Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Spike-and-Slab Prior Distributions in Bayesian Logistic Meta-analysis

Thomas A. Gibson* | Robert E. Weiss

[1]Department of Biostatistics, University of California, Los Angeles, California, USA

Correspondence
*Thomas Gibson Email: tommy.a.gibson@gmail.com

Summary

This is sample abstract text this is sample abstract text this is sample abstract text this is sample abstract text this is sample abstract text this is sample abstract text this is sample abstract tex
t this is sample abstract text.

KEYWORDS:
keyword1, keyword2, keyword3, keyword4

1 | INTRODUCTION

Data from medical studies can often be tabulated appropriately in a 2×2 contingency table. Usually the table has columns stratified by a dichotomous outcome and rows stratified by a dichotomous covariate. There are many summary statistics one might calculate from a 2×2 contingency table, such as positive/negative predictive value (PPV/NPV), sensitivity and specificity, and positive and negative likelihood ratios (LR+/LR-), among others. We will refer to any statistic that can be calculated using some or all of the four observed values in a 2×2 contingency table as a *contingency table statistic* (CTS). For an individual study's table, this would mean calculating statistics using the counts in each cell. For a population it would mean calculating statistics using the underlying multinomial cell probabilities. CTSs describe the relationship between the outcome and the covariate, and calculating most CTSs requires conditioning on either rows or columns. Meta-analysis methods for contingency table data reflect this conditioning, and differ based on the statistics we are interested in.

An area of medical literature particularly suited to generating 2×2 data is emergency department (ED) visits for syncope (fainting), where around 5-10% of older syncope patients experience an adverse event in the 30 days after their initial ED visit[1].

[0]**Abbreviations:** ANA, anti-nuclear antibodies; APC, antigen-presenting cells; IRF, interferon regulatory factor

There are many studies that provide information for 2×2 tables for dichotomous covariates that are regularly collected during an ED visit for syncope patients, including comorbidities, symptoms, and test characteristics. Each covariate is a potential risk factor (RF) for an adverse event, and we want to use meta-analysis to determine how much (if any) diagnostic value each covariate has in predicting short-term adverse events. We would also like to "weed out" those covariates that do not have diagnostic value in predicting adverse events. More generally, we are interested in classifying subjects into higher- and lower-risk categories, which is a ubiquitous task in medicine.

One class of meta-analysis models, originally developed for randomized controlled trials (RCTs), models CTSs that condition on the presence or absence of the RF (rows). Modeled CTSs include relative risks (RRs), risk differences (RD), and positive/negative predictive values (PPV/NPV), though these last two are generally referred to as the absolute risk for each treatment group. A standard random effects model for these types of statistics is given in **(author?)** [2], where the main inference target is the log-odds ratio log(OR). **(author?)** [3] provides a model to explicitly model both RRs and RDs. These methods condition on group $j = 0, 1$ where $j = 1$ when the RF is present and $j = 0$ when it is absent, and use binomial likelihoods, where the number of events y_{ij} in study i group j depends on the number of people n_{ij} in group j, and the probability of an event in that group, π_{ij}. Another class of meta-analysis models focuses on statistics that condition on event status (columns), such as sensitivity, specificity, and positive/negative likelihood ratios (LR+/LR-). **(author?)** [4] provides a review of models designed for this type of data, including the summary receiver operating characteristic (SROC) curve [5,6] and bivariate random effects models where estimates of sensitivity and specificity are the data [7,8,9].

With syncope data we are interested in both classes of statistics mentioned above, as is true for *observational data* where neither row totals or column totals of the 2×2 tables are fixed by study investigators. We propose a novel Bayesian meta-analysis model with 3 random effects (3 RE) as an extension of the model in **(author?)** [2], with random effects on the log-odds of an event, the log(OR), and the log-odds of having the risk factor. By modeling these three factors together we can obtain the posterior distribution of any CTS regardless of whether it conditions on rows or columns. The model proposed by **(author?)** [10] models the probability of a positive diagnostic test simultaneously with sensitivity and specificity, which enables the user to obtain estimates of PPV and NPV. We will use a fully Bayesian approach which has advantages in interpretation and flexibility. Additionally, the model in **(author?)** [10] estimates medians for both study-specific and global parameters. Within our Bayesian approach we can easily calculate means of study-specific parameters and we define a novel estimand for global parameters, the expected value of a given statistic for a new study, which accounts for all appropriate variability, and we outline a procedure to get the posterior distribution of the estimand.

Because there are a large number of regularly measured covariates for syncope patients, we would like to determine which covariates have no diagnostic value in predicting adverse events. A natural scientific question in random effects meta-analysis is whether or not the mean effect size for a given covariate (here a log(OR)) is different from zero [11]. Classical meta-analyses answer

this question using Wald statistics [11,12], or with a method proposed by **(author?)** [13] using maximum likelihood, which accounts for uncertainty in the estimation of heterogeneity parameters. We introduce a spike-and-slab prior distribution [14,15,16,17] on the mean parameter for random effects on the log(OR) in the 3 RE model, which allows us to calculate the posterior probability that the null hypothesis is true. We will borrow the spike-and-slab prior outlined in **(author?)** [16], which was originally developed for variable selection models. In regression, it multiplies regression coefficients by indicator variables, so that when the indicator is 0 the regression coefficient is 0 (the spike), and when the indicator is 1 the regression coefficient is non-zero and is modeled a priori with a continuous prior distribution (the slab). To our knowledge, spike-and-slab priors have not been used in the Bayesian meta-analysis literature.

