Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in futureworks (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Date

Assessing the Impact of Misclassification in Case-Control Studies: An Examination of Bayesian Credible Intervals

By

Jiali Lu
Master of Science in Public Health

Biostatistics

_____

Dr. Robert H. Lyles
Thesis Advisor

_____

Dr. Michael J. Haber
Reader

Assessing the Impact of Misclassification in Case-Control Studies: An Examination of Bayesian Credible Intervals

Jiali Lu

B.S.
Northwest A&F University
2021

Advisor: Robert H. Lyles, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Biostatistics
2024

Abstract

Assessing the Impact of Misclassification in Case-Control Studies: An Examination of Bayesian Credible Intervals

By Jiali Lu

Objectives: Case-control studies are fundamental in identifying associations between exposures and rare diseases, yet the accuracy of odds ratio (OR) estimates can be compromised by misclassification. Traditional Wald confidence intervals have limitations, especially with small sample sizes or extreme proportions. Bayesian methods utilizing Jeffreys priors have been proposed as a robust alternative, yet empirical evidence on their comparative efficacy is sparse when misclassification is an issue. This study aimed to evaluate the performance of Bayesian methods with Jeffreys priors against traditional Wald approaches in interval estimation for the exposure odds ratio, focusing on interval width and frequentist coverage in the context of case-control studies with main and internal validation data.

Methods: We analyzed a real dataset to compare a naïve log odds ratio estimate versus an adjusted (via maximum likelihood) estimate, and to compare the Wald-type confidence interval with a proposed credible interval based on a Jeffreys Dirichlet prior. Then, we conducted simulations with smaller case and control sample sizes to compare the performance of Wald-based and Bayesian credible intervals in terms of interval width and coverage probability. The simulations were designed to reflect a realistic range of exposure misclassification scenarios encountered in epidemiological research.

Results: Our simulations demonstrated that the proposed Bayesian credible interval, compared to the Wald interval, offered significantly narrower interval widths while maintaining near-nominal frequentist coverage across a variety of exposure odds ratios and misclassification scenarios. Specifically, Bayesian methods provided more precise interval estimates and favorable coverage probabilities, underlining their potential for more accurate and reliable statistical inference in case-control studies.

Conclusions: The findings suggest that Bayesian methods utilizing Jeffreys priors represent a significant advancement over traditional Wald intervals for the estimation of exposure odds ratios in case-control studies featuring main and internal validation study data. This study supports the adoption of Bayesian approaches in epidemiological research, especially in the presence of misclassification and when precise interval estimation is critical.

Keywords: Case-control studies, Bayesian methods, Jeffreys priors, Misclassification.

Assessing the Impact of Misclassification in Case-Control Studies: An Examination of Bayesian Credible Intervals

By

Jiali Lu B.S.
Northwest A&F University
2021

Advisor: Robert H. Lyles, Ph.D.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Biostatistics
2024

Acknowledgment

One year ago, I was deeply struggling with my life in biostatistics, nearly resigning myself to return home for a common job with a modest monthly earning of ¥6,000. My parents were ready to support me, which was a comfort—though I'm unsure if that was entirely a good thing. I was prepared to leave it all behind and venture into the unknown, trying things I'd never considered. But then, things changed dramatically.

I've traveled to many places, enjoyed delicious food, and met a variety of people. Notably, I visited 80 coffee shops in Atlanta—my absolute favorite pursuit. I've also dabbled in academic experiences outside of biostatistics, an area where I used to feel out of depth, as if I were the least knowledgeable person in the room. My research interests have unexpectedly veered into health service research, a field I never thought I would explore. However, with the support of wonderful faculty and friends, I began to consider continuing my academic journey.

After an agonizing four-month wait and a fair share of struggles, I received an offer from Health Policy & Management—the only offer I received this cycle. Now, I am ready to continue my studies at Emory, in the field of Public Health. It's a departure from biostatistics, but I am excited to spend another five years in the same building that has become a second home.

I owe immense thanks to Dr. Bob Lyles for his willingness to guide me through my graduation thesis. His encouragement and belief in me have been a beacon during my darker days. When I confided in him about my lack of self-belief and intention to seek a job after graduation, he helped me to see this thesis not as an end, but as a new beginning. My upcoming Ph.D. will be in a different area, but I aim to incorporate Bayesian methods and missing data into my research endeavors. Dr. Lyles is not only a careful and experienced biostatistician but also an outstanding professor. I am truly grateful to collaborate with him. My gratitude also extends to Dr. Haber for being my reader and for his prompt and insightful suggestions. A heartfelt thank you to all the faculty and staff in BIOS. It has been an honor to be part of such a wonderful department.

Lastly, I thank myself for not giving up.

1. Introduction

Case-control studies serve as a cornerstone in understanding the connections between exposures and rare diseases, revealing potential risk factors through the analysis of exposure distributions between cases and controls[1]. It is imperative to acknowledge that choosing cases and controls reflects a deliberate design decision rather than a reflection of inherent biological significance. Despite this, the odds ratio (OR) stands as a reliable measure of risk ratio, rooted in methodology rather than biology[2]. However, the accuracy of OR estimates can be compromised by misclassification—where participants are erroneously categorized due to observational or measurement inaccuracies—obscuring the true relationship between exposure and outcomes[2]. The introduction of a misclassification matrix by Barron accentuates the gravity of this error by showcasing its ability to distort relative risk estimation[3]. Additionally, Lyles' demonstration of a closed-form ML estimator for the OR in the milieu of exposure misclassification when internal validation data are available furthers our understanding of its effects[4]. An examination of Greenland's analysis of antibiotic use during pregnancy and its correlation with sudden infant death syndrome provides a practical illustration of these principles within case-control studies[5].

