

ARTICLE TYPE

Bayesian Meta-analysis of Observational Contingency Table Data with a Nested Monte Carlo Procedure for Estimating Global Effects

Thomas A. Gibson* | Robert E. Weiss

¹Department of Biostatistics, University of California, Los Angeles, California, USA**Correspondence***Thomas A. Gibson Email:
tommy.a.gibson@gmail.com**Summary**

Observational contingency table data arises when study investigators fix neither row nor column totals. Meta-analysis methods for 2×2 contingency table data often tend to focus on estimating select subsets of statistics that one can calculate from values in a contingency table, and methods that attempt to measure “global” statistics often use naive estimators that do not take into account heterogeneity parameters for random effects. In this paper we present a novel three-random effect (3RE) model, define an estimand for global statistics that takes into account all appropriate variability, and outline a Monte Carlo-within-Markov chain Monte Carlo procedure for sampling from the posterior distribution of the estimand. In an extensive simulation study we show how the proposed Monte Carlo procedure has superior performance to naive methods, and we apply our method to a dataset of observational studies on patients who present to the emergency department with syncope.

KEYWORDS:

Meta-analysis, diagnostic test, Monte Carlo simulation, syncope

1 | INTRODUCTION

Data from medical studies can often be tabulated in a 2×2 contingency table. The tables have columns stratified by a dichotomous outcome and rows stratified by a dichotomous covariate. Summary statistics from a 2×2 contingency table include positive/negative predictive value (PPV/NPV), sensitivity and specificity (Sens and Spec), and positive and negative likelihood ratios (LR+ and LR-), among others. We will refer to statistics that can be calculated as functions of some or all of the four values in a 2×2 contingency table as a *contingency table statistic* (CTS). For an individual study’s table, this would mean using the counts in each cell to calculate an observed CTS, and for a population it would mean using the underlying multinomial cell probabilities to calculate a population CTS. CTSs describe the relationship between the outcome and the covariate, and calculating most CTSs

⁰**Abbreviations:** CTS, contingency table statistic; PPV, positive predictive value; NPV, negative predictive value; LR, likelihood ratio; RF, risk factor; RE, random effect, E, event.

requires conditioning on either rows or columns. Meta-analysis methods for contingency table data reflect this conditioning, and can generally be segregated into two groups that allow inference for different CTSs.

Meta-analysis models for randomized controlled trials (RCTs) allow inference for *treatment* CTSs (T-CTSs), and naturally condition on the dichotomous covariate treatment/placebo. T-CTSs include the odds ratio (OR), relative risk (RR), risk difference (RD), and positive/negative predictive values (PPV/NPV). A standard random effects model for T-CTSs is given in Smith et al.¹, with the log-odds ratio $\log(\text{OR})$ as the main inference target. The model of Smith et al.¹ has been extended to network meta-analysis with K treatment groups^{2,3,4}. Models for diagnostic tests allow inference for *diagnostic* CTSs (D-CTSs) that condition on the presence or absence of an adverse event, denoted by E or \bar{E} . D-CTSs include sensitivity (Sens), specificity (Spec), ORs, and positive/negative likelihood ratios (LR+/LR-). Ma et al.⁵ reviews meta-analysis models that condition on event status, including the summary receiver operating characteristic (SROC) curve^{6,7,8}, bivariate random effects models^{9,10,11,12,13,14}, and trivariate random effects models^{15,16,17}. Models for T-CTSs and D-CTSs are similar in that they use binomial likelihoods conditioning on rows or columns, respectively. Multiple models^{15,16} aim to estimate *global* statistics, synonymously referred to as “overall”, “summary”, or “population” statistics that are not study-specific.

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 visit¹⁸. Many studies provide 2×2 tables of counts for dichotomous *risk factors* (RFs) that are regularly collected during an ED visit for syncope patients, including demographics, comorbidities, symptoms, and test characteristics. The syncope data is unique in that 1) it is *observational*, with neither row totals nor column totals fixed by study investigators, and for which we are interested in both T-CTSs and D-CTSs, and 2) given the large number (> 30) of regularly-measured covariates in the ED for syncope patients, we would like to “weed out” those covariates that are unrelated to the probability of 30-day adverse events.

To make inference on T- and D-CTSs together we propose a novel 3 random effect (3RE) Bayesian meta-analysis model as an extension of the model in Smith et al.¹, with study-level random effects on the average log-odds of an event, the $\log(\text{OR})$ of the event, and the log-odds of having the risk factor. Existing 3RE models^{15,16,17} model the probability of a positive diagnostic test simultaneously with sensitivity and specificity, but their methods use *plug-in estimators*, plugging in hyperparameters to calculate global statistics, to provide *median* estimates of global T-CTSs and D-CTSs. We instead posit that global *mean* effects are more desirable and define a novel estimand, the *expected value of a given statistic for a new study*, which accounts for all appropriate variability, and we outline a procedure to sample from the posterior distribution of the estimand. We use a fully Bayesian approach which has advantages in interpretation and flexibility, and our parameterization is different in that it builds off of Smith et al.¹ with the $\log(\text{OR})$ as the natural parameter. Whether or not a risk factor is related to 30-day adverse events corresponds to a natural scientific question in random effects meta-analysis of whether or not the mean $\log(\text{OR})$ for a given risk factor is different from zero¹⁹. We introduce a mixture spike-and-slab prior distribution^{20,21,22,23} on the mean parameter for the

random effects on the $\log(\text{OR})$ in the 3RE model, which allows us to calculate the posterior probability that the null hypothesis is true. The mixture prior places point mass on the probability that the mean $\log(\text{OR}) = 0$ (the spike), and if not 0, models uncertainty in the mean $\log(\text{OR})$ with a continuous prior distribution (the slab).

We present the 3RE meta-analysis model, a nested Monte Carlo procedure for calculating global CTSs, and a spike-and-slab prior on the global $\log(\text{OR})$ in Section 2. Section 3 presents a simulation to show how well the model identifies true zero and non-zero effects and how accurately and precisely the nested Monte Carlo procedure estimates common CTSs as compared with plug-in estimators. Section 4 applies the model to our motivating syncope data. The paper closes with discussion.

2 | THREE RE META-ANALYSIS MODEL

In the usual meta-analysis, each study $i = 1, \dots, S$ reports a 2×2 table of counts n_{ijk} with rows $j = 0, 1$ defined by the absence or presence of a risk factor (RF), denoted $\overline{\text{RF}}$ and RF , and columns $k = 0, 1$ defined by no adverse event ($\overline{\text{E}}$) or adverse event (E). Let $n_{i1} = n_{i10} + n_{i11}$ and $n_{i0} = n_{i00} + n_{i01}$ be the number of people with or without the risk factor, respectively, $N_i = n_{i1} + n_{i0}$ be the total sample size in study i , and π_{ij} is the probability of an adverse event for a patient in study i , group j as illustrated in Table 1. Assuming binomial sampling, the standard Bayesian random effects meta-analysis model is

$$n_{ij1} | \pi_{ij} \sim \text{Bin}(n_{ij}, \pi_{ij}) \quad (1)$$

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

where $\text{logit}(a) = \log(a/(1-a))$, $0 < a < 1$, β_i is a random intercept term for the log-odds of the event for study i and δ_i is a random effect for the $\log(\text{OR})$ of the event in study i . We model δ_i and β_i as normal with unknown mean and variance

$$\beta_i | \beta_0, \sigma_\beta^2 \sim \text{N}(\beta_0, \sigma_\beta^2), \quad (3)$$

$$\delta_i | \delta_0, \sigma_\delta^2 \sim \text{N}(\delta_0, \sigma_\delta^2), \quad (4)$$

where population means β_0 and δ_0 and variances σ_β^2 and σ_δ^2 have priors $p(\beta_0)$, $p(\delta_0)$, $p(\sigma_\beta)$, and $p(\sigma_\delta)$ which we discuss in Section 2.3.

With this model we can make inference on T-CTSs. For observational data where we also want to make inference on D-CTSs, we expand model (1) - (4) to include a random effect $\psi_i = \text{P}(\text{RF in study } i)$, the probability of a subject having the risk factor in study i . Assuming binomial sampling of n_{i1} from N_i as $\text{Bin}(N_i, \psi_i)$, we model $\nu_i = \text{logit}(\psi_i)$ as normal with unknown mean

and cross-study variance σ_v^2 on the logit scale

$$n_{i1}|\psi_i \sim \text{Bin}(N_i, \psi_i) \quad (5)$$

$$v_i|\nu_0, \sigma_v^2 \sim N(\nu_0, \sigma_v^2), \quad (6)$$

where the unknown population parameters ν_0 and σ_v^2 have priors $f(\nu_0)$ and $f(\sigma_v^2)$.