The rest of the paper is organized as follows. Section 2 presents the 3 RE meta-analysis model and details where a spike-and-slab prior can be used. Section 3 presents an example using data from an existing meta-analysis [1] of patients presenting to the emergency department (ED) with syncope. Section 4 presents a simulation.

2 | METHODS

2.1 | Meta-analysis model

In the usual meta-analysis, each paper $i = 1, \ldots, S$ in the meta-analysis provides a 2×2 table with rows defined by the presence (RF), $j = 1$, or absence ($\overline{\text{RF}}$), $j = 0$, of a risk factor and columns defined by adverse event (E) or no adverse event ($\overline{\text{E}}$) as illustrated in Table 1. Let n_{ij} be the number of people in study i, group $j = 0, 1$, y_{ij} be the number of people with an adverse event in study i, group j, and π_{ij} be the probability of an adverse event for a patient in study i, group j. Assuming binomial sampling, the standard Bayesian random effects meta-analysis model is

$$y_{ij}|\pi_{ij} \sim \text{Bin}(n_{ij}, \pi_{ij}) \tag{1}$$

$$\text{logit}(\pi_{ij}) = \begin{cases} \beta_i - \frac{\delta_i}{2} & j = 0 \\ \beta_i + \frac{\delta_i}{2} & j = 1, \end{cases} \tag{2}$$

where $\text{logit}(a) = \log(a/(1-a))$, β_i is a random intercept term for the average log-odds of the event for study i unweighted by group size, and δ_i is a random effect for the log(OR) of the event in study i. Giving each study its own β_i allows the log-odds of an event to be different for each study i, as we might expect due to population and methodology differences between studies, and giving each study its own δ_i models a random study by RF interaction effect. We model δ_i and β_i as normal

$$\delta_i|\delta_0, \sigma_\delta^2 \sim \text{N}(\delta_0, \sigma_\delta^2) \tag{3}$$

$$\beta_i|\beta_0, \sigma_\beta^2 \sim \text{N}(\beta_0, \sigma_\beta^2), \tag{4}$$

and unknown hyperparameters δ_0, β_0, σ_δ^2, and σ_β^2 have appropriate prior distributions which we discuss later.

We now expand the usual Bayesian meta-analysis and add a random effect $\psi_i = \text{P(RF)}$ for risk probability in study i. Let N_i be the total sample size for study i, $N_i = n_{i0} + n_{i1}$. Defining $\nu_i = \text{logit}(\psi_i)$, we assume binomial sampling for n_{i1} and model ν_i as normal

$$n_{ij}|\psi_i \sim \text{Bin}(N_i, \psi_i) \tag{5}$$

$$\nu_i|\nu_0, \sigma_\nu^2 \sim \text{N}(\nu_0, \sigma_\nu^2), \tag{6}$$

where the unknown hyperpriors ν_0 and σ_ν^2 have given appropriate prior distributions $f(\nu_0)$ and $f(\sigma_\nu^2)$.

2.2 | Estimating CTSs

We can calculate two versions of each CTS: study-specific and *global*. Let $\boldsymbol{Y}$ be our data and let $\boldsymbol{\theta}_i = (\beta_i, \delta_i, \nu_i)$ be the vector of study-specific parameters. Any study-specific CTS_i can be calculated by first using the expit function $\text{expit}(x) = 1/(1+\exp(-x))$ to calculate π_{i1}, π_{i0}, and ψ_i as

$$\begin{aligned}\pi_{i1} &= \text{expit}(\beta_i + \delta_i/2)\\ \pi_{i0} &= \text{expit}(\beta_i - \delta_i/2).\\ \psi_i &= \text{expit}(\nu_i),\end{aligned}$$

and CTSs that condition on rows (RF), such as PPV, NPV, RD and RR for each study i are

$$\begin{aligned}\text{PPV}_i &= \text{P(E|RF)} = \pi_{i1} \qquad & \text{NPV}_i &= \text{P}(\overline{\text{E}}|\overline{\text{RF}}) = 1 - \pi_{i0}\\ \text{RR}_i &= \frac{\text{P(E|RF)}}{\text{P(E}|\overline{\text{RF}})} = \frac{\pi_{i1}}{\pi_{i0}} \qquad & \text{RD}_i &= \text{P(E|RF)} - \text{P(E}|\overline{\text{RF}}) = \pi_{i1} - \pi_{i0},\end{aligned}$$

and CTSs that condition on columns (E), such as sensitivity (Sens), specificity (Spec), and LR+/− are

$$\begin{aligned}\text{Sens}_i &= \text{P(RF|E)} = \frac{\pi_{i1}\psi_i}{\pi_{i1}\psi_i + \pi_{i0}(1-\psi_i)} \qquad & \text{LR}-_i &= \frac{1-\text{Sens}_i}{\text{Spec}_i}\\ \text{Spec}_i &= \text{P}(\overline{\text{RF}}|\overline{\text{E}}) = \frac{(1-\pi_{i0})(1-\psi_i)}{(1-\pi_{i0})(1-\psi_i) + (1-\pi_{i1})\psi_i} \qquad & \text{LR}+_i &= \frac{\text{Sens}_i}{1-\text{Spec}_i}.\end{aligned}$$

Each CTS_i is then a function of the study-specific parameters $g(\boldsymbol{\theta}_i)$, and we will use $g(\cdot)$ as shorthand for a function to calculate a CTS.