Wald-based confidence intervals are widely used in statistical analysis, but they are not without their limitations, which become particularly pronounced with small sample sizes or when dealing with extreme proportions[6]. Moreover, Brown and Cai have highlighted that Wald intervals can yield unreliable inferential results about binomial proportions, even with large sample sizes, and especially in scenarios involving extreme proportions. They furthered this critique by studying the benefits of alternative methodologies grounded in Jeffreys prior[7], providing valuable

advancements and advice for analysis. These alternatives promise more robust statistical inference, potentially rectifying the deficits inherent in the conventional use of Wald intervals.

The Bayesian method offers a sophisticated alternative for statistical estimation, akin to the concept of shrinkage, which refines the precision of estimates[9]. Research conducted by Agresti and Min underscores the effectiveness of the Jeffreys prior in the development of Bayesian credible intervals, particularly for 2x2 contingency tables where significant effects are present[10]. Lyles, Weiss and Waller further this discussion when estimating a single binomial proportion, advocating for a balanced approach that combines the conservatism of the Clopper-Pearson method with the flexibility of Jeffreys prior and thereby optimizing with respect to criteria incorporating aspects of both coverage and interval width[11]. Extending these methodologies, Beavers and Stamey have advanced the field by developing a Bayesian framework for sample size determination within a covariate misclassification model, even in the absence of a gold standard[12].

This study builds upon these foundational insights to scrutinize whether Bayesian methods utilizing Dirichlet priors can surpass traditional Wald-based approaches for interval estimation for the exposure odds ratio in the context of a main/internal validation study design as considered by Greenland[5], Morrissey and Spiegelman[13], and Lyles[4], among others. Through extensive simulations, we meticulously compare these methods, focusing on interval width and frequentist coverage. Our goal is to discern whether the Bayesian framework, with its Dirichlet priors, represents a significant advancement over standard methodologies for point and interval estimation of estimation of a particular nonlinear function of multinomial parameters (i.e., the natural log of the true exposure odds ratio). We conducted simulation studies to examine the coverage and width

properties associated with each interval in the case-control setting with main and internal validation data. Of particular interest in this project are the variations in relative performance of the Wald-based and Dirichlet-based Bayesian credible intervals, as conditions (such as the true exposure odds ratio) are varied.

2. Methods

2.1 A case-control study with misclassification

We provide the data layout for a case-control study based on a main/internal validation study design in Table 1. Here, D (= 0,1) denotes disease status (case), and X (= 0,1) represents the surrogate (i.e., potentially mismeasured) exposure status. The study's design ensures sampling conditional on D, with the assumption of accurate classification for disease status. To augment the primary study data, which comprises pairs of (D, X), a subset of participants is selected for an internal validation process, which involves the use of a gold-standard exposure variable E (= 0,1).

Here, we define:

The true probability of exposure among cases and controls:

$$\pi_d = \Pr(E = 1 \mid D = d) \ \ (d = 0,1) \tag{1}$$

The probability of exposure as measured via the error-prone approach among cases and controls:

(2)

$$\pi_d^* = \Pr(X = 1 \mid D = d) \quad (d = 0,1)$$

The Odds Ratio of interest:

$$OR = \frac{\pi_1(1-\pi_0)}{\pi_0(1-\pi_1)} \tag{3}$$

The sensitivity and specificity under potentially differential misclassification:

$$SE_d = \Pr(X = 1 \mid E = 1, D = d) \quad (d = 0,1) \tag{4}$$

$$SP_d = \Pr(X = 0 \mid E = 0, D = d) \quad (d = 0,1) \tag{5}$$

The positive and negative predictive value under potentially differential misclassification:

$$PPV_d = \Pr(E = 1 \mid X = 1, D = d) \quad (d = 0,1) \tag{6}$$

$$NPV_d = \Pr(E = 0 \mid X = 0, D = d) \quad (d = 0,1) \tag{7}$$

			Internal Validation Study				
	Main Study			D=1		D=0	
D	X=1	X=0	X	E=1	E=0	E=1	E=0
1	n_{11}	n_{12}	1	n_{13}	n_{14}	n_{03}	n_{04}
0	n_{01}	n_{02}	0	n_{15}	n_{16}	n_{05}	n_{06}

TABLE 1: Main and Internal Validation Study Data
D: case status; E: gold standard exposure measure; X, surrogate exposure measure (D, X, E = 0,1).

As an example, we explore a previously published dataset through a dual framework of main and internal validation. We examine the potential link between sudden infant death syndrome (SIDS) and maternal antibiotic use during pregnancy[5,14]. Our investigation is supported by both primary and validation datasets, as delineated in Table 1. Specifically, we consider 775 cases ($D_1 = 775$) and 797 controls ($D_0 = 797$), yielding estimated sensitivity for cases as 0.61 ($SE_1 = 0.61$)and estimated specificity for cases as 0.88 ($SP_1 = 0.88$). The control group has estimated sensitivity of 0.60 ($SE_0 = 0.60$) and specificity of 0.93 ($SP_0 = 0.93$). After adjusting for misclassification via methods reviewed below, the estimated true exposure probabilities for cases and controls are 0.212 and 0.179, respectively. A subset, comprising approximately 25% of the subjects, was randomly chosen for the internal validation set. Table 2 presents the cell counts for the SIDS study data, reflecting the layout of Table 1. Exposure assessments labeled as the gold standard (E) derive from medical records, whereas the surrogate (X) assessments rely on self-reported antibiotic use.