2.1 | Predictive contingency table statistics

There are study-specific and *global* versions of each CTS. Let \mathbf{Y} be the data from all S studies, define $\theta_i = (\beta_i, \delta_i, \nu_i)'$, the parameter vector for the i^{th} study, and let $\gamma = (\beta_0, \sigma_\beta, \delta_0, \sigma_\delta, \nu_0, \sigma_v)'$ be the vector of hyperparameters. The unknown study-specific CTS _{i} 's are functions of

$$\begin{aligned} \pi_{i1} &= \text{expit}(\beta_i + \delta_i/2) \\ \pi_{i0} &= \text{expit}(\beta_i - \delta_i/2) \\ \psi_i &= \text{expit}(\nu_i), \end{aligned} \quad (7)$$

where $\text{expit}(x) = 1/(1 + \exp(-x))$. The T-CTSs PPV, NPV, RD, and RR for study i are

$$\begin{aligned} \text{PPV}_i &= P(E|RF) = \pi_{i1} & \text{NPV}_i &= P(\bar{E}|\bar{RF}) = 1 - \pi_{i0} \\ \text{RR}_i &= \frac{P(E|RF)}{P(E|\bar{RF})} = \frac{\pi_{i1}}{\pi_{i0}} & \text{RD}_i &= P(E|RF) - P(E|\bar{RF}) = \pi_{i1} - \pi_{i0}, \end{aligned} \quad (8)$$

and the D-CTSs Sens, Spec, LR+, and LR− are

$$\begin{aligned} \text{Sens}_i &= P(RF|E) = \frac{\pi_{i1}\psi_i}{\pi_{i1}\psi_i + \pi_{i0}(1 - \psi_i)} & \text{LR}_{-i} &= \frac{1 - \text{Sens}_i}{\text{Spec}_i} \\ \text{Spec}_i &= P(\bar{RF}|\bar{E}) = \frac{(1 - \pi_{i0})(1 - \psi_i)}{(1 - \pi_{i0})(1 - \psi_i) + (1 - \pi_{i1})\psi_i} & \text{LR}_{+i} &= \frac{\text{Sens}_i}{1 - \text{Spec}_i}. \end{aligned} \quad (9)$$

Each CTS _{i} is then a function $g(\theta_i)$ of the study-specific parameters θ_i for appropriate choice of $g(\cdot)$.

Usually the purpose of meta-analysis is to consolidate information from multiple studies, and we are interested in *global* rather than study-specific CTSs. Existing methods use *plug-in estimators*, CTS_{plug}(β_0, δ_0, ν_0) as estimates of global CTSs, by plugging in

$$\begin{aligned} \pi_1 &= \text{expit}(\beta_0 + \delta_0/2) \\ \pi_0 &= \text{expit}(\beta_0 - \delta_0/2). \\ \psi &= \text{expit}(\nu_0), \end{aligned}$$

in equations (8) - (9) in place of study-specific π_{i1} , π_{i0} , and ψ_i . The plug-in method ignores 1) the nonlinear relationship between the mean hyperparameters (β_0, δ_0, ν_0) and the global CTSs, and 2) the heterogeneity present across studies represented by $\sigma_\beta, \sigma_\delta$

and σ_v . In contrast, we define the target estimand as the *predictive mean* $\text{CTS}_0(\gamma)$ of a CTS for a new study given γ , $\text{CTS}_0(\gamma) = E[g(\theta_{S+1})|\gamma]$, where $\theta_{S+1} = (\beta_{S+1}, \delta_{S+1}, \nu_{S+1})$ and β_{S+1} , δ_{S+1} , and ν_{S+1} are distributed as in (3), (4), and (6), and study index i can take on value $S + 1$. For brevity, we use $\text{CTS}_0 \equiv \text{CTS}_0(\gamma)$ and $\text{CTS}_{\text{plug}} \equiv \text{CTS}_{\text{plug}}(\beta_0, \delta_0, \nu_0)$ for the rest of this paper.

For given γ , we approximate $\text{CTS}_0 = E[g(\theta_{S+1})|\gamma]$ with a Monte Carlo estimate

$$\begin{aligned} E[g(\theta_{S+1})|\gamma] &= \int g(\theta_{S+1})P(\theta_{S+1}|\gamma)d\theta_{S+1} \\ &\approx \frac{1}{L} \sum_{l=1}^L g(\theta_{S+1}^{(l)}) \end{aligned} \quad (10)$$

where L is a pre-specified integer chosen so that Monte Carlo error is desirably small, and $\theta_{S+1}^{(l)}$ are drawn from the predictive distribution $P(\theta_{S+1}|\gamma)$. To draw samples $m = 1, \dots, M$ from the posterior distribution $P(\text{CTS}_0|\mathbf{Y})$ within a Markov chain Monte Carlo (MCMC) algorithm we approximate the integral $\int g(\theta_{S+1})P(\theta_{S+1}|\gamma)d\theta_{S+1}$ in each iteration m with a Monte Carlo calculation. Given M MCMC samples $\gamma^{(m)}$, $m = 1, \dots, M$ from the posterior of $P(\gamma|\mathbf{Y})$, for each m we

1. Take L draws $\theta_{S+1}^{(m,l)}$, $l = 1, \dots, L$ from the predictive distribution $P(\theta_{S+1}|\gamma^{(m)})$ and calculate the CTS $g(\theta_{S+1}^{(m,l)})$ for each of the L draws,
2. Estimate $\text{CTS}_0(\gamma^{(m)}) = E[g(\theta_{S+1})|\gamma^{(m)}] \approx \frac{1}{L} \sum_{l=1}^L g(\theta_{S+1}^{(m,l)})$.

Sampling $\gamma^{(m)}$ and $\text{CTS}_0(\gamma^{(m)})$ in each iteration of MCMC sampling yields approximate samples from the posterior distribution for the expectation of the CTS given data \mathbf{Y} and γ , $p(\text{CTS}_0|\mathbf{Y})$, where uncertainty in CTS_0 is due to uncertainty in the parameters γ . We refer to this method as the MC procedure.

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

We want to formally test the null hypothesis $H_0: \delta_0 = 0$ that the log(OR) is 0 against $H_A: \delta_0 \neq 0$. The Bayesian approach to testing builds an encompassing model where both H_0 and H_A have positive probability, for example, with a spike-and-slab (SAS) prior $p_1(\delta_0)$ for δ_0

$$\delta_0 = \begin{cases} \delta & \rho = 1 \\ 0 & \rho = 0 \end{cases} \quad (11)$$

$$\rho \sim \text{Bernoulli}(p) \quad (12)$$

$$\delta \sim N(0, b_\delta^2), \quad (13)$$

where p is the prior probability that $\delta_0 = 0$. In the absence of other prior information we usually choose $p = 0.5$. We call the 3RE model (1) - (6) with prior (11) - (13) the 3RE-SAS model. In the absence of prior information one can set prior standard

deviation $b_\delta = 2$ to give support to values of $\delta_0 \in (-4, 4)$, where a log(OR) δ of -4 or 4 corresponds to an OR of 0.02 or 55 . A special case has $\rho = 1$, and $\delta_0 \equiv \delta$, which is a continuous prior for δ_0 .

2.3 | Prior distributions for other parameters

For the mean parameters β_0 and ν_0 we propose normal distributions with known means a_β , a_ν and variances b_β^2 , b_ν^2

$$\beta_0 \sim N(a_\beta, b_\beta^2) \quad (14)$$

$$\nu_0 \sim N(a_\nu, b_\nu^2). \quad (15)$$

Prior means a_β and a_ν are prior guesses at the mean log-odds of an event and log-odds of the risk factor, respectively, and the standard deviations b_β and b_ν are 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

$$\sigma_\beta \sim \text{half-Cauchy}(A_\beta) \quad (16)$$

$$\sigma_\delta \sim \text{half-Cauchy}(A_\delta) \quad (17)$$

$$\sigma_\nu \sim \text{half-Cauchy}(A_\nu) \quad (18)$$

where $\sigma \sim \text{half-Cauchy}(A)$ with scale parameter $A > 0$ has density $p(\sigma) \propto (A^2 + \sigma^2)^{-1} \mathbb{1}_{[\sigma > 0]}$ ²⁴. The scale parameters A_β , A_δ , and A_ν should generally be set between 0.25 and 1 . We should expect standard deviations σ_β , σ_δ , and σ_ν to be below 1.5 , as values above 1.5 may signal problems with model fit/appropriateness or the data because heterogeneity that large is unlikely to occur naturally. Taking $A = 0.25$ yields a prior probability $P(\sigma < 1.5) \approx 0.9$, while $A = 1$ yields $P(\sigma < 1.5) \approx 0.65$. The choice of A matters less with more studies in the meta-analysis. With fewer than 10 studies we recommend $A \in (0.25, 0.5)$. With very few studies (2 or 3), large standard deviations can lead to specificity having posterior mass very close to 1 , inducing calculation problems for the statistic $\text{LR+} = \text{Sens}/(1 - \text{Spec})$.

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

If we have evidence that the standard deviation σ_δ of random effects δ_i is small, i.e. $\sigma_\delta \approx 0$, we can instead model a fixed effect for the log(OR) δ_0 where $\delta_1 = \dots = \delta_S = \delta_0$. Modify (2) to

$$\text{logit}(\pi_{ij}) = \begin{cases} \beta_i - \frac{\delta_0}{2} & j = 0 \\ \beta_i + \frac{\delta_0}{2} & j = 1, \end{cases} \quad (19)$$

where δ_0 has the SAS prior (11) - (13). We call this the 2RE-SAS model. The posterior probability $P(\delta_0 = 0|Y)$ is then the probability that $\delta_i = 0$ for every study i in the analysis, as well as for a future study $S + 1$. Because $\delta_{S+1} = 0$ implies

$\pi_{[S+1]1} = \pi_{[S+1]0}$, then $RD_{[S+1]} = 0$, $RR_{[S+1]} = 1$, $LR+_{[S+1]} = 1$, and $LR-_{[S+1]} = 1$, and the posterior probability $P(\delta_0 = 0|Y)$ is also the posterior probability that $RD_0 = 0$, $RR_0 = 1$, $LR+_0 = 1$, and $LR-_0 = 1$, where 0 and 1 are the null values for the respective CTSs. This is in contrast to the 3RE-SAS model, where $\delta_0 = 0$ does not imply that RR , RD , $LR+$ and $LR-$ are exactly equal to their null values, as $\sigma_\delta > 0$.

