Transformation bias in mixed effects models of meta-analysis

Ilyas Bakbergenuly

Doctor of Philosophy

July 4, 2017

Transformation bias in mixed effects models of meta-analysis

Ilyas Bakbergenuly

Doctor of Philosophy

University of East Anglia

School of Computing Sciences

July 4, 2017

Abstract

When binary data exhibit the greater variation than expected, the statistical methods have to account for extra-binomial variation. Possible explanations for extra-binomial variation include intra-cluster dependence or the variability of binomial probabilities. Both of these reasons lead to overdispersion of binomial counts and the resulting heterogeneity in their meta-analysis. Variance stabilizing or normalizing transformations are often applied to binomial counts to enable the use of standard methods based on normality. In meta-analysis, this is routinely done for the inference on overall effect measure. However, these transformations might result in biases in the presence of overdispersion. We study biases arising in the result of transformations of binary variables in the random or mixed effects models. We demonstrate considerable biases arising from standard log-odds and arcsine transformations both for single studies and in meta-analysis. We also explore possibilities of bias correction. In meta-analysis, the heterogeneity of the log odds ratios across the studies is usually incorporated by standard (additive) random effects model (REM). An alternative, multiplicative random effects model is based on the concept of an overdispersion. The multiplicative factor in this overdispersed random effects model can be interpreted as an intra-class correlation parameter. This model arises when one or both binomial distributions in the 2 by 2 tables are changed to beta-binomial distributions. The Mantel-Haenzsel and inverse-variance approaches are extended to this setting. The estimation of the random effect parameter is based on profiling the modified Breslow-Day test and improving the approximation for distribution of Q statistic in Mandel-Paule method. The biases and coverages from new methods are compared to standard methods through simulation studies. The misspecification of the REM in respect to the mechanism of its generation is an important issue which is also discussed in this thesis.

Table of Contents

List of Tables

List of Figures

List of Publications

- Bakbergenuly, I., Kulinskaya, E. and Morgenthaler, S., 2016. Inference for binomial probability based on dependent Bernoulli random variables with applications to meta-analysis and group level studies. Biometrical Journal, 58(4), pp.896-914.

- Bakbergenuly, I. and Kulinskaya, E., 2017. Beta-binomial model for meta-analysis of odds ratios. Statistics in Medicine, 36(11), pp.1715-1734.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Prof. Elena Kulinskaya for the continuous support of my Ph.D. study and related research, for her patience, motivation, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor and mentor for my Ph.D. study. Also, I would like to thank my mom for supporting me spiritually throughout writing this thesis and my life in general.

Chapter 1

Introduction

Meta-analysis of binary data stratified into 2×2 contingency tables plays a significant role in combining both retrospective and prospective studies such as case-control and randomized controlled trials respectively. Randomized controlled trials are often used to assess the effectiveness of a particular treatment within a given population by so-called effect measure. For example in clinical trials, the main goal is to assess the effectiveness of a medical intervention. In the meta-analysis of Randomized Controlled Trials (RCT), the effect measure for the binary data is often based on two-sample statistics such as odds ratio, risk ratio, and risk difference. Additionally, Cohen (1988) suggests an effect measure based on the difference between arcsine transformations of the binomial proportions. Among these effect measures, the most popular measure is the odds ratio (OR). The relative risk is also common, especially in observational studies. The difference of arcsine transformations is an example of the effect measure with an attractive property of stable variance. A good summary of these effect measures is presented in the paper by Sánchez-Meca et al. (2003). All these measures are functions of probabilities of success in binary data. The standard methods for combining such effect measures are well described in the literature (Sutton et al., 2000).

The standard methods of meta-analysis are subdivided into fixed and random effects models. Fixed effect model assumes a homogeneity of outcomes across the studies. As a generalization of the fixed effect model, random effects model assumes some heterogeneity of outcomes across studies. The choice of model and associated methods can be determined by prior assumptions

of heterogeneity between studies or by tests of homogeneity. However, it is important to identify the sources of heterogeneity and attempt to explain the source by explanatory variables using the meta-regression methods.

This research is motivated by problems arising in overdispersed binomial data. Assume that in practice, there exist a correlation between Bernoulli variables that has to be taken into account. This correlation is called an intra-cluster correlation. Previously, the intra-cluster correlation has been studied in meta-analysis for both fixed and random effect models. In fixed effects model, the within-study variance is adjusted for intra-cluster correlation by a correction factor and the homogeneity of effects is assumed. In random effects model, in addition to adjustment for intra-cluster correlation, an extra between-study variance of a random effect is added to the fixed effect model. When random effects model is assumed, the model has two unknown variance components which have to be estimated. One corresponds to intra-cluster correlation and another to the between-study variance. The heterogeneity induced by intra-cluster correlation itself without an assumption of additional between-study variance has not been discussed in previous studies throughout the literature.

In this research, we consider the influence of intra-cluster correlation on the inference in a meta-analysis of binomial proportions in 2×2 contingency tables. Specifically, we concentrate on transformations of binomial proportions arising from dependent binary data.

In this thesis, we firstly revise the standard methods for combining binary data from 2×2 contingency tables - Chapter 2. We consider the effect measures based on transformations of proportions. Particularly, we are interested in normalizing and variance-stabilizing transformations of proportions. Secondly, we examine the heterogeneity induced by intra-cluster correlation between Bernoulli observations in each study. This results in the overdispersion of summary binomial outcomes and therefore in a heterogeneity of estimates of effect measures which are the functions of summary binomial outcomes.

Chapter 3 concentrates on the transformation bias in a single study and a meta-analysis of binomial proportions. In Chapter 3, we show that the introduction of dependency between

Bernoulli variables results in a bias of order $1/n$ for any non-linear transformation, where n is a sample size. Bias corrections are proposed for the arcsine and log-odds transformations. However, for the log-odds transformation, the bias correction itself is biased due to the dependence of weights on estimated proportions. Furthermore, we show that such biases of order $1/n$ have an impact on the inference in a meta-analysis of transformed binary proportions. Chapter 4 introduces a multiplicative random effects model for the logarithmic transformation of odds ratios. This model is based on the concept of overdispersion introduced by Kulinskaya and Olkin (2014). In standard additive random effects model, the overdispersion is quantified by an additional variance of the random effect. In our study, the intra-cluster correlation explains the overdispersion. In addition, new methods for estimation of the random effect parameter are developed. The validity of the new methods is studied through simulation with data generated from standard additive and proposed multiplicative random effects models. Based on the findings of Chapter 3 and Chapter 4, Chapter 5 studies the generalized linear mixed effect models in meta-analysis. The main models of interest are the conditional generalized linear mixed-effects models with an exact and an approximate likelihood. In these models, we discovered the problems of misspecification resulting from assumptions of a standard additive and a proposed multiplicative random effects models. Wrong assumptions lead to the wrong estimation of the random effect component in methods of standard random effects model. Therefore, in Chapter 6, we study the robustness of standard and new methods for estimation of between-study variance through a simulation study. The simulation study is performed across different types of generated data. In our simulation, we also include the recent bias-corrected maximum likelihood estimate of between-study variance proposed by Kosmidis et al. (2017). Chapter 7 summarizes the findings of Chapters 3-6, discusses the practical issues and provides suggestions for future research on transformation biases in meta-analysis.

1.1 Thesis outline and research objectives

- Describe the normalising and variance stabilizing transformations of proportions

- Review of the methods for fixed and random effects models in meta-analysis
- Discuss the issues in univariate meta-analysis of binary data
- Review the "state of the art" methods for what have been done to resolve the present issues
- Propose the multiplicative random effects model with a pair of beta-binomial distributions for overdispersed binary data
- Adjust the standard methods based on the additive random effects model for estimation of intra-cluster correlation
- Develop new methods for an estimation of intra-cluster correlation in the multiplicative random effects model
- Develop new methods for an estimation of between-study variance in the additive random effects model
- Generalize the Mantel-Haenzsel and Inverse-Variance methods for combining odds ratios to the setting of the multiplicative random effects model
- Study the impact of intra-cluster correlation in binary data causing heterogeneity in contingency tables
- Study the effect of intra-cluster correlation on transformations of proportions
- Propose the bias correction for arcsine and log-odds transformations
- Study the transformation bias in generalized linear mixed effects models for meta-analysis
- Study the transformation bias in maximum likelihood inference in meta-analysis

- Study the robustness of standard and new methods for an estimation of between-study variance in the meta-analysis of binary data
- Study the misspecification of random effects models

1.2 Terminology

Effect measure, bias of order $1/n$, arcsine transformation, log-odds transformation, random effects model (REM), intra-cluster correlation, heterogeneity, mantel-haenzsel method, breslow-day test, mandel-paule method, overdispersion model (ODM), beta-binomial distribution, generalized linear mixed effects models (GLMM), maximum likelihood (ML), penalized maximum likelihood (PML) score function.

