0.020 in PEM), when combining two wrongs makes a right because the analysis of neglecting the left-censored or right-censored data tends to be biased in the opposite direction. However, as illustrated in Table B.1 in the *Supplementary Material*, the smaller error cannot be warranted across different sample size settings. Overall, the MAGEC model yields not only more stable and superior coverage, but also unbiased estimates of the incidence probabilities in all six scenarios especially with censored event data. The proposed method, handling both observed and censored data took, around 20 seconds in the absence of censoring and 50-70 seconds with an increasing percentage of censoring. In contrast, the other four methods, which only consider observe data, completed the task in less than one second.

Across all scenarios considered above, MAGEC outperforms the other four methods in estimating incidence probabilities 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 [8]. The larger the variance, the lower the efficiency of an estimator. Here, the asymptotic relative efficiency (RE), defined by the reciprocal of the ratio of the asymptotic variances of two unbiased estimators, is given to examine the amount of information loss from informative censoring. There is a 10%-20% inflation in the RMSE from 0% to 40% censored data, suggesting limited efficiency loss of MAGEC in incidence estimation due to censoring. In summary, the proposed method consistently achieves a reasonable performance in estimating the incidence rates at different probability levels of AE incidence in the presence of censored data.

4. APPLICATION

In this section, we apply the proposed Bayesian method to modeling a meta-analysis of Grade 3 or higher AEs with censored information [54]. The goal is to evaluate the incidence probabilities of pneumonitis (referring to inflammation of lung tissue) in two PD-1 and three PD-L1 inhibitors in a meta-analysis of 125 clinical studies. Pneumonitis is a typical serious AE not only leading to hospital admission and potentially permanent lung damage or death, but also resulting in treatment discontinuation [9, 59]. Such kind of inflammatory or immune-related AEs are of special interest for cancer immunotherapy. Specifically, around 35% of grade 3 or higher AEs of pneumonitis across studies were incompletely reported.

We compare six models to determine the best model for grade 3 or higher pneumonitis data. The linear combination in (2.2) could entail one or more of the following terms: the between-study random effects (α), drug or drug-dose random effects (η), and cancer type random effects (ζ). Model 1 (M1) is the simplest model with only study effects. Model 2–4 (M2–M4) consider adding cancer effect, drug effect, and drug-dose effect to the M1, respectively. With both cancer effect and drug effect, Model 6 (M6) is more complex than Model 5 (M5) because it also takes dose level into consideration. To specify the priors for each random effect in the proposed logistic model, we assume normal distributions

$$\alpha \sim N(0, \sigma_\alpha^2), \quad \eta \sim N(0, \sigma_d^2) \text{ or } \sim N(0, \sigma_{ad}^2),$$
$$\zeta \sim N(0, \sigma_c^2),$$

with half-Cauchy prior distributions for the standard deviation parameters and the scale parameter $A = 25$ [16].

All six models are implemented in the statistical software R and JAGS [40] using MCMC algorithms to generate samples from the posterior distribution of the parameters of interest. We ran 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 obtained 10,000 posterior samples per chain by retaining one sample out of three. Based upon the traceplot for each parameter of interest, all chains are close to each other and the target distribution. Furthermore, in terms of the potential scaler reduction statistic, also known as `Rhat`, the values of all key parameters are at 1, suggesting convergence. The 30,000 posterior samples of model parameters such as incidence probabilities of cancer effects and drug effects were saved for inference.