Most often we are more interested in global means of parameters rather than study-specific parameters. We might try replacing the study-specific parameters $\boldsymbol{\theta}_i$ with the hyperparameters $\text{E}[\boldsymbol{\theta}_i] = \boldsymbol{\theta}_0 = (\beta_0, \delta_0, \nu_0)$ in calculations. However, this ignores that

every CTS_i is a nonlinear transformation of the parameters $\boldsymbol{\theta}_i$, and $\text{E}[g(\boldsymbol{\theta}_i)|\boldsymbol{Y}] \neq g(\text{E}[\boldsymbol{\theta}_i|\boldsymbol{Y}])$, and $g(\boldsymbol{\theta_0})$ actually represents the *median* global parameter value.

Say $\boldsymbol{\gamma} = (\beta_0, \delta_0, \nu_0, \sigma^2_\beta, \sigma^2_\delta, \sigma^2_\nu)$ is the vector of hyperparameters. We then define our target estimand as $\text{CTS}_0 = \text{E}[g(\boldsymbol{\theta}_{S+1})|\boldsymbol{\gamma}, \boldsymbol{Y}]$, the *expected value of a CTS for a new study*. We need to obtain the posterior distribution of $\text{E}[g(\boldsymbol{\theta}_{S+1})|\boldsymbol{\gamma}, \boldsymbol{Y}] = \int g(\boldsymbol{\theta}_{S+1}) P(\boldsymbol{\theta}_{S+1}|\boldsymbol{\gamma}, \boldsymbol{Y}) d\boldsymbol{\theta}_{S+1}$ by integrating out $\boldsymbol{\theta}_{S+1}$, where $\boldsymbol{\theta}_{S+1} = (\beta_{S+1}, \delta_{S+1}, \nu_{S+1})$ and

$$\beta_{S+1}|\beta_0, \sigma^2_\beta \sim \text{N}(\beta_0, \sigma^2_\beta)$$

$$\delta_{S+1}|\delta_0, \sigma^2_\delta \sim \text{N}(\delta_0, \sigma^2_\delta)$$

$$\nu_{S+1}|\nu_0, \sigma^2_\nu \sim \text{N}(\nu_0, \sigma^2_\nu)$$

are the posterior predictive distributions for a new study. We can approximate the integral with a Monte Carlo calculation within Markov chain Monte Carlo (MCMC). Say we have M iterations of MCMC, and in each iteration we sample $\boldsymbol{\gamma}^{(m)}$, $m = 1, \ldots, M$. For each m, we

1. Take L draws $\boldsymbol{\theta}^{(m,l)}_{S+1}$, $l = 1, \ldots, L$ from the predictive distribution $P(\boldsymbol{\theta}_{S+1}|\boldsymbol{\gamma}^{(m)}, \boldsymbol{Y})$ and calculate the CTS $g(\boldsymbol{\theta}^{(m,l)}_{S+1})$ for each of the L draws,
2. Estimate $\text{E}[g(\boldsymbol{\theta}_{S+1})|\boldsymbol{\gamma}^{(m)}, \boldsymbol{Y}] \approx \frac{1}{L} \sum_{l=1}^{L} g(\boldsymbol{\theta}^{(m,l)}_{S+1})$

Generating this estimate in each iteration of MCMC sampling gives us the posterior distribution for the expected value of the CTS for a new study, $\text{E}[g(\boldsymbol{\theta}_{S+1})|\boldsymbol{Y}, \boldsymbol{\gamma}]$. By obtaining the posterior distribution of the expected value of a CTS, we can then generate CIs, along with summary measures such as the probability that a CTS_0 is greater than or less than a certain threshold value.

2.3 | Spike-and-slab prior for the log-odds ratio

A common question in random effects meta-analysis is whether or not the global mean effect size δ_0 is zero. We propose a spike-and-slab prior$p_1(\delta_0)$ (SS prior) for δ_0

$$\delta_0 = \begin{cases} \delta & \rho = 1 \\ 0 & \rho = 0 \end{cases} \tag{7}$$

$$\rho \sim \text{Bernoulli}(p) \tag{8}$$

$$\delta \sim \text{N}(0, b^2_\delta), \tag{9}$$

and refer to the 3RE model with this prior as the SS prior model. We can estimate the posterior probability that δ_0 is exactly zero by looking at the posterior mean of $1 - \rho$. The prior standard deviation b_δ should be set to give support to values $\in (-4, 4)$,