Chapter 2

Meta-analysis of binary data

2.1 Introduction

Meta-analysis is a statistical technique for synthesizing the outcomes from several studies. Originally, the purpose of meta-analysis was to combine the evidence from published observational or experimental studies. At the present day, meta-analysis is also concerned with exploring and explaining the heterogeneity between different studies. The earliest example of meta-analysis is given by Simpson and Pearson (1904). They studied the effectiveness of typhoid vaccine on mortality. In total 11 studies were included in their meta-analysis. The correlation between the vaccine inoculation and typhoid mortality was the main parameter of interest (the effect measure). Later, Tippett (1931) and Fisher (1932) discussed the method for combining p-values from significance tests of a null hypothesis common to all studies. The term meta-analysis was introduced to statistical theory by Glass (1976). Statistical foundations of combining studies in a meta-analysis are discussed by Hedges and Olkin (1985). Apart from combining summary statistics at the study level, the regression-based methods can be used in a meta-analysis to explain the heterogeneity using possible predictors at the study level.

The methods of meta-analysis are generally divided into the fixed or random effects models. The fixed effect model assumes homogeneous effect measures across all studies. The random effects model allow expected effect measures to differ from one study to another according to some distribution. Cochran (1937) and Yates and Cochran (1938) discuss the similarities of

meta-analysis with the analysis of variance (1-way ANOVA) developed by R.A. Fisher, where the combined effect measure is the overall mean and within-study variance of each effect is measured by residual variation. The only difference between the one-way analysis of variance and meta-analysis is that in meta-analysis each study has its own within-study variance and in the one-way analysis of variance, the variances are assumed to be homogeneous across responses.

A combined estimate of any effect measures and its estimated variance are primary interest in a meta-analysis. For continuous data, an effect measure can be the difference between means of control and treatment groups on testing a particular treatment. Absolute or relative differences in units of standard deviation can be used. The later is called the standardized mean difference. In a meta-analysis of binary data, the effect measures are the functions of probabilities for a particular event. The probabilities are known as "risks". Using these risks, the aim of meta-analysis is to compare the responses between control and treatment groups. Risk difference, relative risk and the odds ratio are the main statistics for comparing two binary outcomes from control and treatment arms in randomized controlled trials. Relative risk and odds ratio statistics can be log transformed for normalization. Another popular transformation of risks is the arcsine transformation. The arcsine transformation helps to stabilize the variance and therefore simplifies the weights in a meta-analysis. Comparing two outcomes, the difference between arcsine transformed risks can be used as an effect measure. Rücker et al. (2009) discusses the use of arcsine difference as a measure of intervention effect in the meta-analysis of binary data.

One of the issues in the use of these measures are the studies reporting zero events. In this case, the use of continuity corrections is usually suggested. The use of continuity corrections should be undertaken with care since the choice of continuity correction might influence the conclusions about the overall effect measure.

Meta-analysis can be performed for comparative or non-comparative experiments. For comparative experiments, a comparative effect measure such as odds ratio is required. For non-comparative experiments, each study may provide an estimate of the probability or the odds of an event from a single treatment group and these probabilities or odds may be combined on arcsine scale or on the log-scale.

Meta-analysis of the binary data has become popular in various research areas including epidemiology, medical and social research. The binary data from each study can be presented as a contingency table. Meta-analytic methods allow combining several studies taking their accuracy into account. Often, a single study does not have enough power to make general inference about the tested intervention for dichotomous outcomes. Meta-analysis is about to combine separate effects from underpowered studies and construct an overall effect measure with adequate power.

In this chapter, we provide a review of statistical methods for combining data from contingency tables. The log-odds and arcsine transformations are of the main interest to us.

2.2 Generation of binary data

Let X_{ijk} be a Bernoulli random variable which takes on the values of 1 (success) or 0 (failure). The subscript ijk refers to the Bernoulli variable k in particular group j of the study i in the meta-analysis of binary data in the following subsections. Let the probability distribution of X_{ijk} be $P(X_{ijk} = 1) = 1 - P(X_{ijk} = 0) = p_{ij}$ which can be re-expressed in general form $P(X_{ijk} = r) = p_{ij}^r(1 - p_{ij})^{1-r}$ for $r = 0, 1$. First two standard moments of X_{ijk} are

$$\mathrm{E}(X_{ijk}) = p_{ij} \qquad \text{and} \qquad \mathrm{Var}(X_{ijk}) = p_{ij}(1 - p_{ij}).$$

Now assume that there are n_{ij} Bernoulli variables $< X_{ijk}, k = 1, \ldots, n_{ij} >$ with X_{ij} successes, i.e $X_{ij} = \sum_{k=1}^{n_{ij}} X_{ijk}$. Then, the expected value of X_{ij} is

$$\mathrm{E}(X_{ij}) = \mathrm{E}(\sum_{k=1}^{n_{ij}} X_{ijk}) = \sum_{k=1}^{n_{ij}} \mathrm{E}(X_{ijk}) = n_{ij}p_{ij},$$

and the variance is

$$\mathrm{Var}(X_{ij}) = \mathrm{Var}(\sum_{k=1}^{n_{ij}} X_{ijk}) = \sum_{k=1}^{n_{ij}} \mathrm{Var}(X_{ijk}) + \sum_{k=1}^{n_{ij}} \sum_{k \neq k'} \mathrm{Cov}(X_{ijk} X_{ijk'}). \tag{2.2.1}$$

When $\mathrm{Cov}(X_{ijk} X_{ijk'}) = 0$, Bernoulli variables $< X_{ijk}, k = 1, \ldots, n_{ij} >$ are independent and variance (2.2.1) reduces to

$$\mathrm{Var}(X_{ij}) = \sum_{k=1}^{n_{ij}} p_{ij}(1 - p_{ij}) = n_{ij} p_{ij}(1 - p_{ij}) \tag{2.2.2}$$

and the probability distribution is the standard binomial distribution $B(n_{ij}, p_{ij})$:

$$P(X_{ij} = k) = \binom{n_{ij}}{k} p_{ij}^{k} (1 - p_{ij})^{n_{ij} - k}.$$

The variance of X_{ij} given by (2.2.2) depends on the mean with quadratic function which is concave and symmetrical around $p_{ij} = 0.5$. The distribution of X_{ij} is symmetric for $p_{ij} = 0.5$ and the variance is greatest in this case. Whenever the probability p_{ij} is either very high or very low the distribution of X_{ij} becomes negatively or positively skewed with low variance. Considering the estimates for proportions of successes $\hat{p}_{ij}$, the expectation and variance of these proportions are

$$\mathrm{E}(\hat{p}_{ij}) = p_{ij} \qquad \text{and} \qquad \mathrm{Var}(\hat{p}_{ij}) = \frac{p_{ij}(1 - p_{ij})}{n_{ij}}.$$

One of the difficulties with estimating proportions is that the responses are bounded between 0 and 1. Also, the variance of a binomially distributed random variable is not constant over the range of $P \in (0, 1)$. Due to these issues, variance stabilizing or normalizing transformations are often applied prior to modelling the binomial counts with the good of using the methods aimed at analysis of normally distributed data. The main objective of this research is to assess the behaviour of these transformations in the case of dependent Bernoulli responses. Another question that we consider is how this dependence induces the heterogeneity or overdispersion in the standard analysis. We develop meta-analytic methods appropriate for dependent binary data.

2.2.1 Overdispersion in binary data

Binary data might exhibit a greater variation then expected. One of the possible explanations for extra-binomial variation lies in intra-cluster dependence. This leads to overdispersion of binomial counts and the resulting heterogeneity in their meta-analysis. Williams (1982), Williams (1996) introduced a class of binomial mixture models to incorporate the overdispersion in binary data. Collett (1991) provides a practical introduction to models considered by Williams (1982) and Williams (1996). The latest overview of overdispersion theory is summarized in the paper by Xekalaki (2014).

Dependent Bernoulli variables are widely used in analysis of studies with repeated measurements on binary data. The most general assumptions about dependent Bernoulli variables $< X_{ijk}, k = 1, \ldots, n_{ij} >$ specify only the first two moments:

$$\mathrm{E}(X_{ijk}) = p_{ijk}, \quad \mathrm{Var}(X_{ijk}) = p_{ijk}(1 - p_{ijk}) \tag{2.2.3}$$

and

$$\mathrm{Cov}(X_{ijk}, X_{ijk'}) = \rho_{kk'}\sqrt{p_{ijk}(1 - p_{ijk})p_{ijk'}(1 - p_{ijk'})}, \tag{2.2.4}$$

where p_{ijk} is the probability of success for $X_{ijk}, k = 1, \ldots, n_{ij}$, and $\rho_{kk'} = \mathrm{corr}(X_{ijk}, X_{ijk'})$ is the correlation of X_{ijk} and $X_{ijk'}$, $k \neq k'$. The joint probability mass function for any two Bernoulli variables X_{ijk} and $X_{ijk'}$ is

$$P(X_{ijk} = 1, X_{ijk'} = 1) = p_{ijk}p_{ijk'} + \rho_{kk'}\sqrt{p_{ijk}(1 - p_{ijk})p_{ijk'}(1 - p_{ijk'})};$$

$$P(X_{ijk} = 1, X_{ijk'} = 0) = p_{ijk}(1 - p_{ijk'}) - \rho_{kk'}\sqrt{p_{ijk}(1 - p_{ijk})p_{ijk'}(1 - p_{ijk'})};$$

$$P(X_{ijk} = 0, X_{ijk'} = 1) = (1 - p_{ijk})p_{ijk'} - \rho_{kk'}\sqrt{p_{ijk}(1 - p_{ijk})p_{ijk'}(1 - p_{ijk'})};$$

$$P(X_{ijk} = 0, X_{ijk'} = 0) = (1 - p_{ijk})(1 - p_{ijk'}) + \rho_{kk'}\sqrt{p_{ijk}(1 - p_{ijk})p_{ijk'}(1 - p_{ijk'})}.$$

Since $\mathrm{E}(X_{ijk}X_{ijk'}) = P(X_{ijk} = 1, X_{ijk'} = 1)$ and $E(X_{ijk}) = p_{ijk}$,

$$\mathrm{Cov}(X_{ijk}, X_{ijk'}) = P(X_{ijk} = 1, X_{ijk'} = 1) - p_{ijk}p_{ijk'},$$

and the correlation is

$$\rho_{kk'} = \frac{P(X_{ijk} = 1, X_{ijk'} = 1) - p_{ijk}p_{ijk'}}{\sqrt{p_{ijk}(1 - p_{ijk})p_{ijk'}(1 - p_{ijk'})}}.$$

The probability mass functions above clearly show the relationship between the probabilities and the intra-cluster correlation.

