Winter 12-18-2021

Exact Inference for Meta-Analysis of Rare Events and Its Application in Human Genetics

Yanqiu Shao
Southern Methodist University, yanqius@smu.edu

EXACT INFERENCE FOR META-ANALYSIS OF RARE EVENTS
AND ITS APPLICATION IN HUMAN GENETICS

Approved by:

Dr. Chao Xing
Associate Professor of Statistical
Genetics, UTSW

Dr. Daniel F. Heitjan
Professor of Biostatistics, SMU & UTSW

Dr. Xinlei (Sherry) Wang
Professor of Statistics, SMU

Dr. Zahid Ahmad
Associate Professor of Endocrinology,
UTSW

EXACT INFERENCE FOR META-ANALYSIS OF RARE EVENTS AND ITS APPLICATION IN HUMAN GENETICS

A Dissertation Presented to the Graduate Faculty of the

Dedman College

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

with a

Major in Biostatistics

by

Yanqiu Shao

B.S., Mathematics, Zhejiang Gongshang University
M.S., Applied Statistics and Data Analytics, Southern Methodist University

December 18, 2021

ACKNOWLEDGMENTS

First of all, I would like to express my deepest thankfulness to my advisor, Dr. Chao Xing, for his valuable guidance and patient support in my dissertation work. His passion and expertise in both statistics and genetics encourage me to be an enthusiastic statistician. It's my pleasure working with him and I have learned a lot in practical and theoretical studies during the past three years.

Second, I would like to thank the rest of my dissertation committee members, Dr. Daniel F. Heitjan, Dr. Xinlei (Sherry) Wang, and Dr. Zahid Ahmad, for their generous support and insightful comments in my dissertation. Special thanks to Dr. Heitjan for teaching me in biostatistics courses and guiding me to be a biostatistician. I am also grateful to Dr. Wang for her consistent supports since my Master's study. I also appreciate working on clinical projects with Dr. Ahmad at UT Southwestern Medical Center.

In addition, I would like to extend my thanks to all the professors and staff in the Department of Statistical Science at Southern Methodist University for their support and help in my Master's and Ph.D. studies during the past six years.

Finally, I would like to sincerely thank my parents and my husband for their unconditional support and encouragement.

Shao, Yanqiu B.S., Mathematics, Zhejiang Gongshang University
M.S., Applied Statistics and Data Analytics, Southern Methodist University

Exact Inference for Meta-Analysis of Rare Events
and Its Application in Human Genetics

Advisor: Dr. Chao Xing

Doctor of Philosophy degree conferred December 18, 2021
Dissertation completed August 27, 2021

Meta-analysis is a statistical approach that integrates data from multiple studies. By aggregating information, it enhances the power to detect the effects of interest and provides an estimate of the effect size with both accuracy and precision. Both fixed-effect and random-effect models are developed and widely used in biomedical research including clinical trials and genomic studies. In the case of rare events data, conventional meta-analysis methods that rely on large sample approximation may not be able to make reliable inferences. There have been various approaches proposed to deal with this situation, in particular, rare binary adverse events in clinical studies.

Genome-wide association studies (GWAS) is the most popular study design of human genetic mapping. Large consortia are organized to increase the power of association detection, and therefore meta-analysis becomes a necessity in GWAS. Advances in sequencing technology enable a complete survey of both common and rare variants. Although numerous statistical methods to analyze rare variants are developed for a single study, there is no meta-analysis approach developed to specifically deal with rare variants. In this dissertation we aim to develop methods to make exact inferences in meta-analysis of rare variants association. The exact methods are based on exact distribution not approximate distribution so it is derived from all known parameters.

We first adapt and implement a fixed-effect exact meta-analysis approach that is based on the concept of p-value function with the specific aim of performing rare variants genetic

association studies. It can conduct robust inference on risk difference (RD) and construct a reliable confidence interval (CI) without ignoring studies with zero event, adding arbitrary continuity corrections, or using large sample approximation. We compare the exact method with the commonly used Mantel-Haenszel method in terms of CI coverage probability, CI length, type I error rate, statistical power, and absolute bias in various scenarios of balanced and unbalanced study sample sizes. Simulation results show that the exact methods are more stable when the event rates are extremely rare and sample sizes are unbalanced between case and control groups.

We then extend the exact meta-analysis approach to a random-effect model, which, compared with the fixed-effect model, can handle between-study variances. The proposed method enables an unbiased estimate of odds ratio (OR) and makes exact inferences of CI. We propose a method to shrink the parameter search region, which can substantially reduce the computational cost. Simulation studies are conducted to investigate the performance of the proposed method in terms of CI coverage probability and CI length. The proposed method maintains stable coverage probabilities under various settings of heterogeneity, number of studies, and magnitude of ORs.

We further consider a special study design that there are multiple case groups but there is only one common control group. This may happen when researchers only recruit patients but not controls, and use some large survey database from the general population as the common control group. We propose an exact method to construct CI for the event rate in the case groups and then make inferences on the pooled effect size measures, e.g. RD and OR, compared with the common controls. The proposed exact method shows stable performance regardless of the parameter settings. We particularly recommend applying it when the number of studies is small and the event rate is rare.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1. Background

Meta-analysis synthesizes information from multiple independent yet "similar" studies to make inferences on a common parameter of interest. By aggregating information from multiple studies meta-analysis can lead to higher statistical power and more accurate point estimates. Meta-analysis is a popular statistical approach to make reliable inferences in many disciplines, especially in public health and medical research where the effect size measures for binary outcomes, e.g., risk difference (RD), odds ratio (OR), and relative risk (RR), are of interest.

Depending on the assumptions, there are two major types of modeling in meta-analysis approaches: the fixed-effect model and the random-effect model. The fixed-effect model assumes that effect sizes are fixed across studies. This assumption is strict and often unrealistic as it requires the individual studies to share the same sources of variability, e.g., measurements definition and patient demographics. The random-effect model is more flexible that it can accommodate variability from different sources. The random-effect model assumes that measures of effect size across studies can be different but follow an underlying distribution instead of being a fixed value. The underlying distribution thereby accounts for between-study variance, which is also known as heterogeneity.

A major goal of meta-analysis is to estimate the pooled effect size measure. A commonly used estimate for effect size is a weighted estimate, where the weights are inverse of variance of each estimate, so this weighted estimate is known as the inverse-variance (INV) estimate. Denote by θ the true effect size. Suppose there are K studies, and we observe the effect size measure $\hat{\theta}_k$ in the kth study, where $k = 1, \ldots, K$. The INV estimate of the pooled effect size $\hat{\theta}$ is computed as

$$\hat{\theta}_w = \frac{\sum_{k=1}^{K} w_k \hat{\theta}_k}{\sum_{k=1}^{K} w_k}, \tag{1.1}$$
$$w_k = [\mathrm{Var}(\hat{\theta}_k)]^{-1}.$$

The INV estimate tends to assign more weights to studies with larger sample size as they have smaller variance. Under the fixed-effect model where the variability only comes from sampling errors, the INV weight only contains within-study variance, denoted by s_k^2, thus $w_k = [s_k^2]^{-1}$. Under the random-effect model where the between-study variance, denoted as τ^2, is taken into account, the weight becomes $w_k = [s_k^2 + \tau^2]^{-1}$.

For binary data, when the event rate is rare or the sample size is small, the variance estimate used in the INV method can be problematic [18]. An alternative method to estimate the overall effect size is the Mantel-Haenszel (MH) method [30]. The data can be summarized into multiple 2×2 tables with each table representing data from the kth study as in table 1.1.

kth study	Case	Control
Exposed	a_k	b_k
Unexposed	c_k	d_k

Table 1.1: An example of a 2×2 table for the kth study.

Assume the sample size for the kth study is n_k. The MH type estimates [15, 30, 38] are

$$\text{RD:} \quad \hat{\theta}_{MH} = \frac{\sum_{k=1}^{K} a_k(\frac{c_k+d_k}{n_k}) - c_k(\frac{a_k+b_k}{n_k})}{\sum_{k=1}^{K} \frac{(a_k+b_k)(c_k+d_k)}{n_k}},$$

$$\text{OR:} \quad \hat{\theta}_{MH} = \frac{\sum_{k=1}^{K} \frac{a_k d_k}{n_k}}{\sum_{k=1}^{K} \frac{b_k c_k}{n_k}},$$

$$\text{RR:} \quad \hat{\theta}_{MH} = \frac{\sum_{k=1}^{K} a_k(\frac{c_k+d_k}{n_k})}{\sum_{k=1}^{K} c_k(\frac{a_k+b_k}{n_k})}.$$

The MH method and the INV method can both be applied to estimate RD, OR and RR. However, many studies [4, 18, 39] point out that the INV method are not appropriate when the event rate is rare. The INV method works best when limited studies are collected but each has a large sample size; in contrast the MH method is preferable when we observe large number of studies and each study has a small sample size [44].

The Peto method [55] is another way to estimate OR under the assumption of a fixed-effect model. It pools OR by the observed and expected counts in 2×2 tables.

$$\text{OR:} \quad \hat{\theta}_{Peto} = \exp \frac{\sum_{k=1}^{K}(O_k - E_k)}{\sum_{k=1}^{K} V_k}$$

$$\text{where} \begin{cases} O_k = a_k \\ E_k = \frac{(a_k+b_k)(a_k+c_k)}{n_k} \\ V_k = \frac{(a_k+b_k)(b_k+d_k)(a_k+c_k)(c_k+d_k)}{n_k^2(n_k-1)} \end{cases}$$

The Peto OR estimate does not require continuity correction when we observe zero event in a study. It works well when ORs are around 1, event rates are rare (less than 1%), and sample sizes in case and control groups are balanced [4, 18]. Otherwise the Peto estimate can be severely biased. For example, it underestimates OR for large risk ratios and it overestimates OR when the risk ratio is below 1 [4, 44].

In meta-analysis, besides estimating effect size measures, estimating heterogeneity is another study aim if a random-effect model is assumed. Cochran's Q statistic is used to measure whether heterogeneity exists or not [11, 12]. It is defined as

$$Q = \sum_{k=1}^{K} w_k(\hat{\theta}_k - \hat{\theta}_w)^2,$$

where $\hat{\theta}_w$ is the INV estimate defined in equation (1.1). Derived from the Q statistic, the I^2 statistic [19] is designed to measure the percentage of variation caused by the between-study variance,

$$I^2 = \Big[\frac{Q - (K-1)}{Q}\Big] \times 100\%. \tag{1.2}$$

The I^2 statistic can be used to describe the extent of heterogeneity compared with sampling errors. An I^2 statistic of 0% means all variability comes from sampling errors, 25%, 50%, 75% suggest mild, moderate, high heterogeneity, respectively, and 100% means all variance is due to heterogeneity. When few studies are involved in a meta-analysis, the Q statistic and the I^2 statistic may have poor statistical power to detect the true between-study variance [21].

A widely used method to estimate the between-study variance (τ^2) is the DerSimmonian and Laird (DL) method [13]. It utilizes the Q statistic to estimate heterogeneity,

$$\hat{\tau}^2_{DL} = \max\left\{0, \frac{Q - (K-1)}{\sum w_k - \frac{\sum w_k^2}{\sum w_k}}\right\},$$

where $w_k = [s_k^2]^{-1}$ is the inversed sampling errors as in the fixed-effect model. The pooled effect size estimate of the DL method is simply an INV estimate with weight $w'_k = [s_k^2 + \hat{\tau}^2_{DL}]^{-1}$. The DL method tends to underestimate heterogeneity when the true between-study variance is large [33, 51].

There are many alternative methods that outperform the DL method in some scenarios. For instance, the Paule and Mandel (PM) method [34] uses an iterative way to estimate τ^2. The PM method is more robust than the DL method when the underlying assumptions are not met [37], and the PM estimate is less biased compared with that of other methods [33]. We will discuss in more details the PM method and its improved version (IPM) in chapter 3. The Sidik and Jonkman (SJ) method uses an non-iterative way to estimate τ^2 and its estimate is less biased than the DL estimate when the number of studies or heterogeneity is large [41, 42]. We apply it for comparison purpose in chapter 4. Besides, there are other methods like maximum likelihood method [17, 48, 52], restricted maximum likelihood method [35, 52], and bayesian methods [9, 10, 32] proposed to estimate heterogeneity.

Besides the commonly used fixed-effect and random-effect models, nowadays many novel techniques have been developed in meta-analysis and can be applied to broad scenarios [45]. For example, network meta-analysis is a technique used for simultaneously assessing multiple treatments [8, 36, 50]; meta-regression utilizes regression ideas to synthesize study information that incorporates both fixed-effect and mixed-effect [20, 43, 47].

1.2. Meta-Analysis Methods for Rare Events Data

In the case of rare events, a single study is inadequate to draw a reliable statistical conclusion. Therefore, meta-analysis is desired to provide strengthened inferences with greater statistical power. When the confidence interval (CI) for the pooled effect size measure is of interest, most existing methods rely on large-sample approximations to estimate CIs. Such approximations may be inaccurate and can lead to invalid conclusions when the study sample sizes are small, the number of studies is limited, or data are sparse [5, 28, 49].

In addition, conventional approaches apply continuity correction when there is no event observed in either case or control group, and some methods may even ignore those studies containing zero event in both groups. It will lead to biased estimates, especially when the

events are extremely rare [4, 24]. For example, if any individual study contains zeros in both case and control groups, the conventional MH method either excludes this study or adds 0.5 continuity correction to all cells in this study. Although both strategies enable the MH method with OR to tolerate zero event, the MH method becomes less robust when event rates are extremely rare [24].

Recently many novel meta-analysis approaches have been developed for the rare events situations, in particular, rare binary adverse events in clinical studies [45]. An exact meta-analysis method [49] was proposed to construct CIs for the common parameter of interest with the advantage of avoiding continuity correction and including available information from all studies. The term 'exact' refers to using probability distributions that do not depend on unknown parameters [1]. Liu *et al.* [28] further proposed a generalized version of the exact method by means of pooling individual p-value functions [14, 54]. Since then, more exact random procedures have been proposed [16, 29, 31, 53]. These exact methods are more reliable when the event rates are rare and thus are preferred over the conventional approximation methods.

Besides the exact methods, there are other methods developed for rare events data. Under a binomial-normal (BN) assumption, Bhaumik *et al.* [3] improved the estimate of between-study variance from the PM method [34] by borrowing information from all studies. The improved method, abbreviated as IPM, is proved to be capable of reducing bias for estimating log OR when the event rates are rare. Zhang *et al.* [57] reviewed several commonly used estimates of heterogeneity and concluded that the IPM estimate is less biased compared with others. Li and Wang [27] proposed a generalized random-effect model based on that in Bhaumik *et al.*. In addition, novel Bayesian methods designed for rare events are also developed. Cai *et al.* [6] proposed a Bayesian approach based on a Poisson model for RR. Bai *et al.* [2] developed a Bayesian estimator based on the model in Bhaumik *et al.*.

1.3. Application of Exact Meta-Analysis Methods in Human Genetics

In genome-wide association studies (GWAS), meta-analysis is widely performed to detect association between traits and genetic variants. However, rare events data, i.e., rare genetic variants, are handled by the conventional methods relying on large sample approximation, which can yield invalid inferences.

In this dissertation, we specifically aim to develop and apply the exact meta-analysis methods in case-control human genetic association studies. We first adapt and implement a fixed-effect exact meta-analysis approach that is based on the concept of p-value function. Then we extend the exact meta-analysis approach to a random-effect model, which, compared with the fixed-effect model, can handle the between-study variances. Further, we consider a special study design that there are multiple case groups but there is only one common control group, which is the norm in human genetic studies.

CHAPTER 2

EXACT INFERENCE FOR META-ANALYSIS OF RARE EVENTS BASED ON THE FIXED-EFFECT MODEL

In this chapter, we adapt an exact meta-analysis method based on the fixed-effect model with algorithm speed-up for case-control genetic association studies of rare variants. This chapter is structured as follows. In section 2.1, we introduce the method. Section 2.2 describes the algorithm improvements. In section 2.3, we conduct simulation studies to compare the exact method with the MH method in different scenarios. In section 2.4, we apply the exact method to two real datasets and convert the estimated RD to an OR. Section 2.5 summarizes the findings and directs the research topic to the random-effect model detailed in the next chapter.

2.1. Methods

2.1.1. Overview

Consider a meta-analysis of K independent case-control genetic association studies of the same trait with overlapping genetic variants. Hereinafter, assume we focus on one specific variant. Consider one particular study k, $k = 1, \ldots, K$. Denote by n_{k1} and n_{k0} the sample size of case and control groups, respectively, and denote by X_{k1} and X_{k0} the random variables of the number of individuals carrying the variants in case and control groups, respectively. X_{k1} and X_{k0} are assumed to follow binomial distributions, $X_{k1} \sim Bin(n_{k1}, p_1)$, and $X_{k0} \sim Bin(n_{k0}, p_0)$, where p_1 and p_0 are the true event rates in case and control groups, respectively.

Here event is defined as carrying the variant of interest.

	Carrier	Non-Carrier	Total
Case	x_{k1}	$n_{k1} - x_{k1}$	n_{k1}
Control	x_{k0}	$n_{k0} - x_{k0}$	n_{k0}

Table 2.1: Notation of the observed counts in the kth study by a 2×2 table.

As shown in table 2.1, x_{k1} and x_{k0} are the observed number of carriers of the genetic variant in the kth study. The estimated event rates are $\hat{p}_{k1} = x_{k1}/n_{k1}$ and $\hat{p}_{k0} = x_{k0}/n_{k0}$, with properties of expectation $\mathrm{E}(\hat{p}_{ki}) = p_i$ and variance $\mathrm{Var}(\hat{p}_{ki}) = p_i(1-p_i)/n_{ki}$, $i = 0, 1$. Under the assumption of a fixed-effect model, the common RD θ between the event rates of case and control groups across studies is fixed and defined as $\theta = p_1 - p_0$, for any k, $k = 1, \ldots, K$. The joint probability of observing x_{k1} and x_{k0} is

$$\begin{aligned} &\Pr(X_{k1} = x_{k1}, X_{k0} = x_{k0} | \theta, p_1, p_0) \\ &\quad = \Pr(X_{k1} = x_{k1} | \theta, p_1) \times \Pr(X_{k0} = x_{k0} | \theta, p_0) \\ &\quad = \binom{n_{k1}}{x_{k1}} \binom{n_{k0}}{x_{k0}} p_1^{x_{k1}} p_0^{x_{k0}} (1-p_1)^{n_{k1}-x_{k1}} (1-p_0)^{n_{k0}-x_{k0}}. \end{aligned} \tag{2.1}$$

2.1.2. Analysis of a Single Study

To make inferences on the RD θ in its parameter space Θ for a single study k, we can start from conducting a right-tailed test, $H_0^{(k)} : \theta = \theta'$ v.s. $H_1^{(k)} : \theta > \theta'$, where θ' can be any fixed value in Θ, to assess if θ' is the true value of θ. At the significant level α, given the observed data from study k, we are able to obtain the right-tailed p-value at θ', denoted as $p^{(k)}(\theta')$, and then reject the null hypothesis if $p^{(k)}(\theta') < \alpha$ or fail to reject the null hypothesis if $p^{(k)}(\theta') \geq \alpha$. As a result, each possible θ' leads to a corresponding right-tailed p-value $p^{(k)}(\theta')$ that can be collected into a p-value function $p^{(k)}(\theta)$ as θ shifts in its parameter space Θ. Thus, given the observed data of study k, the $100(1-\alpha/2)\%$ one-sided CI for θ is

$(\theta_L^{(k)}, \infty)$, where the lower bound $\theta_L^{(k)} = \inf_\theta\{\theta : p_L^{(k)}(\theta) > \alpha/2\}$. Note that the right-tailed p-value function $p_L^{(k)}(\theta)$ is the p-value function from aforementioned right-tailed test and the subscript "L" refers to the lower bound. If θ' falls outside the $100(1-\alpha/2)\%$ CI, we reject the null hypothesis.

Next, we conduct a similar procedure of a left-tailed test, $H_0^{(k)} : \theta = \theta'$ v.s. $H_1^{(k)} : \theta < \theta'$, to obtain the left-tailed p-value. The $100(1-\alpha/2)\%$ one-sided CI for θ becomes $(-\infty, \theta_U^{(k)})$, where the upper bound $\theta_U^{(k)} = \sup_\theta\{\theta : p_U^{(k)}(\theta) > \alpha/2\}$, and the left-tailed p-value function $p_U^{(k)}(\theta)$ is obtained from the left-tailed test with "U" referring to the upper bound. The right-tailed p-value function $p_L^{(k)}(\theta)$ is a non-decreasing function of θ, and the left-tailed p-value function $p_U^{(k)}(\theta)$ is non-increasing. Combining the lower and upper bounds from the two $100(1-\alpha/2)\%$ one-sided CIs, the $100(1-\alpha)\%$ two-sided CI for individual study k is $(\theta_L^{(k)}, \theta_U^{(k)})$.

In practice, we use a grid of parameter space Θ, denoted as Θ^*, to approximate the continuous parameter space, which consists of I grid values including the essential element 0. In the exact test, values of the p-value functions are collected by performing the right-tailed and left-tailed tests over all grid values of Θ^*. Given each pair of θ and p_0, we are able to derive p_1 as $\theta + p_0$. The search range of p_0 can be based on its Wald-type CI with a predefined high coverage, say 99%. Denote by J the number of grids for p_0. Therefore, to generate the two p-value functions for each study, the p-values at each specific θ are computed through I grids of θ and J grids of p_0. Given θ and p_0, the aforementioned joint probability (equation 2.1) can be rewritten as

$$\begin{aligned}&\Pr(X_{k1} = x_{k1}, X_{k0} = x_{k0}|\theta, p_0)\\ &\quad= \binom{n_{k1}}{x_{k1}}\binom{n_{k0}}{x_{k0}}(\theta+p_0)^{x_{k1}}p_0^{x_{k0}}(1-\theta-p_0)^{n_{k1}-x_{k1}}(1-p_0)^{n_{k0}-x_{k0}}.\end{aligned} \tag{2.2}$$

The test statistic T used in the right-tailed and the left-tailed tests is defined as

$$T = \frac{\hat{p}_{k1} - \hat{p}_{k0} - \theta'}{\sqrt{\frac{\hat{p}'_{k1}(1-\hat{p}'_{k1})}{n_{k1}} + \frac{\hat{p}'_{k0}(1-\hat{p}'_{k0})}{n_{k0}}}},$$

where we perform correction in denominator for all studies no matter whether there is zero event or not,

$$\hat{p}'_{ki} = \frac{x_{ki} + 0.5}{n_{ki} + 1}, \qquad i = 0, 1.$$

This statistic T is calculated over all pairs of (X_{k1}, X_{k0}) where $0 \leq X_{k1} \leq n_{k1}, 0 \leq X_{k0} \leq n_{k0}$. Let $T_{obs} = T(x_{k1}, x_{k0})$ denote the observed test statistic.

