Received: 22 August 2017 | Revised: 12 August 2018 | Accepted: 22 August 2018

DOI: 10.1002/sim.7969

RESEARCH ARTICLE

WILEY Statistics in Medicine

A Bayesian approach for correcting exposure misclassification in meta-analysis

Qinshu Lian[1] | James S. Hodges[1] | Richard MacLehose[2] | Haitao Chu[1]

[1]Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455

[2]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN 55455

Correspondence
Qinshu Lian, Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455.
Email: lianx025@umn.edu

Haitao Chu, Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455.
Email: chux0051@umn.edu

Funding information
National Heart, Lung, and Blood Institute, Grant/Award Number: T32HL129956; National Institute of Dental and Craniofacial Research, Grant/Award Number: R03DE024750; National Institute of Diabetes and Digestive and Kidney Diseases, Grant/Award Number: U01DK106786; U.S. National Library of Medicine, Grant/Award Number: R21LM012197 and R21LM012744

In observational studies, misclassification of exposure is ubiquitous and can substantially bias the estimated association between an outcome and an exposure. Although misclassification in a single observational study has been well studied, few papers have considered it in a meta-analysis. Meta-analyses of observational studies provide important evidence for health policy decisions, especially when large randomized controlled trials are unethical or unavailable. It is imperative to account properly for misclassification in a meta-analysis to obtain valid point and interval estimates. In this paper, we propose a novel Bayesian approach to filling this methodological gap. We simultaneously synthesize two (or more) meta-analyses, with one on the association between a misclassified exposure and an outcome (main studies), and the other on the association between the misclassified exposure and the true exposure (validation studies). We extend the current scope for using external validation data by relaxing the "transportability" assumption by means of random effects models. Our model accounts for heterogeneity between studies and can be extended to allow different studies to have different exposure measurements. The proposed model is evaluated through simulations and illustrated using real data from a meta-analysis of the effect of cigarette smoking on diabetic peripheral neuropathy.

KEYWORDS
external validation data, meta-analysis, misclassification, observational studies

1 | INTRODUCTION

Meta-analyses of observational studies on exposure-disease relationships can serve as the motivation and groundwork for large clinical trials. They also provide important evidence to facilitate health policy decision-making and patient management, especially when randomized controlled trials are unethical or unavailable, such as studies of etiology. However, because it is common for each observational study in a meta-analysis to use an imperfect measurement as a surrogate for the true exposure status, observations from each observational study are prone to exposure misclassification. It is widely recognized that biases produced by exposure misclassification pose a potential threat to the validity of statistical inferences and may be more serious than measurement error for a measure on a continuous scale.[1] Ignoring misclassification also gives falsely small credible intervals because it does not take the uncertainty in the classification into account, leading to low coverage probabilities of the true effect size. In a meta-analysis, error-prone evidence from different studies is combined and contrasted. Greenland,[2] Zeisser et al,[3] and Wolpert and Mengersen[4] warned about the deleterious effect of

classification error in meta-analysis. Therefore, it is important to properly account for misclassification in meta-analyses to obtain valid point and interval estimates.

There is extensive literature on parameter estimation from *one* contingency table under misclassification, beginning with seminal papers characterizing its effects on estimation of the odds ratio from one 2×2 table.[5,6] Bross,[7] Barron,[5] and Green and Kleinbraum[6] proposed a matrix method built on sensitivity and specificity, where sensitivity is defined as the probability of a measurement correctly identifying a truly exposed subject and specificity is the probability of a measurement correctly identifying an exposure-free subject. When internal validation information is available, these closed-form estimators based on the matrix method were proved to be inefficient.[8] Alternatively, one can use an inverse-variance weighted estimator[9,10] or an inverse matrix method.[11,12] Several authors have explored likelihood-based estimators, including standard missing data methods that take the true exposure status as a partially missing variable, double-sampling likelihood, and weighted estimation equations.[13-15] Recently, developments in computational techniques and the Bayesian paradigm have fueled investigations in sensitivity analysis when misclassification rates were not precisely known. Besides varying the misclassification rates across a set of fixed values, one can randomly draw them from a joint prior distribution or integrate them out in a Bayesian analysis.[1,2,16-18]

Although inference under misclassification for one observational study has been well studied, little attention has been paid to misclassification bias in a meta-analysis. Extensive literature discusses methods for estimating accuracy of medical diagnostic techniques and devices from a meta-analysis with a gold standard test,[19-24] and applications of these methods are not limited to the area of diagnostic tests. Many meta-analyses have used these approaches to investigate the accuracy of a surrogate exposure measurement in practice, which can serve as an external validation source to correct for potential misclassification in a meta-analysis of observational studies using the surrogate measurement. However, no existing approach is able to simultaneously synthesize a meta-analysis of observational studies on an exposure-disease relationship and a meta-analysis or a network meta-analysis on validity of surrogate exposure measurements. As a result, benefits of existing evidence on misclassification rates have not been fully reaped to estimate the association between an exposure and a disease in meta-analyses.

Several technical difficulties have led to such a methodological gap. First of all, current methods that correct for misclassification in one observational study with a single external validation study assume the same fixed misclassification rates in both the main study and the validation study, called the "transportability" assumption.[8,10,11,25] This assumption is problematic even in one observational study and it becomes unrealistic in a meta-analysis with potential heterogeneity. Acknowledging that some degree of heterogeneity is inevitable in a meta-analysis, we are left with a question of how to account for variation in both the target exposure effects and the misclassification rates. Wolpert and Mengersen[4] introduced the idea of adjusting the exposure misclassification bias when synthesizing heterogeneous evidence on exposure effect. However, in their approach, both the overall misclassification rates and variability in the study-specific misclassification rates are primarily based on expert opinions or estimates from existing literature. Using expert opinions may introduce cognitive bias, while directly using the historical estimates fails to properly account for the uncertainty in these estimates. Secondly, internal validation is rare due to cost and ethical issues. Therefore, it is impractical to propose a meta-analysis with internal validation available for each study. The goal of this paper therefore is to develop a flexible statistical framework that can (1) synthesize two or more meta-analyses simultaneously to correct misclassification in meta-analyses, (2) relax the strict assumption of transportability in traditional methods using external validation studies, and (3) allow some studies in the meta-analysis of observational studies to use an error-free exposure measurement, while the rest use an error-prone surrogate measurement. We propose a Bayesian framework that can be formulated in two stages: *model of interest* (MOI) and *model of validation* (MOV). Throughout this paper, we focus on cases in which both the exposure and the outcome are binary.

This paper is organized as follows. We describe a novel Bayesian framework to correct misclassification in a meta-analysis in Section 2. Section 3 applies our method to a meta-analysis for the effect of cigarette smoking on diabetic peripheral neuropathy. Section 4 presents simulation studies to illustrate the performance of our approach under a variety of conditions. Finally, Section 5 discusses our findings and implications for future analyses.

2 | CORRECTING MISCLASSIFICATION IN META-ANALYSES

This section describes a novel framework to account for misclassification in meta-analysis when estimating the association between an exposure and a disease. Suppose we have two separate sets of meta-analyses. One is on the association between a surrogate exposure measure and a disease, which is called the main meta-analysis because it provides direct

evidence on the target exposure-disease relationship. The other meta-analysis is on the validity of the surrogate exposure measure, which is called the external validation meta-analysis. Throughout this paper, we focus on estimating the true log odds ratio describing the exposure-disease relationship. We model exposure prevalence given disease status, because this approach can be applied to all case-control, cohort, and cross-sectional designs (but not to randomized trials) as long as the sampling is not based on the error-prone exposure status. For ease of notation, we assume misclassification is non-differential, ie, the misclassification rates are the same in the diseased and control group of each study. However, it is straightforward to extend the proposed framework to incorporate differential misclassification.

Consider a main meta-analysis with a collection of K observational studies, indexed by $i = 1, \dots, K$. Let $T = \{k_1, \dots, k_T\}$ denote a subset of "contaminated" studies among the main studies, for which we suspect the potential for misclassification. Let $\overline{T} = \{i \notin T; i = 1, 2, \dots, K\}$ be a subset of error-free main studies that use the gold standard exposure measurement. $\overline{T}$ is an empty set when all studies use an error-prone surrogate measurement. Let D_{ij}, E_{ij}, and E^*_{ij} be the true disease status, error-free, and observed exposure measurements of the jth patient from the ith study, respectively, taking the value 1 or 0. In general, the true exposure status E_{ij} is unknown to us, though in a study with error-free exposure measure (ie, $i \in \overline{T}$), we know that $E^*_{ij} = E_{ij}$. Let y_{imd} be the *observed* cell counts from the ith study, with the second subscript m denoting the observed exposure status (1 or 0) and the third subscript d denoting the disease status (1 or 0), respectively. The total numbers of cases and controls in the ith study are denoted by $y_{i\cdot 1}$ and $y_{i\cdot 0}$, respectively. The unobserved true study-level exposure prevalences in the diseased and the non-diseased groups of the ith study are defined as $\pi_{i1} = P(E_{ij} = 1|D_{ij} = 1)$ and $\pi_{i0} = P(E_{ij} = 1|D_{ij} = 0)$, respectively, $i = 1, \dots, K$. We define the corresponding observed misclassified study-level conditional exposure prevalences as $P_{i1} = P(E^*_{ij} = 1|D_{ij} = 1)$ and $P_{i0} = P(E^*_{ij} = 1|D_{ij} = 0)$, $i = 1, \dots, K$.