3 | SIMULATION STUDIES

We perform three simulations to

1. Determine appropriate choices for L in calculating CTS_0 in the MC procedure,
2. Compare the MC procedure CTS_0 to the plug-in estimator CTS_{plug} with known target values CTS_0 ,
3. Assess the posterior probability of the null hypothesis $\delta_0 = 0$ in the 3RE-SAS model.

We keep certain conditions equivalent across the three simulations. In all simulations we fix $\beta_0 = \nu_0 = \log(.15/.85)$ and independently draw the number of subjects for each study from a discrete uniform distribution $\text{Uniform}(250, 2500)$ to match values typical of the syncope data analysis. In each analysis we set 4000 MCMC iterations in each of 4 chains, discard the first 2000 iterations as burn-in and use a thin of 2, leaving 4000 MCMC samples from each posterior. We use prior means $a_\beta = a_\nu = -1$, prior variances $b_\beta = b_\delta = b_\nu = 4$, and $A_\beta = A_\delta = A_\nu = \frac{1}{\sqrt{2}} \approx .707$. For the SAS prior we use a prior probability $P(\delta_0 = 0) = 0.5$. In each analysis, initial values for mean parameters are drawn independently from a $\text{Uniform}(-1, 1)$ distribution, and initial values for random effect standard deviations σ_β , σ_δ , and σ_ν are drawn independently from a $\text{Uniform}(0.2, 1)$ distribution.

Each simulation varies different factors, and we refer to each combination of factors as a scenario.

3.1 | Simulation 1: Choice of L for Monte Carlo procedure

Simulation 1 consists of two smaller simulations. First we compare posterior standard deviations (SDs) and 95% interval lengths for $LR+_0$ for several choices of L . First, we generate $K_{11} = 100$ datasets for two scenarios – one with $(\sigma_\beta, \sigma_\delta, \sigma_\nu) = (0.5, 0.5, 0.5)$ and one with $(\sigma_\beta, \sigma_\delta, \sigma_\nu) = (1, 1, 1)$, and we fix $\delta_0 = 2$. Each dataset has $S = 10$ studies. Table 2 shows average posterior standard deviations (SDs) and 95% interval lengths of $LR+_0$ for $L \in \{1, 10, 100, 1000, 10000\}$ for the two simulation scenarios. As L increases, estimation variance of $CTS_0(\gamma)$ decreases for each γ , and for sufficiently large L the estimation variance is negligible in comparison to the posterior variance. This will be reflected in posterior variance of CTS_0 and the posterior interval width for CTS_0 decreasing to asymptotes.

Posterior SDs and 95% credible interval (CI) lengths decrease with increasing L , and the effect of further increasing L has a decreasing impact upon SD and 95% CI length. If true random effect standard deviations σ_β , σ_δ , and σ_v are moderate (≈ 0.5), then $L = 100$ is often sufficient. However, if the standard deviations σ_β , σ_δ , and σ_v are large then $L = 1000$ may be preferable and could have an impact on inferences. Taking $L = 10000$ becomes computationally burdensome, and offers little additional precision

In a second simulation, we take two fixed values of γ , $\gamma = (\log(\frac{15}{.85}), 0.5, 2, 0.5, \log(\frac{15}{.85}), 0.5)$ or $\gamma = (\log(\frac{15}{.85}), 1, 2, 1, \log(\frac{15}{.85}), 1)$ and calculate $K_{12} = 2000$ replicates of LR+₀ to estimate the *within-MC* standard deviation SD_{within} for LR+₀. Results are shown in Table 3. The bottom row of Table 2 with $L = 10000$ offers a close approximation to the *between-MC* standard deviation SD_{between} for CTS₀. The user's choice of L should depend on the ratio of within-MC SD to between-MC SDs. We recommend at least $SD_{\text{within}}/SD_{\text{between}} < 0.1$. For Simulations 2 and 3 we use $L = 1000$.

3.2 | Simulation 2: estimating CTS₀

Simulation 2 evaluates how accurately the posterior mean of CTS₀, $E[\text{CTS}_0|Y]$, from the 3RE model estimates the true CTS values LR+₀, LR−₀, PPV₀, NPV₀, Sens₀, and Spec₀ for a new study and how far off the naive plug-in estimator CTS_{plug} is from CTS₀. We fix $\delta_0 = 2$, vary $\sigma_\delta \in \{0.1, 0.25, 0.5\}$, and vary the number S of studies per meta analysis with $S \in \{10, 30, 50\}$ in a 3×3 factorial design for 9 scenarios.

Let T_k be the posterior mean of the estimand of interest in simulation iteration k , $k = 1, \dots, K$, and define the simulation mean $\bar{T} = \frac{1}{K} \sum_{k=1}^K T_k$ and variance $V_T = \frac{1}{K-1} \sum_{k=1}^K (T_k - \bar{T})^2$, and let μ be the known target value that T_k is estimating. We choose K to make the Monte Carlo standard error (MCSE) of relative bias, $\text{MCSE}(\text{rBias}) = \sqrt{V_T/(K\mu^2)}$, sufficiently small for each scenario. In this simulation we define T_k as the posterior mean $T_k = E[\text{CTS}_0|Y_k]$ for the CTSs LR+, LR−, NPV, PPV, Sens, and Spec. We calculate target values μ for each scenario by generating 100,000 probability tables from (2) - (7) using the known hyperparameters γ , calculating the desired CTSs for each probability table, and averaging each CTS across tables. Using a preliminary set of 100 simulations, we calculate $K_2 = 2500$ such that $\text{MCSE}(\text{rBias}) < 0.0025$ for all scenarios and CTSs.

We record the posterior mean, SD, and 95% CI for every CTS for both CTS₀ and CTS_{plug} and calculate bias, 95% CI coverage, average 95% CI length, and root mean-squared error (RMSE) for the CTSs LR+, LR−, NPV, PPV, Sens, and Spec.

Figure 1 shows bar plots of bias for each CTS, combination of S and σ_δ , and CTS₀ or CTS_{plug} . Dotted bars plot the bias for CTS_{plug} and solid bars plot the bias for CTS₀. For sample sizes $S = 30$ and 50, CTS_{plug} has bias that is both significantly different from zero and larger than bias for CTS₀, with CTS_{plug} bias surprisingly increasing as sample size increases for NPV and Spec, which we comment on in Section 5. There is a similar pattern in Figure 2 for RMSE, where RMSE for CTS₀ is smaller than for CTS_{plug} when $S \in \{30, 50\}$, although not always significantly. We expect CTS₀ to have larger uncertainty than CTS_{plug} because CTS₀ accounts for variation in random effects, so an unexpected result in Figure 4 is that the 95% credible intervals using CTS₀

are shorter on average than those of CTS_{plug} for LR– and Sens. The 95% interval coverage of CTS_0 is always greater than or equal to 95%, while CTS_{plug} coverage falls significantly below 95% for NPV, PPV, and Spec as S increases for all values of σ_δ , and for LR+ when $\sigma_\delta = 0.5$.

Overall, CTS_0 from the MC procedure tends to have lower bias and lower RMSE than the plug-in estimator as S increases. It also maintains at least nominal coverage as S increases, while the plug-in estimator sees coverage probabilities fall below nominal levels for multiple CTSs.

3.3 | Simulation 3: true zero and non-zero effects

For simulation 3, we vary $\delta_0 \in (0, 1, 2)$ and $\sigma_\delta \in (0.1, 0.25, 0.5)$ in a 3×3 factorial experiment with 9 scenarios. The number of simulation iterations K_3 is set equal to 1000 so that the posterior probability $T_k = P(\delta_0 = 0|Y_k)$ has $\text{MCSE}(\bar{T}) < 0.005$. For each scenario we generate $K_3 = 1000$ datasets, where every dataset Y_k has $S = 10$ studies. For each iteration $k = 1, \dots, 1000$ we fit the 3RE-SAS model to data Y_k and calculate $P(\delta_0 = 0|Y_k)$ as the proportion of MCMC samples in which $\rho = 0$.

Figure 5 presents boxplots of the distribution of $P(\delta_0 = 0|Y_k)$ for each combination of $(\delta_0, \sigma_\delta)$, and we report the mean and standard deviation of $P(\delta_0 = 0|Y_k)$ for each situation in Table 4. There is a clear distinction in Figure 5 between simulations with $\delta_0 = 0$ (left-most boxplot in each panel) and simulations with $\delta_0 \neq 0$ (middle and right-most boxplot in each panel). The simulation mean of $P(\delta_0 = 0|Y_k)$ is near zero for true non-zero effects $\delta_0 \in (1, 2)$, and ranges from 0.78 to 0.87 for true mean zero effects $\delta_0 = 0$.

