BayesSenMC: an **R** package for Bayesian Sensitivity Analysis of Misclassification

Jinhui Yang
Carleton College

Lifeng Lin
Florida State University

Haitao Chu
University of Minnesota
at Twin Cities

Abstract

In case - control studies, the odds ratio is commonly used to summarize the association between a binary exposure and a dichotomous outcome. However, exposure misclassification frequently appears in case-control studies due to inaccurate data reporting, which can produce bias in measures of association. In this article, we implement a Bayesian sensitivity analysis of misclassification to provide a full posterior inference on the corrected odds ratio under both non-differential and differential misclassification. We present an R (R Core Team 2018) package **BayesSenMC**, which provides user-friendly functions for its implementation. The usage is illustrated by a real data analysis on the association between bipolar disorder and rheumatoid arthritis.

Keywords: Bayesian analysis, misclassification, odds ratio, sensitivity analysis, case-control study.

---

1. Introduction

Many epidemiological studies are concerned with assessing the risk of an outcome between exposed and non-exposed subjects. For example, in a case-control study, researchers first identify subjects who have the disease of interest (the case group) and subjects who do not (the control group), and then ascertain the exposure status of the subjects in each group. The odds ratio is typically used to assess the association between the exposure and disease in the case–control study; it describes the ratio of the exposure odds in the case group to that in the control group.

However, misclassification of exposure, disease outcome, or covariates appears frequently in observational studies of epidemiological or medical research (Rothman *et al.* 2008; Brakenhoff *et al.* 2018). In a case-control study, misclassification is often due to inaccurate reporting of the exposure status (e.g., self-reported data). This can consequently lead to biased exposure probabilities and odds ratio. To adjust for such biases, we can correct the odds ratio using the observed data from the case-control study and the sensitivity and specificity of correctly classifying exposure status from external data. Here, the sensitivity is the proportion of exposed subjects that are correctly classified as exposed (i.e., true positive), and the specificity is the proportion of non-exposed subjects that are correctly classified as non-exposed (i.e. true negative).

Quantitative assessment of misclassification bias is necessary to estimate uncertainty in study

results. There are many statistical methods for misclassification correction; nearly all of them use information that maps observed measurements to true values (Greenland 2005). These methods include regression calibration and multiple imputation (Rosner *et al.* 1989; Spiegelman *et al.* 2001; Cole *et al.* 2006), in which the mapping is based on a validation study. Also, sensitivity analysis can be used to evaluate the effects of uncertainties on observed results (Greenland 1996; Lash and Flink 2003; Chu *et al.* 2006), in which the mapping may be based on prior information or expert opinion. However, when such information is lacking, researchers may over- or under-adjust for misclassification with an inaccurate guess, which may, in turn, produce a poor estimate (Gustafson *et al.* 2006).

Moreover, despite the ubiquity of measurement error, these methods remain rarely used due to complexity of statistical approaches, especially the complexity of prior specifications as well as lack of software package (Lash and Flink 2003). For example, in a random sample survey of 57 epidemiological studies (Jurek *et al.* 2006), the investigators found only one used quantitative corrections. Sensitivity analysis is simple but limited insofar as it does not provide formal interval estimates that combine uncertainty due to random error with misclassification. Several authors have addressed this deficiency by using probabilistic (Monte Carlo) sensitivity analyses (Greenland 2005); for example, Fox *et al.* (2005) proposed a probabilistic sensitivity analysis of misclassified binary variables based on multiple imputation, and provided SAS code and Excel macro for the approach. Such methods can be viewed as means of summarizing over sensitivity analyses using a prior distribution (Greenland 2005). Many of them have been implemented in the R (R Core Team 2018) package **episensr** (Haine 2018). Specifically, **episensr** allows for specifications of prior distributions for sensitivity and specificity, such as uniform and logit normal, as well as sequential bias modeling that can be applied to more than 1 type of bias, such as for both misclassification and selection biases.

Other methods employ a Bayesian framework (Greenland 2005; Chu *et al.* 2006; MacLehose and Gustafson 2012; Gustafson *et al.* 2006). The Bayesian analysis is substantially different from conventional probabilistic sensitivity analysis and much more flexible for incorporating different types of priors; however, it is often computationally expensive and more difficult to conduct by general users without statistical background. Gustafson *et al.* (2006) accounted for the prior uncertainties of sensitivity and specificity in the evaluation of the results of a case-control study. MacLehose and Gustafson (2012) compared Bayesian approach with probabilistic bias analysis based on a case–control study of congenital defects, concluding that the two approaches are mostly similar if using similar prior data admissibility as well as uniform priors on exposure probabilities.

This article focuses on an R package correcting for exposure misclassification in a case-control study. Extending from the Bayesian approach introduced by Gustafson *et al.* (2006), we implement the methods outlined in Chu *et al.* (2006), which account for the correlation between the sensitivity and specificity in the model specification. The methods can be applied for both non-differential and differential misclassification; that is, the degree of misclassification can be the same across the case and control groups or distinctly different. Furthermore, we use the generalized linear mixed bivariate effects model introduced by Chu *et al.* (2010) to jointly model the sensitivity and specificity that may be informed by an external meta-analysis on the diagnostic accuracy of the exposure factor.