The model assessment results are summarized in Table 3. Model 5 containing study, cancer and drug effects was selected based upon the smallest DIC and WAIC among all models considered. According to the subgroup analysis of incidence probability of AE by drug, there were no significant differences in the incidence among different dosing schedules for PD-1/PD-L1 drugs. Figure 3 displays the incidence probabilities, along with their 95% credible intervals, for grade 3 or higher pneumonitis across 49 clinical trials of Pembrolizumab. The comprehensive forest plot (Figure C.1 in the *Supplementary Material*) illustrates the estimated incidence probability for grade 3 or higher pneumonitis across all 125 clinical trials. The vertical dashed line in Figure C.1 is the overall incidence probability of grade 3 or higher pneumonitis (0.48%; 95% CI, 0.30%–0.70%) across all studies. By contrast, if the censored outcomes were treated as missing at random, the estimated incidence rate would be biased and overestimated by 12.5%. If we further ignored the covariate information and analyzed the case study using four other frequentist methods in simulation study, the overall incidence probability estimates would increase more than

Table 3. Model Comparison: heterogeneity between study (σ_α), cancer type (σ_c), and drug (σ_d) or drug-dose (σ_{ad}), deviance information criterion (DIC), and widely applicable information criterion (WAIC) from modeling observed and censored grade 3 or higher AE (pneumonitis) data.

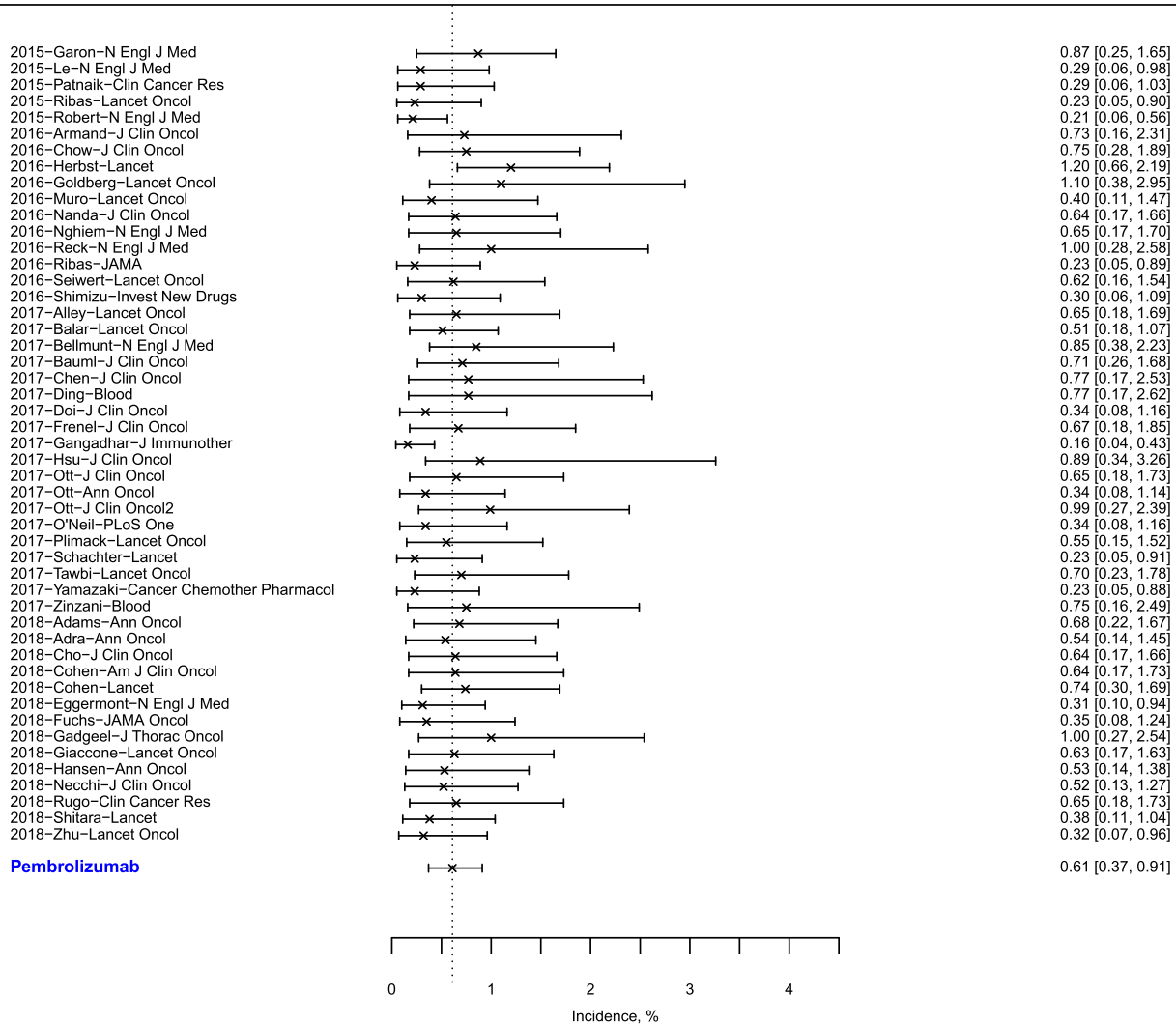| Model	σ_α	σ_c	σ_d	σ_{ad}	DIC	WAIC
M1	0.714 [0.415, 1.100]	−	−	−	243.7	223.2
M2	0.494 [0.060, 0.927]	0.697 [0.186, 1.864]	−	−	248.5	227.2
M3	0.701 [0.401, 1.080]	−	0.357 [0.021, 1.650]	−	241.1	222.0
M4	0.704 [0.396, 1.084]	−	−	0.223 [0.010, 0.822]	244.0	223.9
M5	0.425 [0.037, 0.865]	0.739 [0.240, 1.977]	0.458 [0.038, 1.904]	−	**237.0**	**217.1**
M6	0.410 [0.032, 0.876]	0.739 [0.232, 1.959]	−	0.321 [0.020, 0.899]	238.7	218.2