4 | SYNCOPe DATA ANALYSIS: ASSESSING DIAGNOSTIC UTILITY OF REGULARLY MEASURED COVARIATES

Syncope, defined as transient loss of consciousness with rapid and spontaneous recovery, accounts for approximately 1.3 million emergency department (ED) visits every year in the United States²⁵. Syncope is often harmless, but may be a harbinger of an impending serious cardiac event. ED physicians have difficulty determining which patients are at high risk for an event, and as a result admit up to 85% of older adults presenting with syncope²⁶ even though only 5-10% of those presenting to the ED will have an event in the ensuing 30 days¹⁸. Given the difficulty of predicting serious cardiac events, we wish to measure the diagnostic value of regularly measured risk factors and determine which risk factors may have zero diagnostic value in predicting 30-day adverse events. Potential risk factors include demographics/comorbidities, symptoms, physical findings, and biomarkers.

There are 17 studies which each report information on some but not all risk factors; we meta-analyze 31 risk factors for which at least 2 studies provided a 2×2 table. For each analysis we first assess $P(\delta_0 = 0|Y)$ using the 3RE-SAS model and then we re-run the model with continuous prior on δ_0 and summarize posteriors of LR+₀, LR–₀, PPV₀, NPV₀, Sens₀, and Spec₀. If the

posterior probability of the risk factor having no effect is > 0.5 , one would usually forego computing CTS_0 estimates because they would be close to their null values and would be unlikely to provide any diagnostic value.

Table 5 details results. Rows are sorted by $P(\delta_0 = 0|Y)$, labeled ‘‘Spike’’, from lowest to highest. Columns give posterior means and 95% credible intervals for the CTSs LR+, LR−, NPV, PPV, Sens, and Spec. Of the 31 risk factors analyzed 20 had posterior probability $P(\delta_0 = 0) \leq 0.50$ and 11 had $P(\delta_0 = 0) > 0.50$. Risk factors with more studies have smaller posterior SDs, and risk factors with high posterior probability of specificity near 1 tend to have wider CIs for $\text{LR}+_{\delta_0}$. In many circumstances, a threshold of $\text{LR}+ > 10$ or $\text{LR}- < 0.1$ are used to either rule in or rule out an impending adverse event with the presence or absence of a risk factor, respectively^{27,28}. Here, none of the variables have a posterior mean $\text{LR}+_{\delta_0} > 6.50$ or $\text{LR}-_{\delta_0} < 0.44$, 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 have the fewest studies in this meta-analysis, but appear to be promising diagnostic predictors of adverse events with large mean values of $\text{LR}+_{\delta_0}$. The risk factors age > 75 , male gender, and abnormal electrocardiogram (ECG) have the highest sensitivities. Figure 6 shows contour plots of $\text{LR}+_{\delta_0}$ vs $\text{LR}-_{\delta_0}$ for the four risk factors with the smallest posterior $P(\delta_0 = 0|Y)$, which are age, male gender, history of congestive heart failure (CHF), and history of heart disease. We see varying degrees of correlation between $\text{LR}+$ and $\text{LR}-$, with a correlation of -0.15 for age and of -0.94 for male gender.

5 | DISCUSSION

Given information about a possible future study $S + 1$, such as the probability of the risk factor ψ_{S+1} , one can incorporate the information to estimate $\text{CTS}_{[S+1]}$ ’s more accurately and precisely. For example, say we know that 5%, 10%, or 25% of people in some new population have elevated blood troponin. Table 6 lists posterior summaries for CTS_0 s given known values of $P(\text{elevated troponin in new study}) = \psi_{S+1} \in \{.05, .10, .25\}$. We see that rising prevalence of elevated blood troponin corresponds with increasing sensitivity, and decreasing specificity, $\text{LR}-$ and $\text{LR}+$, while there is no effect on NPV and PPV because NPV and PPV condition on the presence or absence of the risk factor.

The Monte Carlo-within MCMC procedure to calculate CTS_0 is much more quickly implemented with a post-processing procedure rather than being done directly within MCMC. Instead of estimating CTS_0 with a Monte Carlo estimate in each iteration of MCMC, extract posterior draws $\gamma^{(m)}, m = 1, \dots, M$, and estimate $\text{CTS}_0(\gamma^{(m)})$ for each m outside of the MCMC algorithm.

In Section 3.2 we noted that positive bias for NPV_{plug} and $\text{Spec}_{\text{plug}}$ increased slightly as the number of studies S per meta-analysis increased. While this is surprising, we believe it is due to the set of true hyperparameters γ used to generate data. We used values of γ which correspond roughly with those seen in the syncope data analysis, which tend to yield values of NPV and

Spec close to 1. For a given set of hyperparameters γ , long left tails for the distributions of NPV_{new} and Spec_{new} are induced through the random effect variation σ_β , σ_δ , and σ_v . Larger posterior variance for the mean parameters β_0 , δ_0 , and v_0 can help CTS_{plug} reproduce some of the long left tail in this scenario. However, as S increases and posterior variances of β_0 , δ_0 , and v_0 decrease, less of the skew is captured.

One can incorporate information from studies that only report a $\log(\text{OR})$ $\hat{\delta}_i$ and its standard error $\text{SE}(\hat{\delta}_i)$. We say $\hat{\delta}_i$ was drawn from a normal distribution centered around its true $\log(\text{OR})$ δ_i with variance equal to $\text{SE}(\hat{\delta}_i)^2$

$$\hat{\delta}_i | \delta_i \sim N(\delta_i, \text{SE}(\hat{\delta}_i)^2), \quad (20)$$

where the true $\log(\text{OR})$ δ_i is distributed as in (4). Studies reporting pairs $(\hat{\delta}_i, \text{SE}(\hat{\delta}_i))$ provide information on the mean and variance $(\delta_0, \sigma_\delta^2)$ of the random effects for the $\log(\text{OR})$. If there is a set of additional studies reporting 2×2 tables *with fixed row-totals*, we can model them using (1) - (3). They provide information on $(\delta_0, \sigma_\delta^2)$ and the mean and variance $(\beta_0, \sigma_\beta^2)$ of the random effects for the log-odds of the event but not on the value of v_i . If the set of additional studies instead have *fixed column-totals*, one can reformulate the model using the “opposite” parameterization, where the parameters (β_i, δ_i, v_i) represent the log-odds of the risk factor, the diagnostic $\log(\text{OR})$, and the probability of the event respectively in study i . Column totals are often fixed for case-control or diagnostic studies.

If there is suspected correlation between the random effects β_i , δ_i , and v_i , they may be modeled with a multivariate normal distribution

$$\begin{pmatrix} \beta_i \\ \delta_i \\ v_i \end{pmatrix} \sim N \left(\begin{pmatrix} \beta_0 \\ \delta_0 \\ v_0 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_\beta^2 & \rho_{\beta\delta}\sigma_\beta\sigma_\delta & \rho_{\beta v}\sigma_\beta\sigma_v \\ \rho_{\beta\delta}\sigma_\beta\sigma_\delta & \sigma_\delta^2 & \rho_{\delta v}\sigma_\delta\sigma_v \\ \rho_{\beta v}\sigma_\beta\sigma_v & \rho_{\delta v}\sigma_\delta\sigma_v & \sigma_v^2 \end{pmatrix} \right) \quad (21)$$

with an inverse-Wishart prior on the covariance matrix. In this paper we model the random effects as independent because the syncope data analysis showed nearly identical results.

There is also potential for meta-analysis results to be used for prior specification in the design and analysis of a future study. The posterior probability $P(\delta_0 = 0|Y)$ can be used to either screen out variables in future analyses, or can be used as prior probabilities in a Bayesian variable selection model using spike-and-slab priors for regression coefficients, as was done in Probst et al.²⁹.

References

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

2. Lu G, Ades A. Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine* 2004; 23(20): 3105–3124.
3. Dias S, Sutton AJ, Ades A, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Medical Decision Making* 2013; 33(5): 607–617.
4. Zhang J, Carlin BP, Neaton JD, et al. Network meta-analysis of randomized clinical trials: reporting the proper summaries. *Clinical Trials* 2014; 11(2): 246–262.
5. 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.
6. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine* 2001; 20(19): 2865–2884.
7. 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.
8. Lian Q, Hodges JS, Chu H. A Bayesian Hierarchical Summary Receiver Operating Characteristic Model for network meta-analysis of diagnostic tests. *Journal of the American Statistical Association* 2019.
9. 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.
10. 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.
11. 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.
12. Chu H, Nie L, Chen Y, Huang Y, Sun W. Bivariate random effects models for meta-analysis of comparative studies with binary outcomes: methods for the absolute risk difference and relative risk. *Statistical Methods in Medical Research* 2012; 21(6): 621–633.
13. Guo J, Riebler A, Rue H. Bayesian bivariate meta-analysis of diagnostic test studies with interpretable priors. *Statistics in Medicine* 2017; 36(19): 3039–3058.