Given θ and p_0, the right-tailed p-value $p_L^{(k)}(\theta, p_0)$ and the left-tailed p-value $p_U^{(k)}(\theta, p_0)$ can be computed by the mid p-value method [7, 22, 26], where

$$\begin{aligned} p_L^{(k)}(\theta, p_0) &= \Pr(T > T_{obs}|\theta, p_0) + \frac{1}{2}\Pr(T = T_{obs}|\theta, p_0), \\ p_U^{(k)}(\theta, p_0) &= \Pr(T < T_{obs}|\theta, p_0) + \frac{1}{2}\Pr(T = T_{obs}|\theta, p_0). \end{aligned} \tag{2.3}$$

Note that for a given θ, there are J right-tailed and J left-tailed p-values since we have J grid values of p_0. The ultimate right and left tailed p-values at θ are determined by the "best" p_0 candidate that provides the largest p-value. Consequently, for study k, the right-tailed p-value function $p_L^{(k)}(\theta)$ and the left-tailed p-value function $p_U^{(k)}(\theta)$ are

$$\begin{aligned} p_L^{(k)}(\theta) &= \max_{p_0}\{p_L^{(k)}(\theta, p_0)\}, \\ p_U^{(k)}(\theta) &= \max_{p_0}\{p_U^{(k)}(\theta, p_0)\}. \end{aligned} \tag{2.4}$$

2.1.3. Meta-Analysis

Once the right-tailed and left-tailed p-value functions are computed for all K studies, we are able to synthesize these individual p-value functions to estimate the true common event

rate θ. It is accomplished by constructing two new test statistics, $t_L^{(c)}(\theta)$ for the right-tailed test and $t_U^{(c)}(\theta)$ for the left-tailed test,

$$\begin{aligned} t_L^{(c)}(\theta) &= \sum_{k=1}^{K} w_k h[p_L^{(k)}(\theta)], \\ t_U^{(c)}(\theta) &= \sum_{k=1}^{K} w_k h[p_U^{(k)}(\theta)], \end{aligned} \tag{2.5}$$

where w_k's are study-specific positive weights subject to $\sum_{k=1}^{K} w_k = 1$, and $h[\cdot]$ is a non-decreasing transformation function. Accordingly, the combined p-value functions are

$$\begin{aligned} p_L^{(c)}(\theta) &= \Pr\Big(t_L^{(c)}(\theta) \geq \sum_{k=1}^{K} w_k h[U_k]\Big), \\ p_U^{(c)}(\theta) &= \Pr\Big(t_U^{(c)}(\theta) \geq \sum_{k=1}^{K} w_k h[U_k]\Big), \end{aligned} \tag{2.6}$$

where U_k is independent uniformly distributed random variable from $U(0,1)$.

Denote the combined $100(1-\alpha)\%$ two-sided CI for θ as $(\theta_L^{(c)}, \theta_U^{(c)})$. Its lower and upper bounds are derived in terms of the combined p-value functions,

$$\begin{aligned} \theta_L^{(c)} &= \inf_{\theta}\Big\{\theta : p_L^{(c)}(\theta) \geq \frac{\alpha}{2}\Big\}, \\ \theta_U^{(c)} &= \sup_{\theta}\Big\{\theta : p_U^{(c)}(\theta) \geq \frac{\alpha}{2}\Big\}. \end{aligned}$$

The empirical point estimate is then $\hat{\theta} = \inf_{\theta}\{\theta : |t_L^{(c)}(\theta) - t_U^{(c)}(\theta)|\}$. A weak support two-sided p-value for the exact test $H_0^{(c)} : \theta = \theta'$ v.s. $H_1^{(c)} : \theta \neq \theta'$ is defined as $p^{(c)}(\theta') = 2\min\{p_L^{(c)}(\theta'), p_U^{(c)}(\theta')\}$ and should be restricted in the range of $[0,1]$.

Furthermore, one can estimate the corresponding common odds ratio $\hat{\lambda}$ in terms of $\hat{\theta}$ as

$$\hat{\lambda} = \frac{\sum_{k=1}^{K}\{(1-\hat{p}_{k0})(\hat{\theta}+\hat{p}_{k0})\frac{n_{k0}n_{k1}}{n_{k0}+n_{k1}}\}}{\sum_{k=1}^{K}\{(\hat{p}_{k0})(1-\hat{\theta}-\hat{p}_{k0})\frac{n_{k0}n_{k1}}{n_{k0}+n_{k1}}\}}, \tag{2.7}$$

which is derived from the MH method of estimating the common odds ratio for a $2 \times 2 \times K$ table. This allows us to transfer the estimated RD to OR if the point estimate of OR is desired.

2.2. Algorithm Improvement

The exact algorithm starts from evaluating all grid values of the two p-value functions for each pair of θ and p_0, and then "smoothes" them to be monotonic functions. One main drawback of the exact method is that it is computationally intensive because it needs to calculate the joint probability for all the entries in the $(n_{k0}+1) \times (n_{k1}+1)$ matrix (equation 2.2), denoted as $\mathbf{P}$, and repeats calculating such large matrices for $I \times J$ possible combinations of θ and p_0. The intuitive way to carry out the calculation of each $\mathbf{P}$ is

1. Generate a probability vector $\Pr(X_{k1} = x_{k1}|\theta, p_0)$ as x_{k1} varies from 0 to n_{k1};
2. Generate a probability vector $\Pr(X_{k0} = x_{k0}|\theta, p_0)$ as x_{k0} varies from 0 to n_{k0};
3. Generate the outer product of the two probability vectors to have the joint probability matrix $\mathbf{P}$.

Since $\mathbf{P}$ is the outer product of the $1 \times (n_{k1}+1)$ probability vector of $\Pr(X_{k1} = x_{k1}|\theta, p_0)$ and the $1 \times (n_{k0}+1)$ probability vector of $\Pr(X_{k0} = x_{k0}|\theta, p_0)$. The algorithm can be improved by screening out smallest probabilities that have a cumulative sum less than a cutpoint (e.g., 10^{-6}) in each of the two probability vectors before performing outer product. We term this algorithm the naive algorithm. It splits step 3 into two sub-steps:

3.1. Screen out smallest values of the two probability vectors;

3.2. Generate the outer product of the two trimmed probability vectors to have the critical part of matrix $\mathbf{P}$.

Note that the binomial probability density functionfor has only one mode, and it is extremely right-skewed in case of rare event rates with large n_{k0} and n_{k1}. Its right tail has many small values that are surly to be excluded in step 3.1 above. We accordingly propose an algorithm improvement to avoid generating these infinitesimals in probability vectors. Figure 2.1 is an example of how we accelerate generating vectors of right-skewed binomial distribution.

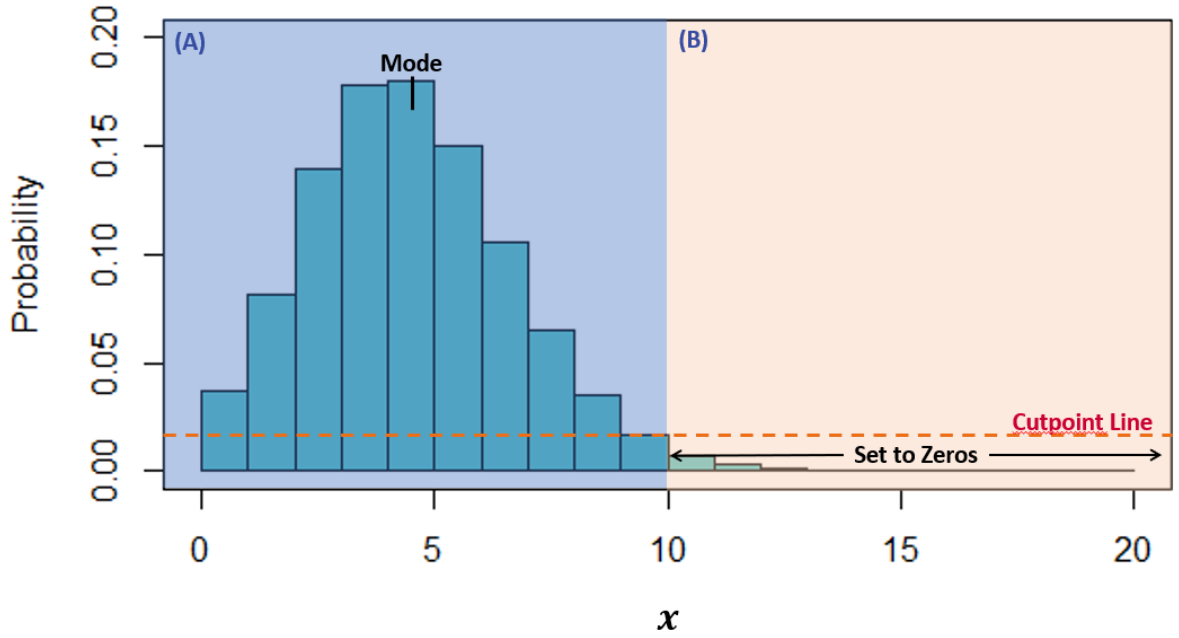

Figure 2.1: Example of accelerating procedure of generating probability vectors $\Pr(X_{ki} = x_{ki}|\theta, p_0), i = 0, 1$.

First, we find the quantile value that has a density probability equals to a pre-defined cutpoint value (e.g. 10^{-15}), as the dashed red line shown in figure 2.1. Second, we compute all probabilities until we reach the quantile value. To be more specific, all probabilities located in the blue area, region (A), in figure 2.1 will be calculated. Last, set zeros to those elements to be omitted. That is, all x's located in the orange area, region (B), are automatically set to be zeros.

The whole procedure is summarized as follows:

1. Generate the essential part of the probability vector $\Pr(X_{k1} = x_{k1}|\theta, p_0)$:

  1.1. Find the quantile value with the density probability equals to a cutpoint;

  1.2. Compute probabilities as x_{k1} varies from 0 to the quantile value;

  1.3. Set zeros to those right tail elements.

2. Repeat step 1 for $\Pr(X_{k0} = x_{k0}|\theta, p_0)$;

3. Generate the outer product of the two trimmed probability vectors to have the critical part of matrix $\mathbf{P}$.

We term this algorithm the accelerated algorithm. This algorithm speeds up via assigning zeros to small probabilities instead of generating their precise values that will be excluded later. We choose 10^{-15} as the probability cutpoint because this number is small enough that it will need 10^7 such elements to reach the cumulative sum 10^{-6} previously mentioned. Note that in the case of rare events, a majority of probabilities in the vectors are smaller than this cutpoint. Therefore, the accelerated algorithm is much faster comparing to the naive algorithm, especially when sample sizes are large.

Figure 2.2 compares the average running time (in seconds per study) for 20 replicates of the setting $\{\theta = 0, K = 5, p_0 \sim U(0, 0.01), n_{k1} \sim U(100, n_{max}), n_{k0} \sim U(100, n_{max}), n_{max} = \{500, 1000, \ldots, 9500, 10000\}\}$ using (1) the naive algorithm, and (2) the accelerated algorithm. Comparing to the naive algorithm, the accelerated algorithm becomes more and more efficient with the increase of sample sizes.

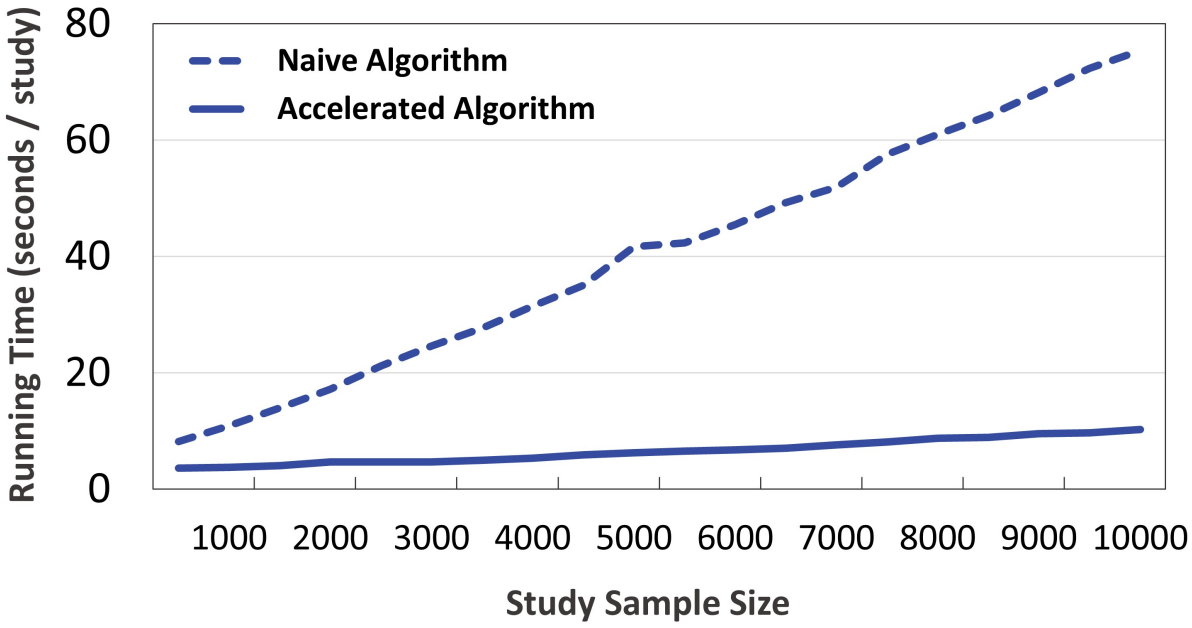

Figure 2.2: Comparison of time consumption between the naive algorithm (dashed line) and the accelerated algorithm (solid line).

2.3. Simulation

2.3.1. Settings

In this section, we conduct simulation studies to compare performance of the exact method and the MH method in both balanced and unbalanced study design scenarios with various parameter settings. For the MH method, we consider both without continuity correction (MH) and with 0.5 continuity correction (MH-cc).

We generate K, $K = \{5, 10, 20\}$, case-control genetic assocation studies for one genetic variant. The sample size of the case group n_{k1} is generated from $U(100, 1000)$ and the corresponding control group size is generated as a multiplication of the case group $n_{k0} = r \times n_{k1}$, where ratio $r = \{1, 10\}$. Data structure is balanced if the sample size ratio $r = 1$, and

is unbalanced if the sample size ratio $r = 10$. Set RD to $\theta = \{0, 0.01\%, 0.05\%, 0.1\%, 0.5\%\}$ and randomly generate the event rate in the control group p_0 from a uniform distribution with the lower bound min $= 0$ and the upper bound max $= \{0.01\%, 0.1\%, 1\%\}$. Denoted by p_0^u the upper bound. Then we generate the number of observed events x_{k0} and x_{k1} from the binomial distribution $Bin(n_{k0}, p_0)$ and $Bin(n_{k1}, p_0 + \theta)$. The number of grid values of θ is set to be $I = 1000$ and the number of grid values of p_0 is $J = 20$. We also generate $B = 1,000,000$ independent random variables $U_k \sim U(0, 1)$. The weight in equation (2.5) is set to be $w_k = (\frac{n_{k0}n_{k1}}{n_{k0}+n_{k1}})/(\sum_{k=1}^{K} \frac{n_{k0}n_{k1}}{n_{k0}+n_{k1}})$. We consider three non-decreasing transformation functions $h[\cdot]$: identity, inverse normal cumulative distribution function (IN-CDF), and arcsin of square root (Asin_sqrt). The second one is used in [28]. Hence we compare the performance of five models: three of the exact method and two of the MH method.

For each simulation setting 1000 replicates are generated. Each run generates a corresponding CI for each model. Four measurements are used to assess CIs' performance: (1) empirical coverage probability, (2) average length (AL) of CI, (3) probability of rejecting null hypothesis $H_0 : \theta = 0$, and (4) average absolute bias (AAB) where AAB $= |\hat{\theta} - \theta|$. Coverage probability is calculated as the probability of covering the underlying θ. Probability of rejecting null hypothesis $H_0 : \theta = 0$ can be divided into two parts, type I error rate if the true RD $\theta = 0$, and power otherwise.

2.3.2. Simulation Results

2.3.2.1. CI Coverage and Length

The 95% CI of an appropriate method is expected to have a 95% coverage probability of containing the true RD. Figure 2.3 presents coverage probabilities (left) and AL $\times 1000$ (right) for the upper bound $p_0^u = 0.01\%$ when sample sizes are balanced, and figure 2.4 shows the performance in the unbalanced study design. More figures with regard to $p_0^u =$

$\{0.1\%, 1\%\}$ are listed in appendix where figures A.1 and A.3 are for the balanced scenarios and figures A.2 and A.4 are for the unbalanced scenarios. The three exact models are overall conservative that they constantly have coverage probabilities greater than the nominal coverage probability 95%. The two MH models closely reach the nominal level of 95% when p_0^u is 0.1% and 1% in both balanced and unbalanced sample sizes. However, when p_0^u decreases to 0.01%, MH loses its power of holding coverage probabilities over the nominal level in the balanced scenario, and the coverage probabilities of both MH and MH-cc drop dramatically in the unbalanced scenarios.

It is notable that, in the case of unbalanced design and rare event rate, MH-cc has a severe problem of low coverage and is highly anticonservative, especially when $p_0^u = 0.01\%$ and K is large. MH-cc hence is not a stable model to be used in unbalanced studies. Overall, the simulation results suggest that, although the exact models are conservative, their coverage are close to 0.95 with relatively larger upper bound p_0^u and RD θ. Both MH methods has stable coverage only when p_0^u increases to 1%.

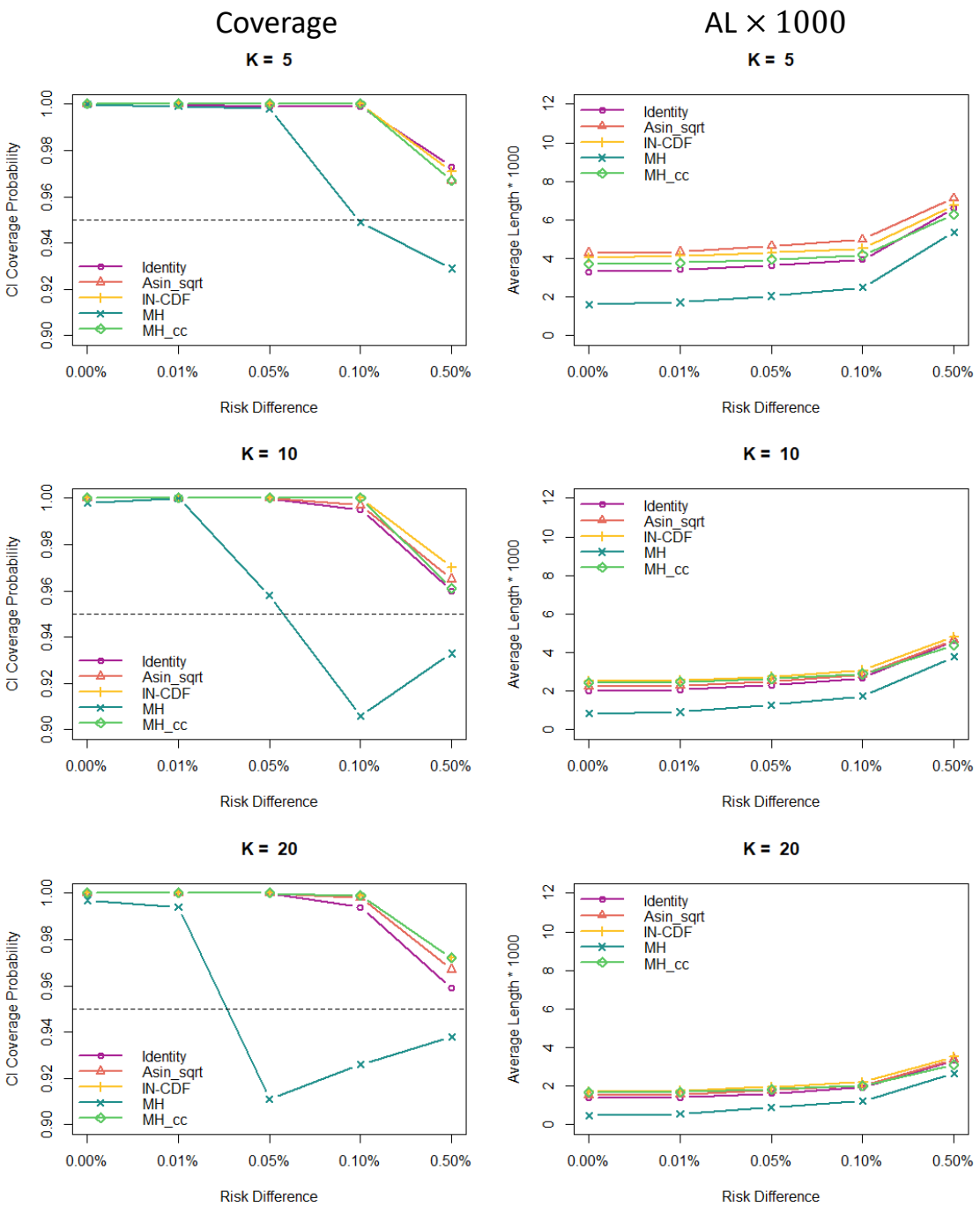

Figure 2.3: Coverage and average length when sample sizes in control and case groups are balanced and $p_0^u = 0.01\%$.

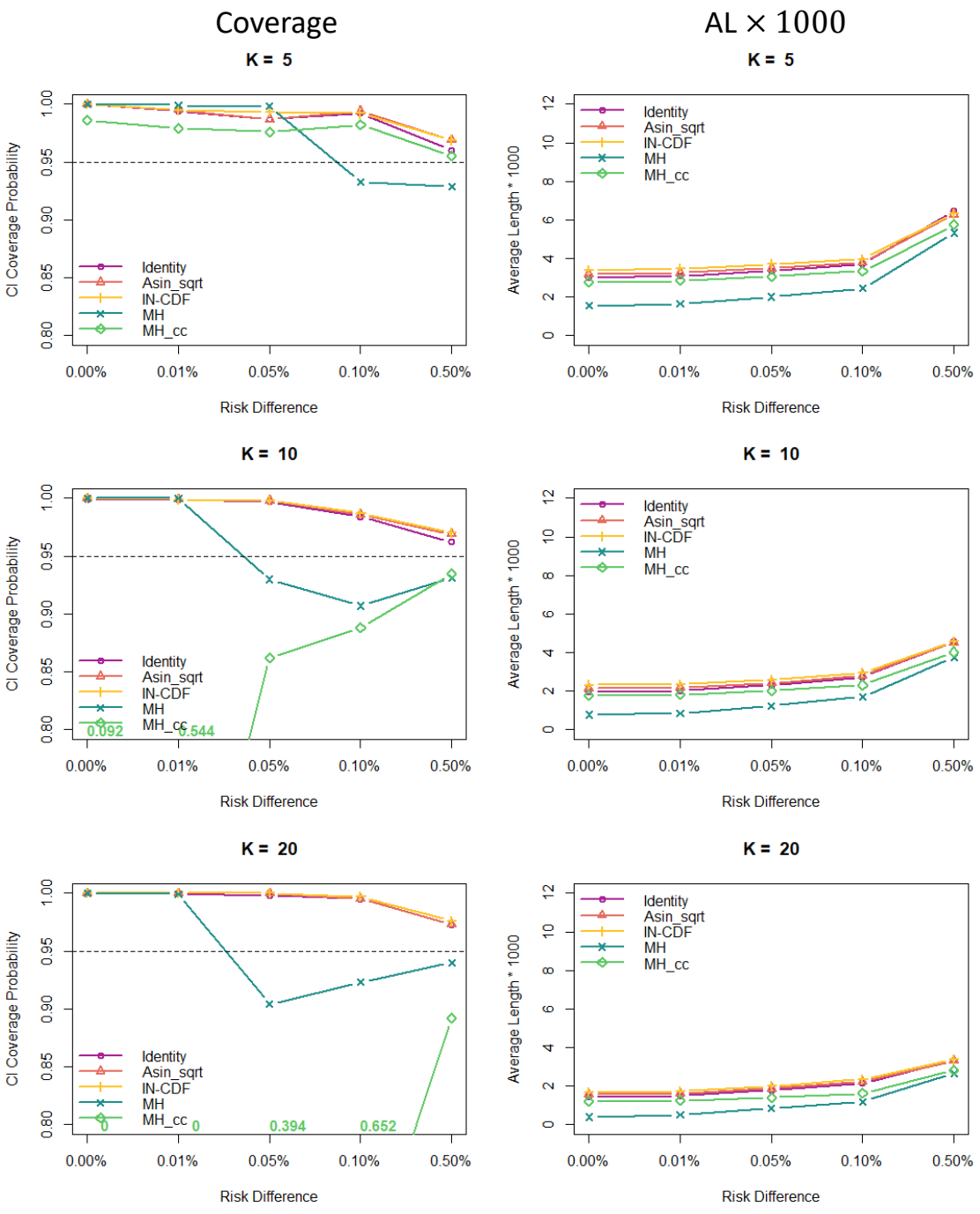

Figure 2.4: Coverage and average length when sample sizes in control and case groups are unbalanced and $p_0^u = 0.01\%$.