Figure 3: Incidence of grade 3 or higher AE (Pneumonitis) by study for Pembrolizumab.

60%. The case study results are consistent with our simulation studies, in that existing methods which rely only on reported counts (which by definition were not subject to censoring) tend to overestimate the incidence rates. Our proposed method avoids this upward bias, suggesting a more favorable safety profile for the PD-1/PD-L1 drugs. These findings demonstrate that our proposed method can play a critical role in providing a more accurate understanding of the risk/benefit ratio for new therapies with serious but low-frequency adverse events.

5. DISCUSSION

In this paper, we presented a Bayesian hierarchical modeling framework in the meta-analysis setting when the study-level event rates could be incompletely reported. Our Bayesian approach, named MAGEC, is efficient to handle censored safety data without requiring the multiple imputation. Further, we demonstrated the superior performance of our proposed method in simulations. Specifically, simulation results showed that the proposed Bayesian approach generated unbiased estimations of drug effects with limited information loss. Moreover, the proposed method was robust to rare events, which are commonly encountered in data analysis of AEs. Note that the definition of rare events here is in a statistical sense following the previous works in the meta-analysis literature [2, 32], rather than in the pharmacovigilance setting [10]. Finally, we illustrated the implementation with a real data application, while employing the proposed method can effectively reduce the over-estimation bias towards the incidence rate of pneumonitis in cancer patients with the treatment of immunotherapy.

Special consideration must be given to incomplete reporting of AEs to avoid incorrect statistical inference. In general, there are mainly three concerns when conducting safety meta-analyses: (1) AE incidences are heterogeneous across studies, (2) AEs of interest could be rare, and (3) AEs are incompletely or inconsistently reported [1, 50]. While various researchers have proposed different methods to address the first two concerns, the third concern is largely overlooked in methodology development. Naively applying standard methods in the presence of missing or censored data can result in bias or incorrect conclusions [31]. In our motivating application, clinical trial reporting standards only require that AEs be included in the final trial report if they exceed a pre-specified cut-off that is specific to that study; this means that rare AEs that occur less frequently than the threshold are omitted and can be considered as left-censored data. For some older clinical trials, the trial cutoff could be (implicitly) as large as the sample size, while only efficacy outcomes were reported in the publication with no mention of AEs [17]. Particularly for newly developed drugs, we believe that such completely censored reporting or related publication bias should not be a major concern. Since April 18, 2017, it is mandated that all studies registered with ClinicalTrials.gov must report AE counts that exceed a pre-specified cutoff (e.g. 5%) for each study arm within one year after the primary completion date. Nevertheless, the problem of incomplete AE reporting with study-specific cutoff, as this paper attempts to address, is still unavoidable for meta-analysis of drug safety with this recent reporting requirement.