14. Hoyer A, Kuss O. Meta-analysis for the comparison of two diagnostic tests to a common gold standard: a generalized linear mixed model approach. *Statistical Methods in Medical Research* 2018; 27(5): 1410–1421.
15. 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.
16. Ma X, Lian Q, Chu H, Ibrahim JG, Chen Y. A Bayesian hierarchical model for network meta-analysis of multiple diagnostic tests. *Biostatistics* 2018; 19(1): 87–102.
17. Wynants L, Riley R, Timmerman D, Van Calster B. Random-effects meta-analysis of the clinical utility of tests and prediction models. *Statistics in Medicine* 2018; 37(12): 2034–2052.
18. 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.
19. 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.
20. George EI, McCulloch RE. Variable selection via Gibbs sampling. *Journal of the American Statistical Association* 1993; 88(423): 881–889.
21. George EI, McCulloch RE. Approaches for Bayesian variable selection. *Statistica Sinica* 1997; 7(2): 339–373.
22. Kuo L, Mallick B. Variable selection for regression models. *Sankhyā: The Indian Journal of Statistics, Series B* 1998; 60(1): 65–81.
23. Ishwaran H, Rao JS, others . Spike and slab variable selection: frequentist and Bayesian strategies. *Annals of Statistics* 2005; 33(2): 730–773.
24. 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.
25. Probst MA, Kanzaria HK, Gbedemah M, Richardson LD, Sun BC. National trends in resource utilization associated with ED visits for syncope. *The American Journal of Emergency Medicine* 2015; 33(8): 998–1001.
26. Birnbaum A, Esses D, Bijur P, Wollowitz A, Gallagher EJ. Failure to validate the San Francisco Syncope Rule in an independent emergency department population. *Annals of Emergency Medicine* 2008; 52(2): 151–159.
27. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ* 2004; 329(7458): 168–169.

28. Ranganathan P, Aggarwal R. Understanding the properties of diagnostic tests–Part 2: Likelihood ratios. *Perspectives in Clinical Research* 2018; 9(2): 99.
29. Probst MA, Gibson T, Weiss RE, et al. Risk stratification of older adults who present to the emergency department with syncope: the FAINT score. *Annals of Emergency Medicine* 2020; 75(2): 147–158.

	No Event	Event	Total
j = 0, RF Absent	n_{i00}	n_{i01}	n_{i0}
j = 1, RF Present	n_{i10}	n_{i11}	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) and a probability representation conditional on presence or absence of the risk factor (bottom).

L	$\sigma_\beta = \sigma_\delta = \sigma_v = 0.5$		$\sigma_\beta = \sigma_\delta = \sigma_v = 1$	
	SD	95% Length	SD	95% Length
1	1.969	7.502	3.869	14.876
10	0.868	3.360	2.295	8.646
100	0.642	2.508	1.730	6.472
1000	0.615	2.400	1.640	6.138
10000	0.612	2.390	1.633	6.099

TABLE 2 Simulation 1: Average posterior standard deviation (SD) and 95% interval length of $\text{LR}+_0$ for different values of L across $K_{11} = 100$ simulation iterations.

Second and third columns are results when data was generated with $\sigma_\beta = \sigma_\delta = \sigma_v = 0.5$, and in fourth and fifth columns $\sigma_\beta = \sigma_\delta = \sigma_v = 1$.

SDs and 95% interval lengths decrease with increasing L with diminishing returns.

L	SD	
	$\sigma_\beta = \sigma_\delta = \sigma_v = 0.5$	$\sigma_\beta = \sigma_\delta = \sigma_v = 1$
1	1.682	4.009
10	0.537	1.480
100	0.173	0.500
1000	0.054	0.155
10000	0.017	0.051

TABLE 3 Standard deviation across $K_{12} = 2000$ Monte Carlo approximations for LR+₀ for two fixed values of γ . We call these values *within-MC* SDs.

The mean parameters $\beta_0 = \nu_0 = \log(.15/.85)$ and $\delta_0 = 2$ are constant across the two scenarios. In the second column, random effect standard deviations are $\sigma_\beta = \sigma_\delta = \sigma_v = 0.5$ and in the third column $\sigma_\beta = \sigma_\delta = \sigma_v = 1$.

		Posterior $P(\delta_0 = 0 Y)$			
		σ_δ	$\delta_0 = 0$	$\delta_0 = 1$	$\delta_0 = 2$
Mean	0.10	0.9113	0.0002	0.0000	
	0.25	0.8536	0.0023	0.0000	
	0.50	0.7917	0.0307	0.0003	
SD	0.10	0.0795	0.0009	0.0000	
	0.25	0.1156	0.0078	0.0001	
	0.50	0.1438	0.0582	0.0009	

TABLE 4 Simulation mean and SD of posterior $P(\delta_0 = 0|Y)$ across 1000 simulation iterations for each combination of $\sigma_\delta \in \{0.1, 0.25, 0.5\}$ and $\delta_0 \in \{0, 1, 2\}$. The first three rows are the mean posterior $P(\delta_0 = 0|Y)$, and the last three rows are the SD of posterior $P(\delta_0 = 0|Y)$.

RF	Num. Papers	Spike	LR+	LR-	NPV	PPV	Sens	Spec
Age	6	0.002	2.08 (1.66, 2.65)	0.44 (0.32, 0.61)	0.94 (0.85, 0.98)	0.18 (0.08, 0.34)	0.71 (0.58, 0.81)	0.62 (0.51, 0.72)
Male Gender	7	0.002	1.39 (1.30, 1.50)	0.72 (0.65, 0.78)	0.93 (0.83, 0.97)	0.12 (0.06, 0.25)	0.58 (0.54, 0.62)	0.58 (0.57, 0.60)
CHF	8	0.004	3.43 (2.44, 4.84)	0.82 (0.74, 0.90)	0.89 (0.79, 0.95)	0.26 (0.15, 0.42)	0.24 (0.16, 0.32)	0.93 (0.90, 0.95)
Heart Disease	9	0.009	2.24 (1.69, 2.94)	0.79 (0.69, 0.88)	0.92 (0.84, 0.96)	0.17 (0.10, 0.29)	0.33 (0.24, 0.44)	0.84 (0.77, 0.88)
Arrhythmia	6	0.010	3.09 (2.08, 4.40)	0.78 (0.63, 0.89)	0.93 (0.84, 0.97)	0.18 (0.09, 0.33)	0.30 (0.17, 0.47)	0.88 (0.76, 0.94)
Dyspnea	6	0.020	2.75 (1.88, 4.43)	0.88 (0.79, 0.94)	0.87 (0.78, 0.92)	0.28 (0.18, 0.42)	0.18 (0.12, 0.28)	0.93 (0.89, 0.95)
Troponin	3	0.056	3.90 (1.94, 8.02)	0.63 (0.38, 0.88)	0.95 (0.90, 0.97)	0.21 (0.14, 0.33)	0.46 (0.22, 0.74)	0.79 (0.51, 0.93)
White Race	3	0.060	1.26 (1.10, 1.54)	0.60 (0.46, 0.80)	0.94 (0.82, 0.98)	0.10 (0.04, 0.27)	0.77 (0.55, 0.89)	0.36 (0.20, 0.58)
ECG	6	0.067	2.36 (1.55, 3.68)	0.60 (0.40, 0.87)	0.91 (0.82, 0.96)	0.24 (0.14, 0.38)	0.57 (0.40, 0.73)	0.71 (0.58, 0.80)
Resp. Rate	2	0.069	3.80 (1.80, 8.10)	0.93 (0.82, 0.98)	0.90 (0.68, 0.97)	0.24 (0.09, 0.51)	0.10 (0.05, 0.22)	0.97 (0.94, 0.98)
Urea	2	0.143	6.52 (2.56, 17.09)	0.82 (0.61, 0.96)	0.90 (0.68, 0.97)	0.29 (0.11, 0.56)	0.21 (0.09, 0.43)	0.96 (0.90, 0.98)
Trauma	2	0.147	2.03 (1.22, 3.41)	0.81 (0.56, 0.97)	0.93 (0.86, 0.96)	0.14 (0.08, 0.27)	0.34 (0.11, 0.69)	0.77 (0.41, 0.94)
Creatinine	2	0.148	4.74 (2.05, 12.42)	0.86 (0.67, 0.96)	0.90 (0.69, 0.97)	0.25 (0.09, 0.52)	0.18 (0.08, 0.38)	0.95 (0.89, 0.97)
CVD	3	0.171	1.70 (1.27, 2.38)	0.92 (0.84, 0.97)	0.95 (0.82, 0.98)	0.07 (0.03, 0.24)	0.17 (0.10, 0.30)	0.90 (0.80, 0.94)
Diabetes	6	0.207	1.70 (1.24, 2.29)	0.88 (0.78, 0.96)	0.93 (0.83, 0.97)	0.12 (0.06, 0.24)	0.26 (0.17, 0.37)	0.84 (0.77, 0.89)
Hematocrit	7	0.211	2.56 (1.52, 4.59)	0.92 (0.82, 0.97)	0.87 (0.78, 0.93)	0.23 (0.14, 0.37)	0.14 (0.07, 0.25)	0.93 (0.87, 0.96)
Hypotension	5	0.232	4.92 (1.98, 13.06)	0.91 (0.78, 0.98)	0.86 (0.74, 0.93)	0.32 (0.18, 0.49)	0.13 (0.06, 0.27)	0.96 (0.92, 0.98)
Pacemaker	3	0.299	2.33 (1.21, 4.04)	0.95 (0.88, 1.00)	0.87 (0.65, 0.96)	0.21 (0.07, 0.47)	0.09 (0.05, 0.16)	0.96 (0.94, 0.97)
Hypertension	5	0.309	1.42 (1.08, 1.85)	0.72 (0.51, 1.00)	0.93 (0.79, 0.98)	0.12 (0.05, 0.30)	0.60 (0.43, 0.75)	0.55 (0.41, 0.68)
Murmur	4	0.311	2.29 (1.19, 5.87)	0.90 (0.72, 1.00)	0.84 (0.68, 0.92)	0.26 (0.13, 0.46)	0.21 (0.08, 0.47)	0.87 (0.64, 0.96)
Oxygen	2	0.508	1.69 (1.01, 2.74)	0.88 (0.69, 1.01)	0.89 (0.81, 0.93)	0.17 (0.10, 0.29)	0.27 (0.13, 0.53)	0.82 (0.60, 0.91)
Seizure	3	0.658	1.77 (0.74, 4.64)	0.97 (0.86, 1.02)	0.93 (0.75, 0.98)	0.09 (0.02, 0.29)	0.08 (0.03, 0.25)	0.95 (0.84, 0.97)
Arr. Rx	2	0.703	2.72 (0.55, 11.89)	0.96 (0.79, 1.06)	0.81 (0.63, 0.91)	0.24 (0.09, 0.45)	0.10 (0.03, 0.29)	0.94 (0.85, 0.97)
Effort	3	0.724	1.76 (0.86, 3.73)	0.95 (0.84, 1.02)	0.87 (0.76, 0.92)	0.19 (0.10, 0.33)	0.12 (0.06, 0.25)	0.92 (0.86, 0.95)
Palpitations	4	0.760	1.85 (0.82, 5.22)	0.96 (0.83, 1.03)	0.83 (0.69, 0.91)	0.22 (0.11, 0.41)	0.12 (0.04, 0.32)	0.91 (0.75, 0.97)
Prev. Syncope	3	0.767	0.88 (0.43, 1.82)	1.07 (0.93, 1.34)	0.84 (0.68, 0.92)	0.12 (0.05, 0.29)	0.15 (0.04, 0.44)	0.82 (0.53, 0.95)
Chest Pain	3	0.785	1.65 (0.72, 3.96)	0.96 (0.85, 1.04)	0.85 (0.69, 0.93)	0.18 (0.08, 0.36)	0.11 (0.05, 0.22)	0.92 (0.89, 0.94)
Stroke	2	0.785	1.06 (0.38, 3.11)	1.06 (0.87, 1.48)	0.89 (0.75, 0.95)	0.08 (0.03, 0.22)	0.16 (0.03, 0.50)	0.83 (0.48, 0.95)
Supine	2	0.786	2.06 (0.70, 7.89)	1.05 (0.56, 2.08)	0.82 (0.59, 0.93)	0.23 (0.09, 0.51)	0.37 (0.11, 0.71)	0.69 (0.35, 0.91)
No Prodromes	6	0.827	1.25 (0.85, 1.90)	0.94 (0.71, 1.27)	0.90 (0.82, 0.94)	0.12 (0.07, 0.21)	0.45 (0.27, 0.64)	0.60 (0.42, 0.75)
Hispanic	2	0.861	0.96 (0.50, 1.99)	1.01 (0.90, 1.09)	0.89 (0.73, 0.95)	0.10 (0.04, 0.25)	0.11 (0.06, 0.24)	0.88 (0.80, 0.92)