As Emrich and Piedmonte (1991) show, the correlation values are restricted by the interval

$$\big(\max\big(-\sqrt{\frac{p_{ijk}p_{ijk'}}{q_{ijk}q_{ijk'}}}, -\sqrt{\frac{q_{ijk}q_{ijk'}}{p_{ijk}p_{ijk'}}}\big), \min\big(\sqrt{\frac{p_{ijk}q_{ijk'}}{q_{ijk}p_{ijk'}}}, \sqrt{\frac{q_{ijk}p_{ijk'}}{p_{ijk}q_{ijk'}}}\big)\big), \tag{2.2.5}$$

where $q_{ij} = 1 - p_{ij}$. The above mentioned bounds for correlations do not need to apply if the interest lies in an overdispersed binomial distribution for the number of successes, where a generation mechanism is not necessarily restricted to a sum of dependent Bernoullis. The intra-cluster correlation $\rho_{kk'}$ is a measure for the degree of similarity between any two Bernoulli responses within the same group of Bernoulli variables.

We concentrate on the simple model with constant probabilities $p_{ijk} = p_{ij}$ and correlations $\rho_{kk'} = \rho$. Assume that all pairs of responses have the same joint distribution for $X_{ijk} \cdot X_{ijk'}$, then for $k \neq k'$

$$\mathrm{E}(X_{ijk}) = \mathrm{E}(X_{ijk'}) = p_{ij}, \quad P(X_{ijk} = X_{ijk'} = 1) = p_{ij}^2 + p_{ij}(1 - p_{ij})\rho$$

$$P(X_{ijk} = X_{ijk'} = 0) = (1 - p_{ij})^2 + p_{ij}(1 - p_{ij})\rho$$

$$P(X_{ijk} = 0, X_{ijk'} = 1) = P(X_{ijk} = 1, X_{ijk'} = 0) = p_{ij}(1 - p_{ij})(1 - \rho).$$

The restriction on ρ then reduces to $\max(-p_{ij}/q_{ij}, -q_{ij}/p_{ij}) < \rho < 1$. Further restrictions may arise from particular data-generating distribution and/or latent variables.

Gulliford et al. (2005) discusses a relationship between the intra-cluster correlation ρ and the overall prevalence in dichotomous outcomes. Eldridge et al. (2009) provides a clear review of intra-cluster correlation coefficient for continuous and dichotomous outcomes in cluster randomized trials. In our case, dichotomous responses are our primary interest.

2.2.2 Generation of dependent Bernoullis

Available generation mechanisms of dependent Bernoulli random variables (r.v.'s) include normal or Archimedian copulas (Demirtas et al., 2009; Emrich and Piedmonte, 1991; Madsen and Birkes, 2013), and the method by Lunn and Davies (1998). In addition to summing dependent Bernoulli r.v.'s, generation mechanisms of overdispersed Binomial r.v.'s include the beta-binomial distribution, other Binomial mixtures (Qaqish et al., 2012) and methods based on sums of Poisson random variables (Demirtas et al., 2009).

Without further specification (beyond the first two moments), likelihood-based methods are not available, and Generalised Estimating Equations can be used for analysis, (Chaganty and Joe, 2004). A distribution needs to be specified to the third and forth moments to compare asymptotic efficiency of various methods.

The three methods used in the rest of this thesis to generate the overdispersed binomial data are described in more details below. These are the Gaussian Copula (GC) method by Emrich and Piedmonte (1991), the method by Lunn and Davies (1998) and the beta-binomial distribution.

Our main interest is not in the Bernoulli variables, but rather in their sums, i.e. total numbers of successes out of n trials. The total numbers of successes out of n trials is contingent on probabilities and correlations between Bernoulli variables. The dependence between variables which takes values either 0 or 1 is described through the intra-cluster correlation. Assuming $\mathrm{Cov}(X_{ijk}, X_{ijk'}) = p_{ij}(1-p_{ij})\rho$ and using (2.2.1) the variance of X_{ij} is

$$\mathrm{Var}(X_{ij}) = \mathrm{Var}(\sum_{k=1}^{n_{ij}} X_{ijk}) = n_{ij}p_{ij}(1-p_{ij}) + 2\binom{n_{ij}}{2}p_{ij}(1-p_{ij})\rho =$$

$$= n_{ij}p_{ij}(1-p_{ij}) + n_{ij}(n_{ij}-1)p_{ij}(1-p_{ij})\rho = n_{ij}p_{ij}(1-p_{ij})(1+(n_{ij}-1)\rho).$$

Comparing this variance to binomial variance (2.2.2), there is an additional inflation term $(1+(n_{ij}-1)\rho)$ for representation of overdispersion. The same results follows from the mixture of distributions. The distributions are usually mixed in two stages. Consider random variables P_{ij} which has a continuous distribution on $[0,1]$ with $\mathrm{E}(P_{ij}) = p_{ij}$ and $\mathrm{Var}(P_{ij}) = \phi p_{ij}(1-p_{ij})$.

Conditional on P_{ij},

$$X_{ij}|P_{ij} \sim Bin(n_{ij}, P_{ij}).$$

This is a first stage of two stage model. The quantity P_{ij} is known as unobservable random variable or latent variable. Unconditionally,

$$\mathrm{E}(X_{ij}) = n_{ij}p_{ij} \qquad \text{and} \qquad \mathrm{Var}(X_{ij}) = n_{ij}p_{ij}(1-p_{ij})(1+(n_{ij}-1)\phi), \tag{2.2.6}$$

where ϕ is in inflation parameter. This parameter may vary for different mixtures of distributions. For example, the beta-binomial (BetaBinom) distribution can be obtained as a Binomial-Beta mixed distribution. Assuming P_{ij} has a beta distribution with shape parameters α_{ij} and β_{ij}. Then X_{ij} has beta-binomial distribution with ρ as a function of these shape parameters of beta distribution. Beta is the conjugate prior distribution for the parameter p if the data are binomial. When $X_{ij} \sim Bin(n_{ij}, p_{ij})$ and $p_{ij} \sim Beta(\alpha, \beta)$, then unconditionally, X_{ij} follows a beta-binomial distribution with parameters α_{ij}, β_{ij} and n_{ij}. The expected value and variance of X_{ij} are

$$\mathrm{E}(X_{ij}) = \frac{n_{ij}\alpha_{ij}}{\alpha_{ij}+\beta_{ij}}, \qquad \mathrm{Var}(X_{ij}) = \frac{n_{ij}\alpha\beta(n_{ij}+\alpha_{ij}+\beta_{ij})}{(\alpha_{ij}+\beta_{ij})^2(\alpha_{ij}+\beta_{ij}+1)}.$$

It is more convenient to re-parametrize this distribution as $BetaBinom(n_{ij}, p_{ij}, \rho_{ij})$ for $p_{ij} = \alpha_{ij}/(\alpha_{ij}+\beta_{ij})$ and $\rho_{ij} = 1/(\alpha_{ij}+\beta_{ij}+1)$. Then the moments are

$$\mathrm{E}(X_{ij}) = n_{ij}p_{ij}, \qquad \mathrm{Var}(X_{ij}) = n_{ij}p_{ij}(1-p_{ij})(1+(n_{ij}-1)\rho_{ij}), \tag{2.2.7}$$

so the beta-binomial is an overdispersed binomial distribution. The density function for beta-distribution is

$$f(p_{ij}) = \frac{\Gamma(\alpha_{ij}+\beta_{ij})}{\Gamma(\alpha_{ij})\Gamma(\beta_{ij})} p_{ij}^{\alpha_{ij}-1}(1-p_{ij})^{\beta_{ij}-1}$$

and for binomial distribution conditional on p_{ij} is

$$f(X_{ij}|p_{ij}; n_{ij}) = \binom{n_{ij}}{X_{ij}} p_{ij}^{X_{ij}}(1-p_{ij})^{(n_{ij}-X_{ij})}.$$

The joint density function for X_{ij} and p_{ij} is

$$f(X_{ij}, p_{ij}; n_{ij}) = f(X_{ij}|p_{ij}; n_{ij}) f(p_{ij})$$

$$f(X_{ij}, p_{ij}; n_{ij}) = \binom{n_{ij}}{X_{ij}} p_{ij}^{X_{ij}} (1-p_{ij})^{n_{ij}-X_{ij}} \frac{\Gamma(\alpha_{ij}+\beta_{ij})}{\Gamma(\alpha_{ij})\Gamma(\beta_{ij})} p_{ij}^{\alpha_{ij}-1} (1-p_{ij})^{\beta_{ij}-1}$$