The proposed Bayesian MAGEC modeling framework estimates the incidence probabilities using a one-stage approach. In contrast, a meta-analysis of binary data is usually conducted using a two-stage approach in the literature [12, 48] – the summary statistics are first separately calculated for each trial and then combined by an appropriate meta-analysis model. However, this two-stage approach is likely to perform poorly in the first stage of each study due to the rarity of events [5]. Alternatively, a one-stage approach is preferred as it delivers more exact statistical inference [11].

Although caveats of incomplete reporting of AE data and its impact on biasing meta-analysis results has been noted by many researchers, no statistical approach has yet been proposed to systematically address this issue. Our Bayesian MAGEC model represents the first such attempt, and we present it in a form of simplicity to illustrate the concept. One limitation of our proposed method could be that MAGEC requires a reasonable number of eligible studies to be included in meta-analysis of drug safety. In cases with very few studies, it becomes important to apply proper adjustments to enable stable estimation for random effects [35, 61]. In the Bayesian modeling setting, a robust prior for the standard deviation parameter, such as the half-Cauchy prior with heavy tails [14], is preferred for the between-study random effect to address this limitation. Another limitation of MAGEC is the assumption that the censored safety data exhibit a non-random missingness pattern. This assumption is analogous to the censoring data often encountered in survival analysis, where the data are censored in a non-informative manner. In the real-data application, we only demonstrated the case of left-censored grade 3 or higher AEs when they happened at a frequency lower than pre-specified cutoff values. However, the proposed framework can be extended to joint modeling of AEs of different severity grades also with right-censoring using multinomial distribution. Additionally, as many AE types are clustered in nature, multivariate modeling, rather than univariate modeling, could address the AE correlation and explore high-risk subgroups. Furthermore, it's worth noting that our proposed method can be applied to the meta-analysis of high-dimensional genomic data [21], in which a large number of genes are evaluated to estimate the mutation rate in the panels across studies. For such an extension, information on mutations could be also censored in some studies due to low frequency, which should be considered in the model using a pre-specified cutoff value determined by gene selection criteria. From the modeling perspective, the high-dimensional outcome can be modeled directly using a multivariate normal distribution with a covariance matrix describing the genome-wide association. In implementation can be simplified with pre-existing genomic knowledge or specific covariance structure such as compound symmetry or first-order autoregressive structure. Additional parameters in the covariance matrix can be sampled from their conditional distributions using MH algorithm. Compared to the current simulation setting, computation for high-dimensional censored data could demand significant computational resources. However, the Bayesian framework's ability to borrow information across dimensions

could potentially lead to better performance. Lastly, our model implementation strategy could also be applied to analyze other censored data structures, including time-to-event data, count data, and ranking data, in addition to binomial data [25]. By providing a statistical framework to systematically address incomplete AE reporting, we hope that the Bayesian MAGEC model will facilitate more accurate and reliable meta-analyses in the future.

DATA AVAILABILITY STATEMENT

The data and code to reproduce real-data analysis are available from https://github.com/xinyue-qi/Meta-analysis-of-Censored-AEs.

SUPPLEMENTARY MATERIAL

The Supplementary Material includes the JAGS model specification (Section A) for the application in Section 4, the additional results of Table 2 (Section B) under Scenarios 5 and 6 in Section 3.2, and a comprehensive forest plot across all 125 clinical studies (Section C).