			Internal Validation Study				
	Main Study			D=1		D=0	
D	X=1	X=0	X	E=1	E=0	E=1	E=0
1	n_{11}=122	n_{12}=442	1	n_{13}=29	n_{14}=22	n_{03}=21	n_{04}=12
0	n_{01}=101	n_{02}=479	0	n_{15}=17	n_{16}=143	n_{05}=16	n_{06}=168

TABLE 2: Main and Internal Validation of SIDS

2.2 Maximum Likelihood Estimators of Parameters

Lyles' work enabled us to ascertain the likelihood for the twelve distinct observation types reflected in Table 1[4]. Morrissey & Spiegelman have developed a method to express this likelihood in terms of the log odds ratio[13]. They clarified how the inverse matrix and maximum likelihood (ML) methods are comparable within the studied context. Based on these two references, we advocate for the predominant use of the closed-form MLE previously introduced as the "inverse matrix" method for the odds ratio when differential misclassification is present, and an internal validation study is conducted[8].

By applying the matrix method by Barron[3] and the inverse matrix method of Marshall we can get the two identities[8]:

$$\pi_d^* = SE_d\,\pi_d + (1 - SP_d)(1 - \pi_d)\ (d = 0,1) \tag{8}$$

$$\pi_d = PPV_d\pi_d^* + (1 - NPV_d)(1 - \pi_d^*)\ (d = 0,1) \tag{9}$$

By solving for π_d and π_d^*, we have convenient forms for estimators of the true and error-prone prevalence:

$$\hat{\pi}_d = \frac{\hat{\pi}_d^* + \widehat{SP}_d - 1}{\widehat{SE}_d + \widehat{SP}_d - 1}\ (d = 0,1) \tag{10}$$

$$\hat{\pi}_d^* = \frac{\hat{\pi}_d + \widehat{NPV}_d - 1}{\widehat{PPV}_d + \widehat{NPV}_d - 1} \ (d = 0,1) \tag{11}$$

We can obtain the MLEs for SE_d, SP_d, PPV_d and NPV_d based on the cell counts in Table 1[4]:

$$\widehat{SE}_d = \frac{\widehat{PPV}_d \hat{\pi}_d^*}{\hat{\pi}_d} \ (d = 0,1) \tag{12}$$

$$\widehat{SP}_d = \frac{\widehat{NPV}_d (1 - \hat{\pi}_d^*)}{1 - \hat{\pi}_d} \ (d = 0,1) \tag{13}$$

$$\widehat{PPV}_d = \frac{n_{d3}}{n_{d3} + n_{d4}} \ (d = 0,1) \tag{14}$$

$$\widehat{NPV}_d = \frac{n_{d6}}{n_{d5} + n_{d6}} \ (d = 0,1) \tag{15}$$

$$\hat{\pi}_d^* = \frac{n_{d1} + n_{d3} + n_{d4}}{n_d} \ (d = 0,1) \tag{16}$$

where:

$$n_d = n_{d1} + n_{d2} + n_{d3} + n_{d4} + n_{d5} + n_{d6} \tag{17}$$

The MLE for π_d (d=0,1) is obtained by inserting the estimates in equations (14)-(16) into equation (9). The MLE for the ln (OR) then follows directly.

2.3 Approximate Frequentist Intervals

The Confidence Interval for a binomial proportion, derived from inverting the Wald test, is a common statistical tool[15]. Agresti and Coull investigate the coverage properties of confidence interval construction for a binomial parameter[15]. They compared score intervals attributed to Wilson[16] with Wald[17] intervals and exact intervals based on binomial probabilities.

Wald:

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{18}$$

Wilson:

$$\frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + \frac{z_{\alpha/2}^2}{n}} \tag{19}$$

where:

n = sample size,

$\hat{p}$ = sample proportion,

$z_{1-\alpha/2} = 100(1-\alpha/2)$ standard normal distribution quantile

However, when estimating a proportion's interval, coverage probabilities can be excessively high for "exact" confidence intervals based on inverting the binomial test and excessively low for those based on the Wald large-sample normal test. The latter discrepancy arises due to reliance on the central limit theorem, which can be unreliable when dealing with small sample sizes or success probabilities near 0 or 1[7].

Quesenberry and Hurst[18] adapted the methods of Wilson[16] for simultaneous construction of multinomial parameters and Goodman[19], invoking a Bonferroni argument. Thus, the Wilson intervals are a special case of the Quesenberry and Hurst[18] and Goodman[19] intervals. Agresti and Coull[15] suggested using the Wilson intervals for the binomial proportion. If we are in a case of multinomial, the Goodman[19] intervals are preferred.

Our research conducts log odds ratio and maximum likelihood estimation of the log odds ratio based on the main/internal validation study setting studied by Morrissey and Spiegelman[13] and Lyles[4]. The log odds ratio estimated without adjusting for misclassification ("Naïve") is the initial estimate of the log odds ratio using the observed (main study only) data, mistakenly assuming that the exposure is measured correctly. The Maximum Likelihood Estimate of the log odds ratio[4] accounts for misclassification. This method uses additional information (e.g., the estimated positive predictive values and negative predictive values based on internal validation data) to adjust the odds ratio estimate. It aims to provide a more accurate estimate of the true association between exposure and outcome by correcting for the bias introduced by misclassification.