Assume further that the external validation meta-analysis investigating the accuracy of the surrogate measurement is a collection of N studies, indexed by $i = K + 1, K + 2, \dots, K + N$. The observed data of each validation study can be summarized in a 2×2 table. The number in each cell of the ith validation study is n_{iml}. The second subscript m indicates observed exposure status, with $m = 1$ for exposed and $m = 0$ otherwise. The third subscript l indicates true exposure status, taking value 1 when truly exposed and 0 otherwise. Let $n_{i\cdot 1} = n_{i11} + n_{i01}$ and $n_{i\cdot 0} = n_{i10} + n_{i00}$ be the total numbers of truly exposed and exposure-free subjects in the ith study, respectively. Let Se_i and Sp_i denote the sensitivity and specificity respectively of misclassification in the ith main or validation study ($i = 1, 2, \dots, N + K$). Tables 1 and 2 summarize the observed data and related notation for a main study and an external validation study, respectively. In each cell, the first row shows the observed cell count and second row shows the corresponding observed conditional probability of the outcome in the ith study.

TABLE 1 Notation for main meta-analysis

		Main Study $i = 1, \dots, K$	
		D	
		+	−
E^*	+	y_{i11} P_{i1}	y_{i10} P_{i0}
	−	y_{i01} $1 - P_{i1}$	y_{i00} $1 - P_{i0}$

TABLE 2 Notation for validation meta-analysis

		External Validation $i = K+1, \dots, N$	
		E	
		+	−
E^*	+	n_{i11} Se_i	n_{i10} $1 - Sp_i$
	−	n_{i01} $1 - Se_i$	n_{i00} Sp_i

2.1 | The likelihood

Because we assume that only exposure is misclassified, the total numbers of cases and controls in each study are not affected by misclassification. With the assumption that the reference tests in the validation studies are "gold standard" with perfect accuracy, the observed cell counts of the validation studies and main studies can be modeled as

$$y_{i11} \sim \text{Bin}(y_{i\cdot 1}, P_{i1}),\ y_{i10} \sim \text{Bin}(y_{i\cdot 0}, P_{i0}),\ i = 1, \ldots, K, \tag{1}$$

$$n_{i11} \sim \text{Bin}(n_{i\cdot 1}, Se_i),\ n_{i00} \sim \text{Bin}(n_{i\cdot 0}, Sp_i),\ i = K+1, \ldots, K+N. \tag{2}$$

Taking the potential misclassification into consideration, an error-free main study contributes a likelihood term $\mathcal{L}_i^C = \prod_j \Pr(E_{ij}|D_{ij}) = \pi_{i1}^{y_{i11}}(1-\pi_{i1})^{y_{i\cdot 1}-y_{i11}}\pi_{i0}^{y_{i10}}(1-\pi_{i0})^{y_{i\cdot 0}-y_{i10}}$, $i \in \overline{T}$, because exposure status is correctly observed. Given the non-differential misclassification assumption, the likelihood contribution from subject j in misclassified main study i is

$$P_{id} = \Pr\left(E_{ij}^*|D_{ij}\right) = \sum_{E_{ij}} P\left(E_{ij}^*|D_{ij}, E_{ij}\right) P(E_{ij}|D_{ij}) = \sum_{E_{ij}} P\left(E_{ij}^*|E_{ij}\right) P(E_{ij}|D_{ij}), \tag{3}$$

where $P(E_{ij}^*|E_{ij})$ in (3) is determined by the study-specific sensitivity or specificity of the misclassification. It then follows that

$$P_{id} = \pi_{id}Se_i + (1-\pi_{id})(1-Sp_i), i = 1, \ldots, K \quad \text{and} \quad d = 0, 1. \tag{4}$$

This formula can be viewed as an extension of the matrix method[5-7] to a meta-analytic setting. As a result, the likelihood contribution of a misclassified main study in terms of π_{id}, Se_i, and Sp_i is given by

$$\begin{aligned}
\mathcal{L}_i^M &= \prod_j \Pr\left(E_{ij}^*|D_{ij}\right) \\
&= P_{i1}^{y_{i11}}(1-P_{i1})^{y_{i\cdot 1}-y_{i11}}P_{i0}^{y_{i10}}(1-P_{i0})^{y_{i\cdot 0}-y_{i10}} \\
&= (\pi_{i1}Se_i + (1-\pi_{i1})(1-Sp_i))^{y_{i11}}(1-(\pi_{i1}Se_i + (1-\pi_{i1})(1-Sp_i)))^{y_{i\cdot 1}-y_{i11}} \\
&\quad \times (\pi_{i0}Se_i + (1-\pi_{i0})(1-Sp_i))^{y_{i10}}(1-(\pi_{i0}Se_i + (1-\pi_{i0})(1-Sp_i)))^{y_{i\cdot 0}-y_{i10}},
\end{aligned}$$

where $i \in T$. Based on the distribution assumptions for n_{i11} and n_{i00} in (2), the likelihood contribution from an external validation study is

$$\mathcal{L}_i^V = \prod_j \Pr\left(E_{ij}^*|E_{ij}\right) = Se_i^{n_{i11}}(1-Se_i)^{n_{i10}}Sp_i^{n_{i00}}(1-Sp_i)^{n_{i01}}, \tag{5}$$

where $i = K+1, \ldots, K+N$. Therefore, the full likelihood is a product of the likelihood contributions from all the studies, including the main studies and validation studies, ie,

$$\mathcal{L}_{\mathcal{M}} = \prod_{i\in T}\mathcal{L}_i^M \prod_{i\in\overline{T}}\mathcal{L}_i^C \prod_{i=K+1}^{K+N}\mathcal{L}_i^V. \tag{6}$$

Note that in the naive approach that ignores misclassification, we assume that $E_{ij}^* = E_{ij}$ for all the main studies (ie, $\pi_{id} = P_{id}, \forall\, i, d$), and the corresponding likelihood is

$$\mathcal{L}_{\mathcal{N}} = \prod_{i=1}^{K}\mathcal{L}_i^C. \tag{7}$$

2.2 | Model specifications

In our approach, we explicitly model the underlying misclassification rates (Se_i, Sp_i) and true exposure prevalence (π_{i1}, π_{i0}). Various models have been proposed to estimate the effect of an exposure on a disease in a meta-analysis of observational studies. We refer to the model for the true exposure prevalences (eg, π_{i1} and π_{i0}) in the main studies as *model of interest* (MOI). In this paper, we adopt a bivariate generalized linear mixed effects model (BGLMM), which can account for heterogeneity between studies and provide flexibility in modeling correlation between the diseased and non-diseased groups of the same study. BGLMM models the true exposure prevalence in each group directly (instead of

the study-specific log odds ratio) and therefore can easily be extended to incorporate validation information. The model can be described as follows:

$$\begin{aligned} g(\pi_{i1}) = a_1 + b_{1i},\ g(\pi_{i0}) = a_0 + b_{0i}, \\ \begin{pmatrix} b_{1i} \\ b_{0i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Omega = \begin{pmatrix} \sigma_1^2 & \sigma_{10} \\ \sigma_{10} & \sigma_0^2 \end{pmatrix} \right), i = 1, 2, \ldots, K, \end{aligned} \tag{8}$$

where $g(\cdot)$ is a known link function such as a logit or a probit function, (a_1, a_0) is a vector of fixed effects, and (b_{1i}, b_{0i}) is a vector of random effects specific to study i. The covariance matrix Ω can be written as $\Omega = SQS$, where S is a 2×2 diagonal matrix with diagonal elements capturing the between study heterogeneities and Q is a correlation matrix, the off-diagonal element of which measures potential correlations between cases and controls in study i. When using the logit link function, the log odds ratio γ can be derived as $a_1 - a_0$.

Existing methods incorporating external validation data rely on the transportability assumption, ie, misclassification rates are the same and fixed in all validation and main studies. This assumption is questionable in meta-analysis given the inherent heterogeneity between studies. It would be surprising if multiple studies, conducted by different researchers in different locations and time windows, all ended up with exactly the same misclassification rates and underlying target parameters.[26] To correct misclassification in meta-analysis through external validation meta-analysis, it is essential to properly address variation not only in the main studies but also in the validation studies.