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

[1] BERLIN, J., CROWE, B., WHALEN, E., XIA, H., KORO, C. and KUEBLER, J. (2013). Meta-analysis of clinical trial safety data in a drug development program: answers to frequently asked questions. *Clinical Trials* **10**(1) 20–31.

[2] BHAUMIK, D. K., AMATYA, A., NORMAND, S. Q. L. T., GREENHOUSE, J., KAIZAR, E., NEELON, B. and GIBBONS, R. D. (2012). Meta-analysis of rare binary adverse event data. *Journal of the American Statistical Association* **107**(498) 555–567. https://doi.org/10.1080/01621459.2012.664484. MR2980067

[3] BRADBURN, M. J., DEEKS, J. J., BERLIN, J. A. and RUSSELL LOCALIO, A. (2007). Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Statistics in Medicine* **26**(1) 53–77. https://doi.org/10.1002/sim.2528. MR2312699

[4] BRAHMIK, J. R., DRAKE, C. G. et al. (2010). 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* **28**(19) 3167.

[5] BURKE, D. L., ENSOR, J. and RILEY, R. D. (2017). Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Statistics in Medicine* **36**(5) 855–875. https://doi.org/10.1002/sim.7141. MR3597661

[6] CAI, T., PARAST, L. and RYAN, L. (2010). Meta-analysis for rare events. *Statistics in Medicine* **29**(20) 2078–2089. https://doi.org/10.1002/sim.3964. MR2756556

[7] CARRIERE, K. C. (2001). How good is a normal approximation for rates and proportions of low incidence events? *Communications in Statistics – Simulation and Computation* **30**(2) 327–337. https://doi.org/10.1081/SAC-100002370. MR1864235

[8] CASELLA, G. and BERGER, R. L. (2002) *Statistical Inference* **2**. Duxbury Pacific Grove, CA. MR1051420

[9] CHUZI, S., TAVORA, F., CRUZ, M., COSTA, R., CHAE, Y. K., CARNEIRO, B. A. and GILES, F. J. (2017). Clinical features, diagnostic challenges, and management strategies in checkpoint inhibitor-related pneumonitis. *Cancer Management and Research* **9** 207.

[10] COUNCIL FOR INTERNATIONAL ORGANIZATIONS OF MEDICAL SCIENCES (1999) *Guidelines for Preparing Core Clinical-safety Information on Drugs: Report of CIOMS Working Groups III and V: Including New Proposals for Investigator's Brochures.* CIOMS.

[11] DEBRAY, T. P., MOONS, K. G., ABO-ZAID, G. M. A., KOFFIJBERG, H. and RILEY, R. D. (2013). Individual participant data meta-analysis for a binary outcome: one-stage or two-stage? *PloS ONE* **8**(4) 60650.

[12] DEEKS, J. J. (2002). Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine* **21**(11) 1575–1600.

[13] FDA (2006). Guidance for Industry: Adverse Reactions Section of Labeling for Human Prescription Drug and Biological Products – Content and Format. Technical Report.

[14] GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**(3) 515–534. https://doi.org/10.1214/06-BA117A. MR2221284

[15] GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (1995) *Bayesian Data Analysis.* Chapman and Hall/CRC. MR1385925

[16] GELMAN, A., JAKULIN, A., PITTAU, M. G. and SU, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics* **2**(4) 1360–1383. https://doi.org/10.1214/08-AOAS191. MR2655663

[17] GOLDER, S., LOKE, Y. K., WRIGHT, K. and NORMAN, G. (2016). Reporting of adverse events in published and unpublished studies of health care interventions: a systematic review. *PLoS Medicine* **13**(9) 1002127.

[18] HAMZA, T. H., VAN HOUWELINGEN, H. C. and STIJNEN, T. (2008). The binomial distribution of meta-analysis was preferred to model within-study variability. *Journal of Clinical Epidemiology* **61**(1) 41–51.