TABLE 5 Results of 31 meta-analyses of syncope studies. Each row is for a different risk factor (RF). The second column lists the number of studies that reported a 2×2 table for each RF. Column 3 is the posterior probability $P(\delta_0 = 0|Y)$ from the 3RE-SAS model, sorted from lowest to highest. Columns 4-9 list posterior means and 95% CIs for LR₊, LR₋, NPV₀, PPV₀, Sens₀, and Spec₀ for each RF.

CHF = congestive heart failure; ECG = abnormal electrocardiogram; Resp. Rate = respiratory rate; CVD = cerebrovascular disease; Arr. Rx = arrhythmic medication.

P(RF)	LR-	LR+	NPV	PPV	Sens	Spec
0.05	0.82 (0.67, 0.93)	5.63 (2.77, 11.25)	0.95 (0.89, 0.97)	0.21 (0.13, 0.33)	0.21 (0.12, 0.36)	0.96 (0.95, 0.96)
0.10	0.71 (0.53, 0.88)	4.47 (2.42, 7.78)	0.95 (0.89, 0.97)	0.21 (0.13, 0.33)	0.35 (0.21, 0.52)	0.92 (0.90, 0.93)
0.25	0.53 (0.33, 0.82)	2.86 (1.80, 4.08)	0.95 (0.89, 0.97)	0.21 (0.13, 0.33)	0.59 (0.39, 0.74)	0.78 (0.76, 0.81)

TABLE 6 Posterior summaries for the RF Troponin given known values of $P(RF) \in \{0.05, 0.10, 0.25\}$ reported as mean (95% CI).

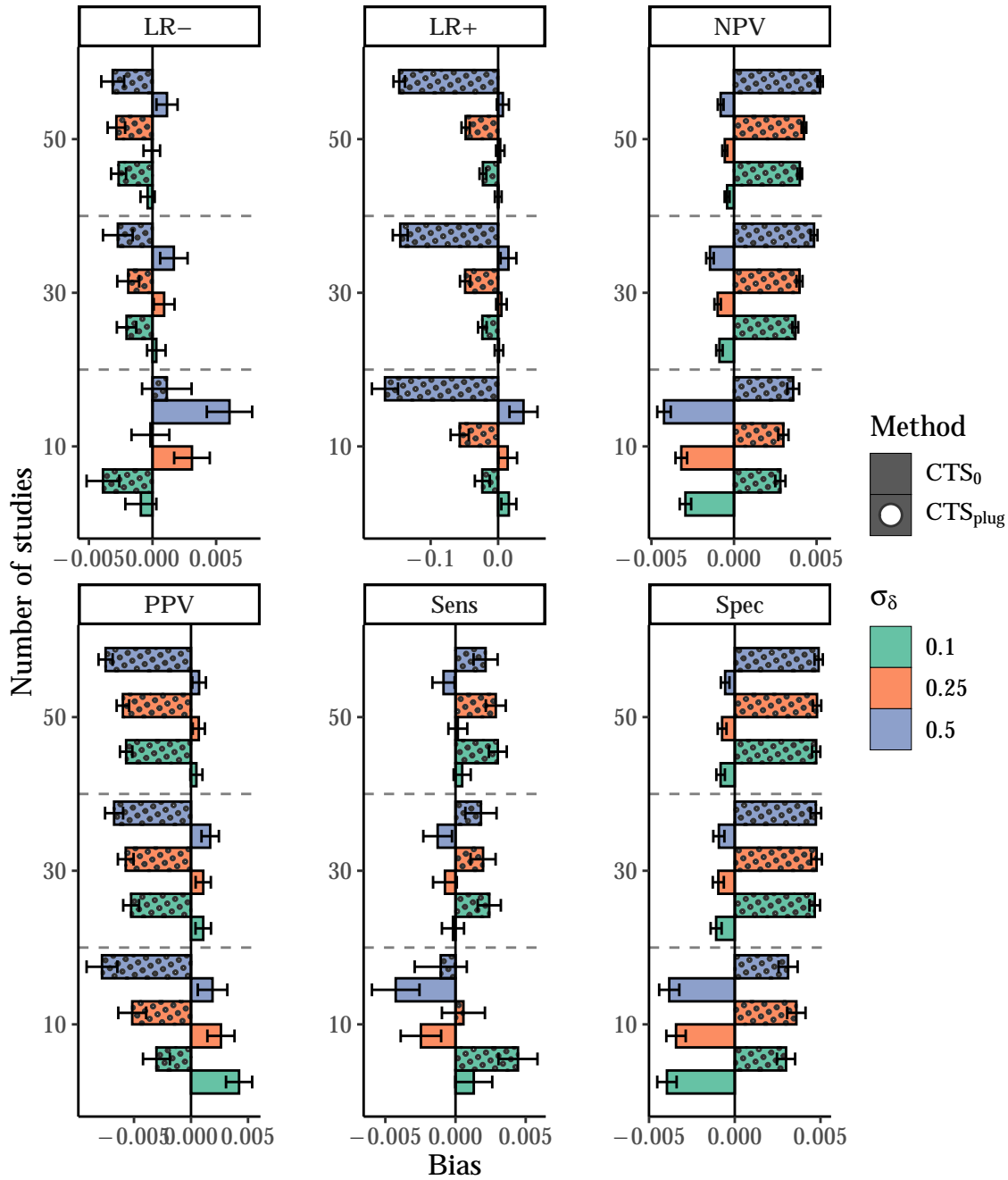


FIGURE 1 Bar plot of bias for each CTS and for each combination of $S \in \{10, 30, 50\}$ on the y-axis and $\sigma_\delta \in \{0.1, 0.25, 0.5\}$ indicated by green, red, and blue, respectively.

Dotted bars plot bias for the plug-in estimator CTS_{plug} and solid bars plot bias for the Monte Carlo estimator CTS_0 . Error bars represent $\pm 1.96 \times \text{MCSE}$.