where a log-odds ratio of 4 corresponds to a change in probability from roughly 0.02 to 0.5 or 0.5 to 0.98. A special case of this prior, where the prior probability p on the spike ρ is 1, has δ_0 following a normal distribution $p_2(\delta_0)$ with zero mean and known variance

$$\delta_0 \sim \text{N}(0, b_\delta^2). \tag{10}$$

In the absence of of heterogeneity in the log(OR) ($\sigma_\delta^2 = 0$), $\delta_0 = 0$ corresponds to a lack of diagnostic value (i.e. $\text{E}[\text{RD}_0] = 0$, $\text{E}[\text{RR}_0] = 1$, $\text{E}[\text{LR}+_0] = 1$ and $\text{E}[\text{LR}-_0] = 1$). In the presence of heterogeneity in the log(OR) ($\sigma_\delta^2 \neq 0$) these expected values will not exactly equal their null values, but will generally be close.

2.4 | Other prior distributions

We must assign the additional hyperparameters $(\beta_0, \nu_0, \sigma_\beta^2, \sigma_\delta^2, \sigma_\nu^2)$ appropriate prior distributions. For the mean parameters β_0 and ν_0 we propose normal distributions with known means and variances

$$\beta_0 \sim \text{N}(a_\beta, b_\beta^2) \tag{11}$$

$$\nu_0 \sim \text{N}(a_\nu, b_\nu^2). \tag{12}$$

Prior means a_β and a_ν are prior guesses at the mean log-odds of an event and log-odds of having the risk factor, respectively, and the standard deviations b_β and b_ν should also be chosen to be large enough to give support to all plausible values of the parameters. As a default we assign each of the prior standard deviations σ_β, σ_δ, and σ_ν weakly informative half-Cauchy prior distributions

$$\sigma_\beta \sim \text{half-Cauchy}(A_\beta) \tag{13}$$

$$\sigma_\delta \sim \text{half-Cauchy}(A_\delta) \tag{14}$$

$$\sigma_\nu \sim \text{half-Cauchy}(A_\nu) \tag{15}$$

where $y \sim \text{half-Cauchy}(A)$ has density $p(y) \propto (A^2 + y^2)^{-1}$ **(author?)** [18]. The scale parameters A_β, A_δ, and A_ν should generally be set between 0.5 and 1, depending on how much information is in the data, to prevent large values of the standard deviations. Standard deviations above 1.5 may signal problems with the model fit or the data. If there are few studies (< 5) then we recommend an informative inverse-gamma (IG) prior on the variances

$$\sigma_\beta^2 \sim \text{IG}(c_\beta, d_\beta)$$

$$\sigma_\delta^2 \sim \text{IG}(c_\delta, d_\delta)$$

$$\sigma_\nu^2 \sim \text{IG}(c_\nu, d_\nu)$$

where $(c_\cdot, d_\cdot)$ are chosen to heavily restrict large values of standard deviations, such as (4, 2) or (3, 2). With very few studies (2 or 3), large standard deviations can lead to estimation problems for the statistics LR+ or LR−.

2.5 | Special case: fixed effect for the log-odds ratio

If we assume a fixed effect for the log-odds ratio δ_0 and use the SS prior then the posterior height of the spike is the probability that the log-odds ratio is zero for every study in the analysis. In the model, we change (2) to

$$\text{logit}(\pi_{ij}) = \begin{cases} \beta_i - \frac{\delta_0}{2} & j = 0 \\ \beta_i + \frac{\delta_0}{2} & j = 1, \end{cases} \tag{16}$$

where δ_0 has the SS prior (7) - (9). Conveniently, the height of the spike is then also the probability that a number of CTS_0's are equal to their null values. These include $\text{RD}_0 = 0$, $\text{RR}_0 = 1$, $\text{LR+}_0 = 1$, and $\text{LR−}_0 = 1$. We need not worry about integrating out the random effects β_{S+1} and ν_{S+1} because $\delta_0 = 0$ implies $\pi_{[S+1]1} = \pi_{[S+1]0}$, and the CTSs no longer depend on any of the random effects.

3 | SYNCOPE DATA EXAMPLE: ASSESSING DIAGNOSTIC UTILITY OF REGULARLY MEASURED COVARIATES

We apply the SS prior model to studies on patients presenting to the emergency department with syncope. The outcome is 30-day adverse events including serious cardiac events and death. The potential risk factors associated with 30-day adverse events are covariates measured in the ED, which include demographics/comorbidities, symptoms, physical findings, and biomarkers. There are 17 studies which each report information on some but not all covariates; we meta-analyze 31 covariates for which at least 2 papers provided a 2×2 table. Each analysis is done in a two-step process, where we first assess the posterior height of the spike at zero; if the posterior probability $\text{P}(\delta_0 = 0) \leq 0.25$ then we re-run the model with a normal prior on δ_0 and estimate CTS_0 for LR+_0, LR−_0, PPV_0, NPV_0, Sens_0, and Spec_0 as described in Section 2.2. If the posterior probability $\text{P}(\delta_0 = 0) > 0.25$ then we forego computing global estimates of CTSs as they will be close to their null values. As we will show in Section 4, the SS prior accumulates evidence in favor of $\delta_0 \neq 0$ more quickly than for $\delta_0 = 0$, which is why we use a cutoff of 0.25 instead of something like 0.5 which is advocated for in variable selection situations [19].