[19] HIGGINS, J. P., WHITE, I. R. and WOOD, A. M. (2008). Imputation methods for missing outcome data in meta-analysis of clinical trials. *Clinical Trials* **5**(3) 225–239.

[20] HIGGINS, J. P., THOMAS, J., CHANDLER, J., CUMPSTON, M., LI, T., PAGE, M. J. and WELCH, V. A. (2019) *Cochrane Handbook for Systematic Reviews of Interventions.* John Wiley & Sons.

[21] HILL, H. A., QI, X., JAIN, P., NOMIE, K., WANG, Y., ZHOU, S. and WANG, M. L. (2020). Genetic mutations and features of mantle cell lymphoma: a systematic review and meta-analysis. *Blood Advances* **4**(13) 2927–2938.

[22] HORN, L., GETTINGER, S. N., GORDON, M. S., HERBST, R. S., GANDHI, L., FELIP, E., SEQUIST, L. V., SPIGEL, D. R., ANTONIA, S. J., BALMANOUKIAN, A. et al. (2018). Safety and clinical activity of atezolizumab monotherapy in metastatic non-small-cell lung cancer: final results from a phase I study. *European Journal of Cancer* **101** 201–209.

[23] HSU, C., LEE, S.-H., EJADI, S., EVEN, C., COHEN, R. B., LE TOURNEAU, C., MEHNERT, J. M., ALGAZI, A., VAN BRUMMELEN, E. M., SARAF, S. et al. (2017). Safety and antitumor activity of pembrolizumab in patients with programmed death-ligand

1–positive nasopharyngeal carcinoma: results of the KEYNOTE-028 study. *Journal of Clinical Oncology* **35**(36) 4050–4056.

[24] JEWELL, N. P. (2003) *Statistics for Epidemiology*. Chapman and Hall/CRC.

[25] JOHNSON, T. R. and KUHN, K. M. (2013). Bayesian Thurstonian models for ranking data using JAGS. *Behavior Research Methods* **45**(3) 857–872.

[26] KALAYCIOGLU, O., COPAS, A., KING, M. and OMAR, R. Z. A comparison of multiple-imputation methods for handling missing data in repeated measurements observational studies. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **179**(3) 683–706 (2016). https://doi.org/10.1111/rssa.12140. MR3501428

[27] KRUSCHKE, J. (2014) *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press.

[28] LANE, P. W. (2013). Meta-analysis of incidence of rare events. *Statistical Methods in Medical Research* **22**(2) 117–132. https://doi.org/10.1177/0962280211432218. MR3190650

[29] LE, D. T., URAM, J. N., WANG, H., BARTLETT, B. R., KEMBERLING, H., EYRING, A. D., SKORA, A. D., LUBER, B. S., AZAD, N. S., LAHERU, D., BIEDRZYCKI, B., DONEHOWER, R. C., ZAHEER, A., FISHER, G. A., CROCENZI, T. S., LEE, J. J., DUFFY, S. M., GOLDBERG, R. M., DE LA CHAPELLE, A., KOSHIJI, M., BHAIJEE, F., HUEBNER, T., HRUBAN, R. H., WOOD, L. D., CUKA, N., PARDOLL, D. M., PAPADOPOULOS, N., KINZLER, K. W., ZHOU, S., CORNISH, T. C., TAUBE, J. M., ANDERS, R. A., ESHLEMAN, J. R., VOGELSTEIN, B. and DIAZ, L. A. (2015). PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *New England Journal of Medicine* **372**(26) 2509–2520. PMID: 26028255. https://doi.org/10.1056/NEJMoa1500596.

[30] LEE, J. S., LEE, K. H., CHO, E. K., KIM, D.-W., KIM, S.-W., KIM, J.-H., CHO, B. C., KANG, J. H., HAN, J.-Y., MIN, Y. J. et al. (2018). Nivolumab in advanced non-small-cell lung cancer patients who failed prior platinum-based chemotherapy. *Lung Cancer* **122** 234–242.