Instead of the transportability assumption, we assume that the misclassification rates are exchangeable across the main and validation studies. Therefore, the misclassification rates of all the main and the validation studies are random draws from an underlying common distribution G_θ, which can be estimated directly from the validation studies and indirectly from the main studies. More specifically, we consider a rescaled BGLMM to model the study-level sensitivity and specificity in the validation and main studies

$$\begin{aligned} Se_i = Se^L + (Se^U - Se^L)\text{expit}(\alpha_1 + \beta_{1i}), \\ Sp_i = Sp^L + (Sp^U - Sp^L)\text{expit}(\alpha_0 + \beta_{0i}), \\ \begin{pmatrix} \beta_{1i} \\ \beta_{0i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} \tau_{\text{Se}}^2 & \tau_{\text{SeSp}} \\ \tau_{\text{SeSp}} & \tau_{\text{Sp}}^2 \end{pmatrix} \right), i = 1, 2, \ldots, N + K, \end{aligned} \tag{9}$$

where $\text{expit}(z) = [1 + \exp(-z)]^{-1}$, $i = 1, \ldots, K$ denote the main studies, $i = K+1, \ldots, K+N$ denote the validation studies, and (Se^L, Sp^L) and (Se^U, Sp^U) are sets of lower and upper bounds, respectively, to exclude improbable values. When $Se^L = Sp^L = 0$ and $Se^U = Sp^U = 1$, (9) corresponds to the regular BGLMM. In this paper, we make a mild assumption that surrogate exposure assessment is at least better than chance, ie, $Se^L = Sp^L = 0.5$ and $Se^U = Sp^U = 1$. We refer to (9) as the *model of validation* (MOV) because it estimates the misclassification rates. When assuming non-differential misclassification, $(\alpha_0, \beta_{0i}, \alpha_1, \beta_{1i}, \Sigma)$ and $(a_0, b_{0i}, a_1, b_{1i}, \Omega)$ are two independent sets of parameters.

2.3 | Model fitting

The likelihood has a complex form (6). Inference using the maximum likelihood method involves integration over multi-dimensional random effects. Let $\boldsymbol{a} = (a_1, a_0)$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_0)$, $\boldsymbol{b}_i = (b_{1i}, b_{0i})$, and $\boldsymbol{\beta}_i = (\beta_{1i}, \beta_{0i})$. The marginal likelihood for the fixed-effect quantities integrated over the random effects for (6) is

$$\begin{aligned} P(\boldsymbol{a}, \boldsymbol{\alpha}|\text{Data}) = \prod_{i \in \overline{T}} \int \int \mathcal{L}_i^C p(\boldsymbol{b}_i|\Omega) d\boldsymbol{b}_i \prod_{i \in T} \int \int \mathcal{L}_i^M p(\boldsymbol{b}_i|\Omega) p(\boldsymbol{\beta}_i|\Sigma) d\boldsymbol{b}_i d\boldsymbol{\beta}_i \\ \prod_{i=K+1}^{K+N} \int \mathcal{L}_i^V p(\boldsymbol{\beta}_i|\Sigma) d\boldsymbol{\beta}_i. \end{aligned} \tag{10}$$

The corresponding asymptotic approximation for standard errors of functional parameters is of unknown accuracy and may not be sufficiently accurate. In this paper, we adopt a full Bayesian approach using Hamiltonian Monte Carlo (HMC) algorithms, which provides a way to simultaneously analyze the main and validation studies. We implement it using the STAN software via the `rstan` package in R. Note that the misclassification rates Se_i and Sp_i of the main studies $i = 1, 2, \ldots, K$ are treated as latent variables and are integrated out using the HMC.

To complete the Bayesian specification of the model, vague normal priors are used for the fixed effects a_0, a_1, α_0, and α_1. Conjugate inverse Wishart priors $\text{IW}(R, k)$ are assumed for the covariance matrices Σ and Ω, where R is the scale matrix

and k is the degrees of freedom. A smaller k corresponds to a more diffuse prior distribution. Although we have used the unstructured covariance matrices throughout the model description, they can be simplified if relevant prior knowledge is available, or through a model selection procedure. For example, Σ and Ω may be reduced to diagonal matrices, which implies independence of the effect sizes. More details on the choice of prior distributions and the model selection procedure will be provided in Section 3. A directed acyclic graph (DAG) can be found in the Supporting Information to help readers understand the model specification and estimation process. Given the prior specifications described above, the joint posterior distribution is proportional to

$$\left(\prod_{i\in\overline{T}}\mathcal{L}_i^C\prod_{i\in T}\mathcal{L}_i^M\prod_{i=K+1}^{K+N}\mathcal{L}_i^V\right)\left(\prod_{i=1}^{K+N}p(\boldsymbol{\beta}_i|\Sigma)\right)\left(\prod_{i=1}^{K}p(\boldsymbol{b}_i|\Omega)\right)f(\boldsymbol{\omega}),$$

where $f(\boldsymbol{\omega})$ is the joint prior distribution for $(\boldsymbol{\alpha}, \boldsymbol{a}, \Sigma, \Omega)$. The overall log odds ratio, exposure prevalences in case and control groups, sensitivity, and specificity of misclassification can be summarized by

$$\begin{aligned}
&\gamma = a_1 - a_0,\ \Pi_1 = \frac{\exp(a_1)}{1+\exp(a_1)},\ \Pi_0 = \frac{\exp(a_0)}{1+\exp(a_0)},\\
&Se = Se^L + (Se^U - Se^L)\text{expit}(\alpha_1),\\
&Sp = Sp^L + (Sp^U - Sp^L)\text{expit}(\alpha_0).
\end{aligned}$$

Specifically, we calculate these statistics in each iteration of the HMC algorithm and inference is based on the medians of their posterior samples. Convergence is assessed using trace plots, Monte Carlo standard error,[27] and Gelman-Rubin statistic.[28] While usually of secondary interest, the estimates for the overall sensitivity and specificity are predictably more efficient since both the direct information from the external validation meta-analysis and the indirect information from the main meta-analysis contribute to their estimates. For a fixed number of participants, each case-control study contributes 2 degrees of freedom, while each cohort study contributes 3 degrees of freedom. As a result, the maximum number of parameters that can be estimated ranges from $2(N + K)$ to $3(N + K)$, depending on the designs of the included studies. The total number of parameters to be estimated is 10. Therefore, we need $N+K > 10/3$ to estimate the parameters in this model reasonably well without informative prior distributions. Note that this calculation provides a rough estimate of the minimum number of studies. It is a necessary condition for identifiability of the cluster-invariant parameters. This number, however, says nothing about the precision or power of the meta-analysis.

3 | A CASE STUDY

Diabetic peripheral neuropathy (DPN), also known as distal symmetrical neuropathology or sensorimotor neuropathy, is the most common complication of diabetes. DPN may cause muscle weakness and loss of reflexes, especially at the ankle, leading to foot deformities among other problems. If an infection occurs and is poorly controlled, the foot may then need to be amputated. Existing studies indicate an increased likelihood of neuropathy in people with diabetes who smoke.[29-31] A meta-analysis of cross-sectional, case-control, prospective, and retrospective cohort studies was conducted to better assess the relationship between smoking and diabetic neuropathy.[32] In this meta-analysis, the prospective cohort studies included 5558 subjects without DPN at baseline. The cross-sectional and case-control studies included 17,594 subjects. Clair et al[32] concluded that smoking and diabetic neuropathy were not significantly associated based on the meta-analysis. However, the included studies relied heavily on self-reported smoking status and were therefore subject to misclassification. The validity of self-reported smoking measurement is often questioned because of the inclination of smokers to underestimate the amount smoked or to deny smoking at all.[33]

This section revisits this dataset and implements the proposed methods (Data is available in the Supporting Information). Throughout this section, we focus solely on comparing the risk of DPN in current smokers (E = 1) vs nonsmokers (former or never, E = 0) for illustration, and we assume that misclassification is non-differential. The purpose of the case study is to illustrate the method we proposed and the result should not be used for decision-making. Fourteen studies in the work of Clair et al[32] compared the risk of DPN in current smokers vs nonsmokers and provided complete sets of cell counts in 2×2 tables. These 14 studies consist of one case-control study, two cohort studies, and 11 cross-sectional studies. In all studies, the measurements of smoking exposure are self-reported. The cardinality of the set T is 14, while $\overline{T}$ is empty.

Based on nine publications that investigated the accuracy of self-reported smoking status using the superior cotinine validation methods,[33] the sensitivity of self-reported smoking status ranges from 0.82 to 1.00, and the specificity ranges

from 0.91 to 1.00. We use these nine case-control studies as the external validation meta-analysis to correct the potential misclassification of self-reported smoking status. The eligibility criteria of the validation and main studies were similar.

We consider various MOIs and MOVs differing in the covariance matrices Ω and Σ in (8) and (9). We consider three models for MOI.