Since, we do not need p_{ij} to be observable, p_{ij} has to be integrated out which produces marginal density

$$f(X_{ij}; \alpha_{ij}, \beta_{ij}, n_{ij}) = \int_0^1 f(X_{ij}|p_{ij}; n_{ij}) f(p_{ij}) dp_{ij}$$

$$f(X_{ij}; \alpha_{ij}, \beta_{ij}, n_{ij}) = \int_0^1 \binom{n_{ij}}{X_{ij}} p_{ij}^{X_{ij}} (1-p_{ij})^{n_{ij}-X_{ij}} \frac{\Gamma(\alpha_{ij}+\beta_{ij})}{\Gamma(\alpha_{ij})\Gamma(\beta_{ij})} p_{ij}^{\alpha_{ij}-1} (1-p_{ij})^{\beta_{ij}-1} dp_{ij}$$

$$f(X_{ij}|n_{ij}, \alpha_{ij}, \beta_{ij}) = \binom{n_{ij}}{X_{ij}} \frac{Beta(\alpha_{ij}+X_{ij}, \beta_{ij}+n_{ij}-X_{ij})}{Beta(\alpha_{ij}, \beta_{ij})} \qquad 0 \le X_{ij} \le n_{ij}. \tag{2.2.8}$$

for beta-binomial distribution.

Williams (1982) discusses another possibility by assuming logistic-normal mixture distribution for X_{ij}. The logistic normal model is a two stage model

$$X_{ij} \sim Bin(n_{ij}, p_{ij}) \quad \text{and} \quad \text{logit}(p_{ij}) \sim N(\mu, \tau^2)$$

where μ is the overall mean and τ^2 represents between-study variance. The general logistic normal model with additive random effect b_j in the linear predictor is

$$\text{logit}(p_{ij}) = \mu + b_j$$

where $\text{logit}(p_{ij})$ is a linear predictor with continuous normal distribution and logit link function (Hinde and Demétrio, 1998). The variance of X_{ij} from logistic-normal mixture is

$$\text{Var}(X_{ij}) = n_{ij} p_{ij}(1-p_{ij})(1+(n_{ij}-1)p_{ij}(1-p_{ij})\tau^2) \tag{2.2.9}$$

Letting $\eta_{ij} = \text{logit}(p_{ij})$, the marginal density function of logistic-normal distribution is the mixture of binomial and normal densities

$$f(X_{ij}, \mu, \tau) = \int_{-\infty}^{\infty} \binom{n_{ij}}{X_{ij}} \frac{\exp(X_{ij}\eta_{ij})}{(1+\exp(\eta_{ij}))^{n_{ij}}} \frac{1}{\tau\sqrt{2p}} \exp(-\frac{(\eta_{ij}-\mu)^2}{2\tau^2}) d\eta_{ij}$$

The logistic normal model is a generalized linear model from linear exponential family. Beta-binomial and logistic normal models behave similarly for p_{ij} between 0.2 and 0.8. However, out of this range when the probability is either close to 0 or 1, there exist some differences

between the skewness of the distributions. This is because the variance (2.2.9) has an extra factor $p_{ij}(1-p_{ij})$ in comparison to (2.2.7).

The variance functions (2.2.7) and (2.2.9) from beta-binomial and logistic normal mixtures are a particular case of a general variance function of an overdispersed binomial distribution

$$\mathrm{Var}(X_{ij}) = n_{ij}p_{ij}(1-p_{ij})(1+\phi(n_{ij}-1)^{\delta_1}(p_{ij}(1-p_{ij}))^{\delta_2}).$$

For $\delta_1 = 1$ and $\delta_2 = 0$, the general variance reduces to beta-binomial variance (2.2.7). For $\delta_1 = 1$ and $\delta_2 = 1$, the general variance reduces to logistic-normal variance (2.2.9).

The estimation of parameters in the binomial mixture models is performed either by maximum likelihood methods or using quasi-likelihood methods. Maximum likelihood methods can be used in two stage models by assigning the distributions to unknown parameters. Quasi-likelihood methods are used in models with general form of the variance function. Crowder (1978) discusses full maximum likelihood estimation for the beta-binomial model. Prentice (1986) overviews the extension of beta-binomial model to handle underdispersion.

Generation through Gaussian copula

A simple method for generation of correlated binary data is proposed by Emrich and Piedmonte (1991). The aim is to generate n_{ij} Bernoulli r.v.'s $< X_{ijk}, k = 1, \ldots, n_{ij} >$ with moments (2.2.3). Let $Z = (Z_{ij1}, ..., Z_{ijn_{ij}})$ be a vector of independent standard normal r.v.'s. Denote by Σ the $n_{ij} \times n_{ij}$ covariance matrix such that

$$\Sigma = (1-\rho^*)I_{n_{ij}} + \rho^* J_{n_{ij}},$$

where ρ^* is the correlation (scalar), $J_{n_{ij}}$ is the matrix of 1's and $I_{n_{ij}}$ is the identity matrix of size $n_{ij} \times n_{ij}$. Let A be the lower triangular matrix resulting from a Cholesky decomposition $\Sigma = AA^T$, and let the random variables $Y_{ij1}, ..., Y_{ijn_{ij}}$ equal to $Y = AZ$. Finally, let $U = (\Phi(Y_{ij1}), ..., \Phi(Y_{ijn_{ij}}))$, where Φ denote the cumulative distribution function of a standard normal distribution. The binomial quantile transformation of the vector U with number of trial 1 and true probability p produces n_{ij} correlated Bernoulli variables with correlation ρ. The latter process is repeated for K studies. According to Emrich and Piedmonte

(1991) and Demirtas et al. (2009), the value for ρ^* can be obtained by solving the equation $\Phi[z(p_{ijk}), z(p_{ijk'}); \rho^*] = \rho(p_{ijk}q_{ijk'}p_{ijk'}q_{ijk'})^{1/2} + p_{ijk}p_{ijk'}$ using the bisection method. Here, $z(p)$ denotes the p^{th} quantile of the standard normal distribution, and $\Phi[x_1, x_2, \rho^*]$ is the standard bivariate normal cumulative distribution function with correlation coefficient ρ^*. This solution is unique as long as the restriction (2.2.5) for the correlation ρ holds. This method of generation is called Gaussian copula (GC) model in subsequent sections.

Generation by the method of Lunn and Davies (1998)

Lunn and Davies (1998) consider the case of clustered binary variables

$$\{X_{ijk},\ i = 1, \cdots, K; j = 1, 2; k\ =\ 1, \cdots, n_{ij}\},$$

where i is the cluster in group j of the size n_{ij}. In order to generate correlated binary data with correlations ρ_{ij} within each cluster i, firstly generate n_{ij} independent Bernoulli random variables $\{Y_{ijk}, k = 1, \cdots, n_{ij}\}$ and Z_k from $B(1, p_k)$. Additionally, generate n_{ij} independent Bernoulli variables U_{ijk} from $B(1, \sqrt{\rho_{ij}})$. The random variables $X_{ijk} = (1 - U_{ijk})Y_{ijk} + U_{ijk}Z_k$ for $k = 1, \cdots, n_{ij}$ are correlated binary random variables such that $P(X_{ijk} = 1) = p_{ijk}$, $\text{Var}(X_{ijk}) = p_{ijk}(1 - p_{ijk})$ and $\text{Cov}(X_{ijk}, X_{ijk'}) = \rho_{ij}p_{ij}(1 - p_{ij})$.

Large sample properties of the overdispersed binomial distributions

To better understand the properties of the overdispersed binomial distributions, we produced the QQ plots exploring large-sample normality or lack thereof, for the arcsine-transformed sample probabilities $\hat{p}$ estimated from the data generated by a beta-binomial distribution, method by Lunn and Davies (1998) and Gaussian Copula for K=1. The QQ plots were generated for observations in a single study. The binomially distributed variables for each plot were generated and arcsine transformed for a single arm. The true probabilities $p = 0.1$ and number of studies $K = 1$ are kept the same for all plots in Figure 2.1. For all three methods, we performed 1000 repetitions with identical combination of sample sizes n and intra-cluster correlation ρ for $n = 20, 1000$ and $\rho = 0.1, 0.9$. These QQ plots for several

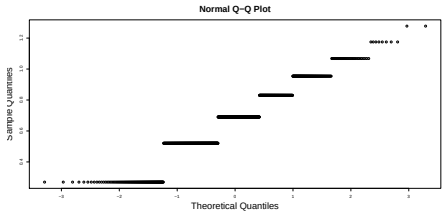
combination of sample sizes and correlation coefficients ρ are given in Figure 2.1. It is clear that the Lunn-Davies model results in a much clumpier distribution. For small values of ρ, the beta-binomial and Gaussian copula models are close to normality, but the Lunn-Davies model is much less normal, it is almost dichotomous for large n. For large n and large ρ all three distributions are almost dichotomous, but once more the Lunn-Davies much more so than the other two models. As we shall see, this results in much worse coverage of the confidence intervals based on the normal quantiles. In our further simulations, the Lunn-Davies model is "the worst case scenario".