2.3.2.2. Type I Error Rate and Power

The left panels of figure 2.5 and figure 2.6 show probability of rejecting null hypothesis $H_0 : \theta = 0$ among 1000 runs when sample sizes are balanced and unbalanced, respectively, with $p_0^u = 0.01\%$. More results of $p_0^u = \{0.1\%, 1\%\}$ are shown in figures A.5-A.8. Each plot is split into two parts by a grey dashed line. When RD equals to 0, rejecting the null hypothesis $H_0 : \theta = 0$ is equivalent to type I error rate. Otherwise, the rejecting probability can be viewed as statistical power.

In the case of a balanced design (figures 2.5, A.5, A.7), all models have similar type I error rates. MH without continuity correction have larger power than others when p_0^u is small. In the case of an unbalanced design (figures 2.6, A.6, A.8), MH models are unstable. MH-cc has higher type I error rate and power than others when p_0^u is small and K is large, particularly, when $p_0^u = 0.01\%$ and $K = 20$. MH-cc constantly yields 100% rejecting probability as RD ranges from 0 to 0.5%. MH without continuity correction also has greater power than the exact models.

In general, the exact models have lower power due to their conservative nature. Power increases as the number of studies K and the underlying RD θ increase. Among the three exact models, the IN-CDF transformation provides lower power than the other two transformations. Furthermore, the fact that MH-cc model yields both high power and high type I error rate indicates that it tends to overestimate RD and generate false positive results.

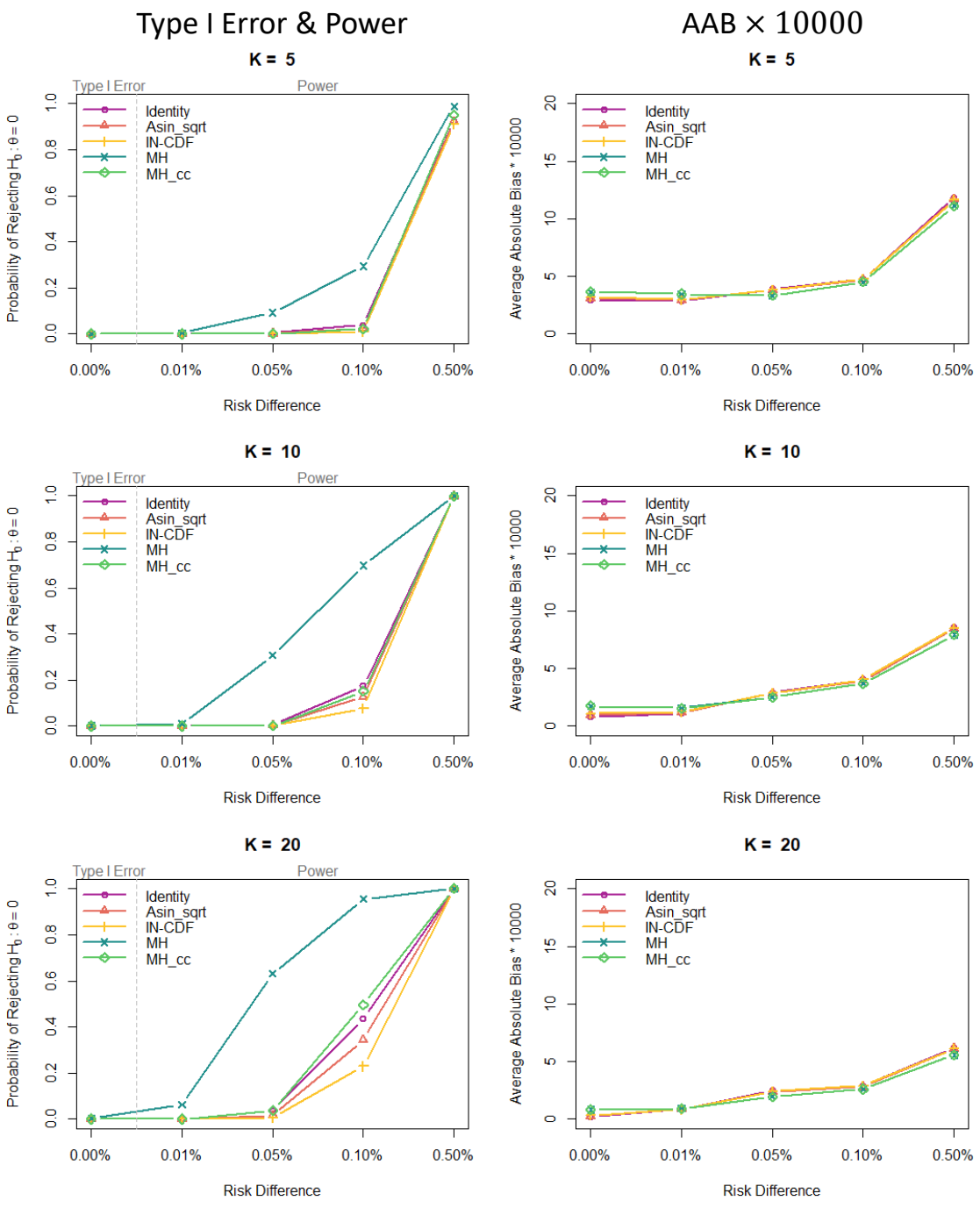

Figure 2.5: Power and AAB when sample sizes in control and case groups are balanced and $p_0^u = 0.01\%$.

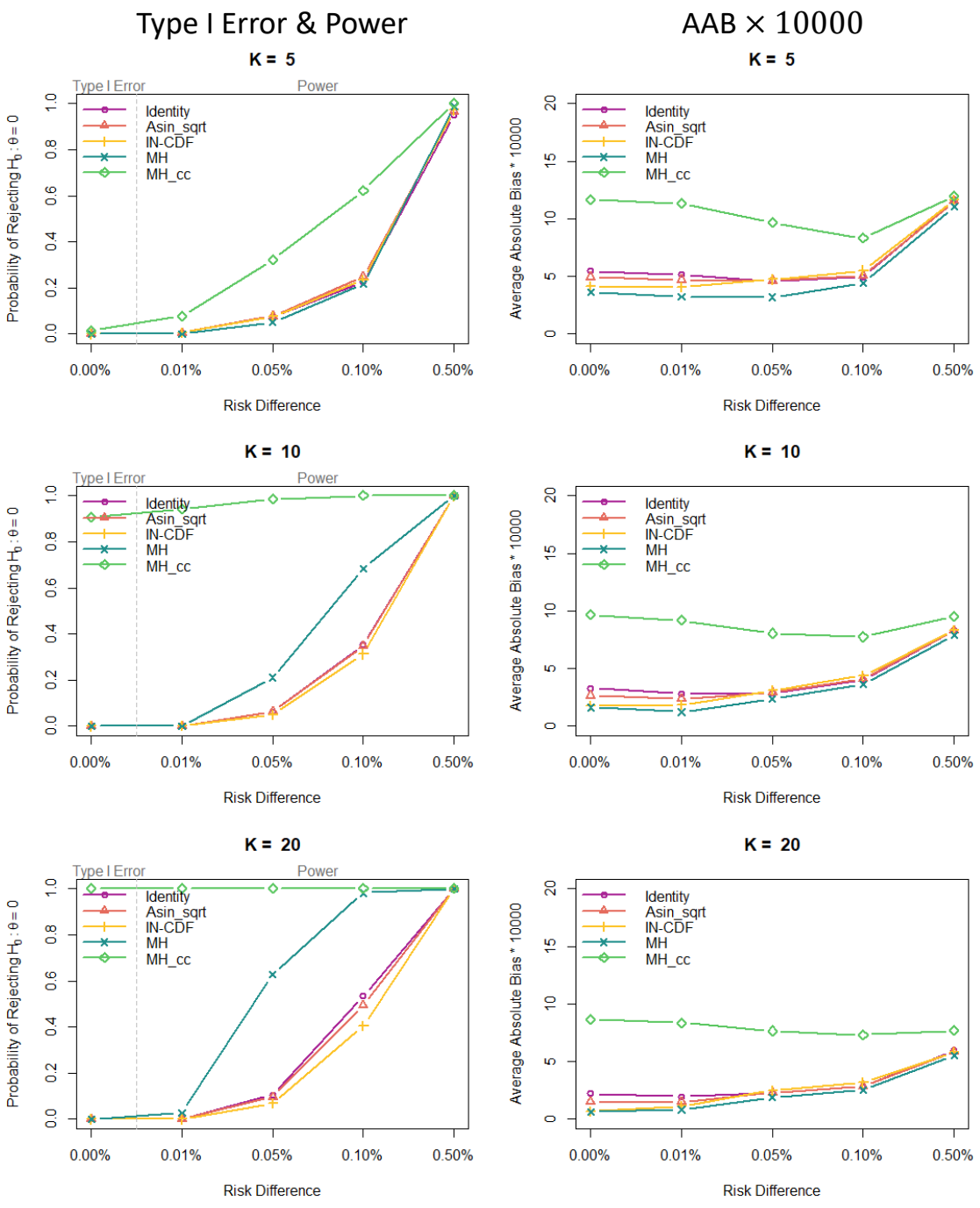

Figure 2.6: Power and AAB when sample sizes in control and case groups are unbalanced and $p_0^u = 0.01\%$.

2.3.2.3. Bias

The right panels of figure 2.5 and figure 2.6 are the estimate bias measured by AAB when $p_0^u = 0.01\%$. Figures A.5-A.8 contain results of $p_0^u = \{0.1\%, 1\%\}$. Note that these figures present AAB in the unit of AAB$\times$10000. When the sample sizes are balanced, MH and MH-cc have similar sizes of AAB. The exact methods yield slightly less biased estimates most of times, especially when RD is small. For unbalanced sample sizes, MH-cc generally produces larger bias when p_0^u is small. Especially when $p_0^u = 0.01\%$ the bias becomes much higher than other models. In contrast, MH without continuity correction gives the least bias when p_0^u is as rare as 0.01%. The three exact models have similar AAB, though the identity transformation produces higher bias than other two models when the true RD is small.

2.4. Real Data Analysis

2.4.1. Association between Gene APOC3 and CHD

High plasma triglyceride levels are associated with the risk of having coronary heart disease (CHD). Variants in gene apolipoprotein C3 (APOC3) was discovered to be strongly associated with plasma triglyceride levels. A meta-analysis [46] was conducted to examine the association of rare variants in APOC3 with CHD. There were a total of 18 studies (K=18), containing 110,970 individuals (34,002 CHD patients and 76,968 controls) from European, African American, and Hispanic ancestries. Details of the APOC3 data structure are listed in appendix table A.2. In their conclusion, carriers of APOC3 rare mutations had lower risk of CHD (OR=0.6; 95% CI=(0.47, 0.75); p-value=4×10^{-6}).

In the original meta-analysis [46], inference on OR was derived by the MH method. There were 4 studies containing zero carriers in the case group and there was no study observing zero events in the control group. Besides, 12 studies had unbalanced sample sizes

ratio (control/case > 2), among which two have extremely unbalanced ratios greater than 15 (8.0 and 27.6). We apply the exact methods and calculate RDs. Figure 2.7 shows the corresponding CIs, point estimates, and p-values of testing if there is difference between the two groups, $H_0^{(c)} : \theta = 0$ v.s. $H_1^{(c)} : \theta \neq 0$.

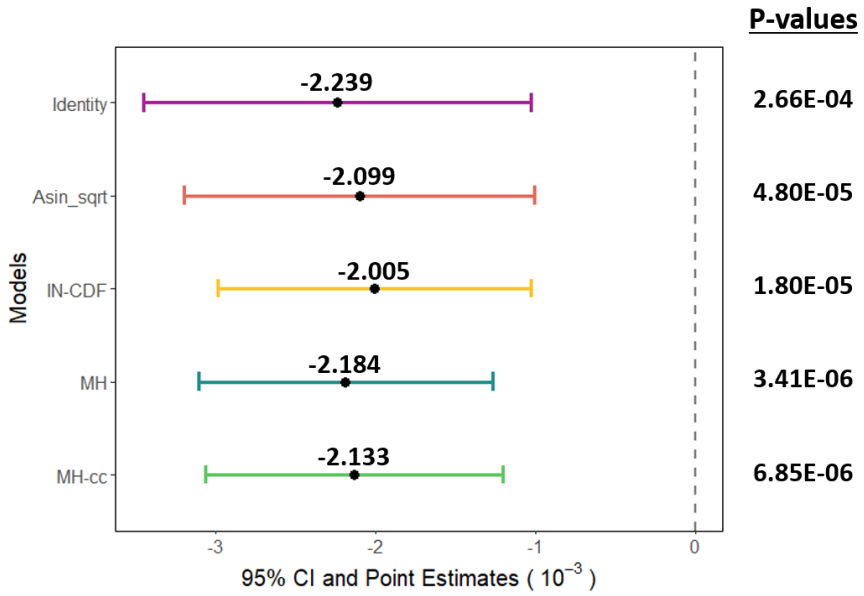

Figure 2.7: Plot of 95% CIs, point estimates, and p-values for the APOC3-CHD data. CIs and point estimates are on the scale of 10^{-3} and p-values are on the original scale.

	Identity	Asin_sqrt	IN-CDF	MH	MH-cc
Estimated RD	-0.002239	-0.002099	-0.002005	-0.002184	-0.002133
Estimated OR	0.612	0.632	0.646	0.620	0.627

Table 2.2: Summary of the estimated OR transferred from the estimated RD for the APOC3-CHD data

All the methods come to a conclusion consistent with the original paper that the rate of APOC3 variants carriers in CHD group is significantly lower than in the control group. This implies that the risk of having CHD is reduced if one carries an APOC3 mutation. Based on equation 2.7, we are able to convert the estimated RD to OR as shown in table 2.2. All the estimated ORs are close to the reported OR of 0.6 in the original paper.

2.4.2. Association between SCARB1 P376L Variant and CHD

The gene scavenger receptor class B member 1 (SCARB1) encodes scavenger receptor BI (SR-BI), which is the major receptor for high-density lipoprotein cholesterol (HDL-C). High levels of HDL-C are associated with a lower risk of CHD. In a meta-analysis [56] of 16 studies, it was found that the loss-of-function variant P376L is associated with lower levels of plasma HDL-C, and P376L carriers have an increased risk of developing CHD (OR=1.79; p-value=0.018) by the MH method.

Among the 16 studies, 6 have unbalanced ratio (control/case > 2) with the largest ratio equal to 5.2. There are 8 studies with zero carrier in either case or control group, and 2 studies with zero carrier in both case and control group. We suspect that ignoring these two studies may bias the results.

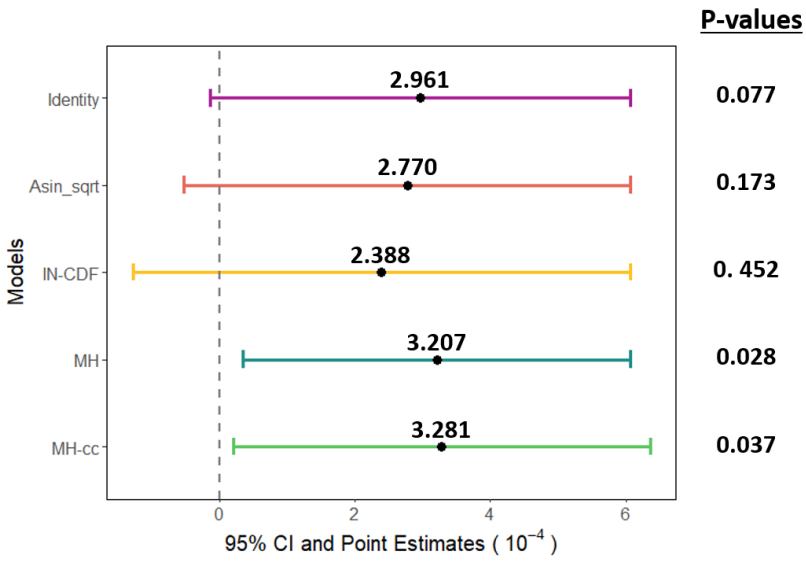

Figure 2.8: Plot of 95% CI, point estimates, and p-values for SCARB1 P376L data. CIs and point estimates are based on the scale of 10^{-4} and p-values are on the original scale.

We perform meta-analysis using RD to investigate the association between SCARB1 P376L and CHD. Figure 2.8 summaries 95% CIs, point estimates and p-values from the three exact models and the two MH models. The estimated OR from the estimated RD is

presented in table 2.3. Note that for this dataset, the exact models conclude differently from the MH models. The exact models fail to reject the null hypothesis with the inference of 95% CI spanning zero and p-values greater than the nominal level $\alpha = 0.05$, whereas the MH models reject the null hypothesis with p-values less than 0.05.

	Identity	Asin_sqrt	IN-CDF	MH	MH-cc
Estimated RD	0.000296	0.000277	0.000239	0.000321	0.000328
Estimated OR	1.727	1.680	1.586	1.787	1.805

Table 2.3: Summary of estimated OR transferred from estimated RD for the SCARB1-CHD data

2.5. Summary and Discussion

For rare events data, conventional OR methods manage zero events by either ignoring them or making continuity correction by adding 0.5, which leads to the risk of biased estimates. Such concern motivates us to use RD instead. It is known the balance of sample sizes in case and control groups plays an important role in the performance of statistical methods [4]. Therefore we conduct simulations of both balanced (1:1) and unbalanced (1:10) sample sizes to compare the exact methods based on Tian *et al.* [49] with identity, arcsin of square root, and IN-CDF (proposed by Liu *et al.* [28]) transformation functions to the MH methods in terms of CI coverage probability, CI length, probability of rejecting null hypothesis $H_0 : \theta = 0$, and AAB of estimates.

The results show that the exact methods have stable but conservative CI coverage. They are in general less biased. Relative bias decreases as the number of studies increases and the control group event rate decreases. Power increases as the number of studies and RD increase. When the sample sizes between groups are balanced, MH without 0.5 continuity correction has comparable bias to the exact methods, and its coverage probability tends to

be lower than the nominal 95% coverage for large number of studies K, big RD θ, and small control group event rates p_0^u. When sample sizes are unbalanced, we do not recommend using MH with 0.5 continuity correction due to its poor performance of unstable coverage and large bias.

The exact method discussed in this chapter is based on the fixed-effect model, which is often an impractical assumption. Further extension to the random-effect model is desired, which allows more variety in the model and is expected to fit real data better.

CHAPTER 3

EXACT INFERENCE FOR META-ANALYSIS OF RARE EVENTS BASED ON THE RANDOM-EFFECT MODEL

In this chapter, we propose an exact meta-analysis method based on the random-effect model. This chapter is organized as follows. In section 3.1 we first develop the exact method and propose an algorithm to construct the CI. In section 3.2 we examine the performance of the proposed method compared with other methods by simulation. In section 3.3 we apply the proposed exact method in real data analysis. Section 3.4 concludes the findings.

3.1. Methods

3.1.1. Estimating Log OR and its Variance

Suppose we have a collection of K independent studies and are interested in inferring a common measure of interest, e.g., OR. The observed data of each study can be summarized as a 2×2 table. Similar to the notations in Chapter 2, for the kth study, $k = 1, \ldots, K$, we define X_{k1} and X_{k0} as the number of observed events in case and control groups with sample sizes n_{k1} and n_{k0}, respectively. Assume random variables X_{k1} and X_{k0} follow binomial distributions,

$$X_{k1} \sim Bin(n_{k1}, p_{k1}), \qquad X_{k0} \sim Bin(n_{k0}, p_{k0}),$$

where p_{k1} and p_{k0} are the underlying event rates. Given the observed number of events equal to x_{k1} and x_{k0} in case and control groups, respectively, we define log OR θ_k in the kth study

as

$$\theta_k = \text{logit}(p_{k1}) - \text{logit}(p_{k0}) \tag{3.1}$$

$$= \ln\left(\frac{x_{k1}}{n_{k1} - x_{k1}}\right) - \ln\left(\frac{x_{k0}}{n_{k0} - x_{k0}}\right). \tag{3.2}$$

Denote by μ_k the logit of event rate in the control group. Assume μ_k's are a constant across all control groups, that is, $\mu_k = \mu$. If θ_k is known, p_{k1} can be written as as a function of μ and θ_k,

$$\begin{cases} \text{logit}(p_{k0}) = \mu, \\ \text{logit}(p_{k1}) = \mu + \theta_k. \end{cases}$$

In the context of a random-effect model, we use a binomial-normal (BN) hierarchical model to make inferences on the common log OR θ. The BN model assumes the log ORs of individual studies θ_k's are normally distributed

$$\theta_k \sim N(\theta, \tau^2),$$

where θ is the underlying common log OR and τ^2 is between-study variance. The binomial fixed-effect model described in chapter 2 can be regarded as a special case of this random BN model when $\tau^2 = 0$.

Bhaumik *et al.* [3] introduced an estimate of individual log OR θ_k by adding a constant a to all cells of a 2×2 table in equation 3.2,

$$\hat{\theta}_{ka} = \ln\left(\frac{x_{k1} + a}{n_{k1} - x_{k1} + a}\right) - \ln\left(\frac{x_{k0} + a}{n_{k0} - x_{k0} + a}\right).$$

They further proposed a simple average estimate of the common log OR θ that borrows information from all studies as

$$\hat{\theta}_{sa} = \sum_{k=1}^{K} \frac{\hat{\theta}_{ka}}{K},$$

It is proved that this estimate is approximately unbiased when $a = 1/2$. Variance of $\hat{\theta}_{ka}$ is the sum of within-study variance σ_k^2 and between-study variance τ^2, that is, $\mathrm{Var}(\hat{\theta}_{ka}) = \sigma_k^2 + \tau^2$. Accordingly variance of the simple average estimate $\hat{\theta}_{sa}$ is $\mathrm{Var}(\hat{\theta}_{sa}) = \sum_{k=1}^{K} \mathrm{Var}(\hat{\theta}_{ka})/K^2$. An estimate of $\mathrm{Var}(\hat{\theta}_{ka})$ is

$$\widehat{\mathrm{Var}}(\hat{\theta}_{ka}) = \frac{1}{n_{k1}\hat{p}_{k1}(1-\hat{p}_{k1})} + \frac{1}{n_{k0}\hat{p}_{k0}(1-\hat{p}_{k0})} + \hat{\tau}^2_{IPM},$$

where $\hat{p}_{ki} = \frac{x_{ki}+1/2}{n_{ki}+1}, i = 0, 1$, and $\hat{\tau}^2_{IPM}$ is the IPM estimate of τ^2. Thus one can derive $\hat{\mu} = \mathrm{logit}(\hat{p}_{k0})$ and $\hat{\theta} = \mathrm{logit}(\hat{p}_{k1}) - \mathrm{logit}(\hat{p}_{k0})$.