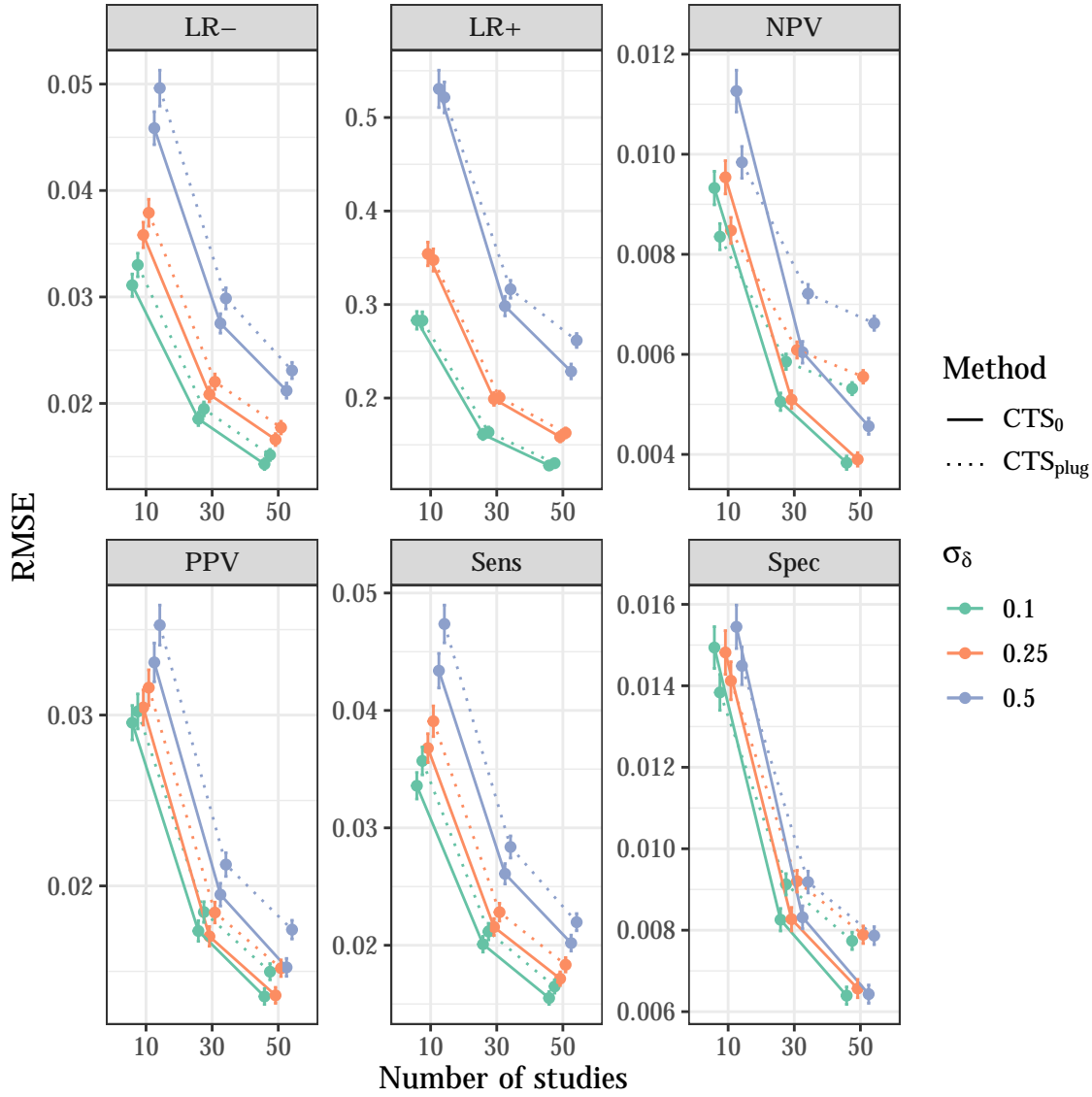


FIGURE 2 Plot of RMSE for CTS_0 (solid lines) and CTS_{plug} (dotted lines). Each panel plots RMSE for a different CTS against sample size $\in \{10, 30, 50\}$ on the x-axis and $\sigma_\delta \in \{0.1, 0.25, 0.5\}$ indicated by green, red, and blue lines, respectively. Vertical error bars plot $\pm 1.96 \times MCSE$.

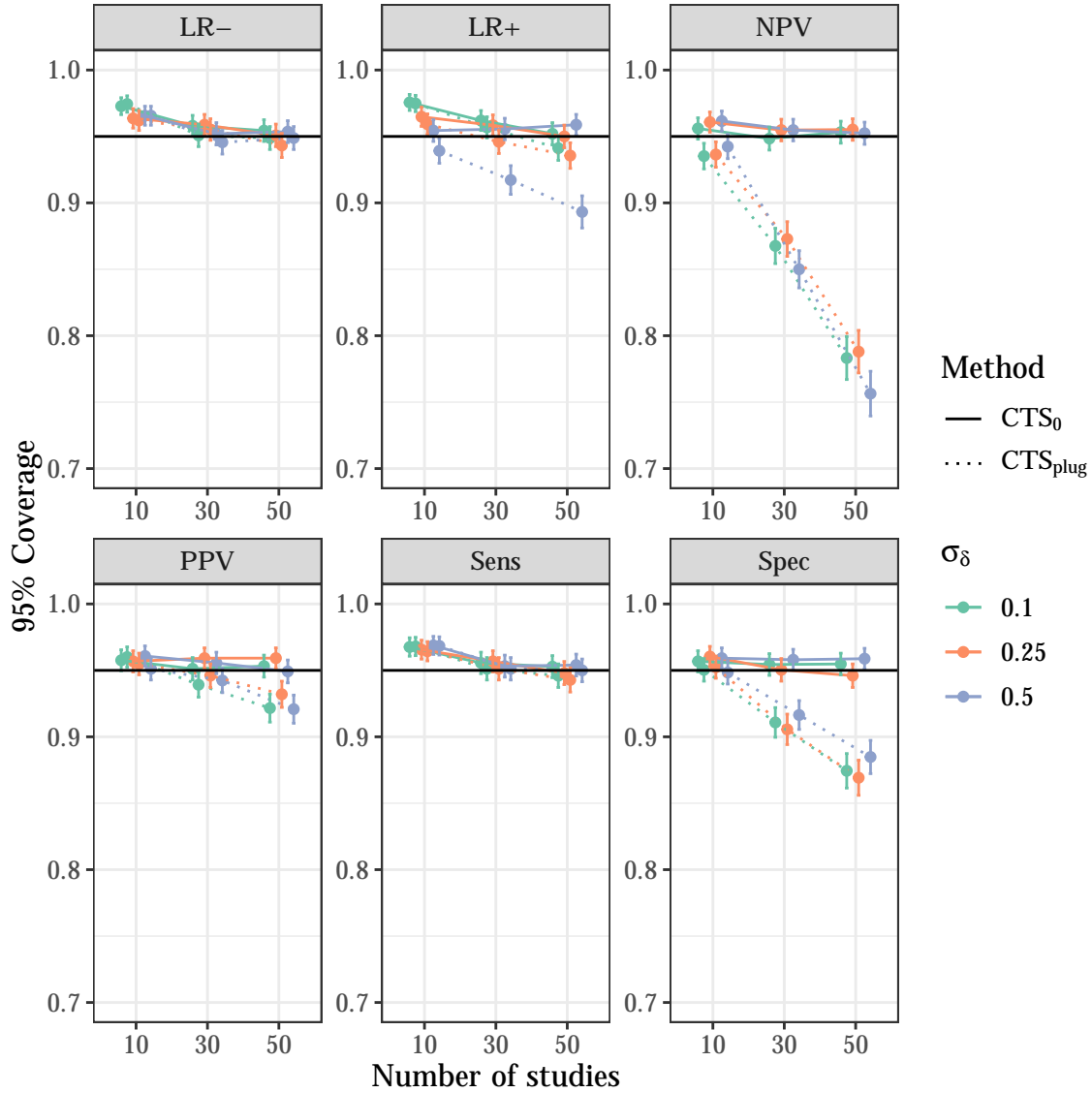


FIGURE 3 Coverage probabilities for 95% posterior intervals (PIs) for CTS_0 (solid lines) and CTS_{plug} (dotted lines) methods. Each panel plots coverage probability for different a different CTS against sample size $\in \{10, 30, 50\}$ on the x-axis, with $\sigma_\delta \in \{0.1, 0.25, 0.5\}$ indicated by green, red, and blue lines respectively. Vertical bars plot $\pm 1.96 \times MCSE$.