The asymptotic variance for the ln (OR) estimator is given as follows[4]:

$$\widehat{Var}\{\ln(\widehat{OR})\} = \sum_{d=0}^{1}\{\hat{\pi}_d(1-\hat{\pi}_d)\}^{-2} \times \left\{\left(\widehat{PPV_d} + \widehat{NPV_d} + 1\right)^2 \widehat{var}(\hat{\pi}_d^*) + (\hat{\pi}_d^*)^2 \widehat{var}\left(\widehat{PPV_d}\right) + (1-\hat{\pi}_d^*)^2 \widehat{var}\left(\widehat{NPV_d}\right)\right\} \quad (20)$$

where:

$$\widehat{var}(\hat{\pi}_d^*) = \frac{\hat{\pi}_d^*(1-\hat{\pi}_d^*)}{\sum_{j=1}^{6} n_{dj}} \quad (21)$$

$$\widehat{var}\left(\widehat{PPV_d}\right) = \frac{\widehat{PPV_d}(1-\widehat{PPV_d})}{n_{d3}+n_{d4}} \quad (22)$$

$$\widehat{var}\left(\widehat{NPV_d}\right) = \frac{\widehat{NPV_d}(1-\widehat{NPV_d})}{n_{d5}+n_{d6}} \quad (23)$$

2.4 Bayesian Credible Intervals with Jeffreys Priors

In Bayesian statistics, credible intervals represent a range of values within which an unobserved parameter is expected to fall with a specified probability. This interval is delineated within the domain of either a posterior or predictive probability distribution[22]. Agresti and Min investigate the performance of Bayesian credible intervals for the difference of proportions and consider beta priors, logit-normal priors, and related correlated priors[10]. The most popular Bayesian credible intervals for p are based on Jeffreys beta priors, which leads to corresponding beta posteriors, The Jeffreys interval comes highly recommended for its satisfactory average coverage properties[7]. In our study, we considered credible intervals for the log odds ratio [$\ln(OR)$] and for the odds ratio[4,13]. The first of these refers to a range of values for the logarithm of the odds ratio (OR) as evidenced

by the data. The logarithmic transformation is often applied because odds ratios can only take positive values, and their distribution tends to be right skewed[23]. The Odds Ratio Credible Interval provides a range of values within which the true odds ratio is expected to lie, with a certain probability, based on the Bayesian analysis. It is directly related to the $\ln(OR)$ credible interval but is expressed in the original odds ratio scale rather than the logarithmic scale[24]. In presenting our results, we focus here on intervals for the $\ln(OR)$.

The Dirichlet distribution is the conjugate family of priors for the multinomial distribution. The parameters of the Dirichlet prior have the same sort of interpretation as those of a beta prior, which of course is a special case of the Dirichlet[20]. The Jeffreys prior is a weakly informative prior distribution commonly applied to parameter spaces[21]. We use $Beta\ (\kappa, 1-\kappa)$ and $Beta\ (1-\kappa, \kappa)$ priors to calculate the lower bound and upper bound of the credible interval. In this study, we use a variant on this approach for multinomial data that is analogous to choosing κ =0.5, to produce credible intervals for association parameters. Specifically, we use $Dirichlet\ (0.5, \ldots, 0.5)$ priors for the six multinomial proportions underlying the cells in Table 1, separately for cases and controls. Our approach is to generate a large number of draws from the corresponding Dirichlet posteriors, converting these to draws of the 12 cell counts in Table 1. For each such draw, we mimic the ML estimand. The credible interval is produced by taking the 2.5th and 97.5th percentiles of the resulting estimands.

2.5 Simulation Process

The SAS statistical package V9.4 was used for generating the simulations. Wald intervals were simulated within the SAS IML procedure. A SAS macro generating Dirichlet random variables was built. Credible intervals were estimated based on the posterior distribution. For each scenario considered, 5000 simulations were generated with the number of Dirichlet draws set to 10,000 per simulation.

We initiate a simulation study in a case-control setting to evaluate the effects of differential misclassification on odds ratio estimates. Based on specified parameters including the sensitivities and specificities, the program calculates other essential parameters like the true odds ratio, its logarithm, and predictive values. It then generates data for both cases and controls using the appropriate multinomial distribution based on the specified parameters. Once the data are generated, the program computes the MLE of the $\ln(OR)$ parameter of interest (equivalent to Marshall's inverse matrix estimator[4,8]), along with its standard error.

The simulation iterates through multiple runs, each time generating data and calculating confidence intervals for both the original (without misclassification) and adjusted odds ratio estimates (with misclassification). Additionally, a macro performs further analysis, employing the Dirichlet distribution for posterior distribution estimation and analyzing Bayesian credible intervals. The final step compiles all results, providing a comprehensive analysis of the credible intervals and their coverage, and reports findings through summary statistics, offering insights into the impact of misclassification in epidemiological studies.

3. Results

3.1 Example of SIDS

In the analysis of the SIDS dataset, we estimated the values for each parameter, some of which are presented in Table 3. The naïve estimate of exposure prevalence for cases (p_1) is 0.223, whereas for controls (p_0) it is 0.168. The derived $\ln(OR)$ estimate of 0.351 reflects the odds of exposure between cases and controls, serving as an initial measure of association without adjusting for misclassification. Furthermore, the maximum likelihood estimates (MLE) suggest that the exposure probability for cases stands at 0.209, and for controls at 0.179. The MLE of the log odds ratio is computed to be 0.193, as detailed in Table 3. We can see that the values of the estimates are close, but the suggestion is that the naïve $\ln(OR)$ estimate may be biased away from the null. Table 4 provides an analytical comparison between Wald intervals and Bayesian Credible Intervals. It is noteworthy that the lower limits (-0.241 vs. -0.236), upper limits (0.626 vs. 0.627), and widths of the intervals (0.867 vs. 0.863) share a marked resemblance.