This article introduces the implementation of the methods for misclassification via our R (R Core Team 2018) package **BayesSenMC** (Bayesian sensitivity analysis by Monte Carlo sampling). The package is mainly implemented in Stan, an imperative probabilistic lan-

guage, which uses Hamilton Monte Carlo (HMC), a form of Markov Chain Monte Carlo (MCMC) sampling. The article is organized as follows. Section 2 gives an actual case–control study that can be used to illustrate the use of the package. Section 3 introduces the specific models and methods to deal with misclassification. Section 4 outlines the implementations in **BayesSenMC**. Section 5 illustrates the capabilities of **BayesSenMC** using the illustrative case–control study and the meta-analysis data on the accuracy of bipolar disorder. Finally, Section 6 concludes this paper with brief discussion.

2. An illustrative Example

This section presents an illustrative case-control study on the association between bipolar disorder and rheumatoid arthritis, originally investigated by Farhi *et al.* (2016); this example will be used to demonstrate the implementation of the methods for misclassification. The exposure is bipolar disorder and the disease outcome is rheumatoid arthritis, which is a chronic autoimmune disorder that primarily affects joints and occurs in nearly 1% of the population in developed countries (McInnes and Schett 2017). Table 1 presents the data.

Table 1: Counts of the case–control study of the association between bipolar disorder and rheumatoid arthritis.

	Bipolar Disorder		
Rheumatoid arthritis	Exposed	Unexposed	Total
Case	66	11,716	11,782
Control	243	57,730	57,973

The unadjusted odds ratio is 1.34 with the 95% confidence interval (CI) (1.02, 1.76), indicating a significant association between rheumatoid arthritis and bipolar disorder. Of note, Farhi *et al.* (2016) acknowledged the limitation that "lack of validation of the diagnosis of bipolar disorder in the subjects cannot be completely excluded". For example, the bipolar disorder can be classified as type I, type II, etc.; it is especially difficulty to diagnose bipolar disorder type II (Phillips and Kupfer 2013).

Assuming certain fixed values or prior distributions of sensitivity and specificity, we can use the method by Chu *et al.* (2006) to correct the odds ratio accounting for the exposure misclassification. The sensitivity and specificity may be either some tentative fixed values or random variables following some prior distributions.

The prior distributions can be estimated from external evidence using a meta-analysis, e.g. using the Bayesian approach by Chu *et al.* (2010). Section 3 presents the details of these methods. This article uses the meta-analysis performed by Carvalho *et al.* (2015) to obtain the prior distributions of the sensitivity and specificity of classifying the exposure status of bipolar disorder. The meta-analysis contains three subgroups of screening detection instruments: bipolar spectrum diagnostic scale (8 studies with the sensitivity between 0.52 and 0.90 and the specificity between 0.51 and 0.97); hypomania checklist (17 studies with the sensitivity between 0.69 and 1.00 and the specificity between 0.36 and 0.98); and mood disorder questionnaire (30 studies with the sensitivity between 0.00 and 0.91 and the specificity between 0.47 and 1.00). The data set from the Clalit Health Services (the largest Health Maintenance Organization in Israel) used in Farhi *et al.* (2016) does not specify the exact screening detection instrument for identifying bipolar disorder; therefore, we will use all three subgroups'

data (55 studies in total) in Carvalho *et al.* (2015) for our analysis on the sensitivity and specificity. A subset of the meta-analysis data is shown in Table 2.

Table 2: The meta-analysis on diagnosis accuracy of bipolar disorder performed by Carvalho *et al.* (2015).

Study ID	True positive	False negative	True negative	False positive
1	81	9	444	427
2	12	3	44	19
3	74	26	97	3
4	52	16	23	4
5	228	113	18	4
⋮	⋮	⋮	⋮	⋮
55	63	6	32	13

According to their definitions, the study-specific sensitivity and specificity can be estimated as the number of true positives / (the number of true positives plus the number of false negatives) and the number of true negatives / (the number of true negatives plus the number of false positives). For example, study 1 gives the sensitivity $81/(81 + 9) = 0.90$ and the specificity $444/(444 + 427) \approx 0.51$. The generalized linear mixed-effects model will be used to synthesize all 55 studies to estimate an overall sensitivity and specificity.

3. Methods

3.1. Bayesian approach to correcting misclassification bias

Consider a case–control study and we are interested in the odds ratio from this study. Table 3 presents the notation of the observed data. The odds ratio is estimated as

$$\widehat{\text{OR}} = \frac{ad}{bc}.$$

When the odds ratio is larger or smaller than 1, the exposure happens more or less likely in the case group, suggesting an association between the disease status and the exposure status. On the other hand, the odds ratio close to 1 suggests that the disease and the exposure are not associated.

Table 3: Observed counts of a case–control study

Group	Exposed	Unexposed	Total
Case	a	b	N_1
Control	c	d	N_0

Assume that the observed exposure probability is P_k, the true exposure probability is π_k, the sensitivity is Se_k and the specificity is Sp_k for group k ($k = 1$ for the case group and 0 for the control group) in the case–control study. Then, we can represent the observed exposure

probability in terms of the true exposure probability, the sensitivity, and the specificity:

$$
\begin{aligned}
P_k &= \mathsf{P}(\text{observed } E \text{ in group } k) \\
&= \mathsf{P}(\text{observed } E \mid \text{true } E \text{ in group } k)\mathsf{P}(\text{true } E \text{ in group } k) \\
&\quad + \mathsf{P}(\text{observed } E \mid \text{true } \overline{E} \text{ in group } k)\mathsf{P}(\text{true } \overline{E} \text{ in group } k) \\
&= \pi_k Se_k + (1 - \pi_k)(1 - Sp_k),
\end{aligned} \tag{1}
$$

where E denotes exposure and $\overline{E}$ denotes non-exposure. This yields

$$
\pi_k = (P_K + Sp_k - 1)/(Se_k + Sp_k - 1).
$$

Consequently, the misclassification-corrected odds ratio can be calculated as

$$
\mathrm{OR_c} = \frac{\pi_1/(1 - \pi_1)}{\pi_0/(1 - \pi_0)} = \frac{(P_1 + Sp_1 - 1)(Se_0 - P_0)}{(P_0 + Sp_0 - 1)(Se_1 - P_1)}. \tag{2}
$$