- MOI1: π_{i1} and π_{i0} are assumed to be mutually independent and have homogeneous variance, ie, the covariance matrix Ω is reduced to a diagonal matrix with diagonal elements (σ^2, σ^2).
- MOI2: π_{i1} and π_{i0} are assumed to be mutually independent but have heterogeneous variances. The covariance matrix Ω is then reduced to a diagonal matrix with diagonal elements (σ_1^2, σ_0^2).
- MOI3: the most general model, which assumes arbitrary unstructured Ω.

We also consider three similar structures for Σ in MOV.

- MOV1: the study-specific sensitivities and specificities are assumed to have homogeneous variance and are mutually independent, ie, Σ is reduced to a diagonal matrix with diagonal elements both equal to τ^2.
- MOV2: the study specific sensitivities and specificities are allowed to have heterogeneous variances but are assumed to be mutually independent, ie, Σ is reduced to a diagonal matrix with diagonal elements $(\tau_{Se}^2, \tau_{Sp}^2)$.
- MOV3: the most general model, assuming arbitrary unstructured covariance matrix Σ.

Combining a MOI and a MOV gives nine possible models. Since non-informative prior distributions can lead to inaccurate posterior estimates,[34] we select proper but diffuse prior distributions for the hyperparameters. We assign IW$(I_2, 3)$ to Ω and Σ, where I_2 is the 2 by 2 identity matrix. Based on simulations in R, this prior distribution corresponds to 95% prior credible interval of (0.368, 4.517) for standard deviations and (−0.951, 0.948) for correlation parameters. A half-Cauchy distribution with scale 0.25 is used as the prior for σ, τ, σ_1, σ_0, τ_{Se}, and τ_{Sp} in MOI1, MOV1, MOI2 and MOV2, respectively. The corresponding 95% prior credible interval is (0.010,6.335). A weakly-informative Normal prior with mean 0 and variance 25 is used for the fixed effects a_0, b_0, α, and β. This corresponds to a prior distribution with 95% prior credible interval (−14.969, 14.586) for the log odds ratio, which is large enough to cover all plausible values. After 5,000 burn-in samples, 10,000 posterior samples are obtained for each model. Model fit is compared using Deviance Information Criterion (DIC).[35] Table 3 presents the median estimates and the 95% credible intervals (CIs) of the log odds ratio, overall exposure prevalence in case and control groups, overall sensitivity and specificity for classification, as well as the DIC values from all the nine models. MOV1+MOI3, MOV2+MOI3, and MOV3+MOI3 have the lowest DIC values among the six models. However, the difference of DIC between any two models is no more than 5 units, so no model has a conclusive advantage over the others (in practice, differences of more than 5 are often considered substantial). The estimates from these models are similar. Take MOV1+ MOI3 as an example. The log odds ratio γ is estimated to be 0.225 with 95% CI (−0.205, 0.703). The CI includes 0, so no significant association between active smoking and DPN is found. Density plots of the overall sensitivity, specificity, and log odds ratio from MOV1+MOI3 are shown in the left panel of Figure 1. Forest plots showing the study-specific estimates are given in the right panel. Figure 2 shows the forest plot for the corrected and uncorrected estimates, using MOV1+MOI3 and the naive approach with likelihood (7) and MOI3, respectively. The point estimates from these two methods are close for most of studies, which is reasonable given the high estimated overall specificity (0.969), but the intervals are rather wider when accounting for the uncertainty caused by misclassification.

We performed sensitivity analyses to investigate the impacts of the prior choices on the results. We consider an alternative Normal prior $N(0, 100)$ for the fixed effects, an alternative half-Cauchy prior with scale 2.5 for the standard deviations, and an alternative inverse Wishart IW$(I_2, 2)$ prior for the covariance matrix. Table 4 shows the results for MOV1+MOI3 under these alternatives. The estimates are not sensitive to prior choices. Similar phenomena are also observed for other models. These indicate that inferences are driven mostly by the data, not the priors.

4 | SIMULATION

4.1 | Setting

In this section, we fit the proposed model to simulated datasets generated from a collection of scenarios that are likely to be encountered in practice. For each scenario, we simulate 1000 hypothetical datasets.

To generate data, we first mimic the conditions of the case study (Scenario I). We generate 14 main studies and nine external validation studies. Sample size in each group is set to be 1000. The overall sensitivity and specificity are 0.94 and

TABLE 3 Case study results: posterior median (95% credible interval). "$\overline{D}$" denotes the deviance, "pD" denotes the effective number of parameters, and DIC = $\overline{D} + pD$

	LogOR	Π_1	Π_0	Se	Sp	$\overline{D}$	pD	DIC
MOV1+MOI1	0.217 (−0.236, 0.748)	0.206 (0.112, 0.273)	0.172 (0.087, 0.230)	0.931 (0.778, 0.979)	0.970 (0.915, 0.994)	263	37	300
MOV2+MOI1	0.226 (−0.333, 0.814)	0.185 (0.105, 0.273)	0.152 (0.080, 0.230)	0.940 (0.832, 0.976)	0.968 (0.913, 0.992)	263	38	301
MOV3+MOI1	0.189 (−0.114, 0.578)	0.217 (0.133, 0.276)	0.187 (0.109, 0.238)	0.932 (0.798, 0.980)	0.976 (0.928, 0.995)	266	33	299
MOV1+MOI2	0.212 (−0.168, 0.604)	0.217 (0.128, 0.278)	0.183 (0.104, 0.229)	0.920 (0.754, 0.979)	0.974 (0.920, 0.995)	264	36	299
MOV2+MOI2	0.211 (−0.250, 0.712)	0.194 (0.114, 0.271)	0.162 (0.092, 0.227)	0.934 (0.784, 0.976)	0.971 (0.917, 0.995)	264	37	302
MOV3+MOI2	0.179 (−0.109, 0.497)	0.222 (0.144, 0.280)	0.193 (0.123, 0.239)	0.927 (0.789, 0.980)	0.978 (0.930, 0.996)	266	33	299
MOV1+MOI3	**0.225 (−0.205, 0.703)**	**0.191 (0.088, 0.303)**	**0.159 (0.065, 0.258)**	**0.942 (0.897, 0.972)**	**0.969 (0.935, 0.986)**	**260**	**36**	**296**
MOV2+MOI3	0.212 (−0.202, 0.651)	0.199 (0.103, 0.316)	0.168 (0.083, 0.273)	0.943 (0.888, 0.977)	0.969 (0.942, 0.985)	261	36	297
MOV3+MOI3	0.222 (−0.185, 0.667)	0.197 (0.102, 0.308)	0.165 (0.080, 0.263)	0.943 (0.887, 0.975)	0.968 (0.938, 0.984)	261	36	296
MOI3 only	0.171 (−0.161, 0.503)	0.226 (0.149.0.316)	0.197 (0.132, 0.274)					

Abbreviation: DIC, Deviance Information Criterion.