The IPM estimate of between-study variance $\hat{\tau}^2_{IPM}$ is an improved version of the one proposed by Paule and Mandel [34] and is the solution to

$$F(\tau^2) = \sum_{k=1}^{K} w_k[\hat{\theta}_{ka} - \hat{\theta}_{wa}]^2 - (K-1) = 0, \tag{3.3}$$

where the weighted average estimate $\hat{\theta}_{wa} = \frac{\sum_{k=1}^{K} w_k \hat{\theta}_{ka}}{\sum_{k=1}^{K} w_k}$, and the inverse variance weights $w_k = (\tau^2 + \hat{\sigma}_k^2)^{-1}$. The estimated within-study variance $\hat{\sigma}_k^2$ is

$$\begin{aligned}\hat{\sigma}_k^2 =& \frac{1}{n_{k1}+1}\Big[\exp(-\hat{\mu} - \hat{\theta}_{sa} + \frac{\tau^2}{2}) + 2 + \exp(\hat{\mu} + \hat{\theta}_{sa} + \frac{\tau^2}{2})\Big] \\ &+ \frac{1}{n_{k0}+1}\Big[\exp(-\hat{\mu}) + 2 + \exp(\hat{\mu})\Big],\end{aligned} \tag{3.4}$$

and $\hat{\mu}$ is a simple average estimate of μ from the control groups. Equation (3.3) can be iteratively solved with a starting point $\tau^2_{(0)} = 0$ and an updating increment d, where d is obtained by the first order Tylor approximation

$$F(\tau^2_{(i+1)}) \approx F(\tau^2_{(i)}) + F'(\tau^2_{(i)}) \times d = 0, \quad i = 0, 1, 2, \ldots$$

such that

$$
\begin{aligned}
d &= -\frac{F(\tau^2_{(i)})}{F'(\tau^2_{(i)})} \\
&= \frac{F(\tau^2_{(i)})}{\sum_{k=1}^{K} w_k^2[\hat{\theta}_{ka} - \hat{\theta}_{wa}]^2\Big[1 + \frac{\exp(-\hat{\mu}-\hat{\theta}_{sa}+\frac{\tau^2}{2})+\exp(\hat{\mu}+\hat{\theta}_{sa}+\frac{\tau^2}{2})}{2n_{kt}+2}\Big]}.
\end{aligned}
$$

Then we update $\tau^2_{(i+1)} = \tau^2_{(i)} + d$ until $d < 10^{-6}$.

3.1.2. Exact CI for Log OR

The procedure of computing the exact $100(1-\alpha)\%$ CI for log OR θ can be split into three major steps:

1. Define parameter search region $[\theta_{SL}, \theta_{SU}] \times [\tau^2_{SL}, \tau^2_{SU}]$, where subscript "SL" stands for the search lower bound and subscript "SU" stands for the search upper bound;

2. For $I \times J$ grid in the search region, calculate the exact p-values $p(\theta_i, \tau_j^2)$ at all grid points (θ_i, τ_j^2), where $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, J$;

3. Project $p(\theta_i, \tau_j^2)$ in the search region onto θ axis to obtain one-dimensional p-values $p(\theta_i)$, and then construct the CI for θ.

In the second step, we define the exact p-value of an observed dataset at θ as the probability of observing the null test statistic $T^{(b)}$ greater than or equal to the observed test statistic T^{obs},

$$
\begin{aligned}
p(\theta_i, \tau_j^2|(x_{k1}, x_{k0})^{obs}) &= \Pr\left\{T^{(b)} \geq T^{obs}\right\} \\
&= \Pr\left\{T[(x_{k1}, x_{k0})^{(b)}] \geq T[(x_{k1}, x_{k0})^{obs}]\right\}, \qquad (3.5)
\end{aligned}
$$

where $(x_{k1}, x_{k0})^{obs}$ is the observed dataset, and $(x_{k1}, x_{k0})^{(b)}, b = 1, 2, \ldots, B$, are random

datasets generated from BN models as

$$\begin{cases} x_{k0}^{(b)} \sim Bin(n_{k0}, p_{k0}), \\ x_{k1}^{(b)} \sim Bin(n_{k1}, p_{k1}^{(b)}). \end{cases}$$

Note that (p_{k0}, n_{k0}, n_{k1}) are from the observed studies, and $p_{k1}^{(b)}$ is generated based on grid point (θ_i, τ_j^2), in particualr,

$$\begin{cases} \theta_k^{(b)} \sim N(\theta_i, \tau_j^2), \\ \text{logit}(p_{k1}^{(b)}) = \text{logit}(p_{k0}) + \theta_k^{(b)}. \end{cases}$$

Size of B can be a trade-off between estimation precision and computation expense. A larger size of B leads to a more precise estimate of $p(\theta_i, \tau_j^2)$, though it requires more time to generate null datasets.

In summary, to calculate p-value at each grid point $(\theta_i,\ \tau_j^2)$ in the search region, the second step can be expanded as the following steps:

2.1 Generate $\theta_k^{(b)}$ from $N(\theta_i, \tau_j^2)$, $k = 1, \ldots, K$;

2.2 Calculate $p_{k1}^{(b)}$ based on $\theta_k^{(b)}$ and p_{k0};

2.3 Generate $x_{k1}^{(b)}$ and $x_{k0}^{(b)}$ from $Bin(n_{k1}, p_{k1}^{(b)})$ and $Bin(n_{k0}, p_{k0})$;

2.4 Given the generated data, esitmate $\hat{\theta}_s^{(b)}$ and $\widehat{\text{Var}}(\hat{\theta}_s^{(b)})$;

2.5 Compute test statistic $T^{(b)}$;

2.6 Repeat the above steps for B times to obtain a null distribution of test statistic T, $b = 1, \ldots, B$;

2.7 Compute p-value as $\Pr(T^{(b)} \geq T^{obs})$.

We use the simple average estimate $\hat{\theta}_{s1/2}$ and its estimated variance $\widehat{\text{Var}}(\hat{\theta}_{s1/2})$ to construct the test statistic T and build the exact CI for log OR θ. For a two-sided test $H_0 : \theta = \theta'$ v.s. $H_1 : \theta \neq \theta'$, a general choice of test statistic T is a Wald-type test statistic

$$T = \frac{(\hat{\theta}_{s1/2} - \theta')^2}{\widehat{\text{Var}}(\hat{\theta}_{s1/2})}.$$

Based on T, one can compute the exact p-values $p(\theta_i, \tau_j^2)$ at all $I \times J$ grid points. Since our goal is to construct CI for θ, we need to transfer the two-dimensional exact p-values $p(\theta_i, \tau_j^2)$ to one-dimensional exact p-values $p(\theta_i)$ by projecting onto θ axis. It can be accomplished by maximizing over all possible τ_j^2's at each θ_i as

$$p(\theta_i) = \max_{\tau_j^2}\{p(\theta_i, \tau_j^2)\}.$$

The lower bound (θ_L) and the upper bound (θ_U) of the exact $100(1-\alpha)\%$ CI for log OR θ are defined as

$$\theta_L = \inf_{\theta}\Big\{\theta : p(\theta_i) \geq \alpha\Big\},$$
$$\theta_U = \sup_{\theta}\Big\{\theta : p(\theta_i) \geq \alpha\Big\}.$$

The corresponding exact $100(1-\alpha)\%$ CI for OR is thus $(\exp(\theta_L), \exp(\theta_U))$.

3.1.3. Efficient Parameter Search Region

To find the initial parameter search region $[\theta_{SL}, \theta_{SU}] \times [\tau_{SL}^2, \tau_{SU}^2]$, we can start with a wide initial range. The initial search bounds for θ and τ^2 can be determined from the 99.9% CI based on the normal approximation. In addition, as τ^2 is non-negative, the search lower bound τ_{SL}^2 is automatically set to be 0. The initial parameter search region therefore becomes $[\theta_{SL}, \theta_{SU}] \times [0, \tau_{SU}^2]$. Note that in the initial parameter search region, θ and τ^2

are independent, which results in a rectangle-shaped search region. For example, as shown in figure 3.1, the initial search region is a rectangle at this stage, and grey points are grid points located in this search region. This parameter search region is too large that it takes a long time to compute the exact p-values at all grid points (e.g., 16 grid points in figure 3.1) located in such large search region. The computation cost issue can be resolved by cutting down the parameter search region.

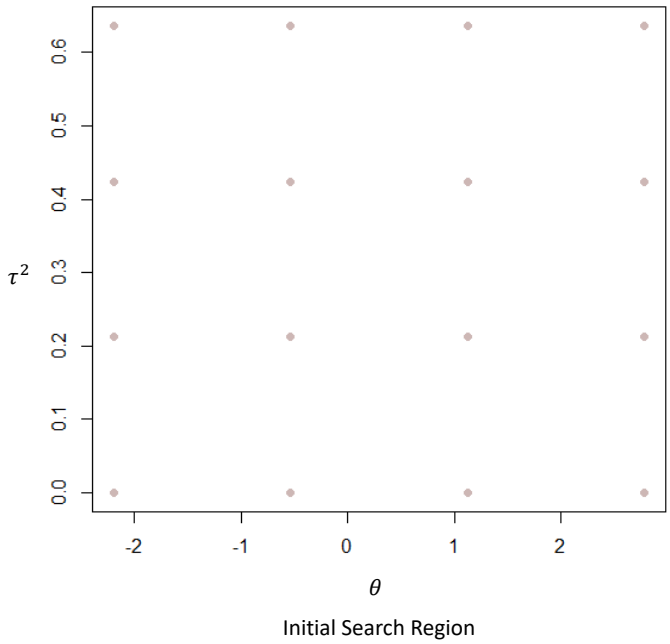

Figure 3.1: Example of initial parameter search region. It is rectangle-shaped due to independence between θ and τ^2, and it has 16 grid points.

A test statistic T_{τ^2} [3] for testing heterogeneity $H_0 : \tau^2 = 0$ designed for rare events is

$$T_{\tau^2} = \frac{\sum_{k=1}^{K}(y_k - A_k)}{\sqrt{\sum_{k=1}^{K} 2A_k}},$$

where

$$y_k = (\hat{\theta}_{k1/2} - \hat{\theta}_{s1/2})^2,$$
$$A_k = \frac{K-2}{K}\mathrm{Var}(\hat{\theta}_{k1/2}) + \frac{\sum_{k=1}^{K}\mathrm{Var}(\hat{\theta}_{k1/2})}{K^2},$$

mean of y_k is A_k and variance of y_k is $2A_k$. To find relationship between θ and τ^2, we can view $\mathrm{Var}(\hat{\theta}_{k1/2})$ as a function of θ and τ^2, denoted as $v_k(\theta, \tau^2)$:

$$\begin{aligned} v_k(\theta,\tau^2) =& \frac{1}{n_{k1}}\Big[\exp(-\hat{\mu}-\theta+\frac{\tau^2}{2}) + 2 + \exp(\hat{\mu}+\theta+\frac{\tau^2}{2})\Big] \\ &+ \frac{1}{n_{k0}}\Big[\exp(-\hat{\mu}) + 2 + \exp(\hat{\mu})\Big] + \tau^2. \end{aligned}$$

Note that, the part of with-in study variance in $v_k(\theta, \tau^2)$ is the expectation of $\frac{1}{n_{k1}p_{k1}(1-p_{k1})}+\frac{1}{n_{k0}p_{k0}(1-p_{k0})}$ under the BN model assumption. Now A_k can be written as

$$A_k = \frac{K-2}{K}v_k(\theta,\tau^2) + \frac{\sum_{k=1}^{K} v_k(\theta,\tau^2)}{K^2},$$

and

$$\begin{aligned} \sum_{k=1}^{K} A_k &= \frac{K-2}{K}\sum_{k=1}^{K} v_k(\theta,\tau^2) + \sum_{k=1}^{K}\frac{\sum_{k=1}^{K} v_k(\theta,\tau^2)}{K^2} \\ &= \frac{K-1}{K}\sum_{k=1}^{K} v_k(\theta,\tau^2). \end{aligned}$$

As test statistic T_{τ^2} is normally distributed, we can find CI bounds of τ^2 by equating test statistic T_{τ^2} to its quantile value $Z_{\alpha(\tau^2)}$,

$$T_{\tau^2} = Z_{\alpha(\tau^2)}. \tag{3.6}$$

The relationship between θ and τ^2 can be identified by solving equation (3.6) as below:

$$\begin{aligned}
& \frac{\sum_{k=1}^{K}(y_k - A_k)}{\sqrt{\sum_{k=1}^{K} 2A_k}} = Z_{\alpha(\tau^2)} \\
\Rightarrow \quad & \sum_{k=1}^{K} y_k - \sum_{k=1}^{K} A_k = Z_{\alpha(\tau^2)} \sqrt{2 \sum_{k=1}^{K} A_k} \\
\Rightarrow \quad & \sum_{k=1}^{K} A_k = \left(\sqrt{\sum_{k=1}^{K} y_k + \frac{1}{2} Z^2_{\alpha(\tau^2)}} - \sqrt{\frac{1}{2} Z^2_{\alpha(\tau^2)}} \right)^2 \\
\Rightarrow \quad & \sum_{k=1}^{K} v_k(\theta, \tau^2) = \frac{K}{K-1} \left(\sqrt{\sum_{k=1}^{K} y_k + \frac{1}{2} Z^2_{\alpha(\tau^2)}} - \sqrt{\frac{1}{2} Z^2_{\alpha(\tau^2)}} \right)^2 . \qquad (3.7)
\end{aligned}$$

The right side of equation (3.7) is a constant part, denoted as $C1$. The left side of equation (3.7) can be expanded as

$$\begin{aligned}
\sum_{k=1}^{K} v_k(\theta, \tau^2) = & \Big[\exp(-\hat{\mu} - \theta + \frac{\tau^2}{2}) + \exp(\hat{\mu} + \theta + \frac{\tau^2}{2}) \Big] \Big(\sum_{k=1}^{K} n_{k1}^{-1} \Big) + K\tau^2 \\
& + 2 \sum_{k=1}^{K} n_{k1}^{-1} + \Big[\exp(-\hat{\mu}) + \exp(\hat{\mu}) + 2 \Big] \Big(\sum_{k=1}^{K} n_{k0}^{-1} \Big). \qquad (3.8)
\end{aligned}$$

Note the second line of equation (3.8) is unrelated to θ or τ^2, and thus can also be viewed as a constant part, denoted as $C2$. Therefore we have two constant parts defined as

$$\begin{aligned}
C1 &= \frac{K}{K-1} \left(\sqrt{\sum_{k=1}^{K} y_k + \frac{1}{2} Z^2_{\alpha(\tau^2)}} - \sqrt{\frac{1}{2} Z^2_{\alpha(\tau^2)}} \right)^2 , \\
C2 &= 2 \sum_{k=1}^{K} n_{k1}^{-1} + \Big[\exp(-\hat{\mu}) + \exp(\hat{\mu}) + 2 \Big] \Big(\sum_{k=1}^{K} n_{k0}^{-1} \Big).
\end{aligned}$$

Now equation (3.7) becomes

$$\Big[\exp(-\hat{\mu} - \theta) + \exp(\hat{\mu} + \theta) \Big] \Big(\sum_{k=1}^{K} n_{k1}^{-1} \Big) \Big(\exp(\frac{\tau^2}{2}) \Big) + K\tau^2 - \Big(C1 - C2 \Big) = 0. \qquad (3.9)$$

Since the lower bound τ^2_{SL} is set to be 0, we are only interested in the upper bound τ^2_{SU}. Given any $\theta_i \in [\theta_{SL}, \theta_{SU}], i = 1, \ldots, I$, the upper bound τ^2_{SU} is the solution to equation

(3.9). Similar to solving equation (3.3), we can solve equation (3.9) from the first order Tylor approximation with a starting point $\tau^2(0) = 0$. Let $F(\tau^2)$ be the left side of equation (3.9), then the updating increment d is

$$
\begin{aligned}
d &= -\frac{F(\tau^2_{(i)})}{F'(\tau^2_{(i)})} \\
&= -\frac{\Big[\exp(-\hat{\mu}-\theta_i)+\exp(\hat{\mu}+\theta_i)\Big]\Big(\sum_{k=1}^{K} n_{k1}^{-1}\Big)\Big(\exp(\frac{\tau^2_{(i)}}{2})\Big)+K\tau^2_{(i)}-\Big(C1-C2\Big)}{\frac{1}{2}\Big[\exp(-\hat{\mu}-\theta_i)+\exp(\hat{\mu}+\theta_i)\Big]\Big(\sum_{k=1}^{K} n_{k1}^{-1}\Big)\Big(\exp(\frac{\tau^2_{(i)}}{2})\Big)+K}.
\end{aligned}
$$

We update $\tau^2_{(i+1)} = \tau^2_{(i)} + d$ until $d < 10^{-6}$.

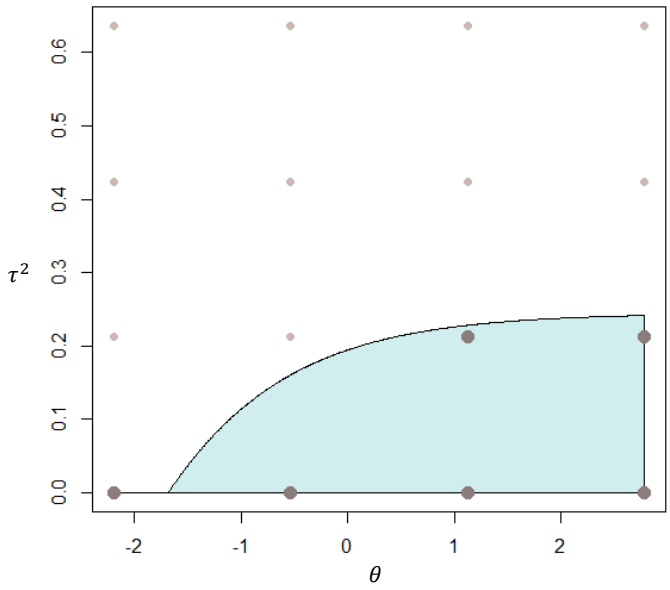

Figure 3.2: Example of the final parameter search region (blue area). The initial rectangle-shaped search region shrinks to the blue area. Now we only need to handle 6 of the 16 grid points.

In this way, we have solution of τ^2 to be related to grid value of θ_i and denoted this solution by $\tau^2_{SU}(\theta_i)$. Now the parameter search region shrinks to $[\theta_{SL}, \theta_{SU}] \times [0, \tau^2_{SU}(\theta_i)]$. For example, in figure 3.2, the blue area is the final parameter search region. Now we only have 6 grid points located in the search region instead of 16 grid points in the initial search region.

This helps speed up the algorithm.

3.2. Simulation

3.2.1. Simulation Settings

We conduct simulation studies to examine the performance of the proposed exact BN method in terms of making inferences on CI. We consider both small and large number of studies, where $K = 5$ mimics a scenario of small size meta-analysis study and $K = 15$ represents a scenario of large size meta-analysis study. We consider a balanced study design with 5000 patients in both case and control groups.

Under a BN model, we generate the number of observed events from a binomial distribution with the event rate $p_{k0} = \{0.5\%, 1\%\}$ in the control group. Such p_{k0} leads to the logit of event rate $\mu = \text{logit}\{p_{k0}\} = \{-5.3, -4.6\}$, respectively. Hereinafter we denote these two event rates as rare and low, respectively. We consider five scales of true mean of log OR, $\theta = \{-1, -0.5, 0, 0.5, 1\}$, which involves situations where the event rate in the case group is less than, equal to, and greater than that in the control group. The underlying between-study variance τ^2 are set according to different sizes of I^2 statistics. It is generated to have I^2 approximately around $0, 25\%, 50\%, 75\%$, which represents none heterogeneity, mild heterogeneity, moderate heterogeneity, and high heterogeneity, respectively [19]. The specific values of τ^2 under different I^2 are listed in table 3.1.

K	μ	τ^2			
		$I^2 = 0$	$I^2 = 25\%$	$I^2 = 50\%$	$I^2 = 75\%$
5	-5.3	0	0.04	0.3	1.2
	-4.6	0	0.012	0.054	0.5
15	-5.3	0	0.04	0.15	0.6
	-4.6	0	0.02	0.008	0.3

Table 3.1: Settings of the underlying τ^2 when $K = \{5, 15\}$, $\mu = \{-5.3, -4.6\}$ for none, mild, moderate, and high heterogeneity.

Simulation for each underlying τ^2 listed in table 3.1 is repeated 500 times. We set the number of grid points for θ, I, to be 1000 and the number of grid point for τ^2, J, to be 10. The number of repeats, B, is set to be 1000. If I, J, B increase, accuracy of the resulting CI will increase at the cost of longer computation time.

We compare the proposed method with six other methods. Among the six methods, three are fixed-effect methods including the Peto method, the MH method, and the inverse variance method (INV). The other three are random-effect methods including the DerSimonian and Laird method (DL) [13], the Paule and Mandel method with INV estimate (INV-PM), and the IPM method. All these six methods construct CIs relying on large sample approximation. Note that the IPM method applies the same method as ours to make the point estimate of θ. The difference between the IPM method and the exact method is that the IPM method builds CI based on normal approximation. Thus we can specifically compare the performance of large sample approximation and the exact calculation when point estimates are same.

Similar to the simulation study in chapter 2, we compare CI coverage probabilities, CI length, and power. Different from chapter 2, estimate bias is no longer considered since it is derived as a simple average estimate instead of being derived from the exact CI.

3.2.2. Simulation Results

Coverage probability is calculated as the probability of covering the true log OR θ. It is expected to be around the nominal level, e.g., 95%. If it is over the nominal level then we consider the CI as conservative, whereas if it is below the nominal level then we regard it as permissive. In this section, we present compare coverage probabilities of the proposed exact method and six other methods in four scenarios:

1. Small size meta-analysis (5 studies) and low event rate in the control group, $p_{k0} = 1\%$;
2. Small size meta-analysis (5 studies) and rare event rate in the control group, $p_{k0} = 0.5\%$;
3. Large size meta-analysis (15 studies) and low event rate in the control group, $p_{k0} = 1\%$;
4. Large size meta-analysis (15 studies) and rare event rate in the control group, $p_{k0} = 0.5\%$.

Performance of the coverage probabilities and CIs length in all scenarios are presented in figures 3.3 - 3.6. Note that $\mu = -4.6$ is equivalent to $p_{k0} = 1\%$ and $\mu = -5.3$ is equivalent to $p_{k0} = 0.5\%$. In each figure, four sub-figures are sorted from left to right presenting results in situations of no heterogeneity, mild heterogeneity, moderate heterogeneity, and high heterogeneity, respectively. In each sub-figure, x-axis is the true log OR ranging from -1 to 1, and y-axis is the coverage probability ranging from 0.8 to 1. Note that in some cases the coverage probabilities are lower than 0.8 and thus are shown as texts in the figure. The black dashed line indicates the nominal level of CI, that is, 95%. Any point below the dashed line means permissive coverage probability and any point above the dashed line means conservative coverage probability. All the fixed-effect methods (Peto, MH, and INV) are highlighted in yellow and all the random-effect methods (DL, INV-PM, and IPM) are in green. The proposed exact method is highlighted in red.