Based on Equation (1), we can specify the following Bayesian hierarchical model to estimate the corrected odds ratio of the case–control study:

$$
\begin{aligned}
\text{Likelihood:} &\quad a \sim Bin(N_1, P_1) \text{ and } c \sim Bin(N_0, P_0); \\
\text{Link:} &\quad P_k = \pi_k Se_k + (1 - \pi_k)(1 - Sp_k), \quad k = 0, 1; \\
&\quad \mathrm{LOR_c} = \mathrm{logit}(\pi_1) - \mathrm{logit}(\pi_0) \text{ and } \mathrm{OR_c} = \exp(\mathrm{LOR_c}); \\
\text{Prior:} &\quad \mathrm{logit}(\pi_0) \sim N(0, 10^2) \quad \text{and} \quad \mathrm{LOR_c} \sim N(0, 2^2); \\
&\quad Se_k, Sp_k \sim f(\cdot).
\end{aligned} \tag{3}
$$

Here, we assume weakly-informative priors for the true exposure probabilities π_0 and π_1, which give a 95% CI of the true odds ratio between $e^{-2\times1.96}$ (≈ 0.02) and $e^{2\times1.96}$ (≈ 50.40), and a binomial distribution for the number of observed exposure in case and control studies. Thus, the Bayesian inference can be formulated as a posterior distribution of π_0, π_1, and the corrected odds ratio.

Additionally, the function $f(\cdot)$ denotes the joint prior for the sensitivity and specificity (for either non-differential or different misclassification). In practice, the sensitivity and specificity are not available from the case-control study, and they may be estimated as certain fixed values by subjective experts' opinions.

Alternatively, we may consider incorporating evidence-based prior information from existing studies on the diagnostic accuracy of the exposure status (e.g., the data in Table 2). This permits us to account for uncertainties in the sensitivity and specificity and the potential correlation between them. Section 3.2 presents methods to obtain the prior information for the sensitivity and specificity from a meta-analysis.

3.2. Estimating prior distributions on the sensitivity and specificity from a meta-analysis

This section briefly discusses the generalized linear mixed-effects model (GLMM) to estimate priors on the sensitivity and specificity. Suppose that a meta-analysis on the diagnostic accuracy of the exposure status is available as an external data to inform the priors of sensitivity and specificity that are needed to correct the odds ratio in the case-control study. Denote the

number of independent studies in the meta-analysis as m, and let n_{i11}, n_{i00}, n_{i01} and n_{i10} be the number of true positives, true negatives, false positives, and false negatives, respectively, in study i ($i = 1, \ldots, m$). Consequently, there are $n_{i11} + n_{i10}$ truly exposed subjects and $n_{i00} + n_{i01}$ truly unexposed subjects.

Assuming that n_{i11} and n_{i00} are binomially distributed given the number of exposed and unexposed subjects, respectively, the bivariate GLMM can be specified as (Chu *et al.* 2010; Ma *et al.* 2016):

$$n_{i11} \sim Bin(n_{i11} + n_{i10}, Se_i) \text{ and } n_{i00} \sim Bin(n_{i00} + n_{i01}, Sp_i), \quad i = 1, \ldots, m;$$

$$g\left(\frac{Se_i - Se^L}{Se^U - Se^L}\right) = u + \mu_i \text{ and } g\left(\frac{Sp_i - Sp^L}{Sp^U - Sp^L}\right) = v + \nu_i; \tag{4}$$

$$\begin{bmatrix} \mu_i \\ \nu_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\mu^2 & \rho\sigma_\mu\sigma_\nu \\ \rho\sigma_\mu\sigma_\nu & \sigma_\nu^2 \end{bmatrix}\right),$$

where u and v are the fixed effects implying the overall sensitivity and specificity in all m studies, and μ_i and ν_i are the study-specific random effects. Also, σ_μ^2 and σ_ν^2 describe the heterogeneity of the underlying sensitivity and specificity across studies, and ρ models the correlation between the sensitivity and specificity. We denote the estimated fixed effects as $\hat{u}$ and $\hat{v}$, the estimated variances as $\hat{\sigma}_\mu^2$ and $\hat{\sigma}_\nu^2$, and the estimated correlation coefficient as $\hat{\rho}$.

The lower and upper bounds Se^L, Se^U, Sp^L, and Sp^U provide constraints on sensitivity and specificity, which are chosen to exclude all improbable values. A smaller difference between Se^L and Se^U (or between Sp^L and Sp^U) indicates higher confidence in the diagnostic accuracy of the exposure status. When there is no confidence for the range, can set $Se^L = Sp^L = 0$ and $Se^U = Sp^U = 1$. Alternatively, setting $Se^L = Sp^L = 0.5$ indicates that the diagnosis of exposure is better than chance. For simplicity of implementation, we only allow the same lower and upper bounds for Se and for Sp in our package **BayesSenMC**. In addition, $g(\cdot)$ is the link function (e.g., the logit, probit, and complementary log-log). The logit link, $\text{logit}(t) = \log\frac{t}{1-t}$, is commonly used in practice, and our package **BayesSenMC** adopts this link.