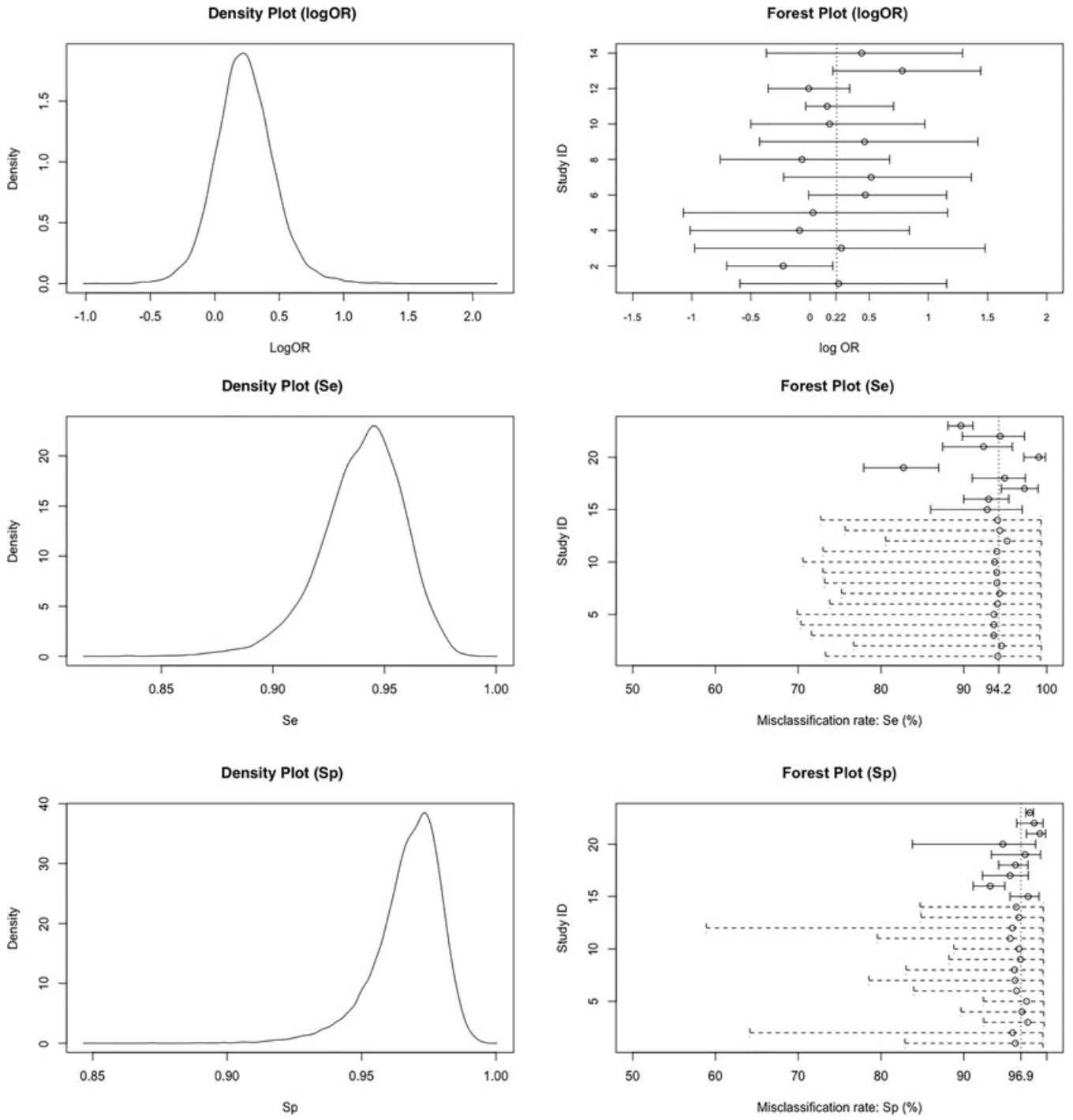

FIGURE 1 Density plots and forest plots for MOV1+MOI3: the left panel is the posterior density plots for log odds ratio, sensitivity, and specificity. The right panel is the forest plots. Circles represent study-specific posterior medians. In the forest plots of sensitivity and specificity, the solid horizontal lines denote the 95% credible intervals of the validation studies, while the dashed horizontal lines denote the 95% credible intervals of the main studies. The vertical dashed lines indicate the overall estimates

0.97, respectively. We assume $\Pi_1 = 0.19$ and $\Pi_0 = 0.16$. The diagonal elements in Σ are set as $\tau^2_{\text{Se}} = 1.44$ and $\tau^2_{\text{Sp}} = 0.81$, and those in Ω are $\sigma^2_1 = \sigma^2_0 = 1$. The off-diagonal elements are -0.54 and 0.8 for Σ and Ω, respectively. These values are equal to the estimates in the case study. Next, we evaluate the proposed model from the following aspects: (A) the impact of the severity of exposure misclassification on the estimates of the exposure effect; (B) the impact of the numbers of the main and validation studies; and (C) the impact of the true overall log odds ratio. To investigate (A), we consider two pairs of overall misclassification rates: Scenario IA-1: $Se = 0.85$, $Sp = 0.85$ and Scenario IA-2: $Se = 0.75$, $Sp = 0.70$. All other

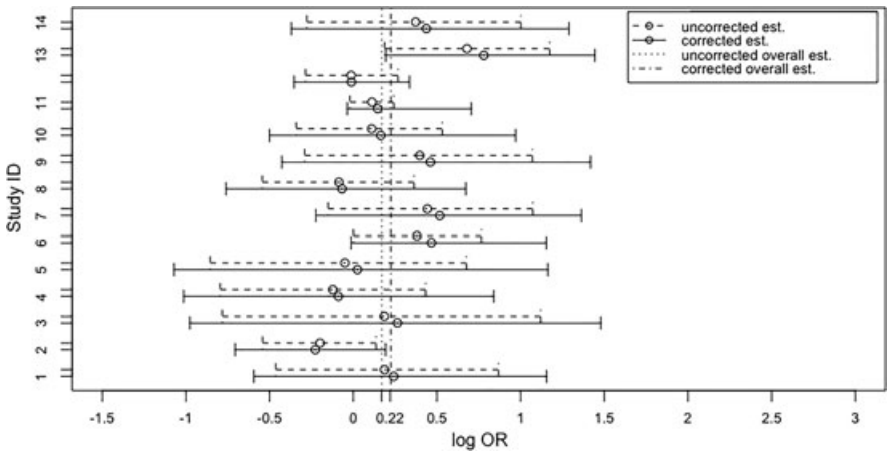

FIGURE 2 Forest plot for the corrected and uncorrected log odds ratios. The solid lines denote 95% credible intervals of the corrected log odds ratios for the main studies. The dashed lines denote the 95% credible intervals of the uncorrected log odds ratios for the main studies

TABLE 4 Sensitivity to the prior distributions for MOV1+MOI3

	Prior	$N(0, 100)$	half-Cauchy(0, 2.5)	IW(I_2, 2)
logOR	Posterior median	0.226	0.221	0.227
	95% CI	(−0.206, 0.683)	(−0.193, 0.686)	(−0.215, 0.728)
Π_1	Posterior median	0.189	0.193	0.188
	95% CI	(0.09, 0.296)	(0.096, 0.306)	(0.082, 0.308)
Π_0	Posterior median	0.158	0.161	0.157
	95% CI	(0.068, 0.255)	(0.071, 0.261)	(0.061, 0.262)
Se	Posterior median	0.942	0.942	0.942
	95% CI	(0.891, 0.974)	(0.893, 0.973)	(0.897, 0.97)
Sp	Posterior median	0.969	0.970	0.968
	95% CI	(0.934, 0.986)	(0.937, 0.986)	(0.935, 0.985)

Abbreviations: CI, credible interval; IW, inverse Wishart.

parameters are set to be the same as in Scenario I. To investigate (B), we generate Scenario IB-1: nine validation studies and 30 main studies; Scenario IB-2: 30 validation studies and 14 main studies; and Scenario IB-3: 30 validation studies and 30 main studies. All other parameters are set to be the same as in Scenario I. To investigate (C), we assume four different pairs of (Π_1, Π_0), while keeping all other parameters identical with Scenario I. Specially, we consider Scenario IC-1: $\Pi_1 = 0.21$ and $\Pi_0 = 0.16$ with corresponding log odds ratio equal to 0.333; Scenario IC-2: $\Pi_1 = 0.25$ and $\Pi_0 = 0.16$ with corresponding log odds ratio equal to 0.560; and Scenario IC-3: $\Pi_1 = 0.30$ and $\Pi_0 = 0.16$ with corresponding log odds ratio equal to 0.811.

Recall that existing methods to correct misclassification bias in a single observational study using external validation assume the transportability assumption. It is easy to construct a method like the proposed new method but making the transportability assumption. Specifically, we assume that the misclassification rates are fixed values and are the same in all the validation and main studies. Equation (9) is then replaced by a fixed effect model, ie,

$$\begin{aligned} m_{ij} &\sim \text{Binomial}(M_{ij}, S_j),\ i = K+1, \ldots, N,\ j = 1, 0 \\ \text{logit}(S_j) &= \alpha_0 + \alpha_1 * X_{ij}, \end{aligned}$$

where $m_{i1} = n_{i11}$, $m_{i0} = n_{i00}$, $M_{i1} = n_{i\cdot 1}$, and $M_{i0} = n_{i\cdot 0}$. The misclassification rate S_j is same in all the main and validation studies, $S_1 = Se$ and $S_0 = Sp$. X_{ij} is an indicator variable, taking 1 for the disease group and 0 otherwise. The

model specification for (π_{i1}, π_{i0}) is given in Equation (8). The likelihood contributions from a validation study and a main study are

$$\begin{aligned}
\mathcal{L}_i^V &\propto S_1^{n_{i11}}(1-S_1)^{n_{i10}} S_0^{n_{i00}}(1-S_0)^{n_{i01}},\\
\mathcal{L}_i^M &\propto (\pi_{i1}S_1 + (1-\pi_{i1})(1-S_0))^{y_{i11}}(1-(\pi_{i1}S_1 + (1-\pi_{i1})(1-S_0)))^{y_{i\cdot1}-y_{i11}}\\
&\quad \times (\pi_{i0}S_1 + (1-\pi_{i0})(1-S_0))^{y_{i10}}(1-(\pi_{i0}S_1 + (1-\pi_{i0})(1-S_0)))^{y_{i\cdot0}-y_{i10}}, i \in T\\
\mathcal{L}_i^C &\propto \pi_{i1}^{y_{i11}}(1-\pi_{i1})^{y_{i\cdot1}-y_{i11}}\pi_{i0}^{y_{i10}}(1-\pi_{i0})^{y_{i\cdot0}-y_{i10}}, i \in \overline{T},
\end{aligned}$$

where T indicates the index set for error-prone validation studies. The total likelihood contribution is $\mathcal{L}_{\mathcal{M}} = \prod_{i\in T}\mathcal{L}_i^M \prod_{i\in\overline{T}}\mathcal{L}_i^C \prod_{i=K+1}^{K+N}\mathcal{L}_i^V$. Compared to this model under the transportability assumption, the proposed framework allows the between-study heterogeneity in the misclassification rates and therefore is more likely to recover the true exposure effect. To demonstrate the advantage of the proposed method given in Section 2, we conduct simulation studies to evaluate the model performance under these assumptions in Scenario II. We generate 30 main studies and 30 external validation studies. In each study, numbers of cases and controls are set as above. The covariance matrix Ω and Σ are set as in Scenario I. The overall exposure prevalences in cases and controls (Π_1 and Π_0) are assumed to be 0.3 and 0.16, respectively, as in Scenario IC-3. The true overall sensitivity and the true overall specificity are 0.85.