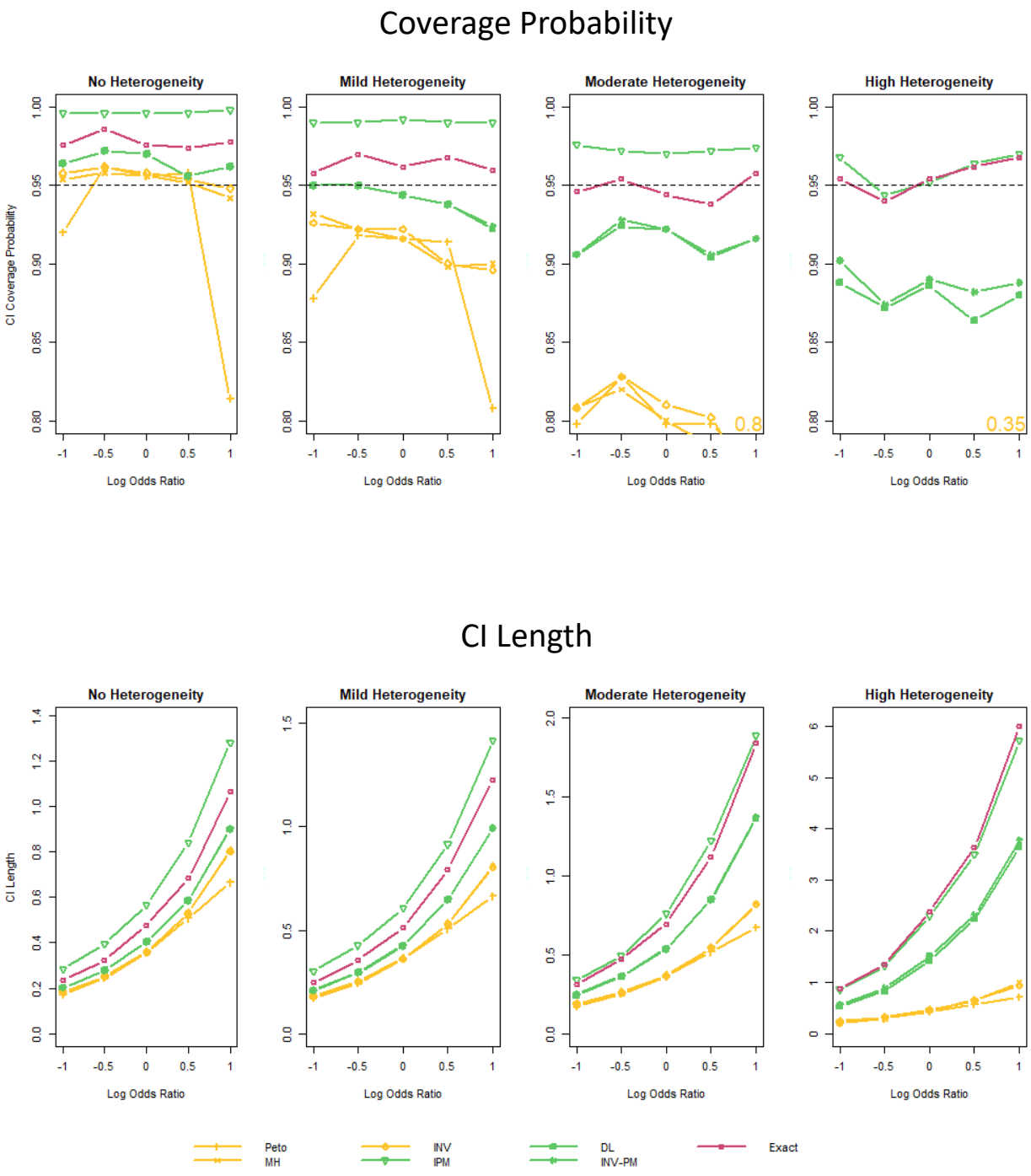

Figure 3.3: Comparison of coverage probabilities and CIs length of the proposed exact method and six other methods when $K = 5$ and $\mu = -4.6$.

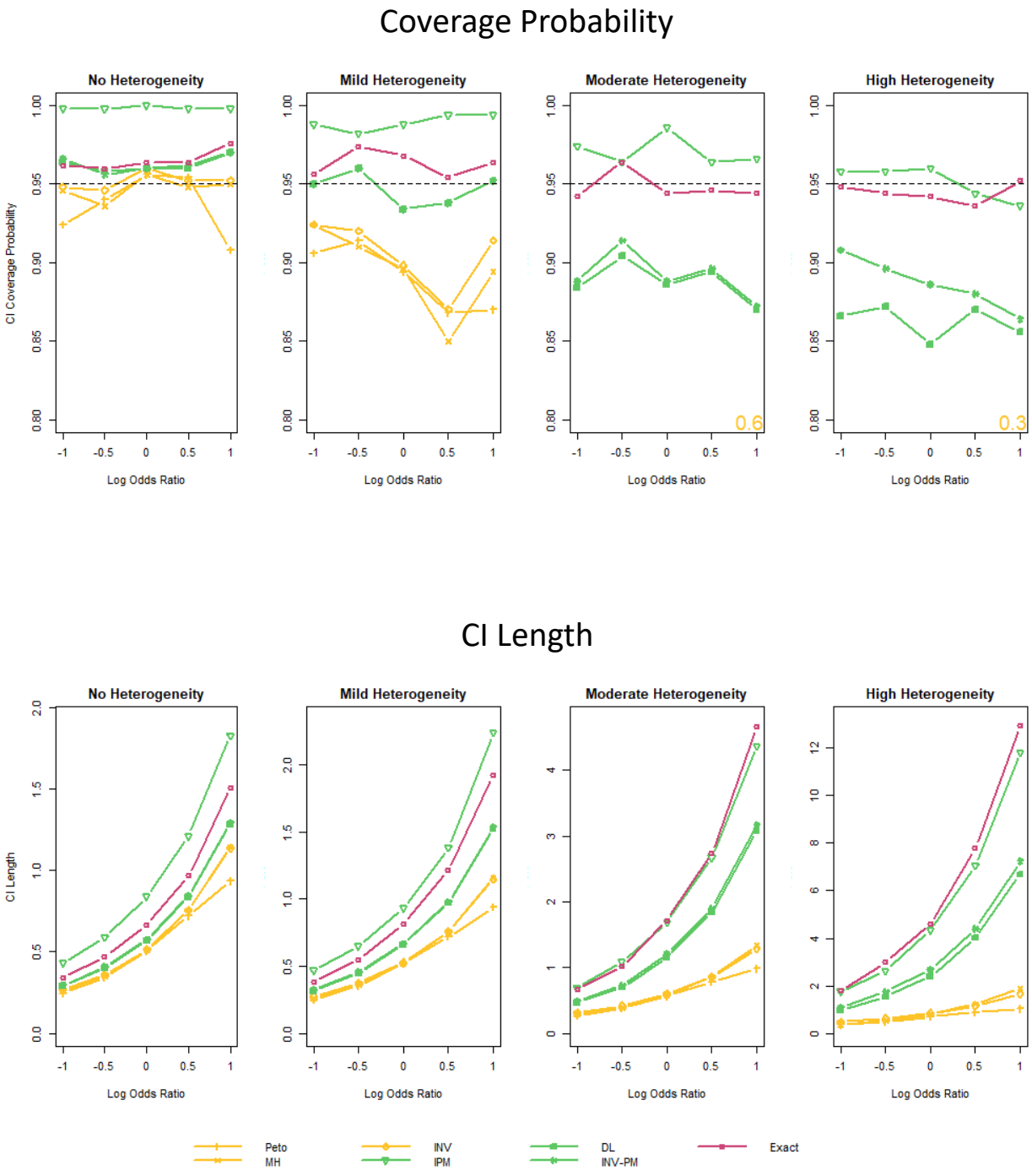

Figure 3.4: Comparison of coverage probabilities and CIs length of the proposed exact method and six other methods when $K = 5$ and $\mu = -5.3$.

Figure 3.3 presents results when there are 5 studies and a low event rate $\mu = -4.6$. When there is no heterogeneity, the MH method and the INV method have the coverage probability constantly around the nominal level. The Peto method shows poor performance when the true log OR is large. In contrast, all random-effect methods and the exact method are relatively conservative, especially the IPM method, which constantly has coverage probability close to 100%. When the between-study variance increases, all methods have decreasing coverage probabilities. Particularly, when high heterogeneity presents, all three fixed-effect methods have coverage probabilities only around 35% for all settings of true log OR. For random-effect methods, the DL and the INV-PM methods have better performance compared to the fixed-effect methods with coverage probabilities around 90%. The IPM method and the exact method have similar performance with the coverage probabilities around the nominal level. Therefore, the fixed-effect methods generally have better performance when the between-study variance is small, and the random-effect methods are better when there is substantial heterogeneity among studies. The proposed exact method outperforms others as it has very stable performance and always yield coverage probabilities around the nominal level.

At the rare event rate, i.e., $\mu = -5.3$, as shown in figure 3.4, the results are similar to that in the case of the low event rate. However, in the presence of moderate to high heterogeneity, all methods except the exact method and the IPM method have even lower coverage probabilities.

As presented in figures 3.5 and 3.6, when the number of studies is large, $K = 15$, all random-effect methods and the exact method yield coverage probabilities close to the nominal level. In particular, when the between-study variance is small, their CIs are no longer conservative. The proposed exact method overall outperforms other methods with its steady performance and accurate coverage probability around 95%.

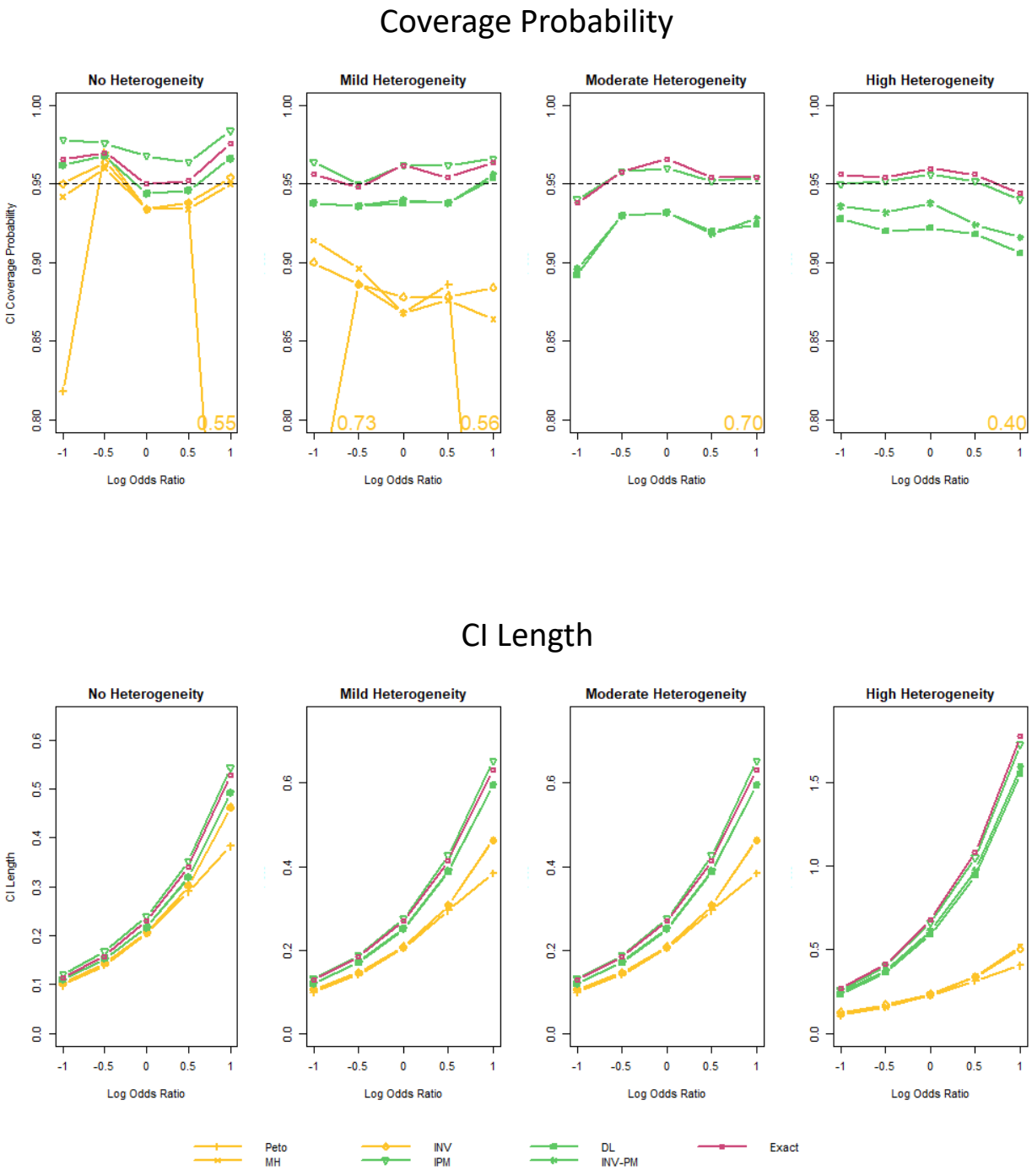

Figure 3.5: Comparison of coverage probabilities and CIs length of the proposed exact method and six other methods when $K = 15$ and $\mu = -4.6$.

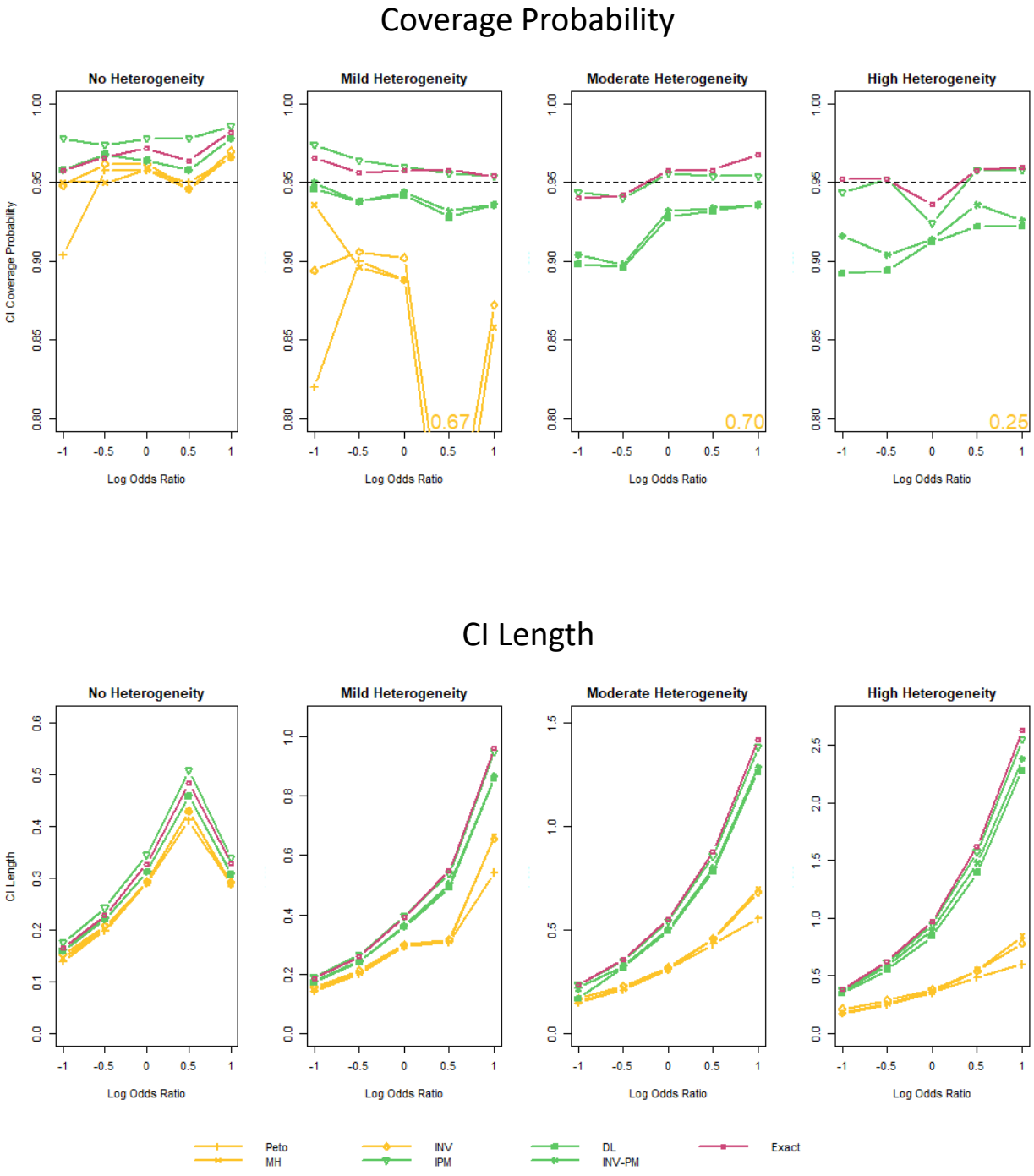

Figure 3.6: Comparison of coverage probabilities and CIs length of proposed exact method and six other methods when $K = 15$ and $\mu = -5.3$.

Statistical power is defined as the probability of CI not covering 0 when the true log OR $\theta \neq 0$. Simulation results are shown in figure B.1 for $K = 5$ and in figure B.2 for $K = 15$ in appendix B. When $K = 15$, the fixed-effect methods have higher power than the random-effect methods; when $K = 15$, all methods have similar power.

3.3. Real Data Analysis

High plasma levels of low-density lipoprotein cholesterol (LDL-C) are associated with high risk of CHD. Inhibiting the function of Niemann-Pick C1-like 1 (NPC1L1) protein can lower LDL-C levels. However, it is unclear whether such inhibition can reduce the risk of CHD. A meta-analysis [23] examined the association between protein-inactivating mutations in NPC1L1 and the risk of CHD. This meta-analysis involved 29,954 patients with CHD and 83,140 healthy controls from 16 studies. Event rates in case and control groups were 0.037% and 0.085%, respectively, and nine studies observed zero event in the case groups. Detail of the data structure is listed in table B.2 in appendix B. In the original study, the MH fixed-effect method without continuity correction was applied, resulting in an estimated OR = 0.47 with the 95% CI = $(0.25, 0.87)$ and p-value = 0.008. It was concluded that NPC1L1 mutations carriers had 53% lower risk of CHD.

We re-analyze the data with the proposed exact method and the other six methods mentioned in the simulation study. The I^2 test statistic equals 9.82%, therefore the extent of heterogeneity is between no and mild heterogeneity. Analysis results are shown in figure 3.7. Different from the original study, we apply the 0.5 continuity correction to all compared methods except the Peto method. Note that CIs presented in figure 3.7 are not symmetric as we calculate CIs first based on a log scale and then convert to CIs of OR. The exact method produces a CI=$(0.184, 0.936)$, which is slightly wider than MH method but still reaches the same conclusion. The point estimate OR = 0.51 with the p-value = 0.026. In contrast, the INV-PM, DL, and INV methods have larger point estimate OR = 0.69 but insignificant

results.

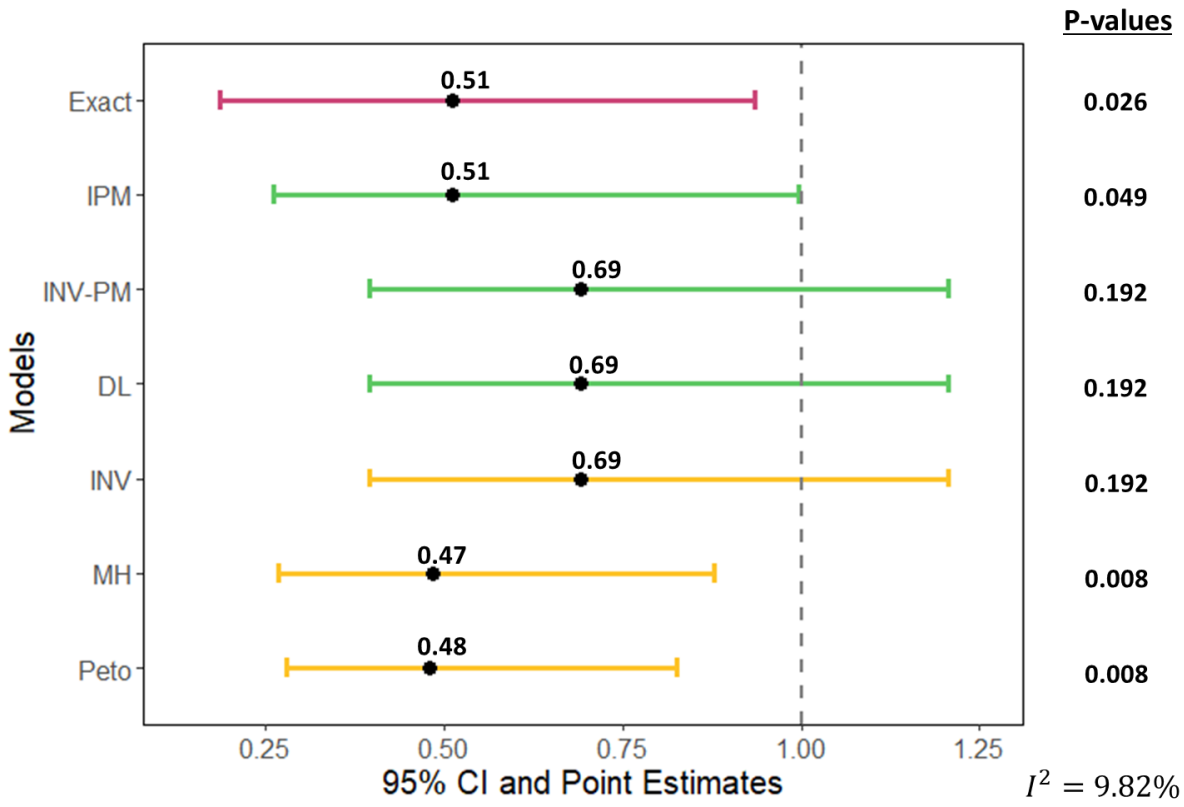

Figure 3.7: Plot of 95% CI, point estimates, and p-values for NPC1L1 data.

3.4. Summary and Discussion

In this chapter, we propose an exact method based on the random-effect model to build CI for OR in the case of rare event rates. It assumes a BN random-effect model where the true log OR in each single study comes from a normal distribution and the number of events in the 2×2 table comes from binomial distributions. The exact CI is constructed as lower bound and upper bound of log OR in the area where all p-values are greater than 0.05. We propose a speed-up algorithm to shrink the parameter search region by utilizing the relationship between log OR and between-study variance.

The simulation study considers various scenarios including different number of studies, various extents of heterogeneity and event rates, and different magnitude of true log ORs. The fixed-effect methods (INV, Peto, MH) have best performance when there is no heterogeneity and the number of studies is large. It is reasonable as they are designed for situation of without between-study variance and they rely on large sample approximation. The random-effect methods (IPM, INV-PM, DL) based on large sample approximation outperform the fixed-effect methods when there is high heterogeneity. The proposed random-effect exact method show stable performance with the coverage probability constantly around the 95% nominal level. Thus, in the case of rare events data, the proposed exact method is preferred to build the CI for log OR regardless of the heterogeneity and sample sizes.

CHAPTER 4

EXACT INFERENCE FOR META-ANALYSIS OF RARE EVENTS USING COMMON CONTROLS

In this chapter, we consider a special study design that there are multiple case groups but there is only one control group, which is usually a large survey database from the general population. We propose an exact method to make inferences on the event rate and effect size measure in the case groups. This chapter is organized as follows: in section 4.1 we introduce the exact method including its estimates and computing algorithm, in section 4.2 we conduct simulation studies, and section 4.3 concludes the findings.

4.1. Methods

4.1.1. Assumptions and Estimates

Assume there are a series of individual studies on the same disease, and in all studies only cases / patients are recruited. A survey dataset of the general population is used as the common control group, e.g., the the Genome Aggregation Database [25]. Denote by K the number of studies involved in meta-analysis except the common control group. X_{k1} is the number of events observed in the kth individual study, where $k = 1, \ldots, K$. Denote by X_0 the number of events observed in the common control group. Sample sizes in the case and control groups are denoted as n_{k1} and n_0, respectively. For the control group, we assume

the number of events X_0 follows a binomial distribution,

$$X_0 \sim Bin(n_0, p_0),$$

where p_0 is the observed event rate.

To estimate the effect size measures RD and OR, we adapt an exact method to estimate and build CI for the pooled effect size from case groups [16]. Assmue a beta-binomial (BB) random-effect model,

$$X_{k1} \sim Bin(n_{k1}, p_{k1}), \quad p_{k1} \sim Beta(\alpha, \beta), \quad k = 1, \ldots, K, \tag{4.1}$$

where p_{k1} is the observed event rate in the kth study, and $\alpha, \beta > 1$. In a beta distribution, parameters α and β are set to be greater than 0. The reason we set them to be greater than one is to ensure the unimodal shape of the distribution. More details of this choice will be discussed later.