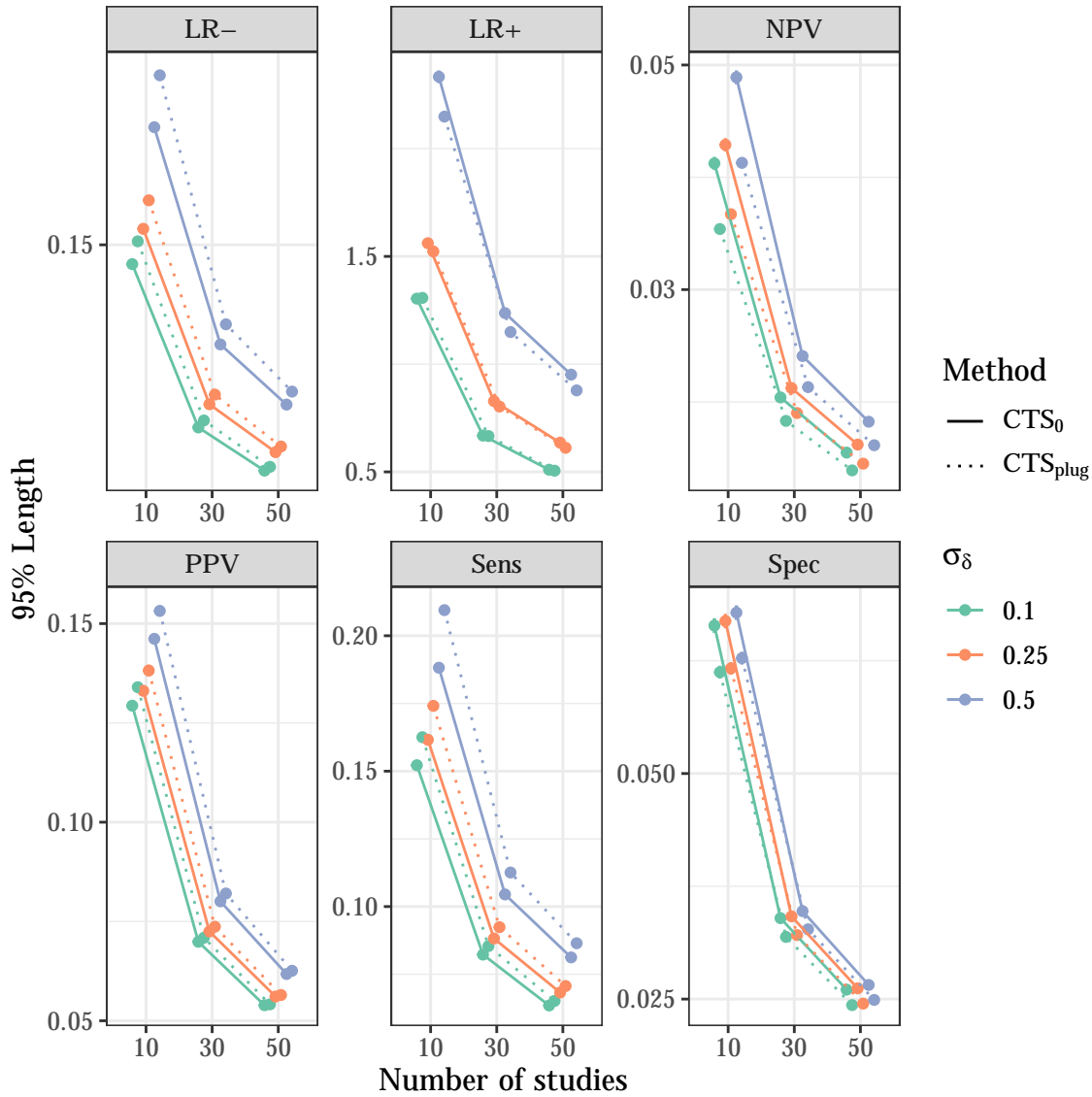


FIGURE 4 Average 95% posterior interval (PI) lengths for CTS_0 (solid lines) and CTS_{plug} (dotted lines). PI endpoints were taken as the 2.5th and 97.5th posterior quantiles. Each panel plots average 95% PI length for a different CTS against sample size $\in \{10, 30, 50\}$ on the x-axis, and $\sigma_\delta \in \{0.1, 0.25, 0.5\}$ indicated by green, red, and blue lines respectively. Vertical bars, though difficult to see for most points, plot $\pm 1.96 \times \text{MCSE}$.

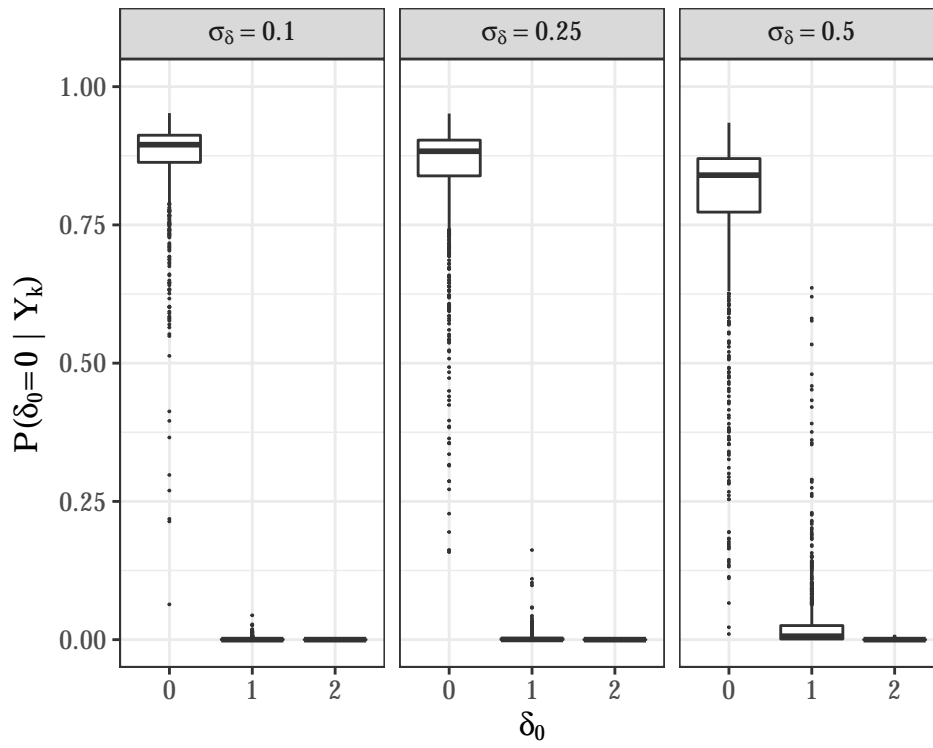


FIGURE 5 Boxplots of posterior $P(\delta_0 = 0)$ for each combination of σ_δ and δ_0 in simulation 3.

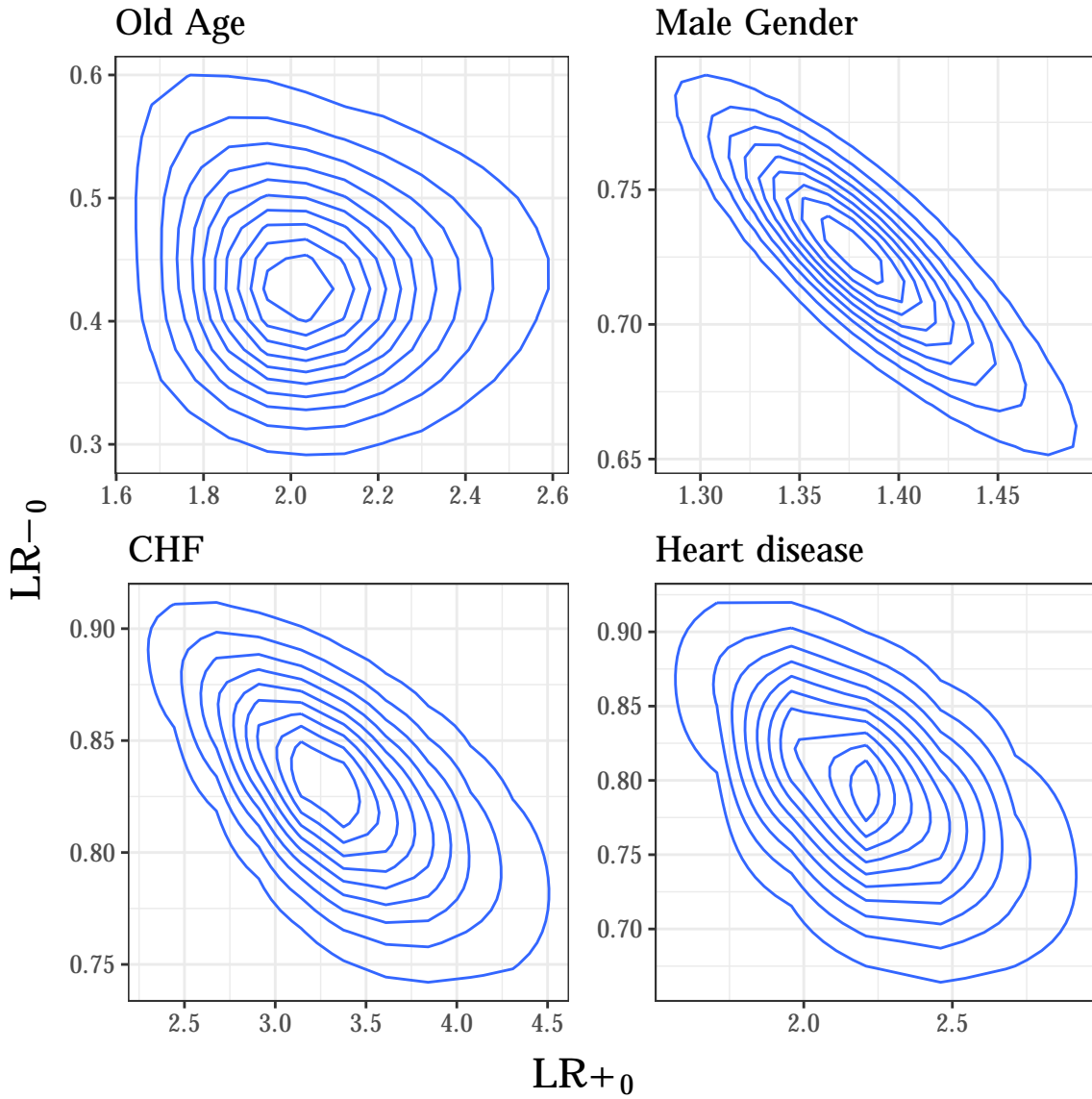


FIGURE 6 Posterior contour plots of $LR-$ on the y-axis against $LR+$ on the x-axis for the four syncope risk factors with the smallest posterior $P(\delta_0 = 0|Y)$. These are old age, male gender, history of congestive heart failure (CHF), and history of heart disease.

Higher (lower) values of $LR+$ ($LR-$) signal stronger diagnostic utility.