Finally, we evaluate the model performance when the main meta-analysis contains studies with an error-free exposure measure. Specially, we create a case in which some main studies use an error-free exposure measure, while the rest do not, ie, the cardinality of T is smaller than the number of studies in the main studies. All the parameters are set to be the same as in Scenario II, except five out of the 30 main studies use a gold standard measurement for exposure status.

4.2 | Results

The performance of the proposed method is compared with the naive uncorrected analysis in terms of absolute bias (AB), relative bias (RB), and coverage probability (CP). Here, relative bias is defined as absolute bias/true value. In the proposed method, we adopt MOV3+MOI3 and list its results under the column "With correction" in Tables 5 to 7. In the naive method, we ignore the misclassification and use only MOI3. The corresponding likelihood is given in (7). The results from the naive model are listed under the column "Without correction" in Tables 5 to 7. To emphasize the differences between the two methods, the estimates for RB are shown in boldface in the tables.

Table 5 shows the simulation results for Scenario I, IA-1, IA-2, IB-1, IB-2, IB-3, IC-1, IC-2, and IC-3. We can see that the absolute bias increases with the decrease of *Se* and *Sp* or the increase of the true log odds ratio in both the proposed method and the naive approach. However, the proposed method is able to effectively control the absolute bias under 0.06, while the absolute bias in the naive method can be as large as −0.157. In addition, the proposed method greatly improves the coverage probabilities. The estimates for the individual parameters in the proposed method are also better than the naive uncorrected method (for conciseness, the results are not shown in the Tables). The results from Scenarios I and IB-1 to IB-3 show that increasing the numbers of main and validation studies helps to improve the accuracy and precision of the estimates of *Se*, *Sp*, Π_1, Π_0. Figure 3 shows the power analysis for Scenario I and Scenario IC-1 to Scenario IC-3. Power increases as the true exposure effect increases. With the same numbers of main and validation studies as in the case study, the power is higher than 0.8 when the true overall log odds ratio is greater than 0.560.

Table 6 shows results for the two correction methods under different assumptions and the results for the naive method. The proposed method under the exchangeability assumption provides estimates with smaller biases, smaller mean square errors (MSE), and better coverage probabilities compared with the other two methods. The naive method does not take the potential misclassification bias into consideration, while the method assuming transportability fails to account for variability in the misclassification rates in different studies. Table 7 compares the simulation results for Scenario II and Scenario III. It shows that including studies with a error-free exposure measure improves the accuracy of the estimates.

By comparing the results in Tables 5 and 6, we can see that the naive method performs poorly when the true exposure effect is large and the overall misclassification rates are low. The coverage probability for the log odds ratio from the naive method is only 0.062 in Scenario II, far lower than in Scenario IA-1 and Scenario IC-3. However, the coverage probabilities from the proposed method are all around 95% in these scenarios. In addition, the naive uncorrected approach is known to provide a falsely narrow credible interval, therefore the MSEs from the proposed method are slightly higher than the naive method when the true overall exposure effect is small. Note that our main purpose is to enhance accuracy rather than efficiency.

TABLE 5 Simulation results I

		With Correction				Without Correction			
	Parameter (True)	**AB**	**MSE**	**CP**	**RB**	**AB**	**MSE**	**CP**	**RB**
Scenario I	Se (0.940)	−0.002	0.001	0.928	**−0.003**	-	-	-	-
	Sp (0.970)	−0.004	0.000	0.941	**−0.004**	-	-	-	-
	Π_1 (0.190)	−0.006	0.003	0.932	**−0.032**	0.025	0.002	0.911	**0.129**
	Π_0 (0.160)	−0.006	0.002	0.930	**−0.039**	0.029	0.002	0.888	**0.178**
	logOR (0.208)	0.016	0.039	0.976	**0.079**	−0.045	0.021	0.985	**−0.217**
Scenario IA-1	Se (0.850)	−0.002	0.002	0.928	**−0.002**	-	-	-	-
	Sp (0.850)	−0.012	0.001	0.940	**−0.014**	-	-	-	-
	Π_1 (0.190)	−0.012	0.005	0.929	**−0.065**	0.113	0.014	0.116	**0.595**
	Π_0 (0.160)	−0.014	0.004	0.930	**−0.086**	0.122	0.016	0.038	**0.764**
	logOR (0.208)	0.054	0.073	0.976	**0.261**	−0.107	0.020	0.970	**−0.514**
Scenario IA-2	Se (0.750)	−0.001	0.002	0.930	**−0.001**	-	-	-	-
	Sp (0.700)	−0.008	0.001	0.946	**−0.011**	-	-	-	-
	Π_1 (0.190)	−0.011	0.007	0.911	**−0.058**	0.202	0.042	**0.000**	**1.062**
	Π_0 (0.160)	−0.013	0.006	0.932	**−0.080**	0.217	0.048	**0.000**	**1.358**
	logOR (0.208)	0.057	0.099	0.975	**0.276**	−0.147	0.025	0.967	**−0.708**
Scenario IB-1	Se (0.940)	−0.004	0.001	0.943	**−0.004**	-	-	-	
	Sp (0.970)	−0.003	0.000	0.953	**−0.003**	-	-	-	-
	Π_1 (0.190)	−0.003	0.001	0.950	**−0.013**	0.024	0.001	0.862	**0.125**
	Π_0 (0.160)	−0.002	0.001	0.936	**−0.015**	0.028	0.001	0.762	**0.175**
	logOR (0.208)	0.008	0.018	0.964	**0.037**	−0.046	0.012	0.941	**−0.223**
Scenario IB-2	Se (0.940)	−0.001	0.000	0.944	**−0.001**	-	-	-	-
	Sp (0.970)	−0.001	0.000	0.946	**−0.001**	-	-	-	-
	Π_1 (0.190)	0.003	0.002	0.946	**0.014**	0.026	0.002	0.915	**0.138**
	Π_0 (0.160)	0.003	0.002	0.940	**0.018**	0.031	0.002	0.860	**0.191**
	logOR (0.208)	0.004	0.037	0.970	**0.017**	−0.048	0.023	0.978	**−0.230**
Scenario IB-3	Se (0.940)	0.001	0.000	0.947	**−0.001**	-	-	-	
	Sp (0.970)	−0.001	0.000	0.942	**−0.001**	-	-	-	-
	Π_1 (0.190)	0.001	0.001	0.949	**0.006**	0.025	0.001	0.850	**0.129**
	Π_0 (0.160)	0.000	0.001	0.954	**0.003**	0.028	0.001	0.785	**0.175**
	logOR (0.208)	0.007	0.016	0.954	**0.035**	−0.043	0.011	0.950	**−0.206**
Scenario IC-1	Se (0.940)	−0.003	0.001	0.927	**−0.003**	-	-	-	-
	Sp (0.970)	−0.004	0.000	0.938	**−0.004**	-	-	-	-
	Π_1 (0.210)	−0.006	0.003	0.937	**−0.027**	0.022	0.002	0.926	**0.103**
	Π_0 (0.160)	−0.006	0.002	0.930	**−0.037**	0.029	0.002	0.888	**0.178**
	logOR (0.333)	0.025	0.039	0.973	**0.074**	−0.071	0.024	0.970	**−0.213**
Scenario IC-2	Se (0.940)	−0.003	0.001	0.930	**−0.003**	-	-	-	-
	Sp (0.970)	−0.004	0.000	0.946	**−0.004**	-	-	-	-
	Π_1 (0.250)	−0.004	0.003	0.933	**−0.018**	0.016	0.002	0.941	**0.064**
	Π_0 (0.160)	−0.005	0.002	0.930	**−0.032**	0.029	0.002	0.891	**0.179**
	logOR (0.560)	0.038	0.041	0.968	**0.068**	−0.112	0.033	0.941	**−0.200**
Scenario IC-3	Se (0.940)	−0.003	0.001	0.929	**−0.003**	-	-	-	-
	Sp (0.970)	−0.003	0.000	0.941	**−0.003**	-	-	-	-
	Π_1 (0.300)	−0.003	0.004	0.932	**−0.011**	0.008	0.002	0.944	**0.026**
	Π_0 (0.160)	−0.004	0.002	0.930	**−0.027**	0.029	0.002	0.888	**0.178**
	logOR (0.811)	0.047	0.043	0.973	**0.058**	−0.157	0.045	0.906	**−0.193**