The between-study variation of p_{k1} comes from the beta distribution, so the expected value of p_{k1}, denoted by p_1, and the between-study variance of p_{k1}, denoted by τ^2, are

$$\begin{aligned} p_1 &= \mathrm{E}(p_{k1}) = \frac{\alpha}{\alpha + \beta}, \\ \tau^2 &= \mathrm{Var}(p_{k1}) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}. \end{aligned}$$

p_1 can be viewed as the true event rate in the case groups. The estimate of p_{k1} is $\hat{p}_{k1} = x_{k1}/n_{k1}$. If there is zero event presenting in a case group then we apply continuity correction by $\hat{p}_{k1} = (x_{k1} + 1)/(n_{k1} + 2)$. The BB distribution in equation (4.1) is also known as

$BB(n_{k1}, \alpha, \beta)$, and the mean and variance of estimate $\hat{p}_{k1}$ are

$$\begin{aligned}
&\mathrm{E}(\hat{p}_{k1}) = p_1, \\
&\mathrm{Var}(\hat{p}_{k1}) = \frac{p_1(1-p_1)}{n_k} + (1 - \frac{1}{n_k})\tau^2.
\end{aligned}$$

The aim is to estimate p_1 and build a CI for it. We use the inverse variance estimate $\hat{p}_1$ to estimate the pooled event rate p_1, where

$$\begin{aligned}
\hat{p}_1 &= \frac{\sum_{k=1}^{K} w_k \hat{p}_{k1}}{\sum_{k=1}^{K} w_k}, \\
w_k &= \left\{\widehat{\mathrm{Var}}(\hat{p}_{k1})\right\}^{-1} = \left\{\frac{\hat{p}_1(1-\hat{p}_1)}{n_{k1}} + (1 - \frac{1}{n_k})\hat{\tau}^2\right\}^{-1}.
\end{aligned}$$

The estimated between-study variance $\hat{\tau}^2$ is a DL-type estimator,

$$\begin{aligned}
\hat{\tau}^2 &= \max\left\{0, \frac{\sum_{k=1}^{K}(\hat{p}_{k1}^2 - \bar{p}_1/n_{k1})}{\sum_{k=1}^{K}(1 - 1/n_{k1})} - \bar{p}_1^2\right\}, \\
\bar{p}_1 &= \frac{\sum_{k=1}^{K} \hat{p}_{k1}}{K},
\end{aligned}$$

where $\bar{p}_1$ is the average event rate estimated from all case groups. Based on these estimates, once we collect data from all cases groups, we can estimate the event rate in the case groups following a beta distribution with the estimated mean $\hat{p}_1$ and variance $\hat{\tau}^2$. Therefore,

$$\begin{aligned}
\hat{\alpha} &= \frac{(1-\hat{p}_1)\hat{p}_1^2}{\hat{\tau}^2} - \hat{p}_1, \\
\hat{\beta} &= \frac{(1-\hat{p}_1)^2\hat{p}_1}{\hat{\tau}^2} - (1-\hat{p}_1).
\end{aligned}$$

Such reparametrization allows us to update estimation of α and β based on the estimated $\hat{p}_1$ and $\hat{\tau}^2$.

4.1.2. Exact Confidence Interval for p_1

As in previous chapters, an exact confidence interval is built based on grid values of parameter search region. Here parameter search region is a two dimensional space $[p_{1L}, p_{1U}] \times [\tau_L^2, \tau_U^2]$, where subscript L and U refer to the lower and upper bounds in the parameter search region. Initial guess of lower and upper bounds can be estimated based on the normal approximated CI.

To accelerate the exact algorithm, the parameter search region can be reduced by expressing τ^2 as a function of p_1. In this way, τ^2 and p_1 become dependent. This can be achieved by setting $\alpha, \beta > 1$, which implies the following inequality

$$\begin{aligned} & \frac{1}{\alpha+\beta+1} \leq \min\{\frac{\alpha}{2\alpha+\beta}, \frac{\beta}{\alpha+2\beta}\}, \\ \Rightarrow \quad & \tau^2 = \frac{p_1(1-p_1)}{\alpha+\beta+1} \leq p_1(1-p_1)\min\{\frac{\alpha}{2\alpha+\beta}, \frac{\beta}{\alpha+2\beta}\}. \end{aligned}$$

Given a specific p_1, the inequality defines that the search upper bound of τ^2 is $\tau_U^2(p_1) = p_1(1-p_1)\min\{\frac{p_1}{1+p_1}, \frac{1-p_1}{2-p_1}\}$.

For a two-sided test $H_0 : p_1 = p_1'$ v.s. $H_1 : p_1 \neq p_1'$, we use a Wald-type test statistic T,

$$T = \frac{(\bar{p}_1 - p_1')^2}{\sum_{k=1}^{K} \widehat{\mathrm{Var}}(\hat{p}_{k1})}.$$

Let $T^{(obs)}$ denote the observed test statistic and $T^{(b)}$ the null test statistic. The null test statistic $T^{(b)}$ is calculated based on generated dataset $p_{k1}^{(b)}, b = 1, \ldots, B$ from a beta distribution with mean $p_1^{(i)}$ and variance $\tau^{2(j)}$, where $p_1^{(i)}, i = 1, \ldots, I$, and $\tau^{2(j)}, j = 1, \ldots, J$, are grid values in the parameter search region. Then we can compute the exact p-value at $(p_1^{(i)}, \tau^{2(j)})$ as $p(p_1^{(i)}, \tau^{2(j)}) = \Pr\{T^{(b)}(p_1^{(i)}, \tau^{2(j)}) \geq T^{(obs)}\}$.

To construct 95% CI for p_1, the idea is to project $p(p_1^{(i)}, \tau^{2(j)})$ onto the p_1 axis to have the final exact p-value $p(p_1^{(i)})$, which is a p-value only regarding to the grid value of p_1. Note

that if τ^2 increases then p-value also increases [16]. This helps accelerate the algoithm that we do not need to go through all grid values of τ^2; instead we can directly compute p-value at the maximum value of $\tau^2_U(p_1^{(i)})$ at each $p_1^{(i)}$. The 95% CI for p_1 is defined as the minimum and maximum grid value of $p_1^{(i)}$ that have p-values greater than or equal to the significant level 0.05,

$$p_{1L} = \inf_{p_1}\Big\{p_1 : p(p_1) \geq 0.05\Big\},$$
$$p_{1U} = \sup_{p_1}\Big\{p_1 : p(p_1) \geq 0.05\Big\}.$$

In practice, the algorithm first computes p-values for all grids and then obtains CI for p_1. The details of each step are:

1. Find an initial search region of $[p_{1L}, p_{1U}]$ and compute p-values on the bounds;

2. Calculate p-values on grids of p_1 and at each $p_1^{(i)}$ use maximum τ^2 value as $\tau^2_U(p_1^{(i)})$:

2.1. If $p(p_{1L}) > 0.05$ then find p-values of outside grid points $p_1^{(i)} < p_{1L}$ until p-value $p(p_1^{(i)}) < 0.05$; otherwise, look for inside grid points $p_1^{(i)} > p_{1L}$;

2.2. If $p(p_{1U}) > 0.05$ then find p-values of outside grid points $p_1^{(i)} > p_{1U}$ until p-value $p(p_1^{(i)}) < 0.05$; otherwise, look for inside grid points $p_1^{(i)} < p_{1U}$;

3. Build 95% CI for p_1 based on the exact p-values.

Once we have the 95% CI for p_1, denoted as $[p_{1L}^{exact}, p_{1U}^{exact}]$, we are able to make inferences on the effect size measures such as RD and OR with the event rate p_0 in the common control group. Let θ represent the effect size measure. Denote by θ_{RD} RD, and denote by θ_{OR} OR. Since RD and OR monotonically increase as p_1 increases, the 95% CI for RD and OR can

be constructed as

$$\begin{aligned} \text{RD:} \quad & [p_{1L}^{exact} - p_0, p_{1U}^{exact} - p_0], \\ \text{OR:} \quad & [\frac{p_{1L}^{exact}(1-p_0)}{p_0(1-p_{1L}^{exact})}, \frac{p_{1U}^{exact}(1-p_0)}{p_0(1-p_{1U}^{exact})}]. \end{aligned}$$

4.2. Simulation

We conduct simulation studies to mimic real data scenarios where we have multiple case groups and one common control group. We perform simulation based on the APOC3-CHD data [46] discussed in chapter 2. The data structure can be found in appendix A.2. We combine the 18 control groups to serve as a large common control group. The event rate in the control group is $p_0 = 0.005$.

For case groups, we use the sample sizes in the APOC3-CHD data and randomly pick up 5, 10, 15 out of 18 groups to mimic small, moderate and large size of meta-analysis. The true event rate in the case group is set to be $p_1 = \{0.001, 0.005, 0.01\}$. To consider between-study variance, p_{1k} in each study is generated from a beta distribution where mean is p_1 and variance is the observed variance in the APOC3-CHD data. Number of events are then generated from a binomial distribution with the sample size n_{k1} and event rate p_{k1}. Each simulation is replicated 1,000 times.

For comparison, we also construct CI for p_1 based on the DerSimonian and Laird method (DL) [13], the Paule and Mandel method (PM) [34], the Sidik-Jonkman method (SJ) [40], and the generalized linear mixed model with maximum likelihood estimator (GLMM) [43]. Figure 4.1 shows that exact methods are overall stable at the nominal level of 95% regardless of different sizes of K and p_1. When K is large, the exact method becomes slightly conser-

vative and its coverage probability is over nominal level. In contrast, other methods have lower coverage probabilities than the nominal level. Among them the SJ method has better performance with higher coverage probability.

We compute CI for the OR of each run of simulation. The average lower and upper bounds are shown in table 4.1. When $p_1 = 0.001$, the exact method, SJ method and GLMM method tend to have wider CI compared with the DL method and the PM method. When K increases, the length of CIs decreases for all methods.

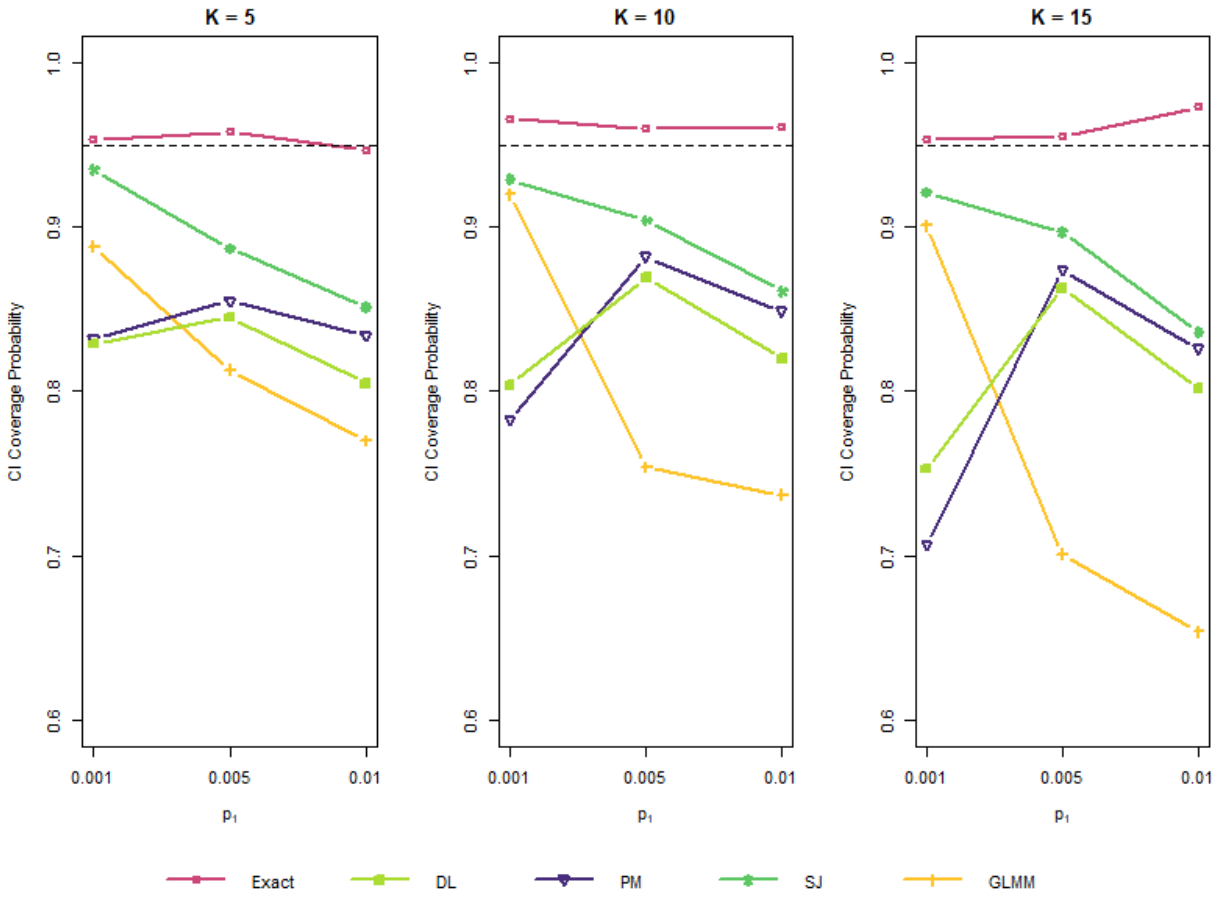

Figure 4.1: Plot of 95% CIs coverage probability for p_1 when $K = \{5, 10, 15\}$ and $p_1 = \{0.001, 0.005, 0.01\}$.

		$K = 5$		$K = 10$		$K = 15$	
		L	U	L	U	L	U
True OR: 0.199 $p_1 = 0.001$	Exact	(0.044,	0.788)	(0.033,	0.536)	(0.038,	0.445)
	DL	(0.130,	0.644)	(0.153,	0.491)	(0.170,	0.439)
	PM	(0.131,	0.639)	(0.158,	0.488)	(0.178,	0.439)
	SJ	(0.104,	0.706)	(0.126,	0.512)	(0.142,	0.448)
	GLMM	(0.064,	0.762)	(0.065,	0.376)	(0.067,	0.305)
True OR: 1 $p_1 = 0.005$	Exact	(0.086,	3.380)	(0.076,	2.134)	(0.100,	1.853)
	DL	(0.455,	2.165)	(0.526,	1.586)	(0.566,	1.430)
	PM	(0.414,	2.223)	(0.480,	1.607)	(0.524,	1.431)
	SJ	(0.379,	2.306)	(0.442,	1.645)	(0.489,	1.451)
	GLMM	(0.332,	1.986)	(0.345,	1.365)	(0.368,	1.190)
True OR: 2.010 $p_1 = 0.01$	Exact	(0.121,	6.171)	(0.105,	4.172)	(0.248,	3.587)
	DL	(0.851,	3.715)	(0.971,	2.855)	(1.056,	2.590)
	PM	(0.711,	4.022)	(0.817,	2.982)	(0.910,	2.646)
	SJ	(0.673,	4.101)	(0.782,	3.029)	(0.874,	2.677)
	GLMM	(0.643,	3.588)	(0.670,	2.616)	(0.732,	2.303)

Table 4.1: Average lower (L) and upper (U) bounds of CIs for OR from 1,000 replications.

4.3. Summary and Discussion

In this chapter, we consider a special study design where researchers of each study only recruit cases but not controls. Instead, a large survey database of the general population is used as the common controls. We propose to first apply the exact method to construct CI for the event rate in the case groups and then make inference on the pooled effect size measures, e.g. RD and OR, compared to the common controls. The exact method introduces beta-binomial distribution to build up a random-effect model where the true event rate in each study can vary from each other but with one underlying mean, and number of observed events are assumed to follow a binomial distribution. The exact method calculates p-values

at all search grids of possible combinations of event rate p_1 and between-study variance τ^2. The algorithm is further accelerated exploiting the relationship between p_1 and τ^2.

We perform simulations to evaluate the performance of the proposed exact method in the case of rare event rates. The simulation settings are based on a real dataset to mimic realistic scenarios. The simulation results suggest that the exact method shows stable performance regardless of the parameter settings. However, it is slightly conservative, in particular, when number of studies K becomes large. In sum, considering its computational cost, we recommend applying it when the number of studies is small and the event rate is rare.

APPENDIX A

APPENDIX of CHAPTER 2

A.1. Simulation Results for Upper Bounds $p^u_{kg0} = \{0.1\%, 1\%\}$

A.1.1. Coverage Probability and Average Length

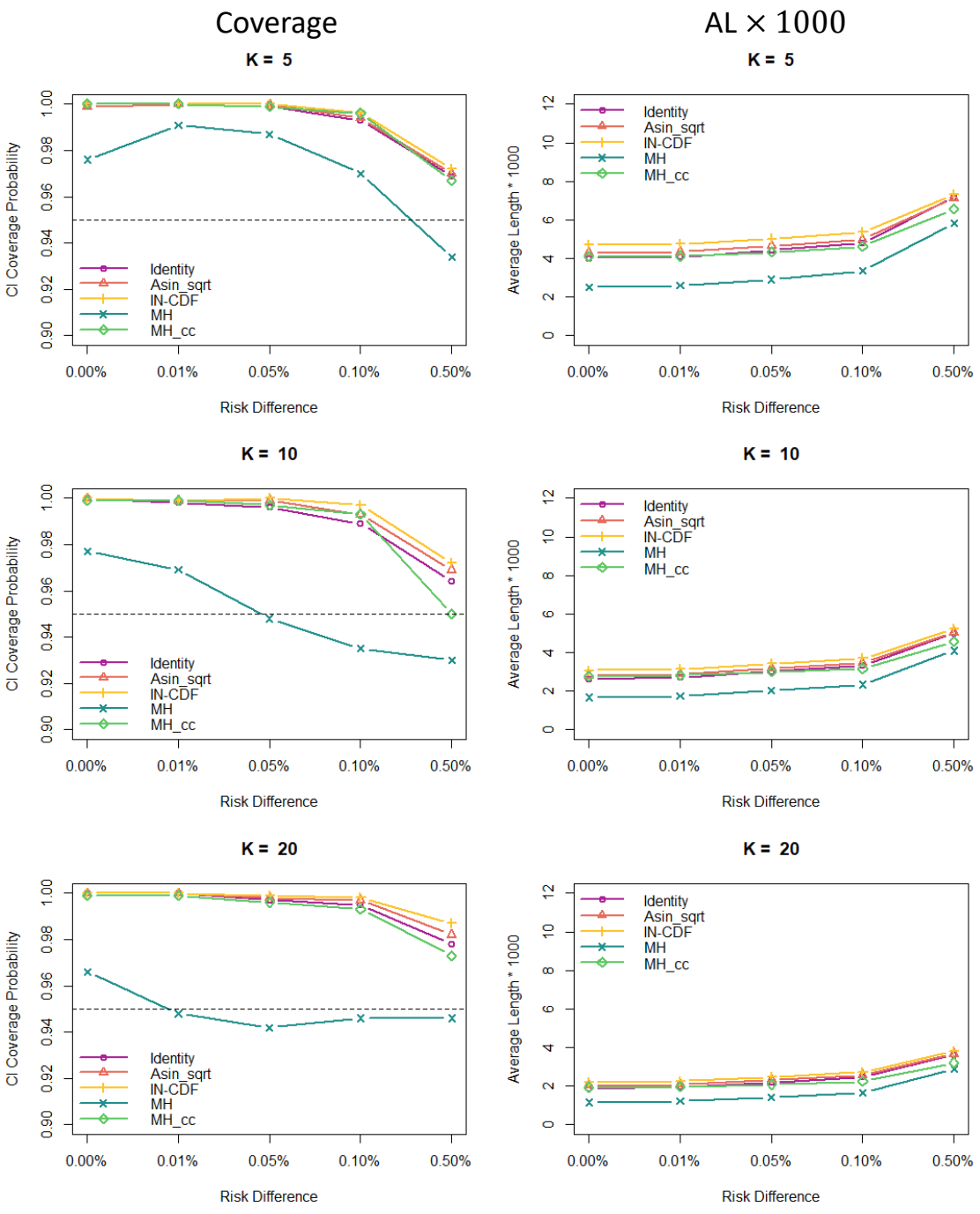

Figure A.1: Coverage and average length when sample sizes in control and case groups are balanced and $p^u_{kg0} = 0.1\%$.

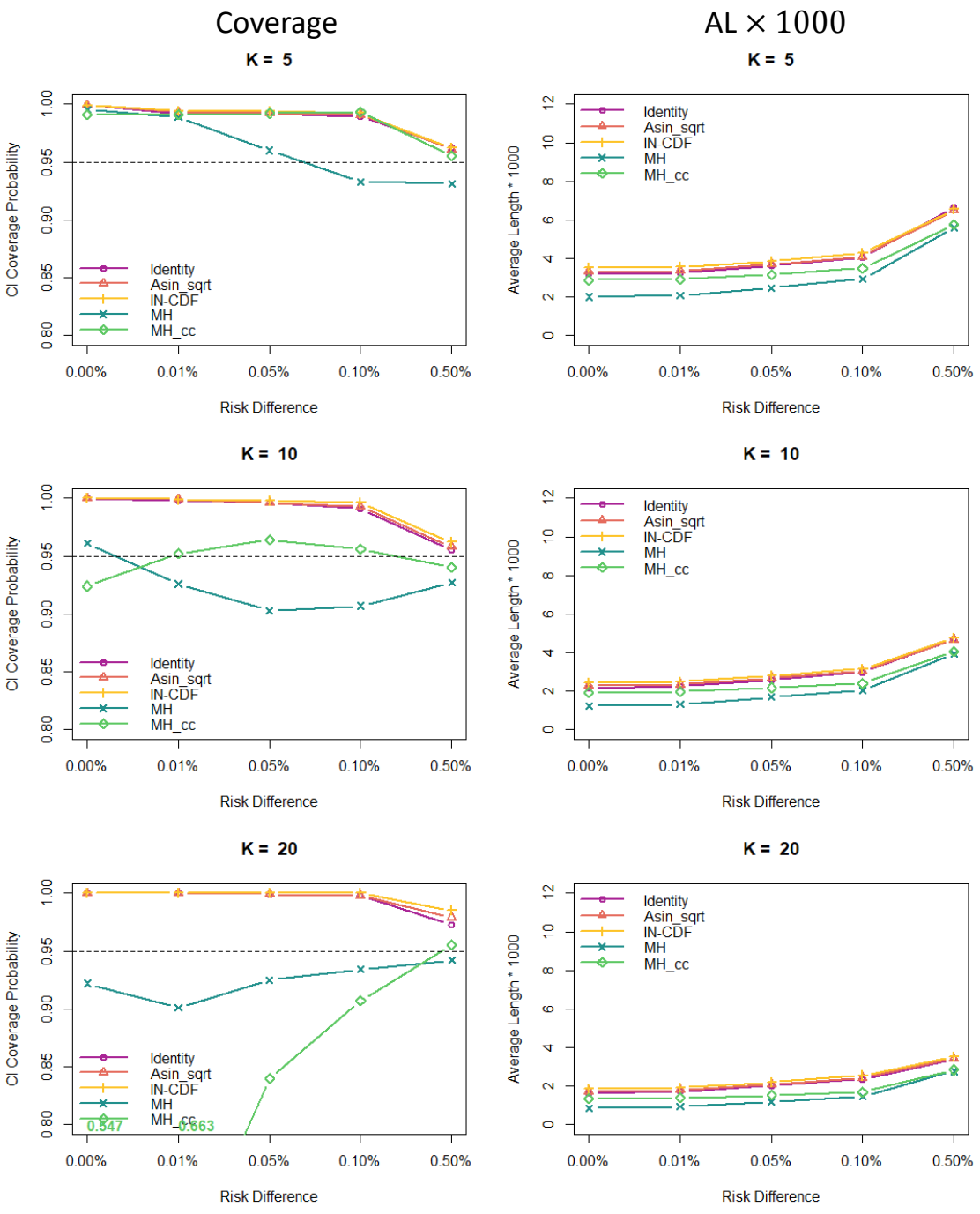

Figure A.2: Coverage and average length when sample sizes in control and case groups are unbalanced and $p^u_{kg0} = 0.1\%$.