Recall that the Bayesian hierarchical model for estimating the corrected odds ratio in the case–control study in Equation (3) specifies a joint prior $f(\cdot)$ for the sensitivity and specificity. We consider five specifications for this prior as follows:

(i) No misclassification. The specification of the prior is equivalent to setting $Se_0 = Se_1 = Sp_0 = Sp_1 = 1$; consequently, $P_1 = \pi_1$ and $P_0 = \pi_0$.

(ii) Misclassification of the exposure status exists, and the sensitivity and specificity for both cases and controls are assumed to be fixed values. These fixed values can be directly plugged in the Bayesian model in Equation (3).

(iii) Non-differential misclassification of the exposure status exists ($Se_0 = Se_1 = Se$ and $Sp_0 = Sp_1 = Sp$), and the uncertainties of the sensitivity and specificity are considered independently by using normal priors on the logit scale. The evidence of the priors comes from the diagnostic meta-analysis performed in the GLMM in Equation (4). Specifically, we can assign

$$\text{logit}\left(\frac{Se - Se^L}{Se^U - Se^L}\right) \sim N(\hat{u}, \hat{\sigma}_\mu^2) \quad \text{and} \quad \text{logit}\left(\frac{Sp - Sp^L}{Sp^U - Sp^L}\right) \sim N(\hat{v}, \hat{\sigma}_\nu^2)$$

as the prior in the Bayesian hierarchical model for the case–control study in Equation (3).

(iv) Non-differential misclassification of the exposure status exists ($Se_0 = Se_1 = Se$ and $Sp_0 = Sp_1 = Sp$), and the sensitivity and specificity have a joint normal prior on the logit scale to account for their correlation. In practice, the sensitivity is very likely correlated with the specificity when dichotomizing a continuous measurement (Chu and Cole 2006). Specifically, we use the following bivariate joint prior

$$
\begin{bmatrix} \text{logit}\left(\frac{Se - Se^L}{Se^U - Se^L}\right) \\ \text{logit}\left(\frac{Sp - Sp^L}{Sp^U - Sp^L}\right) \end{bmatrix} \sim N\left(\begin{bmatrix} \hat{u} \\ \hat{v} \end{bmatrix}, \begin{bmatrix} \hat{\sigma}_\mu^2 & \hat{\rho}\hat{\sigma}_\mu\hat{\sigma}_\nu \\ \hat{\rho}\hat{\sigma}_\mu\hat{\sigma}_\nu & \hat{\sigma}_\nu^2 \end{bmatrix} \right).
$$

Compared with the previous prior specification with independent sensitivity and specificity, the correlation coefficient $\hat{\rho}$ is additionally considered here; it is also estimated from the GLMM in Equation (4).

(v) In addition to the above bivariate joint prior for the non-differential sensitivity and specificity, we can also consider modeling the uncertainties in the estimated correlation coefficient. We consider applying Fisher's z-transformation to the correlation coefficient in the GLMM. Specifically, instead of directly estimating the correlation coefficient ρ in Equation (4), we reparameterize $\rho = \frac{\exp(2z) - 1}{\exp(2z) + 1}$ and obtain the point estimate of z from the GLMM and its standard error, denoted as $\hat{z}$ and s_z, respectively. These estimates can be subsequently used as the prior for the sensitivity and specificity in the case–control study:

$$
\begin{bmatrix} \text{logit}\left(\frac{Se - Se^L}{Se^U - Se^L}\right) \\ \text{logit}\left(\frac{Sp - Sp^L}{Sp^U - Sp^L}\right) \end{bmatrix} \sim N\left(\begin{bmatrix} \hat{u} \\ \hat{v} \end{bmatrix}, \begin{bmatrix} \hat{\sigma}_\mu^2 & \rho\hat{\sigma}_\mu\hat{\sigma}_\nu \\ \rho\hat{\sigma}_\mu\hat{\sigma}_\nu & \hat{\sigma}_\nu^2 \end{bmatrix} \right);
$$
$$
\rho = \frac{\exp(2z) - 1}{\exp(2z) + 1};
$$
$$
z \sim N(\hat{z}, s_z^2).
$$

(vi) Finally, we consider the differential misclassification of the exposure status, i.e., $Se_0 \neq Se_1$ and $Sp_0 \neq Sp_1$. All above choices of prior can be similarly applied the four-variate set $\{Se_0, Sp_0, Se_1, Sp_1\}$. For simplicity, we consider a joint prior that is similar to that in (iv); however, the prior applies to cases and controls separately, and it does not account for the uncertainties in the correlation coefficient as in (v). That is,

$$
\begin{bmatrix} \text{logit}\left(\frac{Se_k - Se^L}{Se^U - Se^L}\right) \\ \text{logit}\left(\frac{Sp_k - Sp^L}{Sp^U - Sp^L}\right) \end{bmatrix} \sim N\left(\begin{bmatrix} \hat{u} \\ \hat{v} \end{bmatrix}, \begin{bmatrix} \hat{\sigma}_\mu^2 & \hat{\rho}\hat{\sigma}_\mu\hat{\sigma}_\nu \\ \hat{\rho}\hat{\sigma}_\mu\hat{\sigma}_\nu & \hat{\sigma}_\nu^2 \end{bmatrix} \right), \quad k = 0, 1.
$$

Because of the complexity of the Bayesian model in Equation (3) with the above various choices of prior for the sensitivity and specificity, we will use Markov chain Monte Carlo (MCMC) sampling to produce the posterior distribution and thus estimate the misclassification-bias-corrected odds ratio in the case–control study and its credible interval.