	Naïve Estimates	MLE Estimates
p_1	0.223	0.209
p_0	0.168	0.179
$\ln(OR)$ [SE]	0.351 [0.130]	0.193 [0.22]

TABLE 3: Naïve vs. ML Estimates for SIDS Data

	Wald Interval	Bayesian Credible Interval
Width	0.867	0.863
Lower	-0.241	-0.236
Upper	0.626	0.627

TABLE 4: Comparing Intervals for $\ln(OR)$ Based on SIDS Data

3.2 Analysis of Naïve and ML Estimates for Log Odds Ratios

We present simulation studies with much smaller sample sizes (100 cases and 100 controls), but otherwise directly mimicking the SIDS data example. Table 5 shows the mean "naïve" and ML $\ln(OR)$ estimates and their empirical SDs. Across 5,000 simulation runs, the Naïve method estimates $\ln(OR)$ at 0.384 on average with a smaller empirical standard deviation, suggesting more precision. In comparison, the ML estimate at 0.213 on average demonstrates a higher standard deviation, indicative of less precision for statistical inference. The sacrifice in precision upon adjusting for misclassification is expected, as generally one expects a bias-variance tradeoff. As we can see, the ML estimate is far less biased, while the naïve estimate is biased away from the null.

	Naïve	ML
$\ln(OR)$	0.384(0.374)	0.213(0.725)
true $\ln(OR)$	0.193	0.193

Table 5. Simulation Comparison of Naïve and ML Estimates

3.3 Brief Analysis of Interval Estimates

Table 6 provides a comparative overview of the 95% Naïve Interval, Wald Confidence Interval (CI), and Bayesian Credible Interval for their width and coverage properties. The Naïve Interval is the narrowest at 1.437 on average, offering a somewhat sub-nominal coverage of 92.50%. The Wald CI is wider at a mean of 2.643 with a coverage of 95.70%. Notably, the Bayesian Credible Interval, while somewhat narrower in width than the Wald CI at 2.572, exhibits favorable

frequentist coverage (96.1%). Thus, while the difference remains small under these simulation conditions, the credible interval is narrower on average than the Wald CI while maintaining near-nominal coverage. This suggests that the Bayesian Credible Interval may be more reliable for encompassing the true parameter value, making it a potentially better choice in cases where maximizing coverage is crucial.

	Naïve Interval	Wald CI	Bayesian Credible Interval
Width	1.437	2.643	2.572
Coverage	92.50%	95.70%	96.10%

Table 6. Simulation Comparison of Interval Widths and Coverages

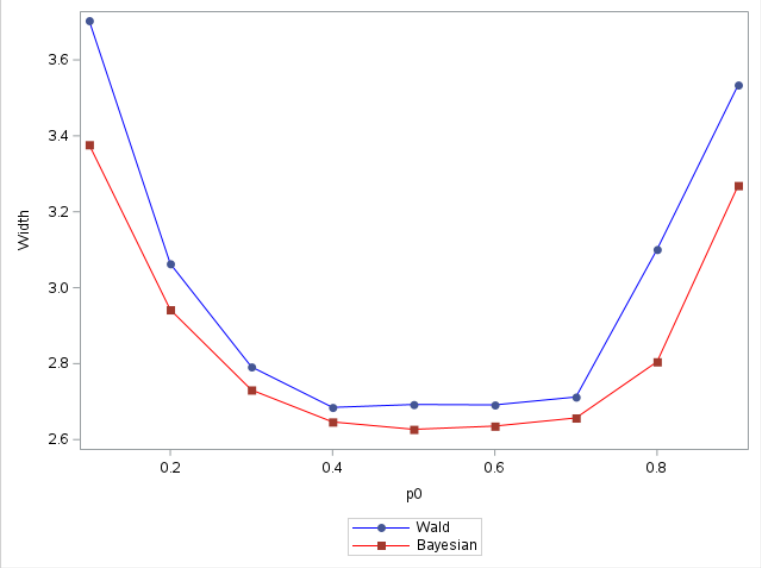
3.3 Varying Exposure Probabilities for Cases and Controls

In an expanded series of simulations (each with 100 cases and 100 controls), we explore the application of Bayesian credible intervals across varying exposure probabilities for cases and controls, with sensitivity and specificity for cases ($SE_1 = 0.6, SP_1 = 0.95$), while for controls, $SE_0 = 0.9$ and $SP_0 = 0.8$. Initially, we established the exposure probability for cases at a fixed rate of 0.1, while the probability for controls was varied between 0.1 and 0.9. Subsequently, we maintained the exposure probability for controls at 0.1 and varied the probability for cases within the same range of 0.1 to 0.9. The outcomes of this variation are initially depicted in Figure 1 and Figure 2, which demonstrate that the Bayesian credible intervals are consistently more precise—indicated by their narrower widths—compared to the Wald intervals. This empirical evidence supports our initial hypothesis regarding the efficiency of Bayesian credible intervals in estimating exposure probability in case-control studies.