2.3 Background information on meta-analysis

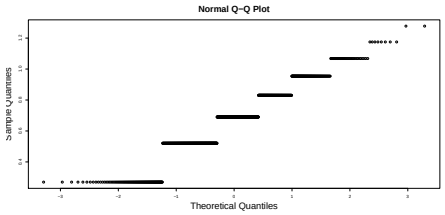
Meta-analysis of binary data combines studies from observational studies and randomized controlled trials. There are different types of observational studies such as case-control studies, cross-sectional studies, cohort studies and etc. Case - control and cohort studies are the two most important study designs in observational studies for evaluation of association between intervention and outcome. Observational studies usually make inference about effect of a treatment from a small sample of large population, where subjects of studies are non-controlled by any experimenter. In contrast, randomized controlled trials differ from observational studies by having each subject being under the control of the experimenter. In randomized controlled trials, experimenter assigns participants of study either to a treatment or a control group. The simplest form of a randomized controlled trial is the trial with a single treatment and a single control arm. However, there may exist a randomized controlled trial comparing a treatment to several control arms. Observational studies and randomized controlled trials usually combine the observed information in 2×2 contingency tables.

Since a single study may not have enough power to make inference about a tested treatment, meta-analytic methods have to be used. Meta-analytical methods assume either the homogeneity or heterogeneity of effect measures across the studies. The methods with assumption of homogeneity are based on fixed effect model. The assumption of homogeneous odds ratios

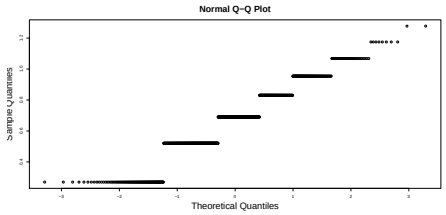
$n = 20, \ \rho = 0$



(a) Beta-Binomial

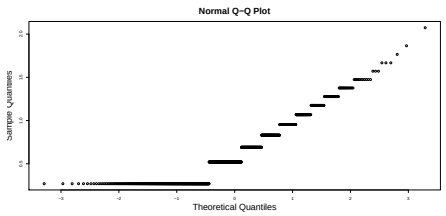


(b) Gaussian Copula model

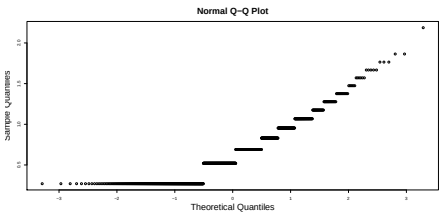


(c) Lunn-Davies model

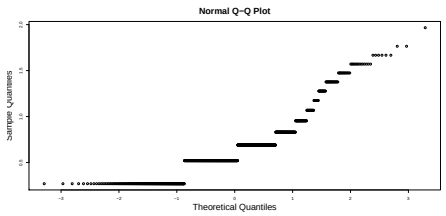
$n = 20, \ \rho = 0.1$



(d) Beta-Binomial

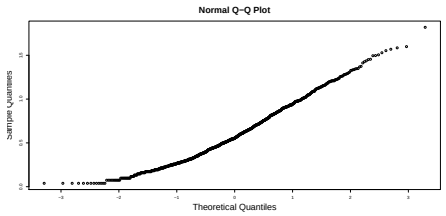


(e) Gaussian Copula model

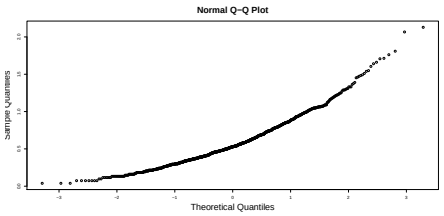


(f) Lunn-Davies model

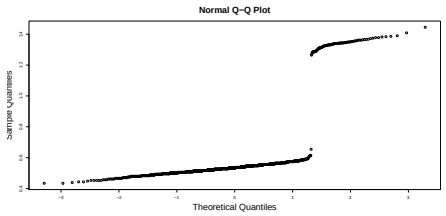
$n = 1000, \ \rho = 0.1$



(g) Beta-Binomial

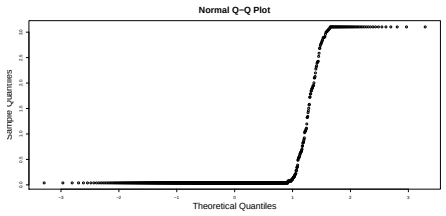


(h) Gaussian Copula model

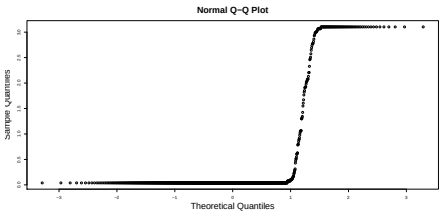


(i) Lunn-Davies model

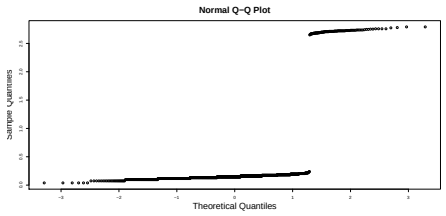
$n = 1000, \ \rho = 0.9$



(j) Beta-Binomial



(k) Gaussian Copula model



(l) Lunn-Davies model

Figure 2.1: QQ plots for arcsine-transformed (with Anscombe (1948) continuity correction) sample probabilities in Beta-Binomial, Gaussian Copula and Lunn-Davies models

or arcsine differences might be violated since each individual study might differ by its design and structure (Higgins et al., 2009; Mosteller and Colditz, 1996). The assumption of heterogeneous effect measures based on either odds ratios or arcsine differences seems to be more realistic. For example in clinical trials, testing the same treatment under different external conditions or for different populations does not produce identical results. The effect measures will differ for different experiments; the differences can be a result of sampling variability or may be due to heterogeneity across studies in meta-analysis. The heterogeneity is usually difficult to explain and quantify both in practice (Higgins and Thompson, 2002) and theory (Thompson and Higgins, 2002). The standard way of explaining heterogeneity is by diversity in populations or interventions. This is assumed to be accounted for in standard random effects model. The drawback of standard random effects model is that it does not clarify the heterogeneity according to individual variables. In general the sources of heterogeneity should be investigated (Thompson, 1994).

In meta-analysis of binary data, before switching from fixed effect model to random effects model, the presence of heterogeneity between the studies is often verified by an appropriate statistical test. The heterogeneity can be tested by a number of well-known methods. These methods include Breslow-Day test (Breslow and Day, 1980), Tarone score test (Tarone, 1985), Cochran's Q statistic (Cochran, 1937), conditional score test (Liang and Self, 1985), likelihood ratio test based on mixed logistic models (Agresti and Hartzel, 2005) and score test statistic based on full likelihood (Liang and Self, 1985). The Cochran's Q statistic and Breslow-Day test are discussed in details in section 2.5. The Breslow-Day and Tarone tests are designed for both odds ratios and relative risk statistics in original non-transformed scale. The Cochran's Q statistic is a general test of heterogeneity applied on log-odds ratio and log-relative risk scale. Likelihood based tests are based on normal or binomial-normal likelihood.

Heterogeneity can be explained by moderators through meta-regression. For example, higher treatment doses, length and intensity may result in a higher treatment effect. Heterogeneity can also be explained by random differences between the effects. In that case, the effect

measures are assumed to follow some distribution. The standard approach to quantify the heterogeneity between studies is to consider some unexplained random effect. The random effect accounts for heterogeneous effect measures by additional variance component across the studies. In random effects model, the between-study variance explains the heterogeneity between effect measures. Estimation of between-study variance has a big influence on inference within random effects model in meta-analysis.

In practice, quantifying the evidence by an effect measure might not always provide precise estimates. Failures in inferences from meta-analysis can be explained by biases. The bias in meta-analysis may arise from different sources. For example, publishing insignificant results may result in publication bias. Studies with small sample sizes and low probabilities may produce sparse data, where the continuity corrections have to be added in case of binary data. This may produce a bias due to continuity corrections. This thesis mainly concentrate on estimation biases. The bias might arise due to transformation of random variables from one scale to another in standard and mixture models. We also consider the bias due to continuity corrections in normalising and variance stabilizing transformations.

It is also possible to avoid the continuity corrections by using generalized linear mixed models for meta-analysis. For the log-odds-ratio,Platt et al. (1999) discusses the generalized linear mixed models (GLMM). Stijnen et al. (2010) proposes a conditional generalized linear mixed-effects model with an exact likelihood for use in meta-analysis of binary data. They suggest using the GLMM's to overcome several potential issues in meta-analysis. These issues are related to non accounting for the variances being estimated or non accounting for the dependence between the estimate of effect and its variance. In generalized linear mixed models, the approximately normal within study likelihood is replaced by exact or approximate likelihood and no continuity corrections are required. Chapter 5 discusses generalized linear mixed models in details.

2.4 Standard Fixed Effect Model

Suppose we have K studies, each estimating the same effect. Assume that θ_i is the true effect measure and $\hat{\theta}_i$ is the estimate of θ_i from study i. The fixed effect model (FEM) assumes a homogeneity of effects across studies such that $\theta_i \equiv \theta$ for $i = 1, ..., K$. Hence, each study is estimating a single true effect.