4. Implementation

The proposed methods in Section 3 have been implemented in the R package **BayesSenMC**. The function `nlmeNDiff` fits a non-differential GLMM and returns a **lme4** (Bates *et al.* 2015) object, for which commands such as `summary` can be used to extract useful statistics from the model (see `methods(class = "merMod")` for more details). Users can also call the `paramEst` function to get a list of specific parameter estimates of the fit that can be directly inputted into the model functions of **BayesSenMC** for Bayesian inferences. In addition, the link function used in `nlmeNDiff` can be modified by specifying `lower` and `upper`, which then changes the lower and upper bounds of Se_k and Sp_k ($k = 1$ for cases and 0 for controls).

There are 6 model functions and 1 graphing function, `plotOR`, in **BayesSenMC**. The model functions return an S4 object of type `stanfit`, an instance of **RStan** (Stan Development Team 2018), which is an interface of Stan Carpenter *et al.* (2017) in R. Users can call methods such as `print` or `extract` to get further information about the posterior sample. All the MCMC procedures are implemented with default 2 chains, each with 1000 iterations of burn-in period and 2000 iterations to estimate the parameters. They are fit using `stan`, and the default Monte Carlo algorithm is the No-U-Turn sampler, a variant of Hamiltonian Monte Carlo (Hoffman and Gelman 2014; Betancourt 2017). Any additional arguments to the model function call will be passed into `stan`. The returned object can then be inputted into `OR_hist` to visualize the posterior distribution of the adjusted odds ratio, plus the probability density lines of odds ratio in the cases of no misclassification and/or constant Se/Sp as a comparison to the posterior distribution. It takes optional argument passed into `geom_histogram`, and returns a **ggplot2** (Wickham *et al.* 2018) object that can be further customized.

The source version of **BayesSenMC** is freely available from the Github repository. The package can be directly installed via the R prompt:

```
R> library("devtools")
R> devtools::install_github("formidify/BayesSenMC", build = TRUE,
+      build_opts = c("--no-resave-data", "--no-manual"))
```

The analyses as well as the reproducible R script can be accessed with the following R prompts:

```
R> library("BayesSenMC")
R> vignette("BayesSenMC_demo", package = "BayesSenMC")
R> edit(vignette("BayesSenMC_demo", package = "BayesSenMC"))
```

5. Example

The data in Table 1 as well as Table 2 of meta analysis data on the diagnosis accuracy of bipolar disorder are used to demonstrate the capabilities of **BayesSenMC**. The analyses are conducted using R version 3.4.2 (2017-09-28).

We first fit the meta-analysis data using the GLMM procedure implemented in our package, assuming non-differential misclassification. Given the range of Se and Sp of the bipolar disorder meta data, we must only assume $Se^L = Sp^L = 0$ and $Se^U = Sp^U = 1$ in order for the GLMM to compute real-value results. However, with more information about the type of diagnoses in Farhi *et al.* (2016), one can find more informed constraint on Se and Sp to fit a more precise model.

```
R> library("BayesSenMC")
R> my.mod <- nlmeNDiff(bd_meta, lower = 0)
```

```
Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [glmerMod]
 Family: binomial  ( logit(0, 1) )
Formula: cbind(Y, N - Y) ~ ((0 + Se + Sp) | sid) + Se
   Data: dat_final
      AIC       BIC    logLik  deviance  df.resid
 851.6825  865.1849 -420.8412  841.6825       105
Random effects:
 Groups Name Std.Dev. Corr
 sid    Se   0.7116
        Sp   0.8935   -0.38
Number of obs: 110, groups:  sid, 55
Fixed Effects:
(Intercept)          Se
    1.12626     -0.05746
```

The indicator variable `Se` has a value of 1 for Se estimates, and 0 for Sp estimates. The random effects are grouped within each study, numbered after `sid`.

The fit has an Akaike Information Criterion (AIC), which can be used to compare across models. The logit means of Se and Sp are given by the fixed effects, 1.069 and 1.126, which translates to a sensitivity of 0.744 and a specificity of 0.755. The values, larger than 0.5, suggest that overall, the diagnostic accuracy for bipolar disorder given our meta-data is better than random, albeit nowhere near perfect. The standard deviations of logit Se and Sp are given by the random effects, as well as the correlation. All of the mentioned parameter estimates can be returned in a list by calling `paramEst`.