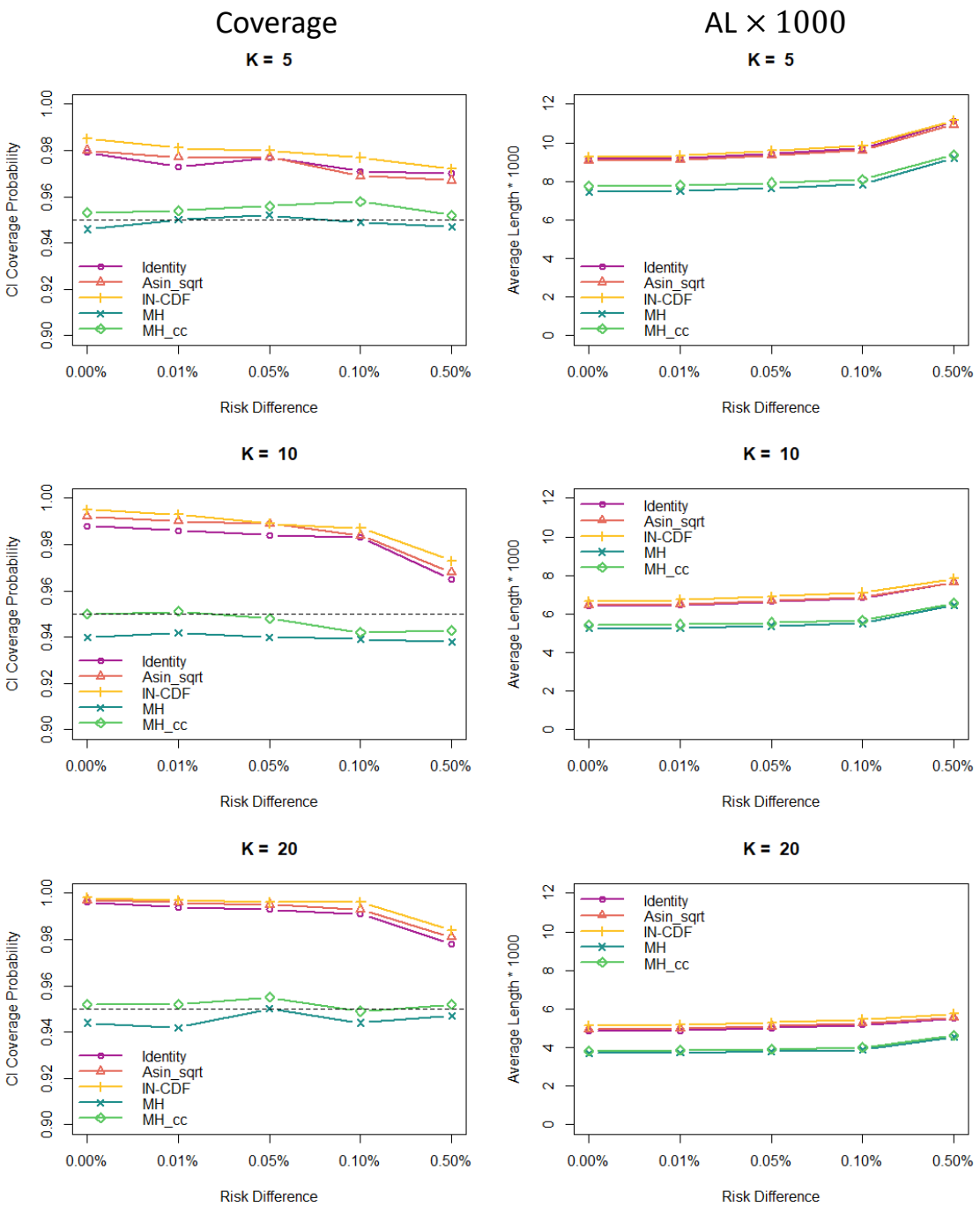

Figure A.3: Coverage and average length when sample sizes in control and case groups are balanced and $p^u_{kg0} = 1\%$.

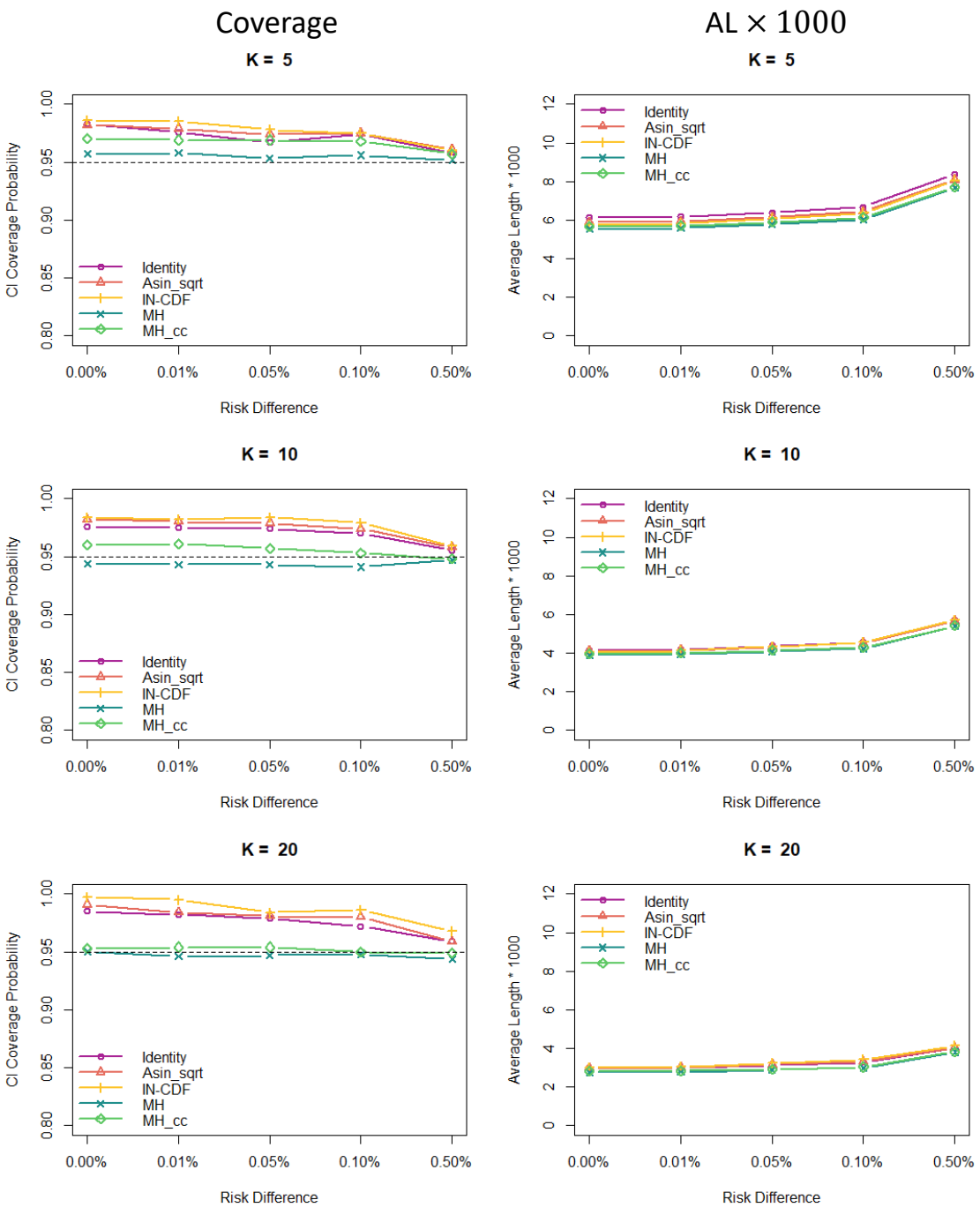

Figure A.4: Coverage and average length when sample sizes in control and case groups are unbalanced and $p^u_{kg0} = 1\%$.

A.1.2. Power and Average Absolute Bias

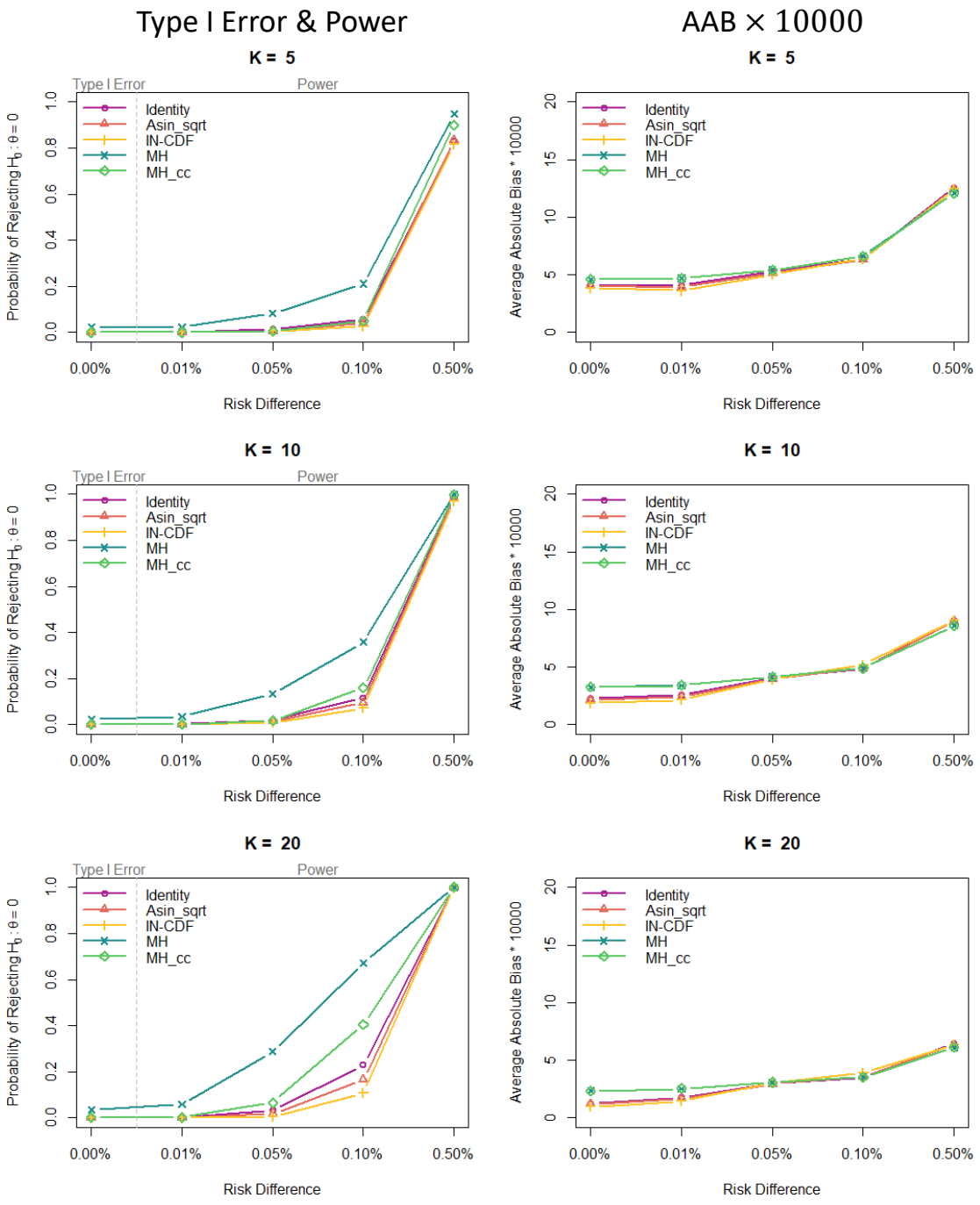

Figure A.5: Power and AAB when sample sizes in control and case groups are balanced and $p^u_{kg0} = 0.1\%$.

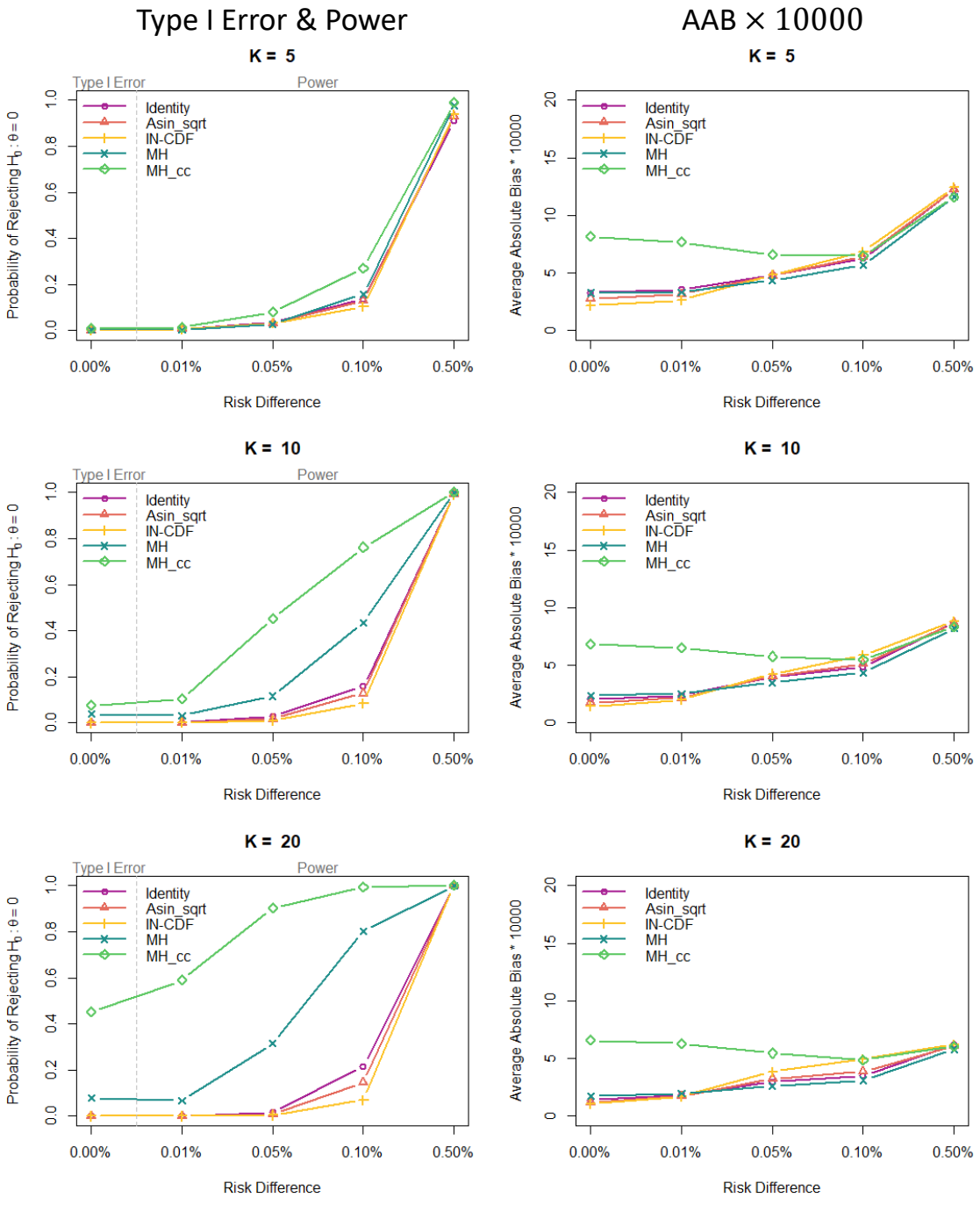

Figure A.6: Power and AAB when sample sizes in control and case groups are unbalanced and $p^u_{kg0} = 0.1\%$.

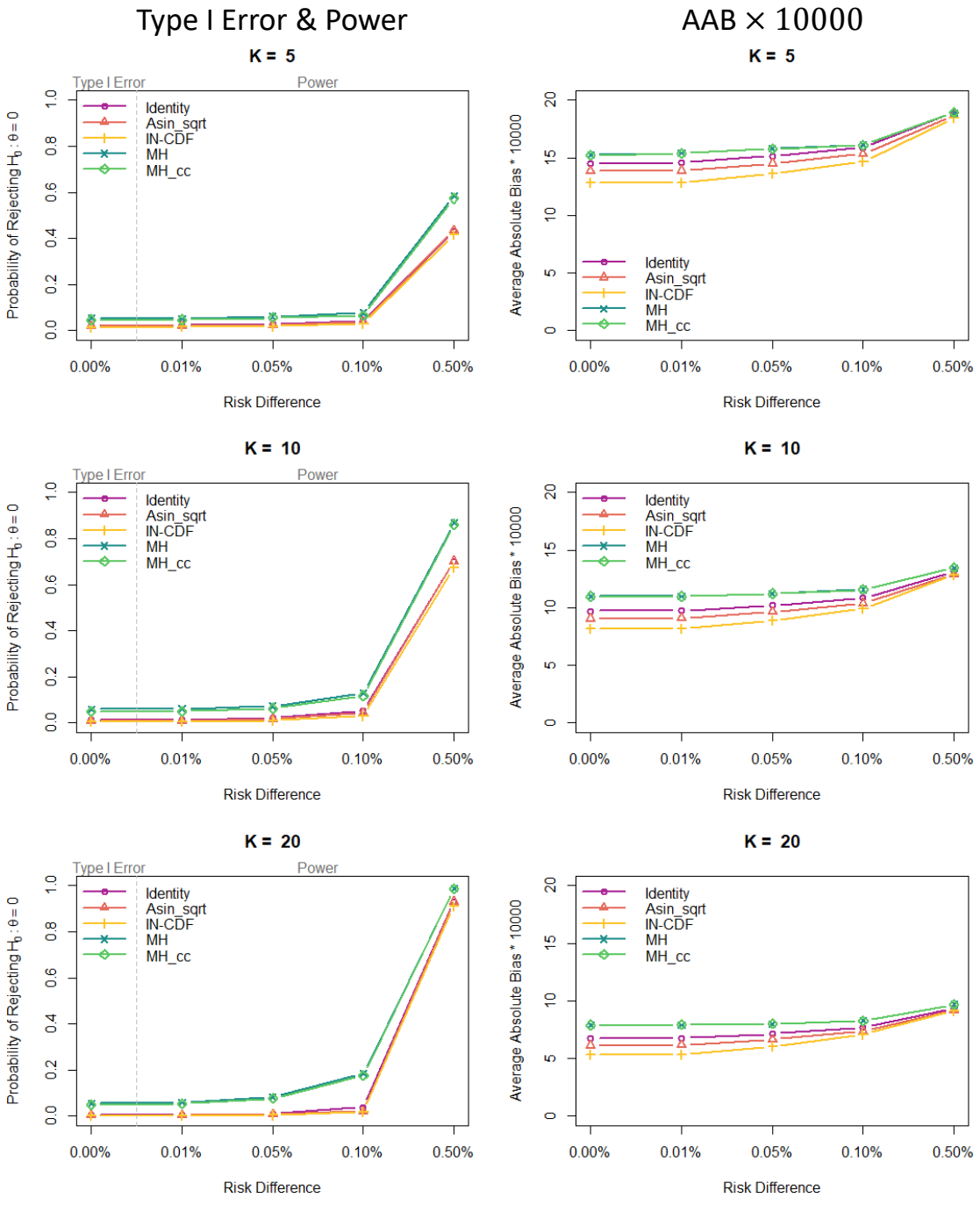

Figure A.7: Power and AAB when sample sizes in control and case groups are balanced and $p^u_{kg0} = 1\%$.

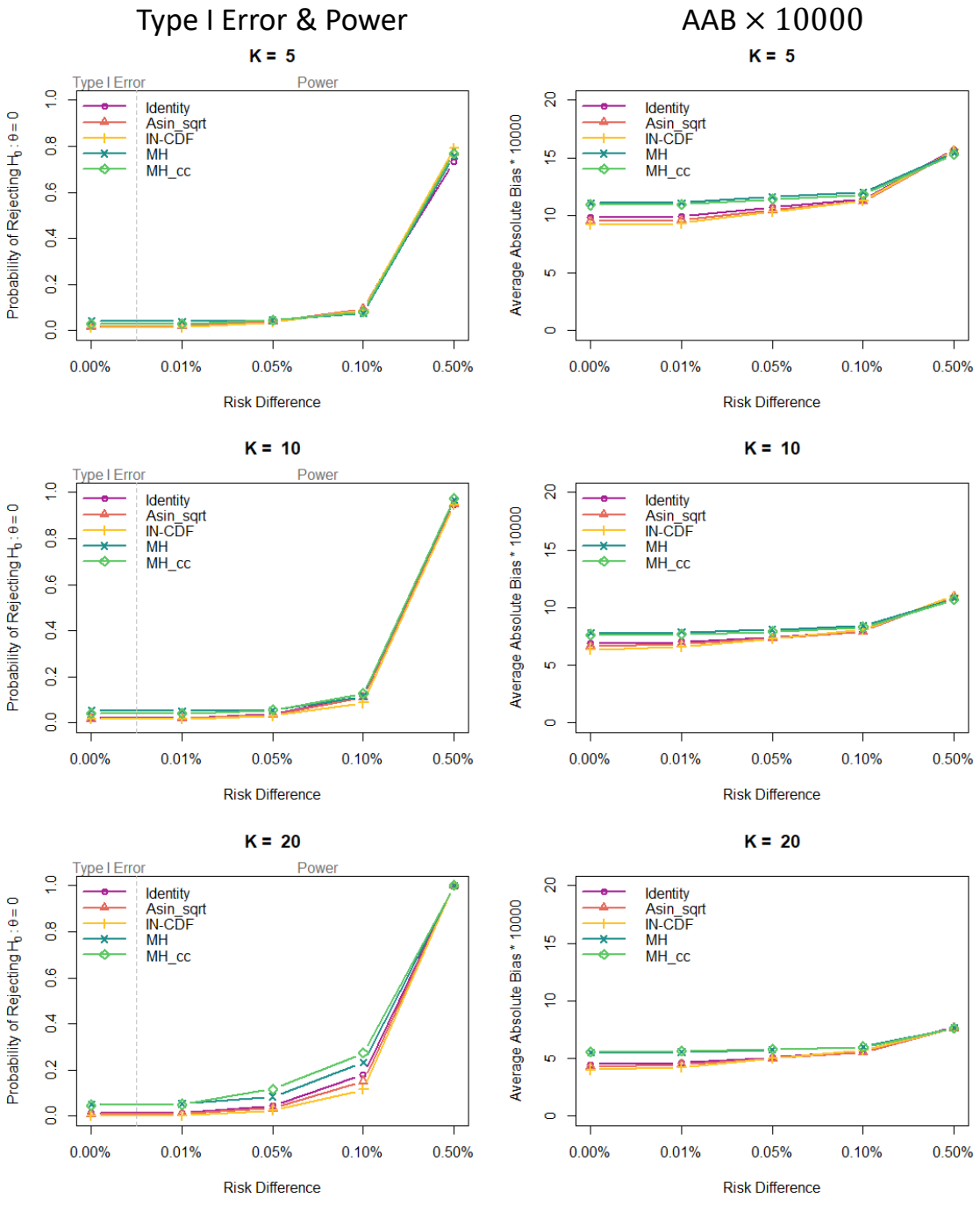

Figure A.8: Power and AAB when sample sizes in control and case groups are unbalanced and $p^u_{kg0} = 1\%$.