AB: absolute bias = estimated value − true value. RB: relative bias = $\frac{\text{AB}}{\text{True value}}$. MSE: mean square error. CP: 95% credible interval coverage probability

TABLE 6 Simulation results for Scenario II: comparing different assumptions

	Method	Assumption	Se (0.85)	Sp (0.85)	Π_1 (0.30)	Π_0 (0.16)	logOR (0.811)
AB	With correction	Exchangeability	0.000	−0.006	−0.007	−0.009	0.098
		Transportability	−0.020	−0.006	−0.020	−0.031	0.226
	Without correction	-	-	-	0.074	0.121	−0.384
MSE	With correction	Exchangeability	0.001	0.000	0.005	0.003	0.081
		Transportability	0.001	0.000	0.006	0.004	0.176
	Without correction	-	-	-	0.007	0.016	0.156
CP	With correction	Exchangeability	0.942	0.944	0.927	0.939	0.963
		Transportability	0.111	0.190	0.924	0.882	0.858
	Without correction	-	-	-	0.487	0.007	0.062
RB	With correction	Exchangeability	0.000	−0.007	−0.024	−0.059	0.121
		Transportability	−0.024	−0.007	−0.067	−0.192	0.279
	Without correction	-	-	-	0.248	0.757	−0.474

AB: absolute bias = estimated value − true value. RB: relative bias = $\frac{\text{AB}}{\text{True value}}$. MSE: mean square error. CP: 95% credible interval coverage probability

TABLE 7 Simulation results: comparing model performance with (Scenario III) versus without (Scenario II) error-free main studies

		Scenario III				Scenario II			
	Parameter (True)	AB	MSE	CP	RB	AB	MSE	CP	RB
With correction	Se (0.850)	−0.001	0.001	0.946	−0.001	0.000	0.001	0.942	0.000
	Sp (0.850)	0.003	0.000	0.947	−0.003	−0.006	0.000	0.944	−0.007
	Π_1 (0.300)	−0.001	0.002	0.930	−0.005	−0.007	0.005	0.927	−0.024
	Π_0 (0.160)	−0.001	0.001	0.927	−0.009	−0.009	0.003	0.939	−0.059
	logOR (0.811)	0.023	0.023	0.958	0.029	0.098	0.081	0.963	0.121

AB: absolute bias = estimated value − true value. RB: relative bias = $\frac{\text{AB}}{\text{True value}}$. MSE: mean square error. CP: 95% credible interval coverage probability

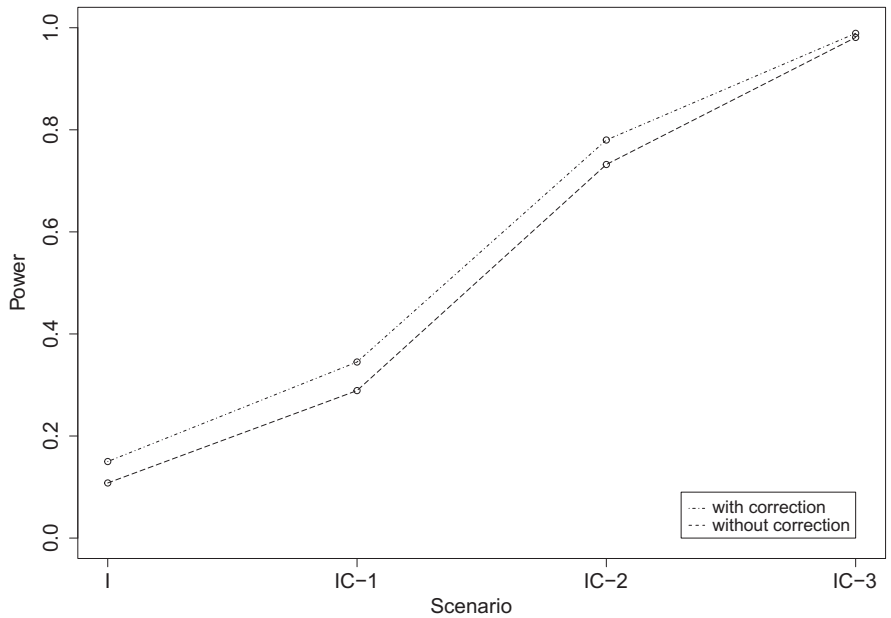

FIGURE 3 Power analysis: power to detect logOR $\neq$ 0

5 | DISCUSSION AND FUTURE WORK

This paper proposed a fully Bayesian approach to accounting for misclassification in meta-analysis. We provide a framework to make a correction based on external validation studies. To the best of our knowledge, this is the first paper to synthesize two separate sets of meta-analyses to correct misclassification bias. It relaxes the strict assumption of transportability, used by existing methods. Both the correlations between the groups and potential heterogeneity of meta-analysis are taken into account. The model consists of two components: a MOI and a MOV. We adopt a BGLMM for the MOI and a rescaled BGLMM for the MOV. Other popular methods of meta-analysis that directly model π_{i1}, π_{i0}, Se_i, and Sp_i,[19,20,36-38] such as the Sarmanov beta-binomial model, also fit in the proposed framework. Although we focus on estimating the log odds ratio throughout this paper, the proposed approach can be easily modified to estimate other popular effect sizes describing the exposure-disease relationship, such as relative risk when the included studies are all cohort studies. In addition, the proposed method can be extended to incorporate studies with different exposure measurements. For example, if the main meta-analysis contains studies using two different surrogate and a error-free measurements (each study only uses one of them), we would need a network validation meta-analytic dataset for the two surrogate measurements and the gold standard test. Se_i and Sp_i in (4) will then be replaced by Se_{it} and Sp_{it}, with the second subscript t indicating the corresponding measurement, $t = 1, 2, 3$. Consequently, the rescaled BGLMM will be replaced by a rescaled multivariate generalized linear mixed effects model (MGLMM)[39,40] for the MOV. In this paper, we discuss only non-differential misclassification. In practice, misclassification rates may differ between two disease groups. To allow differential misclassification, we would need two separate validation meta-analyses, one for misclassification rates in cases and one for misclassification rates in controls. Misclassification rates for each study would then be denoted by Se_{id} and Sp_{id}, with the second subscript d indicating disease status, $d = 1, 0$. MGLMM can be used as MOV to jointly model $(Se_{i1}, Se_{i0}, Sp_{i1}, Se_{i0})$.

Throughout this paper, we assume the validation measurement is a "gold standard". However, in most situations, the assumed error-free measurement in fact is just another fallible surrogate measurement for the true E_{ij}. Some authors have discussed correcting misclassification through dual imperfect exposure assessments in one observational study.[41-45] In meta-analysis of diagnostic tests, several models have been proposed to account for the bias caused by imperfect reference tests.[39,46-49] These methods can be easily incorporated into the MOV of the proposed framework to handle "alloyed gold standard" situations. In addition, a referee pointed out that similar ideas have been discussed in papers that harmonize differently measured constructs in clinical trials.[50]

Existing methods of correcting misclassification through external validation assume transportability. In this paper, we relax this assumption to exchangeability, ie, the misclassification rates of the main studies and validation studies are assumed to be i.i.d. draws from a common distribution. Validation of this assumption is grounded in the underlying requirements of all meta-analysis: the studies being combined share similarities and are assumed to be representative of a target population. When correcting the misclassification bias using the proposed method, the inclusion and exclusion criteria of the meta-analyses combined should be similar in order for the exchangeability assumption to be valid. Ideally, validation studies and main studies should be screened and collected according to the same inclusion criteria by the same research group. When the validation meta-analysis and the main meta-analysis are conducted separately with different study selection rules, and partial internal validation data or error-free main studies are available, a rigorous test, analogous to the test for the transportability,[25] may be developed to further examine the exchangeability assumption. In addition, we focus on the simple case where other covariates like age and race are not considered. If other study-level or patient-level characteristics are suspected to confound the exposure-disease association, meta-regression models with covariates may be used as MOI or MOV to adjust for potential confounding bias. In such a case, patient-level data may be required and more in-depth discussion on the interaction between the confounding variables and the misclassification rates should be considered. We leave these to future work.