We then plug the parameter estimates to get posterior distributions for the corrected odds ratio given different priors of Se and Sp. We run all 6 different models with the case-control study observations in Farhi *et al.* (2016), shown in Table 1.

```
R> m.1 <- correctedOR(a = 66, N1 = 11782, c = 243, N0 = 57973,
+      chains = 3, iter = 10000)
R> m.2 <- crudeOR(a = 66, N1 = 11782, c = 243, N0 = 57973,
+      se = 0.744, sp = 0.755, chains = 3, iter = 10000)
R> m.3 <- logitOR(a = 66, N1 = 11782, c = 243, N0 = 57973,
+      m.lg.se = 1.069, m.lg.sp = 1.126, s.lg.se = 0.893,
+      s.lg.sp = 0.712, chains = 3, iter = 10000)
```

```
R> m.4 <- fixedCorrOR(a = 66, N1 = 11782, c = 243, N0 = 57973,
+      m.lg.se = 1.069, m.lg.sp = 1.126, s.lg.se = 0.893,
+      s.lg.sp = 0.712, rho = -0.379, chains = 3, iter = 10000)
R> m.5 <- randCorrOR(a = 66, N1 = 11782, c = 243, N0 = 57973,
+      m.lg.se = 1.069, m.lg.sp = 1.126, s.lg.se = 0.893,
+      s.lg.sp = 0.712, m.z = -0.399, s.z = 0.139, chains = 3,
+      iter = 10000)
R> m.6 <- diffOR(a = 66, N1 = 11782, c = 243, N0 = 57973,
+      mu = c(1.069, 1.069, 1.126, 1.126), s.lg.se0 = 0.893,
+      s.lg.se1 = 0.893, s.lg.sp0 = 0.712, s.lg.sp1 = 0.712,
+      corr.sesp0 = -0.377, corr.sesp1 = -0.377, corr.group = 0,
+      chains = 3, iter = 10000, traceplot = TRUE)
```
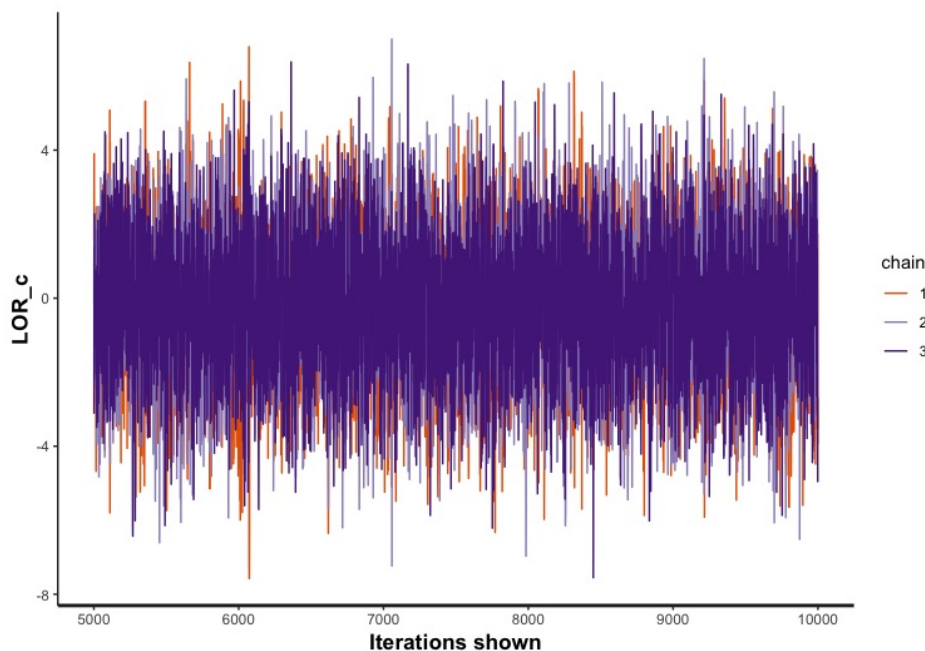

Figure 1: Traceplot of 3 MCMC chains with 10,000 iterations for randomly correlated logit bivariate model

Each model above is ran with 3 Markov chains, and each chain consists of 5000 burn-in samples and 10,000 iterations to estimate the parameters. The posterior mean, median and 95% confidence limits of adjusted odds ratio are as below: 1.34(1.33, 1, 1.72), 5.39(0.84, 0.02, 36.47), 8.39(1.86, 0.03, 51.27), 8.61(1.93, 0.03, 54.18), 9.12(1.89, 0.02, 59.68), 4.87(0.83, 0.02, 33.6). One can also specify `traceplot = TRUE` to display a plot of sampled corrected log odds ratio values over iterations, such as in the above `diffOR` method call.

The above example demonstrates the significance of sensitivity and specificity in a case-control study. We can examine that by the ratio of upper to lower 95% posterior interval: $1.72/1.00 = 1.72$, $36.47/0.02 = 2077.61$, $51.27/0.03 = 1912.59$, $54.18/0.03 = 2002.33$, $59.68/0.02 = 2421.12$ and $33.6/0.02 = 1774.28$. The greatest jump happens when we assume misclassification in the case-control study, and it only increases slightly with more uncertainties in the model. The increase is especially significant in Farhi *et al.* (2016), because as seen from the GLMM,