A.2. Data Structure of APOC3

		APOC3 Carriers		Total		Frequency	
Study	**Ancestry**	**Case**	**Control**	**Case**	**Control**	**Case**	**Control**
Study 1 - WHI	EA	6	89	2418	14098	0.0025	0.0063
Study 1 - WHI	AA	0	14	126	2263	0	0.0062
Study 2 - FHS	EA	0	8	126	3482	0	0.0023
Study 3 - MDC-CDA	EA	2	16	341	4523	0.0059	0.0035
Study 4 - ARIC	EA	2	16	1794	8539	0.0011	0.0019
Study 4 - ARIC	AA	8	19	564	3148	0.0142	0.0060
Study 5 - IPM	EA	10	33	703	1729	0.0142	0.0191
Study 5 - IPM	HA	2	13	1055	3478	0.0019	0.0037
Study 5 - IPM	AA	3	28	556	3240	0.0054	0.0086
Study 6 & 7 - ATVB + VHS	EA	9	16	1604	1233	0.0056	0.0130
Study 8 - Ottawa	EA	3	19	1024	2267	0.0029	0.0084
Study 9 - PROCARDIS	EA	10	16	2436	2179	0.0041	0.0073
Study 10 - HUNT	EA	6	7	2897	2906	0.0021	0.0024
Study 11 - GoDARTS CAD	EA	0	5	1694	2874	0	0.0017
Study 12 - EPIC CAD	EA	2	10	1396	7168	0.0014	0.0014
Study 13 - FIA3	EA	0	8	2657	2120	0	0.0038
Study 14 - German CAD	EA	37	41	9718	5810	0.0038	0.0071
Study 15 - WTCCC	EA	13	27	2893	5911	0.0045	0.0046
Total		**113**	**385**	**34,002**	**76,968**	**0.00332**	**0.00500**

Table A.2: Data Structure of APOC3 data. EA is European ancestry; AA is African American; HA is Hispanic ancestry. Total frequency is calculated as a simple rate of total event over total sample size.

A.3. Data Structure of SCARB1 P376L

	P376L Carriers		Total		Frequency	
Study or Consortium	**CHD Case**	**Control**	**CHD Case**	**Control**	**CHD Case**	**Control**
CARDIoGRAM Exome Consortium						
Study 1 - BioVu	6	10	4587	16546	0.0013	0.0006
Study 2 - BHF	1	0	2833	5912	0.0004	0
Study 3 - GoDARTS-CAD	1	0	1568	2772	0.0006	0
Study 4 - MHI	0	4	2483	8085	0	0.0005
Study 5 - North German	0	1	4464	2886	0	0.0004
Study 6 - Ottawa	0	1	1024	2267	0	0.0004
Study 7 - PAS	1	1	728	808	0.0014	0.0012
Study 8 - Penn	3	0	683	156	0.0044	0
Study 9 - South German	4	0	5255	2921	0.0008	0
Study 10 - WHI-EA	8	29	2860	14929	0.0028	0.0019
CHD Exome+ Consortium						
Study 11 - CCHS	1	1	2020	6087	0.0003	0.0001
Study 12 - CIHDS/CGPS	4	3	8079	10367	0.0003	0.0001
Study 13 - EPIC-CVD	4	2	9810	10970	0.0002	0.0001
Study 14 - MORGAM	0	0	2153	2118	0	0
Study 15 - PROSPER	1	0	640	638	0.0008	0
Study 16 - WOSCOPS	0	0	659	687	0	0
Total	**34**	**52**	**49,846**	**88,149**	**0.00068**	**0.00059**

Table A.4: Data Structure of SCARB1 P376L data. Total frequency is calculated as a simple rate of total event over total sample size.

APPENDIX B

APPENDIX of CHAPTER 3

B.1. Statistical Power Results in Simulation

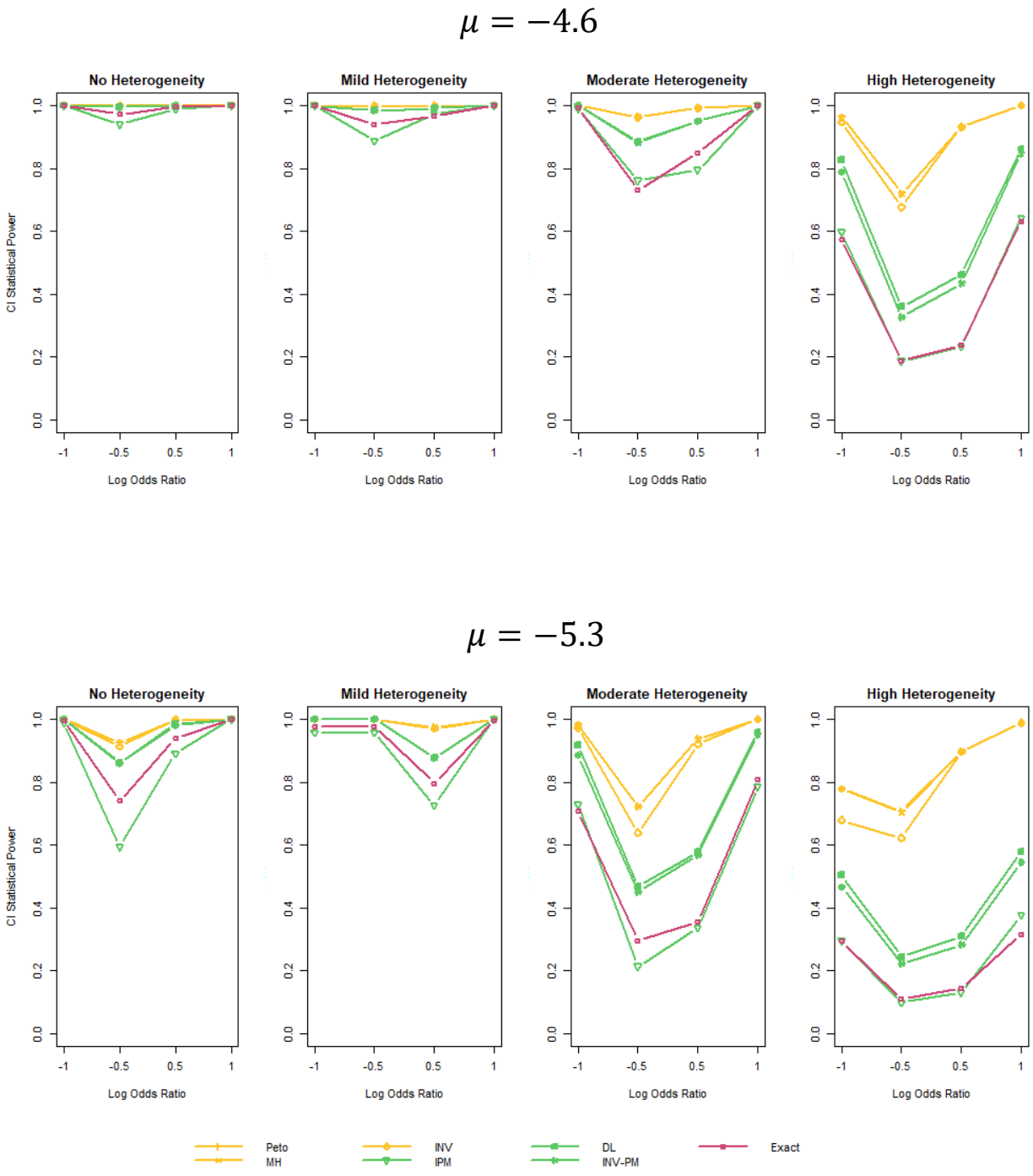

Figure B.1: Comparison of statistical power from proposed exact method and other six methods when $K = 5$.

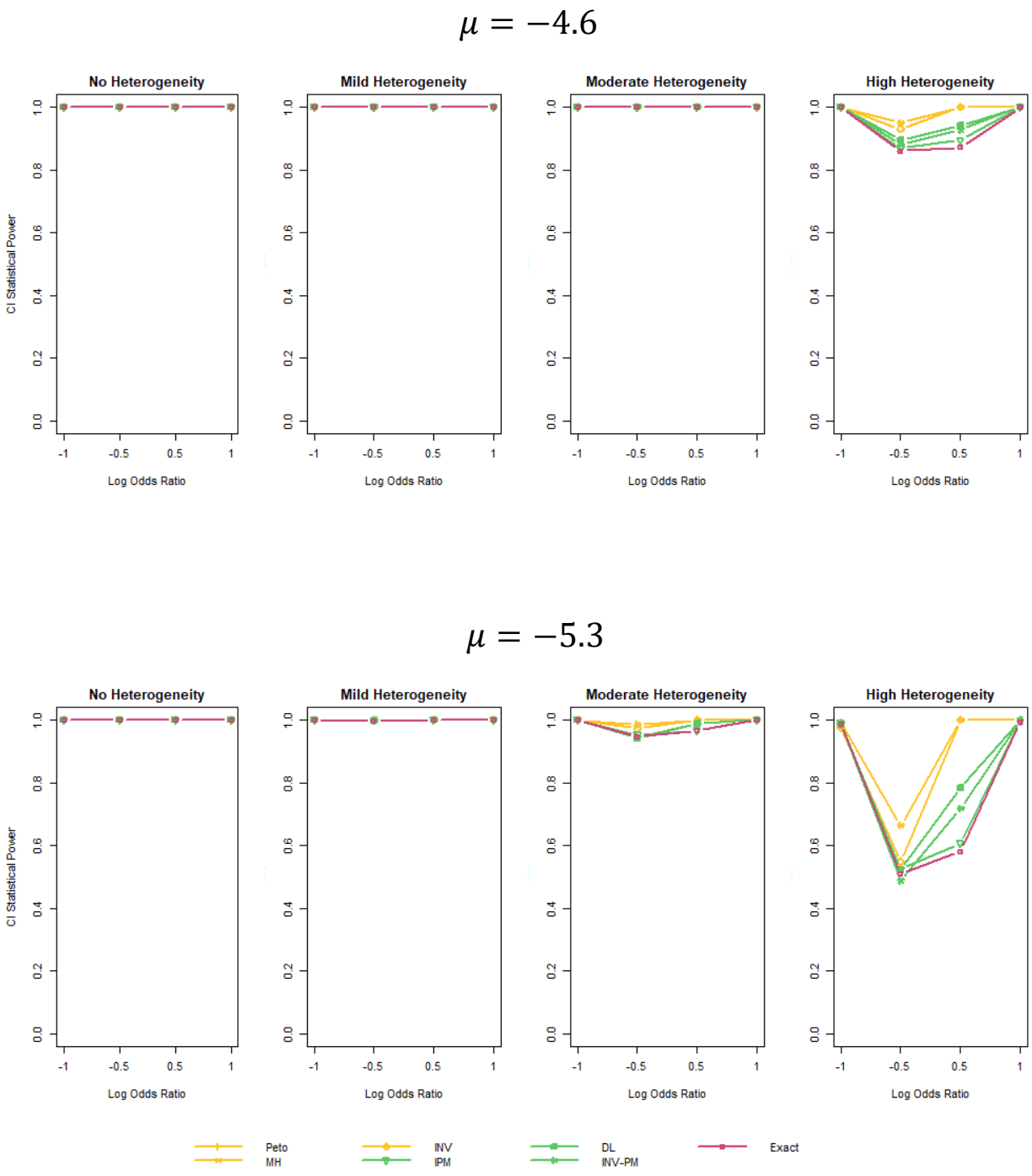

Figure B.2: Comparison of statistical power from proposed exact method and other six methods when $K = 15$.

B.2. Data Structure of NPC1L1

	Carriers		Total		Frequency (%)	
Cohort	**CHD Case**	**Control**	**CHD Case**	**Control**	**CHD Case**	**Control**
Study 1 - ARIC AA	1	7	474	2362	0.21	0.3
Study 1 - ARIC EA	3	7	2299	8656	0.13	0.08
Study 2 - ATVB	1	6	1794	1745	0.06	0.34
Study 3 - ESP EOMI	0	1	178	277	0	0.36
Study 4 - OHS	0	1	966	987	0	0.1
Study 5 - PROCARDIS	0	2	2098	2031	0	0.1
Study 6 - JHS	1	5	235	2016	0.43	0.25
Study 7 - Munich-MI	0	1	368	336	0	0.3
Study 8 - REGICOR	0	2	382	401	0	0.5
Study 9 - PROMIS	0	1	844	1107	0	0.09
Study 10 - BioVU	1	11	4587	16556	0.022	0.066
Study 11 - German North	0	1	4464	2886	0	0.03
Study 12 - German South	1	2	5255	2921	0.02	0.07
Study 13 - GoDARTS	0	4	997	2768	0	0.14
Study 14 - Mayo	0	2	1177	1492	0	0.13
Study 15 - WGHS	2	9	976	21641	0.21	0.042
Study 16 - WHI	1	9	2860	14958	0.03	0.06
Total	**11**	**71**	**29,954**	**83,140**	**0.037**	**0.085**

Table B.2: Data Structure of NPC1L1 data. Total frequency is calculated as a simple rate of total event over total sample size.

BIBLIOGRAPHY

[1] Allan Agresti. Dealing with discreteness: making 'exact' confidence intervals for proportions, differences of proportions, and odds ratios more exact. *Statistical Methods in Medical Research*, 12(1):3–21, 2003.

[2] Ou Bai, Min Chen, and Xinlei Wang. Bayesian estimation and testing in random effects meta-analysis of rare binary adverse events. *Statistics in biopharmaceutical research*, 8(1):49–59, 2016.

[3] Dulal K Bhaumik, Anup Amatya, Sharon-Lise T Normand, Joel Greenhouse, Eloise Kaizar, Brian Neelon, and Robert D Gibbons. Meta-analysis of rare binary adverse event data. *Journal of the American Statistical Association*, 107(498):555–567, 2012.

[4] Michael J Bradburn, Jonathan J Deeks, Jesse A Berlin, and A Russell Localio. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Statistics in medicine*, 26(1):53–77, 2007.

[5] Lawrence D Brown, T Tony Cai, and Anirban DasGupta. Interval estimation for a binomial proportion. *Statistical science*, pages 101–117, 2001.

[6] Tianxi Cai, Layla Parast, and Louise Ryan. Meta-analysis for rare events. *Statistics in medicine*, 29(20):2078–2089, 2010.

[7] Ivan SF Chan and Zhongxin Zhang. Test-based exact confidence intervals for the difference of two binomial proportions. *Biometrics*, 55(4):1202–1209, 1999.

[8] Yuzhou Chen, Yulia R Gel, Vyacheslav Lyubchich, and Todd Winship. Deep ensemble classifiers and peer effects analysis for churn forecasting in retail banking. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 373–385. Springer, 2018.

[9] Yeojin Chung, Sophia Rabe-Hesketh, and In-Hee Choi. Avoiding zero between-study variance estimates in random-effects meta-analysis. *Statistics in medicine*, 32(23):4071–4089, 2013.

[10] Yeojin Chung, Sophia Rabe-Hesketh, Vincent Dorie, Andrew Gelman, and Jingchen Liu. A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika*, 78(4):685–709, 2013.

[11] William G Cochran. Problems arising in the analysis of a series of similar experiments. *Supplement to the Journal of the Royal Statistical Society*, 4(1):102–118, 1937.

[12] William G Cochran. The combination of estimates from different experiments. *Biometrics*, 10(1):101–129, 1954.

[13] Rebecca DerSimonian and Nan Laird. Meta-analysis in clinical trials. *Controlled clinical trials*, 7(3):177–188, 1986.

[14] DAS Fraser. Statistical inference: Likelihood to significance. *Journal of the American Statistical Association*, 86(414):258–265, 1991.

[15] Sander Greenland and James M Robins. Estimation of a common effect parameter from sparse follow-up data. *Biometrics*, pages 55–68, 1985.

[16] Jessica Gronsbell, Chuan Hong, Lei Nie, Ying Lu, and Lu Tian. Exact inference for the random-effect model for meta-analyses with rare events. *Statistics in Medicine*, 39(3):252–264, 2020.

[17] Rebecca J Hardy and Simon G Thompson. A likelihood approach to meta-analysis with random effects. *Statistics in medicine*, 15(6):619–629, 1996.

[18] Julian PT Higgins, James Thomas, Jacqueline Chandler, Miranda Cumpston, Tianjing Li, Matthew J Page, and Vivian A Welch. *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons, 2019.

[19] Julian PT Higgins and Simon G Thompson. Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, 21(11):1539–1558, 2002.

[20] Julian PT Higgins and Simon G Thompson. Controlling the risk of spurious findings from meta-regression. *Statistics in medicine*, 23(11):1663–1682, 2004.

[21] Tania B Huedo-Medina, Julio Sánchez-Meca, Fulgencio Marin-Martinez, and Juan Botella. Assessing heterogeneity in meta-analysis: Q statistic or i^2 index? *Psychological methods*, 11(2):193, 2006.

[22] JT Gene Hwang and Ming-Chung Yang. An optimality theory for mid p–values in 2×2 contingency tables. *Statistica Sinica*, pages 807–826, 2001.

[23] Myocardial Infarction Genetics Consortium Investigators. Inactivating mutations in npc1l1 and protection from coronary heart disease. *New England Journal of Medicine*, 371(22):2072–2082, 2014.

[24] Michael J. Sweeting, Alexander J. Sutton, and Paul C. Lambert. What to add to nothing? use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in medicine*, 23(9):1351–1375, 2004.

[25] Konrad J Karczewski, Laurent C Francioli, Grace Tiao, Beryl B Cummings, Jessica Alföldi, Qingbo Wang, Ryan L Collins, Kristen M Laricchia, Andrea Ganna, Daniel P Birnbaum, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809):434–443, 2020.

[26] H Oliver Lancaster. Significance tests in discrete distributions. *Journal of the American Statistical Association*, 56(294):223–234, 1961.

[27] Lie Li and Xinlei Wang. Meta-analysis of rare binary events in treatment groups with unequal variability. *Statistical methods in medical research*, 28(1):263–274, 2019.

[28] Dungang Liu, Regina Y Liu, and Min-ge Xie. Exact meta-analysis approach for discrete data and its application to 2× 2 tables with rare events. *Journal of the American Statistical Association*, 109(508):1450–1465, 2014.

[29] Sifan Liu, Lu Tian, Steve Lee, and Min-ge Xie. Exact inference on meta-analysis with generalized fixed-effects and random-effects models. *Biostatistics & Epidemiology*, 2(1):1–22, 2018.

[30] Nathan Mantel and William Haenszel. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the national cancer institute*, 22(4):719–748, 1959.

[31] Haben Michael, Suzanne Thornton, Minge Xie, and Lu Tian. Exact inference on the random-effects model for meta-analyses with few studies. *Biometrics*, 75(2):485–493, 2019.

[32] Carl N Morris. Parametric empirical bayes inference: theory and applications. *Journal of the American statistical Association*, 78(381):47–55, 1983.

[33] Putri W Novianti, Kit CB Roes, and Ingeborg van der Tweel. Estimation of between-trial variance in sequential meta-analyses: a simulation study. *Contemporary clinical trials*, 37(1):129–138, 2014.

[34] Robert C Paule and John Mandel. Consensus values and weighting factors. *Journal of Research of the National Bureau of Standards*, 87(5):377–385, 1982.

[35] Stephen W Raudenbush. Analyzing effect sizes: Random-effects models. *The handbook of research synthesis and meta-analysis*, 2:295–316, 2009.

[36] Benjamin Rouse, Anna Chaimani, and Tianjing Li. Network meta-analysis: an introduction for clinicians. *Internal and emergency medicine*, 12(1):103–111, 2017.

[37] Andrew L Rukhin, Brad J Biggerstaff, and Mark G Vangel. Restricted maximum likelihood estimation of a common mean and the mandel–paule algorithm. *Journal of Statistical Planning and Inference*, 83(2):319–330, 2000.

[38] Hardeo Sahai and Anwer Khurshid. Statistics in epidemiology: methods, techniques and applications. 1995.

[39] Jonathan J Shuster and Michael A Walker. Low-event-rate meta-analyses of clinical trials: implementing good practices. *Statistics in medicine*, 35(14):2467–2478, 2016.

[40] Kurex Sidik and Jeffrey N Jonkman. A simple confidence interval for meta-analysis. *Statistics in medicine*, 21(21):3153–3159, 2002.

[41] Kurex Sidik and Jeffrey N Jonkman. Simple heterogeneity variance estimation for meta-analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(2):367–384, 2005.

[42] Kurex Sidik and Jeffrey N Jonkman. A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in medicine*, 26(9):1964–1981, 2007.

[43] Theo Stijnen, Taye H Hamza, and Pinar Özdemir. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in medicine*, 29(29):3046–3067, 2010.

[44] Alex J Sutton, Keith R Abrams, David R Jones, David R Jones, Trevor A Sheldon, and Fujian Song. *Methods for meta-analysis in medical research*, volume 348. Wiley Chichester, 2000.

[45] Alexander J Sutton and Julian PT Higgins. Recent developments in meta-analysis. *Statistics in medicine*, 27(5):625–650, 2008.

[46] TG, Lung HDL Working Group of the Exome Sequencing Project, National Heart, and Blood Institute. Loss-of-function mutations in apoc3, triglycerides, and coronary disease. *New England Journal of Medicine*, 371(1):22–31, 2014.

[47] Simon G Thompson and Julian PT Higgins. How should meta-regression analyses be undertaken and interpreted? *Statistics in medicine*, 21(11):1559–1573, 2002.

[48] Simon G Thompson and Stephen J Sharp. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in medicine*, 18(20):2693–2708, 1999.

[49] Lu Tian, Tianxi Cai, Marc A Pfeffer, Nikita Piankov, Pierre-Yves Cremieux, and LJ Wei. Exact and efficient inference procedure for meta-analysis and its application to the analysis of independent 2× 2 tables with all available data but without artificial continuity correction. *Biostatistics*, 10(2):275–281, 2008.

[50] Gert van Valkenhoef, Guobing Lu, Bert de Brock, Hans Hillege, AE Ades, and Nicky J Welton. Automating network meta-analysis. *Research synthesis methods*, 3(4):285–299, 2012.

[51] Areti Angeliki Veroniki, Dan Jackson, Wolfgang Viechtbauer, Ralf Bender, Jack Bowden, Guido Knapp, Oliver Kuss, Julian PT Higgins, Dean Langan, and Georgia Salanti. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research synthesis methods*, 7(1):55–79, 2016.

[52] Wolfgang Viechtbauer. Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, 30(3):261–293, 2005.

[53] Yan Wang and Lu Tian. An efficient numerical algorithm for exact inference in meta analysis. *Journal of statistical computation and simulation*, 88(4):646–656, 2018.

[54] Min-ge Xie and Kesar Singh. Confidence distribution, the frequentist distribution estimator of a parameter: A review. *International Statistical Review*, 81(1):3–39, 2013.

[55] Salim Yusuf, Richard Peto, John Lewis, Rory Collins, and Peter Sleight. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Progress in cardiovascular diseases*, 27(5):335–371, 1985.

[56] Paolo Zanoni, Sumeet A Khetarpal, Daniel B Larach, William F Hancock-Cerutti, John S Millar, Marina Cuchel, Stephanie DerOhannessian, Anatol Kontush, Praveen Surendran, Danish Saleheen, et al. Rare variant in scavenger receptor bi raises hdl cholesterol and increases risk of coronary heart disease. *Science*, 351(6278):1166–1171, 2016.

[57] Chiyu Zhang, Min Chen, and Xinlei Wang. Statistical methods for quantifying between-study heterogeneity in meta-analysis with focus on rare binary events. *Statistics and Its Interface*, 13(4):449–464, 2020.