In examining the coverage across different values of the exposure probability for controls (p_0), it is evident that both the Wald and Bayesian approaches significantly outperform intervals based on the naïve log odds ratio estimate, particularly when the exposure probability for cases is less than 0.5. As depicted in Figure 3, the coverage for the Naive estimate starts at a lower value and increases sharply as p_0 approaches 0.5, then stabilizes. In contrast, the Wald and Bayesian methods maintain coverage closer to the ideal of 95% throughout, indicating a more consistent performance. Figure 4 highlights coverage for only the Wald and Bayesian intervals, revealing that the Bayesian interval in more conservative (despite its narrower width).

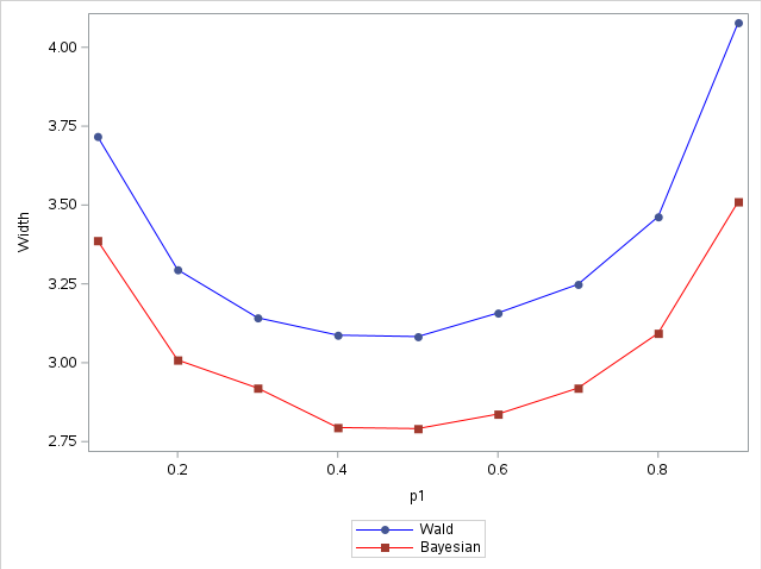
When the exposure probability for the case group (p_1) was varied, our results (as shown in Figure 5 and Figure 6) indicate that neither the Bayesian method nor the Wald method consistently yielded the highest coverage. While both methods generally maintained coverage close to the ideal of 0.95, there are instances where the Wald intervals exhibit slightly higher coverage than the Bayesian credible intervals.

Figure 1: Comparison of Interval Widths When Varying Control Group's Exposure Probability



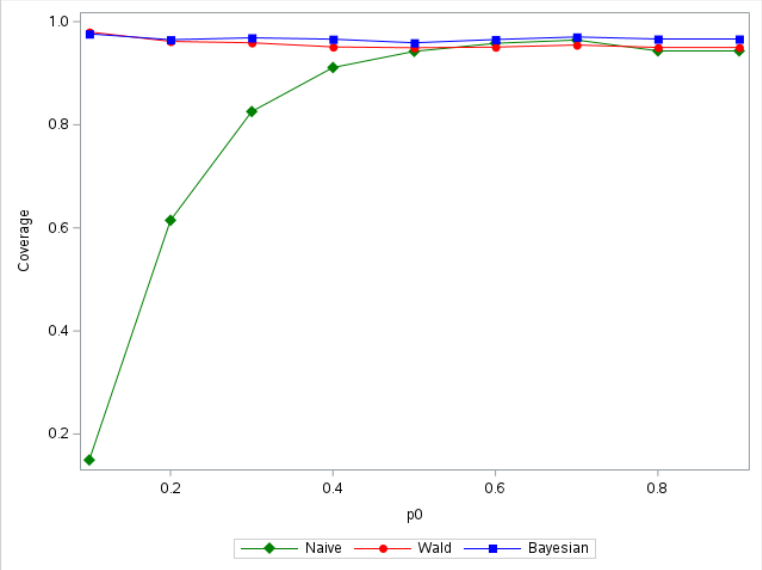
Note: The exposure probability for cases is set at 0.1. Control exposure probability varies from 0.1 to 0.9. The blue line represents the Wald intervals, and the red line represents the Bayesian credible intervals. The narrower Bayesian intervals suggest improved precision.

Figure 2: Comparison of Interval Widths When Varying Case Group's Exposure Probability



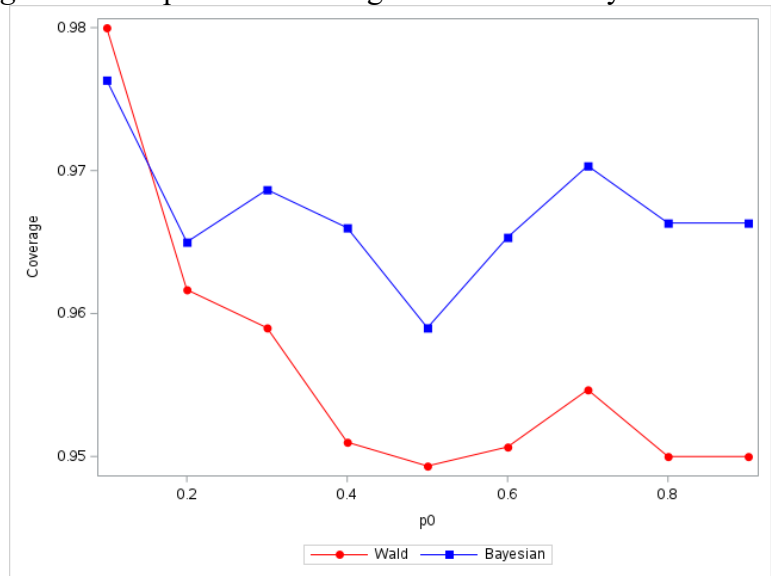
Note: The exposure probability for controls is set at 0.1. Case exposure probability varies from 0.1 to 0.9. The blue line represents the Wald intervals, and the red line represents the Bayesian credible intervals. Consistently narrower Bayesian intervals across most probabilities indicate precision gains.