the estimated mean Se and Sp are around only 0.75.

```
R> library(ggplot2)
R> g1 <- plotOR(m.1, a = 66, N1 = 11782, c = 243, N0 = 57973,
+       se = 0.744, sp = 0.755, x.max = 3, y.max = 5, binwidth = 0.1) +
+       ggtitle("(a)")
```
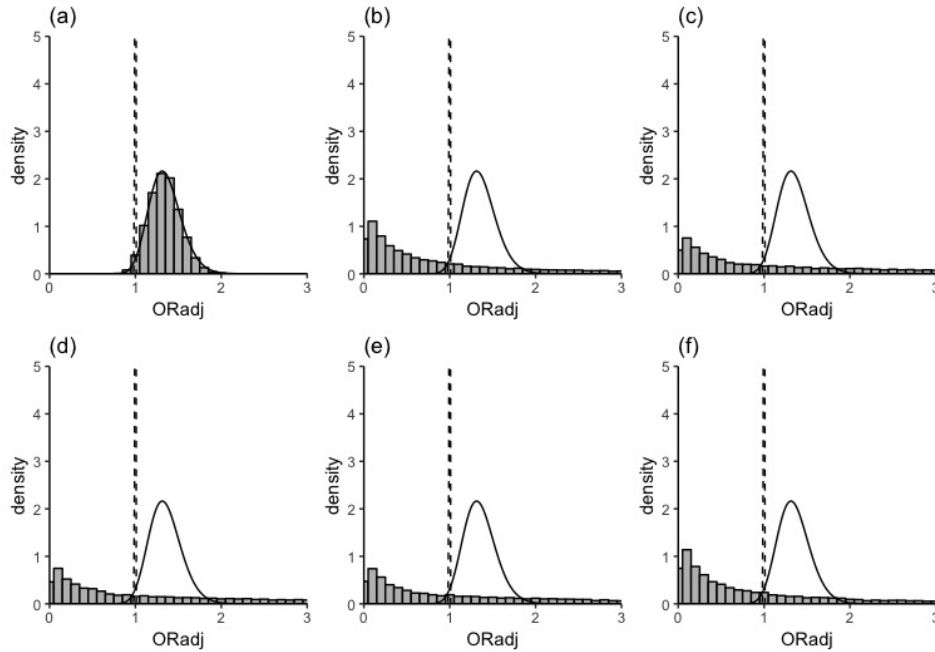

Figure 2: Visualization of Posterior Distributions of corrected odds ratio for all models. (a) no misclassification; (b) constant misclassification with $Se = 0.744$ and $Sp = 0.755$; (c) logit bivariate normal misclassification; (d) extends from (c) but with constant correlation between Se and Sp; (e) extends from (c) but with Fisher's Z transformed correlation; The solid and dotted line are the probability density lines of corrected and crude OR, respectively, assuming log-normality on odds ratio.

We also implement a graphing function, `OR_hist`, which takes input of a model built with one of the above methods, the observations of the same case-control study and the estimated Se and Sp from the GLMM. The method visualizes the posterior distribution of that model while also plotting the probability density line of adjusted odds ratio given no misclassification (corrected OR) and given constant misclassification as specified by `Se` and `Sp` (crude OR). It makes it easy for users to compare the current posterior distribution (especially for models with more uncertainty) with more certain models to visualize the effect of misclassification in a case-control study. In addition, the lines serve as references when comparing across models. The plots and relevant codes are shown in Figure 2.

According to the plot, we observe a drastic change to the posterior distribution after taking non-perfect Se and Sp into account. Then, we observe slightly more uniform distributions as there is more uncertainty in the model. What is also worth noting is that in part (b) of the

plot, the posterior density and MCMC sampling do not share the same shape, even though both assume non-perfect constant Se and Sp. This may be a result of low Se and Sp values, which may affect the log-normality assumption in the MCMC posterior samples.

We now show the effects of number of iterations and chains on the computing speed of our models. For example, `randCorrOR`, which is presumably one of the most complex and time-consuming models to compute, takes about 11.04 seconds to run 3 chains with 5000 warm-up periods and 10,000 iterations each. In comparison, it takes about 3.21 seconds to compute 2 chains with 1000 warm-up periods and 2000 iterations each. In practice, higher number of MCMC chains and iterations results in more stable and accurate results, but the trade-off between the two can be significant depending on the model. Furthermore, we find that models, such as provided by `randCorrOR`, have smaller target posterior distribution regions in a MCMC chain, thus rendering it easy for the algorithm to miss the true distribution and thus result in "divergent transitions", which may return biased estimates. Increasing the value of `adapt_delta` parameter up to 1 (default is 0.99 in our implementation) by including it in the `control` argument of the methods can effectively force **RStan** (Stan Development Team 2018) to take smaller steps to approach the target; however, it may dramatically increase the modeling speed of MCMC chains.

6. Conclusion

In this paper, we introduced and implemented the methods for making posterior inferences on the corrected odds ratio by modeling the uncertainty on both differential and non-differential misclassification through appropriate prior distributions. The specific implementation is publicly available using the R package **BayesSenMC**. The process can be divided into parts. First, one can use the GLMM model with binomial-logit link to estimate prior information on Se and Sp via a meta-analysis on the misclassification of exposure status. Second, the estimates can be plugged into the modeling functions to provide inferences for the corrected odds ratio. The models can also be visualized side-by-side for better comparisons. The validity of the analyses depends highly on the relevance of meta-analysis, in which irrelevant studies may skew the prior estimates of Se and Sp significantly, and consequentially, the corrected odds ratio. In addition, our models assume normal and independent priors on true exposure probabilities, which may be limiting in some cases (Chu *et al.* 2006).

References

Bates D, Mächler M, Bolker B, Walker S (2015). "Fitting Linear Mixed-Effects Models Using **lme4**." *Journal of Statistical Software*, **67**(1), 1–48. `doi:10.18637/jss.v067.i01`.

Betancourt M (2017). "A Conceptual Introduction to Hamiltonian Monte Carlo." *ArXiv e-prints*. `1701.02434`.

Brakenhoff TB, Mitroiu M, Keogh RH, Moons KGM, Groenwold RHH, van Smeden M (2018). "Measurement Error Is Often Neglected in Medical Literature: A Systematic Review." *Journal of Clinical Epidemiology*, **98**, 89–97. `doi:10.1016/j.jclinepi.2018.02.023`.

Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017). "Stan: A Probabilistic Programming Language." *Journal of Statistical Software*, **76**(1), 1–32. doi:10.18637/jss.v076.i01.