[31] LITTLE, R. J. and RUBIN, D. B. (2019) *Statistical Analysis with Missing Data* **793**. John Wiley & Sons. https://doi.org/10.1002/9781119013563. MR1925014

[32] LIU, D., LIU, R. Y. and XIE, M.-G. (2014). Exact meta-analysis approach for discrete data and its application to 2 × 2 tables with rare events. *Journal of the American Statistical Association* **109**(508) 1450–1465. https://doi.org/10.1080/01621459.2014.946318. MR3293603

[33] LUFT, H. S. and BROWN JR, B. (1993). Calculating the probability of rare events: why settle for an approximation? *Health Services Research* **28**(4) 419.

[34] MA, Y., CHU, H. and MAZUMDAR, M. (2016). Meta-analysis of proportions of rare events–a comparison of exact likelihood methods with robust variance estimation. *Communications in Statistics – Simulation and Computation* **45**(8) 3036–3052. https://doi.org/10.1080/03610918.2014.911901. MR3514856

[35] MATHES, T. and KUSS, O. (2018). A comparison of methods for meta-analysis of a small number of studies with binary outcomes. *Research Synthesis Methods* **9**(3) 366–381.

[36] MAVRIDIS, D., CHAIMANI, A., EFTHIMIOU, O., LEUCHT, S. and SALANTI, G. (2014). Addressing missing outcome data in meta-analysis. *Evidence-Based Mental Health* **17**(3) 85–89.

[37] NANDA, R., CHOW, L. Q., DEES, E. C., BERGER, R., GUPTA, S., GEVA, R., PUSZTAI, L., PATHIRAJA, K., AKTAN, G., CHENG, J. D. et al. (2016). Pembrolizumab in patients with advanced triple-negative breast cancer: phase Ib KEYNOTE-012 study. *Journal of Clinical Oncology* **34**(21) 2460.

[38] PEIZER, D. B. and PRATT, J. W. (1968). A normal approximation for binomial, F, beta, and other common, related tail probabilities, I. *Journal of the American Statistical Association* **63**(324) 1416–1456. MR0235650

[39] PIGOTT, T. D. Methods for handling missing data in research synthesis. In *The Handbook of Research Synthesis* 163–176 (1994).

[40] PLUMMER, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing* **124**. Vienna, Austria.

[41] PLUMMER, M. (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics* **9**(3) 523–539.

[42] PLUMMER, M. (2012). SOURCEFORGE JAGS: Just Another Gibbs Sampler *Help Forum: dinterval()*. https://sourceforge.net/p/mcmc-jags/discussion/610037/thread/fd7f3f7e/.

[43] PLUMMER, M. (2017). *JAGS version 4.3. 0 user manual*.

[44] POWLES, T., DURÁN, I., VAN DER HEIJDEN, M. S., LORIOT, Y., VOGELZANG, N. J., DE GIORGI, U., OUDARD, S., RETZ, M. M., CASTELLANO, D., BAMIAS, A. et al. (2018). Atezolizumab versus chemotherapy in patients with platinum-treated locally advanced or metastatic urothelial carcinoma (IMvigor211): a multicentre, open-label, phase 3 randomised controlled trial. *The Lancet* **391**(10122) 748–757.

[45] QI, X., ZHOU, S. and PLUMMER, M. (2022). On Bayesian modeling of censored data in JAGS. *BMC Bioinformatics* **23**(1) 1–13.

[46] ROBERT, C., SCHACHTER, J., LONG, G. V., ARANCE, A., GROB, J. J., MORTIER, L., DAUD, A., CARLINO, M. S., MCNEIL, C., LOTEM, M. et al. (2015). Pembrolizumab versus ipilimumab in advanced melanoma. *New England Journal of Medicine* **372**(26) 2521–2532.