Table 2 details results. Of the 31 variables analyzed 11 had posterior probability $\text{P}(\delta_0 = 0) \leq 0.25$ and 20 had $\text{P}(\delta_0 = 0) > 0.25$. We see that variables with more studies have smaller posterior CIs, and that variables with estimates of specificity nearing 1 tend to have wider CIs for LR+_0. None of the variables have an $\text{LR+}_0 > 6$, which highlights the difficulty physicians face in determining which syncope patients are at high risk of an adverse event. The biomarkers troponin, urea, and creatinine appear to be promising diagnostic predictors of adverse events, but there is plenty of uncertainty around their estimates.

4 | SIMULATION STUDY

We perform a series of three simulations to determine

1. how well the SS prior model identifies true zero effects ($\delta_0 = 0$),
2. how well the SS prior model identifies true non-zero effects ($\delta_0 \neq 0$),
3. how well the 3RE model without the SS prior estimates CTSs.

We fix the values of the hyperparameters $\beta_0 = \nu_0 = \text{logit}(0.15)$ and $\sigma_\beta^2 = \sigma_\nu^2 = 0.4^2$ for all three simulations. These values are loosely based off of the syncope data example.

4.1 | Simulation 1: true zero effects

For Simulation 1, we fix the mean hyperparameter $\delta_0 = 0$. To see how the model reacts to varying levels of heterogeneity in the effect size we do separate simulations with $\sigma_\delta = (0.1, 0.25, 0.5)$. For each of the 3 values of σ_δ we generated $K_1 = 1000$ datasets, where every dataset had 10 studies each with 500 subjects. To generate each study's 2×2 table, we draw the study-specific parameters $(\beta_i, \delta_i, \nu_i)$ from independent normal distributions, which define each study's multinomial probabilities for each cell, and we take random samples of size 500 from the multinomial distributions. For each iteration we run the SS prior model with the 10 generated studies and calculate the posterior height of the spike as the proportion of MCMC samples in which $\rho = 0$.

There are two inference targets: the average posterior height of the spike at zero, and the probability of the spike being above a moving threshold $\in (0, 1)$ for each value of σ_δ. The leftmost boxplot in each panel of Figure 1 shows the distribution of spike heights for each value of σ_δ, and we report the mean spike height for each scenario in Table 3 under the column $\delta_0 = 0$. The upper panel of Figure 2 shows the proportion of K_1 simulation iterations where the posterior spike was higher than corresponding cutoff values on the x-axis. Intuitively, for every value of σ_δ the spike is almost always larger than cutoff values close to zero and almost never larger than cutoff values close to 1. Smaller values of σ_δ make it easier for the model to catch true zeros.

4.2 | Simulation 2: true non-zero effects

In Simulation 2 we vary both $\delta_0 = (1, 2)$ and $\sigma_\delta = (0.1, 0.25, 0.5)$ and use each combination of the two factors for 6 total situations. For each situation we generate $K_2 = 1000$ datasets with 10 studies each with 500 subjects, drawing $(\beta_i, \delta_i, \nu_i)$ from independent normal distributions to determine each study's multinomial probabilities and taking random samples from their multinomial distributions. Again for each iteration we run the SS prior model with the generated dataset and calculate the posterior height of the spike.

Our inference targets are the same as in Simulation 1, assessing the average posterior height of the spike and how often the spike is above a moving threshold $\in (0, 1)$. These will tell us how well the model recognizes true non-zero effects, where small values of the spike signal a likely non-zero mean effect. The middle and right boxplots of each panel in Figure 1 show the distributions for each situation, and the mean spike height for each situation is presented in Table 3. Spike heights for every situation are concentrated near zero in every situation. The lower panel of Figure 2 shows the proportion of K_2 iterations in which the spike at zero was greater than the corresponding cutoff on the x-axis. We see that for $\delta_0 = 2$ the SS prior model nearly always correctly identifies a true nonzero effect, even for minuscule cutoff values and moderate standard deviation of random effects σ_δ. Even in the lowest-powered situation, $\delta_0 = 1$ and $\sigma_\delta = 0.5$, the SS prior model correctly identifies a non-zero mean effect with 98% accuracy using a cutoff of 0.25.

4.3 | Simulation 3: estimating CTS_0

Simulation 3 details how accurately the 3RE model without the SS prior estimates the CTSs LR+, LR−, PPV, NPV, Sens, and Spec. We fix $\delta_0 = 2$ and $\sigma_\delta = 0.4$. We determine the number of simulation iterations K_3 to achieve Monte Carlo standard error (MCSE) of bias for LR+ ≤ 0.01. The formula for MCSE(bias) is

$$\text{MCSE(bias)} = \sqrt{Var(\hat{\theta})/K},$$

where $\hat{\theta}$ corresponds to the point estimate of the estimand (here LR+). A modest number of iterations suggests that $\text{Var}(\widehat{\text{LR}}+) \leq 0.25$, which leads to $K_3 = 2500$ iterations. For each iteration $k = 1, \ldots, K_3$ we generate 10 studies each with 500 subjects as in Simulations 1 and 2, and record the point estimate, SD, and 95% CI for every CTS. Using all 2500 results, we calculate bias, variance, average SD, 95% coverage, average 95% CI length, root mean-squared error (RMSE), and MCSE of bias for every CTS. Results are summarized in Table 4. We see that for every CTS, bias is either not different from zero, or is different from zero in a negligibly small magnitude.