Carvalho AF, Takwoingi Y, Sales PMG, Soczynska JK, Köhler CA, Freitas TH, Quevedo J, Hyphantis TN, McIntyre RS, Vieta E (2015). "Screening for Bipolar Spectrum Disorders: A Comprehensive Meta-Analysis of Accuracy Studies." *Journal of Affective Disorders*, **172**, 337–346. doi:10.1016/j.jad.2014.10.024.

Chu H, Cole SR (2006). "Bivariate Meta-Analysis of Sensitivity and Specificity with Sparse Data: A Generalized Linear Mixed Model Approach." *Journal of Clinical Epidemiology*, **59**(12), 1331–1332. doi:10.1016/j.jclinepi.2006.06.011.

Chu H, Guo H, Zhou Y (2010). "Bivariate Random Effects Meta-Analysis of Diagnostic Studies Using Generalized Linear Mixed Models." *Medical Decision Making*, **30**(4), 499–508. doi:10.1177/0272989X09353452.

Chu H, Wang Z, Cole SR, Greenland S (2006). "Sensitivity Analysis of Misclassification: A Graphical and a Bayesian Approach." *Annals of Epidemiology*, **16**(11), 834–841. doi:10.1016/j.annepidem.2006.04.001.

Cole SR, Chu H, Greenland S (2006). "Multiple-Imputation for Measurement-Error Correction." *International Journal of Epidemiology*, **35**(4), 1074–1081. doi:10.1093/ije/dyl097.

Farhi A, Cohen AD, Shovman O, Comaneshter D, Amital H, Amital D (2016). "Bipolar Disorder Associated with Rheumatoid Arthritis: A Case-Control Study." *Journal of Affective Disorders*, **189**, 287–289. doi:10.1016/j.jad.2015.09.058.

Fox MP, Lash TL, Greenland S (2005). "A Method to Automate Probabilistic Sensitivity Analyses of Misclassified Binary Variables." *International Journal of Epidemiology*, **34**(6), 1370–1376. doi:10.1093/ije/dyi184.

Greenland S (1996). "Basic Methods for Sensitivity Analysis of Biases." *International Journal of Epidemiology*, **25**(6), 1107–1116. doi:10.1093/ije/25.6.1107-a.

Greenland S (2005). "Multiple-Bias Modelling for Analysis of Observational Data." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **168**(2), 267–306. doi:10.1111/j.1467-985X.2004.00349.x.

Gustafson P, Le ND, Saskin R (2006). "Case–Control Analysis with Partial Knowledge of Exposure Misclassification Probabilities." *Biometrics*, **57**(2), 598–609. doi:10.1111/j.0006-341X.2001.00598.x.

Haine D (2018). *episensr: Basic Sensitivity Analysis of Epidemiological Results*. R package version 0.9.3, URL https://CRAN.R-project.org/package=episensr.

Hoffman MD, Gelman A (2014). "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." *Journal of Machine Learning Research*, **15**, 1593–1623.

Jurek AM, Maldonado G, Greenland S, Church TR (2006). "Exposure-Measurement Error Is Frequently Ignored When Interpreting Epidemiologic Study Results." *European Journal of Epidemiology*, **21**(12), 871–876. doi:10.1007/s10654-006-9083-0.

Lash TL, Flink AK (2003). "Semi-Automated Sensitivity Analysis to Assess Systematic Errors in Observational Data." *Epidemiology*, **14**(4), 451–458. doi:10.1097/01.EDE.0000071419.41011.cf.

Ma X, Nie L, Cole SR, Chu H (2016). "Statistical Methods for Multivariate Meta-Analysis of Diagnostic Tests: An Overview and Tutorial." *Statistical Methods in Medical Research*, **25**(4), 1596–1619. doi:10.1177/0962280213492588.

MacLehose RF, Gustafson P (2012). "Is Probabilistic Bias Analysis Approximately Bayesian?" *Epidemiology*, **23**(1), 151–158. doi:10.1097/EDE.0b013e31823b539c.

McInnes IB, Schett G (2017). "Pathogenetic Insights from the Treatment of Rheumatoid Arthritis." *The Lancet*, **389**(10086), 2328–2337. doi:10.1016/S0140-6736(17)31472-1.

Phillips ML, Kupfer DJ (2013). "Bipolar Disorder Diagnosis: Challenges and Future Directions." *The Lancet*, **381**(9878), 1663–1671. doi:10.1016/S0140-6736(13)60989-7.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rosner B, Willett WC, Spiegelman D (1989). "Correction of Logistic Regression Relative Risk Estimates and Confidence Intervals for Systematic Within-Person Measurement Error." *Statistics in Medicine*, **8**(9), 1051–1069. doi:10.1002/sim.4780080905.

Rothman KJ, Greenland S, Lash TL (2008). *Modern Epidemiology*. 3rd edition. Lippincott Williams & Wilkins, Philadelphia, PA.

Spiegelman D, Carroll RJ, Kipnis V (2001). "Efficient Regression Calibration for Logistic Regression in Main Study/Internal Validation Study Designs with an Imperfect Reference Instrument." *Statistics in Medicine*, **20**(1), 139–160. doi:10.1002/1097-0258(20010115)20:1<139::AID-SIM644>3.0.CO;2-K.

Stan Development Team (2018). *RStan: R interface to Stan*. R package version 2.18.1, URL https://CRAN.R-project.org/package=rstan.

Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, Woo K (2018). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. R package version 3.1.0, URL https://CRAN.R-project.org/package=ggplot2.

Affiliation:

Jinhui Yang
Carleton College
Northfield, MN, USA
E-mail: yangj2@carleton.edu