Figure 3: Coverage of Naive Log Odds Ratio, Wald, and Bayesian Intervals



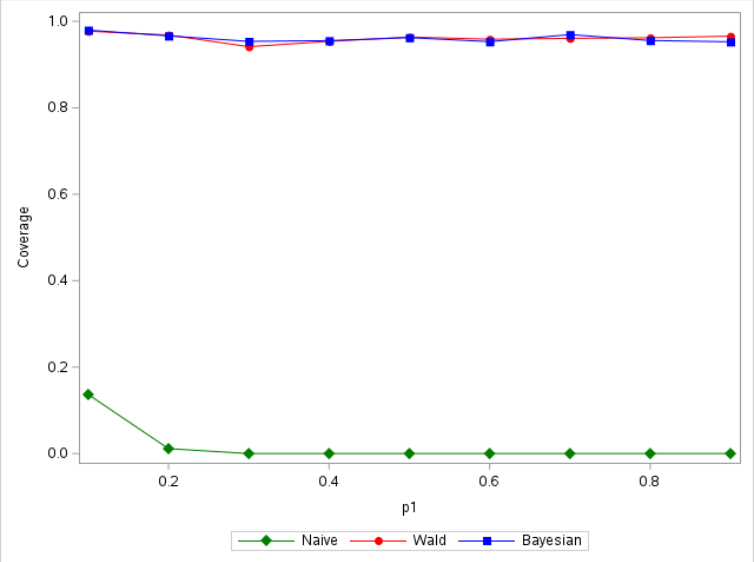
Note: Coverage measures the proportion of times the true parameter value is contained within the interval. Ideal coverage is 0.95. This comprehensive plot includes the Naive method, illustrating its significantly lower coverage compared to Wald and Bayesian methods, which remain closer to the ideal coverage value.

Figure 4: Comparative Coverage of Wald and Bayesian Intervals



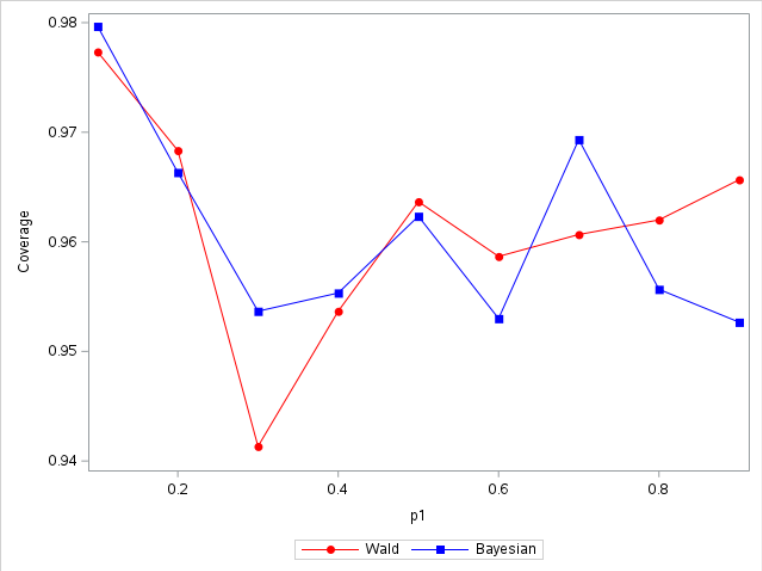
Note: This plot focuses on the comparison between Wald and Bayesian intervals' coverage, highlighting the consistent higher coverage provided by Bayesian intervals as the exposure probability for cases varies. Note that both interval methods produce coverage at or in excess of the targeted 95%.

Figure 5: Coverage of Wald and Bayesian Intervals



Note: Coverage is represented as the proportion of simulations in which the true parameter is captured within the interval. This comprehensive plot includes the Naive method, illustrating its significantly lower coverage compared to Wald and Bayesian methods, which remain closer to the ideal coverage value.

Figure 6: Coverage Comparison of Wald, and Bayesian Intervals



Note: This plot focuses on the comparison between Wald and Bayesian intervals' coverage. While neither approach yields consistently higher coverage in this case, both produce coverage near or in excess of the targeted 95%.

4. Discussion

This study conducted a rigorous examination of the efficacy of Bayesian methods with Jeffrey priors versus traditional Wald intervals for estimating exposure odds ratios in case-control studies with adjustment for nondifferential exposure misclassification by means of internal validation data. The findings reveal that Bayesian credible intervals generally maintained near-nominal coverage but offered narrower widths than Wald CIs, particularly when the exposure probability for cases is below 0.5. This suggests potential for a notable precision advantage of Bayesian intervals in complex data scenarios, such as under misclassification, underscoring their value in statistical estimation within epidemiological research. Upon further analysis, both Bayesian and Wald intervals showed enhanced performance over the naïve CI based on an unadjusted main study-only ln(OR) estimate, maintaining coverage closer to the ideal. However, we noted that it is possible for the naïve interval to perform well while maintaining the narrowest width, such as in our simulations when the exposure probability for controls exceeded 0.5 (Figure 3). Surprisingly, when altering exposure probabilities for controls, Bayesian methods did not always surpass Wald intervals. These findings contradict our initial hypotheses and indicate that the performance of these statistical techniques is more intricate and influenced by specific probabilities.