References

1. Gibson TA, Weiss RE, Sun BC. Predictors of short-term outcomes after syncope: a systematic review and meta-analysis. *Western Journal of Emergency Medicine* 2018; 19(3): 517.

2. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Statistics in Medicine* 1995; 14(24): 2685–2699.

3. 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. *Statistics in medicine* 2002; 21(11): 1601–1623.

4. Ma X, Nie L, Cole SR, Chu H. Statistical methods for multivariate meta-analysis of diagnostic tests: an overview and tutorial. *Statistical Methods in Medical Research* 2016; 25(4): 1596–1619.

5. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine* 2001; 20(19): 2865–2884.

6. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Statistics in Medicine* 1993; 12(14): 1293–1316.

7. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005; 58(10): 982–990.

8. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology* 2006; 59(12): 1331.

9. Arends L, Hamza T, Van Houwelingen J, Heijenbrok-Kal M, Hunink M, Stijnen T. Bivariate random effects meta-analysis of ROC curves. *Medical Decision Making* 2008; 28(5): 621–638.

10. Chu H, Nie L, Cole SR, Poole C. Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: alternative parameterizations and model selection. *Statistics in Medicine* 2009; 28(18): 2384–2399.

11. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2009; 172(1): 137–159.

12. Follmann DA, Proschan MA. Valid inference in random effects meta-analysis. *Biometrics* 1999; 55(3): 732–737.

13. Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* 1996; 15(6): 619–629.

14. George EI, McCulloch RE. Variable selection via Gibbs sampling. *Journal of the American Statistical Association* 1993; 88(423): 881–889.

15. George EI, McCulloch RE. Approaches for Bayesian variable selection. *Statistica Sinica* 1997; 7(2): 339–373.

16. Kuo L, Mallick B. Variable selection for regression models. *Sankhyā: The Indian Journal of Statistics, Series B* 1998; 60(1): 65–81.

17. Ishwaran H, Rao JS, others . Spike and slab variable selection: frequentist and Bayesian strategies. *Annals of Statistics* 2005; 33(2): 730–773.

18. Gelman A, others . Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis* 2006; 1(3): 515–534.

19. Barbieri MM, Berger JO, others . Optimal predictive model selection. *The annals of statistics* 2004; 32(3): 870–897.

	No Event	Event	Total
j = 0, RF Absent	$n_{i0} - y_{i0}$	y_{i0}	n_{i0}
j = 1, RF Present	$n_{i1} - y_{i1}$	y_{i1}	n_{i1}
j = 0, RF Absent	$1 - \pi_{i0}$	π_{i0}	1
j = 1, RF Present	$1 - \pi_{i1}$	π_{i1}	1

TABLE 1 Sample contingency table for study i with subject counts (top), a probability representation conditional on presence or absence of the risk factor (middle), and a joint probability representation (bottom)