[47] SILVA, M. A., SWANSON, A. C., GANDHI, P. J. and TATARONIS, G. R. (2006). Statin-related adverse events: a meta-analysis. *Clinical Therapeutics* **28**(1) 26–35.

[48] SIMMONDS, M. C., HIGGINSA, J. P., STEWARTB, L. A., TIERNEYB, J. F., CLARKE, M. J. and THOMPSON, S. G. (2005). Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clinical Trials* **2**(3) 209–217.

[49] SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. and VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(4) 583–639. https://doi.org/10.1111/1467-9868.00353. MR1979380

[50] STOTO, M. (2015). Drug safety meta-analysis: promises and pitfalls. *Drug Safety* **38**(3) 233–243.

[51] SUTTON, A. J., COOPER, N. J., LAMBERT, P. C., JONES, D. R., ABRAMS, K. R. and SWEETING, M. J. (2002). Meta-analysis of rare and adverse event data. *Expert Review of Pharmacoeconomics & Outcomes Research* **2**(4) 367–379.

[52] SWEETING, M. J., SUTTON, A. J. and LAMBERT, P. C. (2004). What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine* **23**(9) 1351–1375.

[53] VEHTARI, A., GELMAN, A. and GABRY, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* **27**(5) 1413–1432. https://doi.org/10.1007/s11222-016-9696-4. MR3647105

[54] WANG, Y., ZHOU, S., YANG, F., QI, X., WANG, X., GUAN, X., SHEN, C., DUMA, N., AGUILERA, J. V., CHINTAKUNTLAWAR, A. et al. (2019). Treatment-related adverse events of PD-1 and PD-L1 inhibitors in clinical trials: a systematic review and meta-analysis. *JAMA Oncology* **5**(7) 1008–1019.

[55] WATANABE, S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* **11**(12) 3571–3594 (2010). MR2756194

[56] WHITE, I. R., HIGGINS, J. P. and WOOD, A. M. (2008). Allowing for uncertainty due to missing data in meta-analysis—part 1: two-stage methods. *Statistics in Medicine* **27**(5) 711–727. https://doi.org/10.1002/sim.3008. MR2418509

[57] WHITE, I. R., WELTON, N. J., WOOD, A. M., ADES, A. and HIGGINS, J. P. (2008). Allowing for uncertainty due to missing data in meta-analysis—part 2: hierarchical models. *Statistics in Medicine* **27**(5) 728–745. https://doi.org/10.1002/sim.3007. MR2418510

[58] ZHANG, Z. and SUN, J. (2010). Interval censoring. *Statistical Methods in Medical Research* **19**(1) 53–70. https://doi.org/10.1177/0962280209105023. MR2744492

[59] ZHONG, L., ALTAN, M., SHANNON, V. R. and SHESHADRI, A.

(2020). Immune-related adverse events: pneumonitis. *Immunotherapy* **1244** 255–269.

[60] ZHOU, S. (2023). Posterior Averaging Information Criterion. *Entropy* **25**(3) 468. https://doi.org/10.3390/e25030468. MR4575164

[61] ZHOU, S. and SHEN, C. (2022). Statistical concerns for meta-analysis of rare events and small sample sizes. *The Lancet Infectious Diseases* **22**(8) 1111–1112.

Xinyue Qi. Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

Shouhao Zhou. Department of Public Health Sciences, Pennsylvania State University College of Medicine, Hershey, PA, USA. E-mail address: szhou1@pennstatehealth.psu.edu

Christine B. Peterson. Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

Yucai Wang. Department of Hematology, Mayo Clinic, Rochester, MN.

Xinying Fang. Department of Public Health Sciences, Pennsylvania State University College of Medicine, Hershey, PA, USA.

Michael L. Wang. Department of Lymphoma, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

Chan Shen. Department of Surgery, Pennsylvania State University College of Medicine, Hershey, PA, USA.