The main goal of meta-analysis is to combine estimates from all studies to obtain an estimate for true effect. The estimates are asymptotically normal such that $(\hat{\theta}_i - \theta_i) \sim N(0, \sigma_i^2)$ with unknown parameters θ_i and σ_i^2. Large sample confidence interval coefficient $1 - \alpha$ for each estimate is $\hat{\theta}_i \pm z_{1-\alpha/2}(\hat{\sigma}_i^2)^{1/2}$ as $\hat{\sigma}_i^2/\sigma_i^2 \to 1$ in probability with assumption that $\hat{\sigma}_i^2$ is a consistent estimator of σ_i^2 (Kulinskaya et al., 2014). The simplest choice for an overall estimate is the sample mean of $\hat{\theta}_i$'s. However, averaging estimates ignores that each study might not have the same size and/or precision. Variation of study sizes produces estimates with different precision. The general fixed effect model combines effect measures by weighted estimates $\hat{\theta}_i$ using the inverse-variance weights. Weighted average allows to weight each study depending on its precision. Weights can be found as reciprocals of variances σ_i^2. By weighting each estimate, the inverse variance-weighted method for pooling estimates in fixed effect model is

$$\hat{\theta}_{FE} = \frac{\sum_{i=1}^{K} w_i \hat{\theta}_i}{\sum_{i=1}^{K} w_i},$$

where weights w_i are inversely proportional to the variance of individual effect measure such as $w_i = \mathrm{Var}(\hat{\theta}_i)^{-1} = \sigma_i^{-2}$. Assuming that weights $w_i = \sigma_i^{-2}$ are known and $\hat{\theta}_i \sim N(\theta_i, \sigma_i^2)$, the large sample confidence interval for overall estimate is

$$\hat{\theta}_{FE} \pm zW^{-1/2} \qquad \text{with} \qquad W = \sum_{i=1}^{K} w_i.$$

The presence of a non-zero effect, i.e $H_0 : \hat{\theta}_{FE} = 0$ vs alternative $H_1 : \hat{\theta}_{FE} \neq 0$ can be tested by a Wald test by comparing $W^{-1/2}\hat{\theta}_{FE}$ to critical values of $z_{1-\alpha}$ from $N(0, 1)$ distribution.

In reality, the weights of each study are estimated rather than known since the within-study

variances are estimated. Hence an estimator $\hat{w}_i$ is used instead of w_i. Substitution of the estimate $\hat{w}_i$ leads to underestimation of $W^{-1/2}$ as noted by Li et al. (1994) and Rukhin (2009). This results in low coverage for the $\hat{\theta}_{FE} \pm z\hat{W}^{-1/2}$ and liberal Wald test $\hat{W}^{1/2}\hat{\theta}_{FE}$ for testing the hypothesis of no effect. The approximate large sample distribution of the combined effect can be found either increasing the number of studies K or increasing the sample sizes of each study or increasing both of them simultaneously. These three methods may result in different limit distributions (Kulinskaya et al., 2014).

We concentrate on the effect measures based on odds ratio's. Additionally to inverse-variance method for LOR, effects can be combined by using Mantel-Haenszel method (Mantel and Haenszel, 2004), Peto's method (Yusuf et al., 1985), conditional logistic regression (Connolly and Liang, 1988), conditional (Hauck, 1984) and unconditional likelihood- based methods (Emerson, 1994; Hasselblad and McCrory, 1995; Sutton et al., 2000) and Bayesian methods of estimation (Zelen and Parker, 1986). From this list, we concentrate our attention on Mantel-Haenzsel method. The Mantel-Haenzsel method has advantages of fixed weights and it does not require continuity corrections in case of sparse events. This method assumes homogeneity of odds ratio across K studies. We intend to extend this method to random effects settings. Originally, Mantel-Haenzsel method was introduced by Mantel and Haenszel (2004) to combine odds ratios for stratified case-control studies or cohort studies. This method is known for its efficiency and robustness. We will discuss the standard Mantel-Haenzsel method separately in the following subsection (2.4.5).

2.4.1 The structure of 2×2 contingency table

Assume that K studies report comparative binary outcomes. These outcomes can be written as a series of K 2×2 contingency tables as shown in the table 2.1 below. Each of the individual studies reports a pair of independent binomial variables X_{i1} and X_{i2} (numbers of

Table 2.1: Contingency table

	Event	No event	Total
Treatment	X_{i1}	$n_{i1} - X_{i1}$	n_{i1}
Control	X_{i2}	$n_{i2} - X_{i2}$	n_{i2}
Total	x_i	$n_i - x_i$	n_i

harmful outcomes) from two samples of sizes n_{i1} and n_{i2} for treatment and control arms,

$$X_{i1} \sim Binom(n_{i1}, p_{i1}) \qquad \text{and} \qquad X_{i2} \sim Binom(n_{i2}, p_{i2}) \text{ for } i = 1, ..., K$$

where p_{i1} is the risk in the treatment arm and p_{i2} is the baseline risk. Each binomial outcome X_{ij} is a sum of independent Bernoulli variables such that $X_{ij} = \sum_{k=1}^{n_{ij}} X_{ijk}$ for $j = 1, 2$. In reality, the Bernoulli variables X_{ijk} for $k = 1, ..., n_{ij}$ might be dependent within and between arms of the same study so that $\text{corr}(X_{ijk}, X_{ijk'}) \neq 0$ and $\text{corr}(X_{i1}, X_{i2}) \neq 0$ respectively. The dependence between arms was studied by Hwang and Biswas (2008) and Biswas and Hwang (2010).

2.4.2 Odds and odds ratios

The odds of an outcome for group j in study i for $j = 1, 2$ and $i = 1, \ldots, K$, is

$$\varphi_{ij} = \frac{p_{ij}}{1 - p_{ij}} \qquad \text{estimated by} \qquad \hat{\varphi}_{ij} = \frac{X_{ij}}{n_{ij} - X_{ij}},$$

when $X_{ij} \neq 0$. The odds ratio for individual table j is $\psi_i = \varphi_{i1}/\varphi_{i2}$. In terms of probabilities the odds ratio is

$$\psi_i = \frac{p_{i1}(1 - p_{i2})}{p_{i2}(1 - p_{i1})} \qquad \text{estimated by} \qquad \hat{\psi}_i = \frac{X_{i1}(n_{i2} - X_{i2})}{X_{i2}(n_{i1} - X_{i1})} \tag{2.4.1}$$

for $X_{ij} \neq 0$. The natural logarithm of odds ratio denoted by $\log(\hat{\psi}_i)$ is often used since its distribution is approximately normal. Let $\theta_i = \log(\psi_i)$ and $\hat{\theta}_i = \log(\hat{\psi}_i)$. The approximate variance for estimated log-odds-ratio derived by delta method is

$$\sigma_i^2 = \text{Var}(\hat{\theta}_i) = \text{Var}(\log(\hat{\psi}_i)) = \frac{1}{n_{i1} p_{i1}(1 - p_{i1})} + \frac{1}{n_{i2} p_{i2}(1 - p_{i2})},$$

estimated by

$$\hat{\sigma}_i^2 = \frac{1}{X_{i1}} + \frac{1}{X_{i2}} + \frac{1}{n_{i1} - X_{i1}} + \frac{1}{n_{i2} - X_{i2}}. \tag{2.4.2}$$

The variance of $\hat{\theta}_i$ is based on Woolf's variance estimator of the log-odds-ratio (Woolf et al., 1955).

It is important to pay attention to the studies where some cells are empty. When there exist empty cells, the estimate for the effect measure θ_i and its variance (2.4.2) become undefined. This is a case with sparse event data. We will discuss this case in the Subsection 2.4.4.

Note that the estimate $\hat{\theta}_i$ and its estimated variance $\hat{\sigma}_i^2$ are correlated. According to Berkey et al. (1995) and Stijnen et al. (2010), this dependence might lead to estimation bias in combining logarithmic odds ratios in meta-analysis. Berkey et al. (1995) suggested an alternative estimator for σ_j^2, which reduces the correlation between (2.4.1) and (2.4.2), improves the bias and variance properties of (2.4.2). This estimator is

$$\hat{\sigma}_i^2 = \Big[(X_{i1} + X_{i2})\big(\sum_{i=1}^{K}(X_{i1}/(X_{i1} + X_{i2}))/K\big)\Big]^{-1} +$$

$$+\Big[(X_{i1} + X_{i2})\big(1 - \sum_{i=1}^{K}(X_{i1}/(X_{i1} + X_{i2}))/K\big)\Big]^{-1} +$$

$$+\Big[((n_{i1} - X_{i1}) + (n_{i2} - X_{i2}))\big(\sum_{i=1}^{K}((n_{i1} - X_{i1})/((n_{i1} - X_{i1}) + (n_{i2} - X_{i2})))/K\big)\Big]^{-1} +$$

$$+\Big[((n_{i1} - X_{i1}) + (n_{i2} - X_{i2}))\big(1 - \sum_{i=1}^{K}((n_{i1} - X_{i1})/((n_{i1} - X_{i1}) + (n_{i2} - X_{i2})))/K\big)\Big]^{-1}.$$

However, this variance estimator is not popular in practice. Assuming normality of LOR, a 95 percent confidence interval for the log-odds-ratio $\hat{\theta}_i = \log(\hat{\psi}_i)$ is

$$\log(\hat{\psi}_i) \pm 1.96\sqrt{\mathrm{Var}(\log(\hat{\psi}_i))}.$$

Exponentiating the lower and upper bounds of the confidence interval for $\log(\hat{\psi}_i)$, a 95 percent confidence interval for odds ratio ψ_i can be obtained.