Variable	Num. Papers	Spike	LR+	LR-	NPV	PPV	Sens	Spec
ECG	6	0.114	2.34 (1.45, 3.82)	0.63 (0.41, 0.92)	0.90 (0.80, 0.95)	0.24 (0.14, 0.39)	0.56 (0.38, 0.73)	0.70 (0.55, 0.80)
Age	6	0.005	2.07 (1.61, 2.71)	0.46 (0.32, 0.64)	0.94 (0.82, 0.98)	0.18 (0.08, 0.36)	0.71 (0.56, 0.82)	0.62 (0.50, 0.73)
Heart Disease	9	0.013	2.23 (1.66, 2.99)	0.79 (0.68, 0.89)	0.91 (0.83, 0.96)	0.18 (0.10, 0.30)	0.33 (0.23, 0.45)	0.84 (0.76, 0.88)
No Prodromes	6	0.877	–	–	–	–	–	–
Male Gender	7	0.002	1.39 (1.29, 1.53)	0.72 (0.63, 0.79)	0.92 (0.81, 0.97)	0.13 (0.06, 0.26)	0.58 (0.54, 0.63)	0.58 (0.57, 0.61)
Cerebrovascular	3	0.749	–	–	–	–	–	–
Arrhythmia	6	0.039	3.00 (1.94, 4.45)	0.78 (0.61, 0.90)	0.93 (0.82, 0.97)	0.18 (0.09, 0.34)	0.31 (0.17, 0.51)	0.87 (0.71, 0.94)
Trauma	2	0.616	–	–	–	–	–	–
Palpitations	4	0.783	–	–	–	–	–	–
Supine	2	0.744	–	–	–	–	–	–
Effort	3	0.800	–	–	–	–	–	–
CHF	8	0.004	3.41 (2.35, 4.94)	0.82 (0.73, 0.90)	0.89 (0.78, 0.95)	0.26 (0.15, 0.42)	0.24 (0.16, 0.33)	0.92 (0.89, 0.95)
Dyspnea	6	0.062	2.82 (1.78, 4.93)	0.88 (0.77, 0.94)	0.86 (0.76, 0.92)	0.28 (0.18, 0.43)	0.19 (0.12, 0.30)	0.92 (0.88, 0.95)
Hematocrit	7	0.310	–	–	–	–	–	–
Hypotension	5	0.222	4.28 (1.75, 9.41)	0.90 (0.77, 0.98)	0.85 (0.74, 0.93)	0.31 (0.17, 0.49)	0.14 (0.06, 0.29)	0.95 (0.89, 0.98)
Oxygen	2	0.764	–	–	–	–	–	–
Hypertension	5	0.562	–	–	–	–	–	–
Diabetes	6	0.345	–	–	–	–	–	–
Prev. Syncope	3	0.809	–	–	–	–	–	–
Nonwhite Race	3	0.653	–	–	–	–	–	–
Chest Pain	3	0.849	–	–	–	–	–	–
Seizure	3	0.845	–	–	–	–	–	–
Stroke	2	0.781	–	–	–	–	–	–
Arr. Medication	2	0.793	–	–	–	–	–	–
Resp. Rate	2	0.409	–	–	–	–	–	–
Murmur	4	0.571	–	–	–	–	–	–
Hispanic	2	0.824	–	–	–	–	–	–
Pacemaker	3	0.751	–	–	–	–	–	–
Troponin	3	0.076	3.63 (1.70, 6.94)	0.67 (0.40, 0.90)	0.92 (0.83, 0.97)	0.24 (0.11, 0.43)	0.44 (0.21, 0.70)	0.81 (0.60, 0.93)
Urea	2	0.130	5.58 (1.78, 13.68)	0.76 (0.47, 0.96)	0.90 (0.75, 0.97)	0.29 (0.11, 0.54)	0.30 (0.10, 0.61)	0.90 (0.71, 0.97)
Creatinine	2	0.246	4.32 (1.42, 10.41)	0.80 (0.53, 0.99)	0.90 (0.74, 0.97)	0.25 (0.09, 0.49)	0.27 (0.08, 0.57)	0.90 (0.70, 0.97)

TABLE 2 Results of 31 meta-analyses of syncope studies

σ_δ	$\delta_0 = 0$	$\delta_0 = 1$	$\delta_0 = 2$
0.1	0.8706	0.0007	0.0000
0.25	0.8466	0.0026	0.0000
0.5	0.7876	0.0297	0.0001

TABLE 3 Average spike heights from simulations 1 and 2

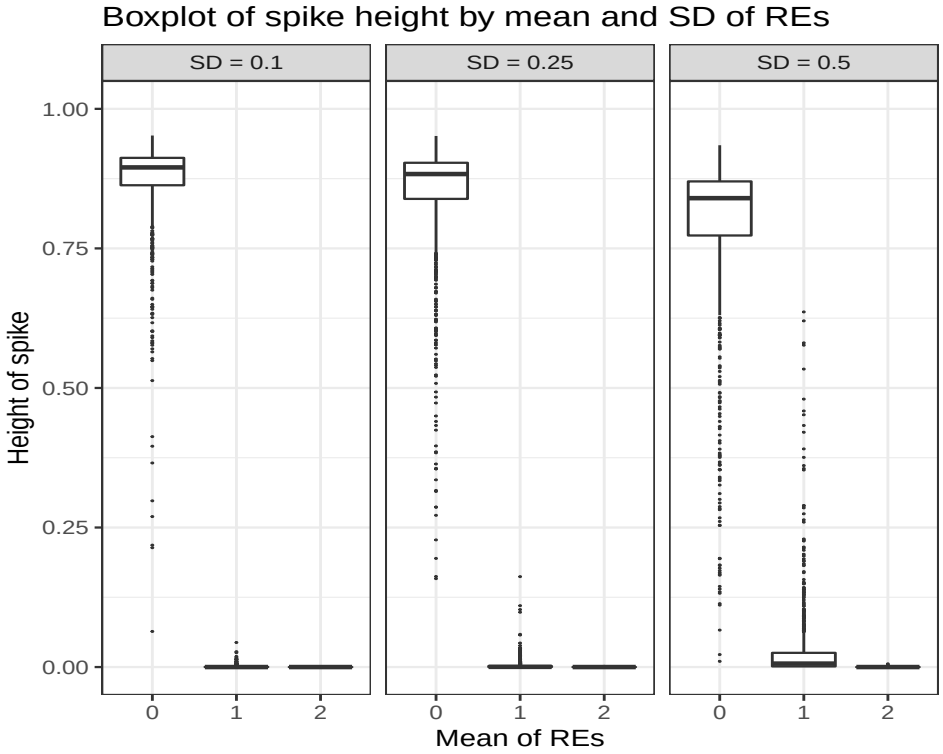

FIGURE 1 Boxplots from Simulations 1 and 2: spike height for each combination of σ_δ and δ_0