The implications of this research could be useful for the field of epidemiology, where accurate risk assessment is paramount. By establishing the efficacy of Bayesian credible intervals, particularly in cases with lower exposure probabilities, we suggest a methodological shift that could influence future epidemiological analyses and public health policies. The narrower interval widths and improved coverage provided by Bayesian methods enhance the statistical validity of studies, which is critical for understanding disease etiology and for designing interventions. The utilization of

Dirichlet priors in Bayesian analysis demonstrated here effectively captured the true variability and uncertainty, offering nuanced insights that can guide targeted public health strategies. Such precision becomes increasingly relevant in the current landscape where epidemiological data often inform immediate and consequential health policy decisions.

However, the study's implications must be contextualized within its limitations. The simulation-based design, while offering a controlled environment to compare methods, may not encapsulate the full spectrum of complexities presented by real-world data. In this sense, the superiority of Bayesian credible intervals over Wald intervals is not a given. As conditions vary, there may be scenarios where the Wald intervals match or even exceed the performance of the proposed Bayesian credible intervals in terms of the joint criteria of coverage and width. The veracity of Bayesian credible intervals and their superiority also hinges on the selection of appropriate priors. In this case, the Dirichlet priors were weakly informative by design, as our goal was to improve upon frequentist inference in terms of traditional interval coverage. The appropriateness of the Jeffreys prior might not hold across various datasets (e.g., as the number of cells increases), and the performance of these intervals might differ with other prior distributions. Additionally, the study's scope was limited to a comparative assessment between Bayesian and Wald methods and did not incorporate other statistical techniques that could potentially provide different insights.

Given the performance disparities observed under various exposure probabilities, further investigation is necessary. Future research should explore the conditional performance of Bayesian intervals in a wider array of scenarios and assess the application of alternative priors. Validating these findings through practical applications in retrospective and prospective studies is essential

to ascertain the utility of Bayesian credible intervals in real-world settings. In conclusion, the findings of this study support the adoption of Bayesian methods with Jeffreys priors in the estimation of exposure odds ratios in the setting considered here, perhaps heralding a new direction for continued exploration in the statistical methodology of epidemiology. While further research is imperative to overcome the study's limitations, our findings offer compelling evidence of the potential for Bayesian approaches to enhance the precision and reliability of public health research outcomes.

References

[1] Schulz, K. F., & Grimes, D. A. (2002). Case-control studies: research in reverse. *The lancet*, *359*(9304), 431-434.

[2] Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological reviews*, *82*(4), 591-605.

[3] Barron, B. A. (1977). The effects of misclassification on the estimation of relative risk. *Biometrics*, 414-418.

[4] Lyles, R. H. (2002). A note on estimating crude odds ratios in case–control studies with differentially misclassified exposure. *Biometrics*, *58*(4), 1034-1036.

[5] Greenland, S. (1988). Variance estimation for epidemiologic effect estimates under misclassification. *Statistics in medicine*, *7*(7), 745-757.

[6] Kodde, D. A., & Palm, F. C. (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica: journal of the Econometric Society*, 1243-1248.

[7] Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical science*, *16*(2), 101-133.

[8] Marshall, R. J. (1990). Validation study methods for estimating exposure proportions and odds ratios with misclassified data. *Journal of clinical epidemiology*, *43*(9), 941-947.

[9] van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., ... & Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, *1*(1), 1.

[10] Agresti, A., & Min, Y. (2005). Frequentist performance of Bayesian confidence intervals for comparing proportions in 2× 2 contingency tables. *Biometrics*, *61*(2), 515-523.

[11] Lyles, R. H., Weiss, P., & Waller, L. A. (2020). Calibrated Bayesian credible intervals for binomial proportions. *Journal of statistical computation and simulation*, *90*(1), 75-89.

[12] Beavers, D. P., & Stamey, J. D. (2012). Bayesian sample size determination for binary regression with a misclassified covariate and no gold standard. *Computational Statistics & Data Analysis*, *56*(8), 2574-2582.

[13] Morrissey, M. J., & Spiegelman, D. (1999). Matrix methods for estimating odds ratios with misclassified exposure data: extensions and comparisons. *Biometrics*, *55*(2), 338-344.

[14] Lyles, R. H., Zhang, F., & Drews-Botsch, C. (2007). Combining internal and external validation data to correct for exposure misclassification: a case study. *Epidemiology*, *18*(3), 321-328.

[15] Agresti, A., & Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, *52*(2), 119-126.

[16] Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, *22*(158), 209-212.

[17] Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical society*, *54*(3), 426-482.

[18] Quesenberry, C. P., & Hurst, D. C. (1964). Large sample simultaneous confidence intervals for multinomial proportions. *Technometrics*, *6*(2), 191-195.

[19] Goodman, L. A. (1965). On simultaneous confidence intervals for multinomial proportions. *Technometrics*, *7*(2), 247-254.

[20] Olkin, I., & Rubin, H. (1964). Multivariate beta distributions and independence properties of the Wishart distribution. *The Annals of Mathematical Statistics*, 261-269.

[21] Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *186*(1007), 453-461.

[22] Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological review*, *70*(3), 193.

[23] Bartlett, M. S. (1947). The use of transformations. *Biometrics*, *3*(1), 39-52.

[24] Bland, J. M., & Altman, D. G. (1996). Transformations, means, and confidence intervals. *BMJ: British Medical Journal*, *312*(7038), 1079.