ACKNOWLEDGEMENTS

We would like to thank James D. Neaton for helpful discussion, which has improved the presentation of this work. The research reported in this publication was supported in part by National Institute of Dental and Craniofacial Research R03DE024750 (H.C.), U.S. National Library of Medicine R21LM012197 (H.C.) and R21LM012744 (H.C., J.H.), National Institute of Diabetes and Digestive and Kidney Diseases U01DK106786 (H.C.), National Heart, Lung, and Blood Institute

T32HL129956 (Q.L.). The content is solely the responsibility of the authors and does not necessarily represent official views of the National Institutes of Health.

ORCID

Qinshu Lian http://orcid.org/0000-0001-9570-5000

REFERENCES

1. Gustafson P. *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*. Boca Raton, FL: Chapman & Hall/CRC; 2003.
2. Greenland S. Multiple-bias modelling for analysis of observational data. *J R Stat Soc Ser A Stat Soc*. 2005;168(2):267-306.
3. Zeisser C, Stockwell TR, Chikritzhs T. Methodological biases in estimating the relationship between alcohol consumption and breast cancer: the role of drinker misclassification errors in meta-analytic results. *Alcohol Clin Exp Res*. 2014;38(8):2297-2306.
4. Wolpert RL, Mengersen KL. Adjusted likelihoods for synthesizing empirical evidence from studies that differ in quality and design: effects of environmental tobacco smoke. *Stat Sci*. 2004;19:450-471.
5. Barron BA. The effects of misclassification on the estimation of relative risk. *Biometrics*. 1977;33(2):414-418.
6. Greenland S, Kleinbaum DG. Correcting for misclassification in two-way tables and matched-pair studies. *Int J Epidemiol*. 1983;12(1):93-97.
7. Bross I. Misclassification in 2×2 tables. *Biometrics*. 1954;10(4):478-486.
8. Morrissey MJ, Spiegelman D. Matrix methods for estimating odds ratios with misclassified exposure data: extensions and comparisons. *Biometrics*. 1999;55(2):338-344.
9. Greenland S. Variance estimation for epidemiologic effect estimates under misclassification. *Stat Med*. 1988;7(7):745-757.
10. Greenland S. Maximum-likelihood and closed-form estimators of epidemiologic measures under misclassification. *J Stat Plan Inference*. 2008;138(2):528-538.
11. Marshall RJ. Validation study methods for estimating exposure proportions and odds ratios with misclassified data. *J Clin Epidemiol*. 1990;43(9):941-947.
12. Lyles RH. A note on estimating crude odds ratios in case–control studies with differentially misclassified exposure. *Biometrics*. 2002;58(4):1034-1036.
13. Chen TT. Log-linear models for categorical data with misclassification and double sampling. *J Am Stat Assoc*. 1979;74(366a):481-488.
14. Espeland MA, Odoroff CL. Log-linear models for doubly sampled categorical data fitted by the EM algorithm. *J Am Stat Assoc*. 1985;80(391):663-670.
15. Espeland MA, Hui SL. A general approach to analyzing epidemiologic data that contain misclassification errors. *Biometrics*. 1987;43(4):1001-1012.
16. Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. *Int J Epidemiol*. 2005;34(6):1370-1376.
17. Chu H, Wang Z, Cole SR, Greenland S. Sensitivity analysis of misclassification: a graphical and a Bayesian approach. *Ann Epidemiol*. 2006;16(11):834-841.
18. Luta G, Ford MB, Bondy M, Shields PG, Stamey JD. Bayesian sensitivity analysis methods to evaluate bias due to misclassification and missing data using informative priors and external validation data. *Cancer Epidemiol*. 2013;37(2):121-126.
19. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med*. 2001;20(19):2865-2884.
20. Ma X, Nie L, Cole SR, Chu H. Statistical methods for multivariate meta-analysis of diagnostic tests: an overview and tutorial. *Stat Methods Med Res*. 2016;25(4):1596-1619.
21. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-990.
22. Chu H, Nie L, Cole SR, Poole C. Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: alternative parameterizations and model selection. *Stat Med*. 2009;28(18):2384-2399.
23. Chu H, Guo H, Zhou Y. Bivariate random effects meta-analysis of diagnostic studies using generalized linear mixed models. *Med Decis Mak*. 2010;30(4):499-508.
24. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol*. 2006;59(12):1331.
25. Lyles RH, Zhang F, Drews-Botsch C. Combining internal and external validation data to correct for exposure misclassification: a case study. *Epidemiology*. 2007;18(3):321-328.
26. Higgins JPT. Commentary: heterogeneity in meta-analysis should be expected and appropriately quantified. *Int J Epidemiol*. 2008;37(5):1158-1160.
27. Gong L, Flegal JM. A practical sequential stopping rule for high-dimensional Markov chain Monte Carlo. *J Comput Graph Stat*. 2016;25(3):684-700.
28. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992;7(4):457-472.

29. Maser RE, Steenkiste AR, Dorman JS, et al. Epidemiological correlates of diabetic neuropathy: report from Pittsburgh epidemiology of diabetes complications study. *Diabetes*. 1989;38(11):1456-1461.
30. Mitchell BD, Hawthorne VM, Vinik AI. Cigarette smoking and neuropathy in diabetic patients. *Diabetes Care*. 1990;13(4):434-437.
31. Tesfaye S, Chaturvedi N, Eaton SEM, et al. Vascular risk factors and diabetic neuropathy. *N Engl J Med*. 2005;352(4):341-350.
32. Clair C, Cohen MJ, Eichler F, Selby KJ, Rigotti NA. The effect of cigarette smoking on diabetic peripheral neuropathy: a systematic review and meta-analysis. *J Gen Intern Med*. 2015;30(8):1193-1203.
33. Patrick DL, Cheadle A, Thompson DC, Diehr P, Koepsell T, Kinne S. The validity of self-reported smoking: a review and meta-analysis. *Am J Public Health*. 1994;84(7):1086-1093.
34. Natarajan R, McCulloch CE. Gibbs sampling with diffuse proper priors: a valid approach to data-driven inference? *J Comput Graph Stat*. 1998;7(3):267-277.
35. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B Stat Methodol*. 2002;64(4):583-639.
36. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat Med*. 1995;14(24):2685-2699.
37. Chen Y, Chu H, Luo S, Nie L, Chen S. Bayesian analysis on meta-analysis of case-control studies accounting for within-study correlation. *Stat Methods Med Res*. 2015;24(6):836-855.
38. Chen Y, Hong C, Ning Y, Su X. Meta-analysis of studies with bivariate binary outcomes: a marginal beta-binomial model approach. *Stat Med*. 2016;35(1):21-40.
39. 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.
40. Lian Q, Hodges JS, Chu H. A Bayesian hierarchical summary receiver operating characteristic model for network meta-analysis of diagnostic tests. *J Am Stat Assoc*. 2018. http://doi:10.1080/01621459.2018.1476239
41. Brenner H. Use and limitations of dual measurements in correcting for nondifferential exposure misclassification. *Epidemiology*. 1992;3(3):216-222.
42. Wacholder S, Armstrong B, Hartge P. Validation studies using an alloyed gold standard. *Am J Epidemiol*. 1993;137(11):1251-1258.
43. Drews CD, Flanders WD, Kosinski AS. Use of two data sources to estimate odds ratios in case-control studies. *Epidemiology*. 1993;4(4):327-335.
44. Flanders WD, Drews CD, Kosinski AS. Methodology to correct for differential misclassification. *Epidemiology*. 1995;6(2):152-156.
45. Zhang J, Cole SR, Richardson DB, Chu H. A Bayesian approach to strengthen inference for case-control studies with multiple error-prone exposure assessments. *Stat Med*. 2013;32(25):4426-4437.
46. Chu H, Chen S, Louis TA. Random effects models in a meta-analysis of the accuracy of two diagnostic tests without a gold standard. *J Am Stat Assoc*. 2009;104(486):512-523.
47. Dendukuri N, Schiller I, Joseph L, Pai M. Bayesian meta-analysis of the accuracy of a test for tuberculous pleuritis in the absence of a gold standard reference. *Biometrics*. 2012;68(4):1285-1293.
48. Liu Y, Chen Y, Chu H. A unification of models for meta-analysis of diagnostic accuracy studies without a gold standard. *Biometrics*. 2015;71(2):538-547.
49. Chu H, Cole SR, Wei Y, Ibrahim JG. Estimation and inference for case–control studies with multiple non–gold standard exposure assessments: with an occupational health application. *Biostatistics*. 2009;10(4):591-602.
50. Siddique J, Reiter JP, Brincks A, Gibbons RD, Crespi CM, Brown CH. Multiple imputation for harmonizing longitudinal non-commensurate measures in individual participant data meta-analysis. *Stat Med*. 2015;34(26):3399-3414.

SUPPORTING INFORMATION

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

How to cite this article: Lian Q, Hodges JS, MacLehose R, Chu H. A Bayesian approach for correcting exposure misclassification in meta-analysis. *Statistics in Medicine*. 2019;38:115–130. https://doi.org/10.1002/sim.7969