	Bias	Variance	Average SD	95% CI Coverage	95% CI Length	root(MSE)	MCSE(bias)
LR−	0.0081	0.0442	0.0511	0.9704	0.2019	0.0449	0.0009
LR+	0.0020	0.4961	0.5815	0.9660	2.2797	0.4960	0.0099
NPV	-0.0052	0.0105	0.0134	0.9616	0.0519	0.0117	0.0002
PPV	0.0026	0.0336	0.0393	0.9732	0.1552	0.0337	0.0007
Sens	-0.0059	0.0423	0.0495	0.9720	0.1950	0.0427	0.0008
Spec	-0.0048	0.0150	0.0188	0.9692	0.0738	0.0158	0.0003

TABLE 4 Results from Simulation 3 with K = 2500 repetitions

How to cite this article: Williams K., B. Hoskins, R. Lee, G. Masato, and T. Woollings (2016), A regime analysis of Atlantic winter jet variability applied to evaluate HadGEM3-GC2, *Q.J.R. Meteorol. Soc.*, *2017;00:1–6*.

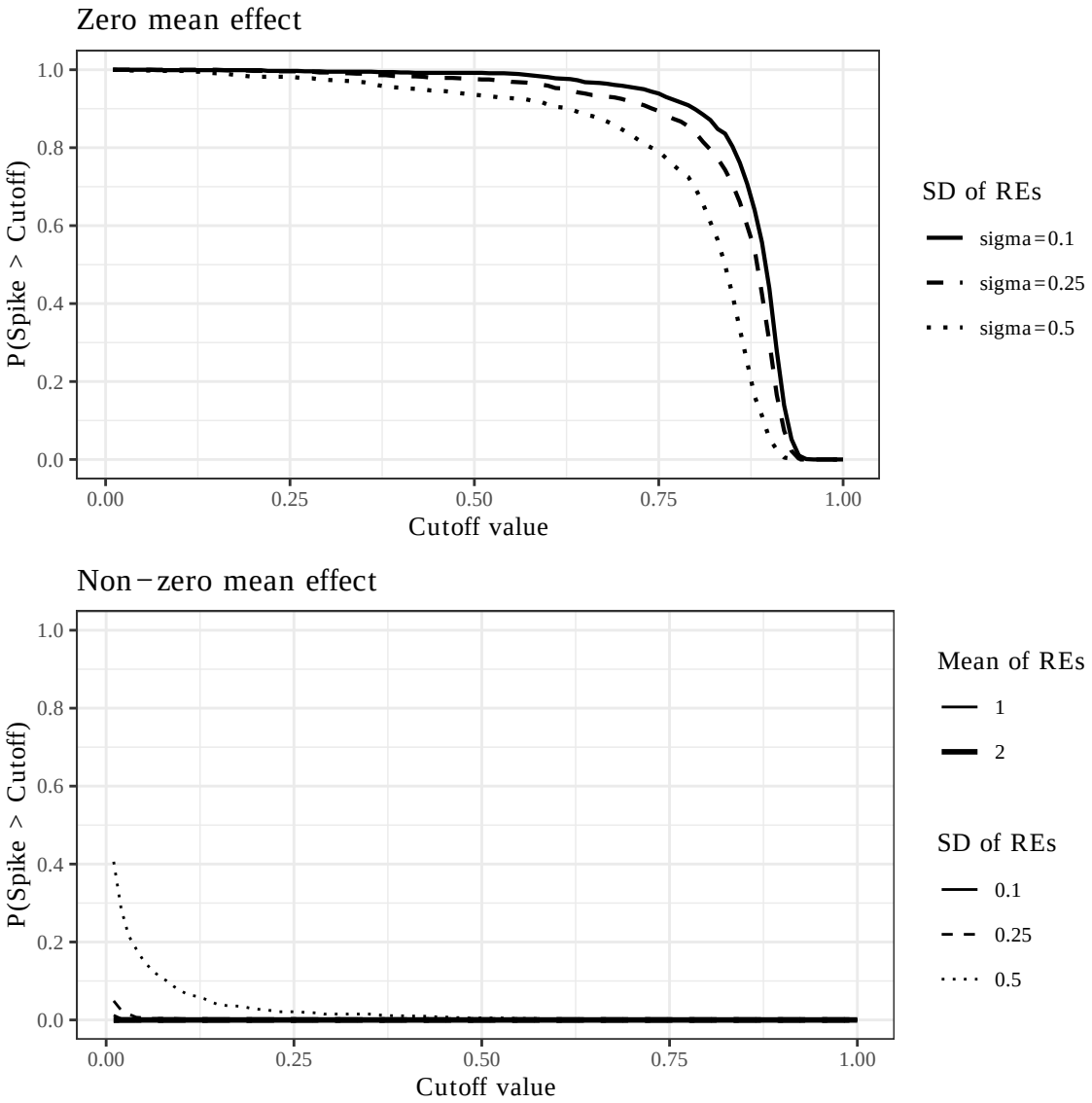

FIGURE 2 Top panel: proportion of Simulation 1 iterations where posterior spike height was higher than cutoff value with true zero mean effect. Bottom panel: proportion of Simuation 2 iterations where posterior spike height was higher than cutoff value with true non-zero mean effect