2.4.3 Arcsine transformation and the Cohen's effect measure

Cohen (1988) proposed an effect measure based on difference between arcsine transformations of the success proportions from treatment and control groups. This effect measure is given by

$$d_i = 2\text{arcsin}(\sqrt{p_{i1}}) - 2\text{arcsin}(\sqrt{p_{i2}}) \tag{2.4.3}$$

and estimated by

$$\hat{d}_i = 2\text{arcsin}(\sqrt{\hat{p}_{i1}}) - 2\text{arcsin}(\sqrt{\hat{p}_{i2}}).$$

The variance of $\hat{d}_i$ is

$$\text{Var}(\hat{d}_i) = \frac{1}{n_{i1}} + \frac{1}{n_{i2}},$$

(Rosenthal, 1994). Arcsine is a variance stabilizing transformation for $\hat{p}_{ij}$, since the variance of arcsine transformed proportions is constant and independent of p_{ij}. Rücker et al. (2008) and Olkin and Gleser (2009) discuss the details on the use of arcsine differences as a measure of intervention effect in meta-analysis of binary data. Sánchez-Meca et al. (2003) concludes that the estimator $\hat{d}_i$ is negatively biased with reference on the book by Lipsey and Wilson (2001).

2.4.4 Continuity corrections in contingency tables

The continuity corrections are added in contingency tables when the probabilities of events in binary endpoints are low. The low probabilities result in studies with zero events. The extensive number of zero events affect the analysis of combining studies and obtaining the overall effect measure. The zero events affect differently on different effect measures. According to Bradburn et al. (2007) and Sweeting et al. (2004), the choice of continuity correction may have a big influence on inference for the common effect measure in meta-analysis, especially in the presence of rare events. When the data are sparse, the estimates of LOR become undefined due to zero events. In order to overcome this problem, the continuity correction c is added to each cell in order to make LORs and their variances estimable and to correct for the possible

bias

$$\hat{p}_{i1} = \frac{X_{i1} + c}{n_{i1} + 2c}, \qquad \hat{p}_{i2} = \frac{X_{i2} + c}{n_{i2} + 2c}. \tag{2.4.4}$$

Adding continuity correction c to number of events X_{ij} for $j = 1, 2$ might cause bias in estimating the probabilities and effect. According to Böhning and Viwatwongkasem (2005), the bias of $\hat{p}_{ij}$ for $j = 1, 2$ is $c((1 - 2p_{ij})/(n_{ij} + 2c))$. Different continuity corrections c might introduce bias in estimation of probabilities $\hat{p}_{i1}$ and $\hat{p}_{i2}$. For example, if c increases, $\hat{p}_{ij}$ obtains positive bias that increases for $p < 1/2$ and negative bias that decreases if $p > 1/2$. For $p = 1/2$, the probabilities $\hat{p}_{i1}$ and $\hat{p}_{i2}$ become unbiased (Böhning and Viwatwongkasem, 2005). The bias in estimators of probabilities might result in estimation bias of effects and its weights. For example in log-odds, the variance of log-odds depends on probabilities p_{ij}. For LOR the most common choice of continuity correction is $c = 1/2$. In log-odds and LOR, the continuity correction $c = 1/2$ eliminates the bias of order $1/n$ (Gart et al., 1985). One of the earlier study for odds ratio is reported by Breslow (1981) when the data is sparse. Other alternatives are discussed in Sweeting et al. (2004).

For arcsine transformation, Anscombe (1948) suggested to use $c = 3/8$ in (2.4.4). According to Anscombe (1948), $\tilde{d}_i$ with $c = 3/8$ should have approximate normal distribution and reduce the bias of $\hat{d}_i$. No changes are applied to the variance of $\tilde{d}_i$ since the arcsine transformation removes the dependence of variance on probabilities.

If the number of studies is large and only few studies have the sparse data, the continuity corrections does not have a strong effect on the combined effect measure. But, in case of majority of studies having the sparse data, the choice of continuity correction is important.

According to Gart et al. (1985) there is no universal continuity correction for $\log(\hat{\psi}_i)$ in weighted regression. For example, sometimes $c = 1/2$ might be the best, other times $c = 1/4$, $c = 0$, c$=-1/2$ or intervening values might work better (Gart et al., 1985). Gart et al. (1985) also states the following: "In the presence of even an optimal bias reducing continuity correction, statistics based on empirical logits tend to be fitted less well by the normal or chi-squared distribution than those based on binomial variates". Friedrich et al. (2007) examined

the changes in risk difference, odds ratio and relative risk when studies with zero events are excluded. Friedrich et al. (2007) believes that for estimation of odds ratio, the inclusion of studies with zero events is essential. This is the result of relatively small changes in the magnitude of the overall effect measures with and without exclusion of studies with zero events. In randomized controlled trials, the zero events might appear either in treatment or control arms. It might be possible that both arms in a single study consist of zero events. In meta-analysis, the studies with zero events in both arms are excluded when estimating the overall effect measure (Whitehead and Whitehead, 1991; Sweeting et al., 2004). Böhning and Mylona (2015) argues that the exclusion of studies with zero events should be avoided due to availability of appropriate statistical methods. Alternatively, the continuity corrections are added to every cell in contingency tables for zero events in one of the arms or both arms. In general, the interpretation of results becomes problematic when either continuity corrections are added to studies or studies with zero events in both arms are excluded from the analysis.

2.4.5 Mantel-Haenzsel method for combining odds ratios

Assume that we have K contingency tables with binary outcomes in the form of Table 2.1 and the natural logarithmic of odds ratio is estimated from each table. The fixed effect model assumes homogeneity of LORs. For each study, the estimate of effect measure i.e the natural logarithm of odds ratio can be written in the form

$$\hat{\theta}_i = \theta + \epsilon_i, \qquad \text{for} \qquad i = 1, ..., K \tag{2.4.5}$$

where θ is a common effect measure (in our case the common log-odds-ratio) and ϵ_i is a sampling error of each estimate $\hat{\theta}_i$ that follows the normal distribution with mean 0 and variance σ_i^2.

For the binomial outcomes, the inverse-variance approach was mentioned earlier as the main method for combining effect measures $\hat{\theta}_i$ from K tables. However, inverse-variance method results in the following issues. Firstly, the estimated effect $\hat{\theta}_w$ and its estimated variance are

biased. Secondly, the weights $\hat{w}_i$ and the effects $\hat{\theta}_i$ are correlated. Also the baseline risks may highly affect the inference. The inverse variance method fails when the numbers or the probabilities are low and the sample sizes are unbalanced. Tang (2000) discusses the bias introduced by weightening binary outcomes according to the inverse-variance.

Mantel and Haenszel (1959) suggested robust alternative to inverse-variance method which combines the estimates of odds ratios $\hat{\psi}_i$ themselves rather than the log transformations of them. This method provides more conservative inference on combined effect measure. Mantel-Haenzsel method assumes common odds ratio across K 2×2 contingency tables, i.e the fixed effect model. The Mantel-Haenzsel estimator for common odds ratio $\hat{\psi}$ is

$$\hat{\psi}_{MH} = \frac{\sum_{i=1}^{K} X_{i1}(n_{i2} - X_{i2})n_i^{-1}}{\sum_{i=1}^{K} X_{i2}(n_{i1} - X_{i1})n_i^{-1}}. \tag{2.4.6}$$

Alternatively, it can be written as a sum of weighted individual odds ratios

$$\hat{\psi}_{MH} = \frac{\sum_{i=1}^{K} w_i^* \hat{\psi}_i}{\sum_{i=1}^{K} w_i^*}$$

with weights $w_i^* = \left[(n_{i1})^{-1} + (n_{i2})^{-1}\right]^{-1}(1 - \hat{p}_{i1})\hat{p}_{i2}$ for the odds ratio $\hat{\psi}_i$ defined by (2.4.1) when $n_{ij} \neq 0$. The advantage of Mantel-Haenzsel odds ratio is the ability to handle cases with empty cells without additional continuity corrections (Bradburn et al., 2007). Also the fixed weights n_i^{-1} in the numerator and denominator result in a better approximation of normality for $\log(\hat{\psi}_{MH})$. The Mantel-Haenzsel odds ratio $\hat{\psi}_{MH}$ is consistent estimator of ψ when K is small and $n_i \to \infty$ and when $K \to \infty$ for fixed n_i (Fleiss et al., 2003). Sutton et al. (2000) recommends the Mantel-Haenzsel odds ratio $\hat{\psi}_{MH}$ for large sample situation with large K and small n_i. One of the main advantage of Mantel-Haenzsel odds ratio is computability in case of zero events. No continuity corrections are required in $\hat{\psi}_{MH}$. The continuity corrections are only added in case of running the simulation study with low probabilities, where zero events may appear in both arms simultaneously.

Before considering the variance of $\hat{\psi}_{MH}$, note that different asymptotics is possible for combining contingency tables. The first option is that K is increasing and n_i is fixed. The second scenario is when K is fixed and n_i is increasing. These different scenarios result in different estimators for the variance of odds ratio. Also, it is possible that both parameters K and n_i increase simultaneously. Different variance estimators for Mantel-Haenzsel odds ratio have been proposed. One of the earliest candidate for the variance of Mantel-Haenzsel odds ratio derived by Hauck (1979) for the case of $n_i \to \infty$ is

$$\mathrm{Var}_H(\hat{\psi}_{MH}) = \psi^2 \frac{\sum_{i=1}^{K} w_i^{*2}/v_i}{(\sum_{i=1}^{K} w_i^*)^2} \tag{2.4.7}$$

where $v_i^{-1} = 1/(n_{i1}p_{i1}(1-p_{i1})) + 1/(n_{i2}p_{i2}(1-p_{i2}))$. Guilbaud and Hauck (1983) revised this estimate for the variance of $\hat{\psi}_{MH}$ correcting for sampling errors of weights, since weights are usually estimated in practice rather than being known. The weakness of the Hauck's variance is that if any of the cell entries are zero, the variance is undefined. Also, the Hauck's variance is not valid for asymptotic cases with increasing K and fixed n_i. Breslow (1981) proposed several variances estimators for increasing K and fixed marginals n_i. Given the conditional distribution of data in each 2×2 table

$$K\mathrm{Var}_C(\hat{\psi}_{MH}) = \frac{\sum_{i=1}^{K} \mathrm{Var}(R_i - \psi S_i | x_i; \psi)}{[\sum_{i=1}^{K} E(S_i|x_i;\psi)/K]^2}, \tag{2.4.8}$$

where ψ is the true common odds ratio, R_i and S_i are the numerator and denominator of Mantel-Haenzsel odds ratio

$$R_i = \sum_{i=1}^{K} \frac{X_{1i}(n_{2i} - X_{2i})}{n_i} \qquad \text{and} \qquad S_i = \sum_{i=1}^{K} \frac{X_{i2}(n_{i1} - X_{i1})}{n_i}.$$

The variance estimator by Breslow (1981) is an empirical variance which estimates (2.4.8) as

$$K\mathrm{Var}_E(\hat{\psi}_{MH}) = \frac{\sum_{i=1}^{K} (R_i - \hat{\psi}_{MH} S_i)^2/K}{(\sum_{i=1}^{K} S_i/K)^2}. \tag{2.4.9}$$

According to Breslow (1981), these variance estimators provide accurate inference about estimated odds ratio when the data is sparse. Breslow and Liang (1982) suggested weighted average of two variances (2.4.7) and (2.4.9). However, all these variance estimators above are limited being not valid for different asymptotics of sample sizes and number of studies. The most commonly used variance estimator for Mantel-Haenzsel odds ratio is suggested by Robins et al. (1986) and Phillips and Holland (1987). They derived the variance for Mantel-Haenzsel odds ratio for both sparse data and large-strata limiting models in the form

$$K\mathrm{Var}(\hat{\psi}_{MH}) = K\Big[\frac{\sum_{i=1}^{K} P_i R_i}{2(\sum_{i=1}^{K} R_i)^2} + \frac{\sum_{i=1}^{K}(P_i S_i + Q_i R_i)}{2(\sum_{i=1}^{K} R_i)(\sum_{i=1}^{K} S_i)} + \frac{\sum_{i=1}^{K} Q_i S_i}{2(\sum_{i=1}^{K} S_i)^2}\Big](\hat{\psi}_{MH})^2$$

where $P_i = (X_{i1} + n_{i2} - X_{i2})/n_i$, $Q_i = (n_{i1} - X_{i1} + X_{i2})/n_i$, $R_i = (X_{i1}(n_{i2} - X_{i2}))/n_i$, and $S_i = ((n_{i1} - X_{i1})X_{i2})/n_i$. Due to correspondence of variances for odds ratio and log-odds-ratio, the variance for $\log(\hat{\psi}_{MH})$ is estimated by

$$\mathrm{Var}(\log(\psi_{MH})) = \frac{\sum_{i=1}^{K} P_i R_i}{2(\sum_{i=1}^{K} R_i)^2} + \frac{\sum_{i=1}^{K}(P_i S_i + Q_i R_i)}{2(\sum_{i=1}^{K} R_i)(\sum_{i=1}^{K} S_i)} + \frac{\sum_{i=1}^{K} Q_i S_i}{2(\sum_{i=1}^{K} S_i)^2} \tag{2.4.10}$$

Using a normal approximation to $\log(\hat{\psi}_{MH})$, the confidence interval for the overall odds ratio can be obtained by

$$\exp[\log(\hat{\psi}_{MH}) - z_{\alpha/2}(\mathrm{Var}(\log(\hat{\psi}_{MH}))^{1/2})] \leq \psi_{MH} \leq \exp[\log(\psi_{MH}) + z_{\alpha/2}(\mathrm{Var}(\log(\hat{\psi}_{MH}))^{1/2})]$$

where $z_{\alpha/2}$ is the $\alpha/2$ percentage point of a standard normal distribution. Recently, the simple proof for the variance (2.4.10) was suggested by Silcocks (2005). Sato (1990) proposed a new approximate confidence limit method for the common odds ratio based on the asymptotic distribution of the Mantel-Haenszel estimator. Leonard and Duffy (2002) describe the Bayesian framework for the Mantel-Haenzsel model.

2.5 Testing for presence of heterogeneity

The tests for homogeneity or equivalently for the presence of heterogeneity are performed to assess the degree of similarity between studies in meta-analysis. These tests are sometimes used to provide an indication about the choice of fixed effect or random effects model (discussed later). The null hypothesis of these tests is

$$H_0 : \theta_1 = ... = \theta_K = \theta \quad \text{or} \quad \psi_1 = ... = \psi_K = \psi$$

versus alternative

$$H_1 : \theta_i \neq \theta \quad \text{or} \quad \psi_i \neq \psi \quad \text{for at least some i=1,. . . ,K.}$$

In order to assess these hypothesises, two popular tests for homogeneity between studies are reviewed in this section. The first test is the popular Cochran's Q statistic (Cochran, 1937). The second test is the Breslow-Day test (Breslow and Day, 1980).

2.5.1 Cochran's Q statistic

Testing the hypothesis of heterogeneity is a common practice in meta-analysis. The null hypothesis is an absence of heterogeneity $H_0 : \theta_i \equiv \theta$ for fixed effect model or $H_0 : \tau^2 = 0$ for random effects model discussed further in the following subsections. In random effects model, τ^2 is the between-study variance to account for heterogeneity across K studies. The alternative hypothesis is the presence of heterogeneity, i.e $\tau^2 \neq 0$. The most popular test statistic for heterogeneity in meta-analysis is the Cochran's Q statistic (Cochran, 1937). The Cochran's Q statistic is given by

$$Q = \sum_{i=1}^{K} \hat{w}_i(\hat{\theta}_i - \hat{\theta}_{FE})^2$$

with inverse variance weights $\hat{w}_i$ and a weighted average $\hat{\theta}_{FE} = \sum_{i=1}^{K} \hat{w}_i\hat{\theta}_i / \sum_{i=1}^{K} \hat{w}_i$. In general, the behaviour of the Q statistic might be not the same for different measures of the effect. Initially, the Q statistic was applied for tests of heterogeneity when the effects were normally

distributed sample means (Cochran, 1937, 1954; Welch, 1951; James, 1951). Later, Woolf et al. (1955) and DerSimonian and Laird (1986) suggested using Q statistic for dichotomous outcomes with effects given by difference of proportions or the logarithmic of odds ratios.

Under H_0, it is conventional to assume that Q follows the chi-square distribution with $K-1$ degrees of freedom without any assumption of the size of studies and the effect measure. If the heterogeneity is present, then the value of Q will be greater than the critical value of the χ^2-distribution and the hypothesis of homogeneity is rejected (Demidenko, 2004). The chi-square distribution is the exact distribution of Q statistic under the condition that the variances, hence weights are assumed to be known and the effects are normally distributed (Kulinskaya et al., 2014). For the logarithmic of odds ratio and other measures of effect, Kulinskaya et al. (2011a); Kulinskaya and Dollinger (2015) show that the chi-square approximation for Q is inaccurate. For the small sample sizes, Biggerstaff et al. (1997) approximated the distribution of the Q statistics by a gamma distribution. Biggerstaff et al. (1997); Biggerstaff and Jackson (2008); Jackson (2006) present results about the distribution for Q statistic with assumption of known weights. Kulinskaya and Dollinger (2015) derive an approximation to make inferences about homogeneity of the logarithmic of odds ratios.

When weights are estimated rather than being known, the distribution of Q statistic does not follow an exact distribution. The known distribution of Q varies with the choice of effect measure. Hedges and Olkin (1985) and Viechtbauer (2007) show in their simulations that Q can be approximated by chi-square distribution for finite K and increasing n_i, however, for the case with increasing K and finite n_i, the normal approximation of θ_i is discussed by Demidenko (2004) for non-normal effect measures. Akritas and Papadatos (2004) discussed an asymptotic approximations for small sample sizes and large number of studies. Several approximations for Q statistic were proposed by James (1951); Welch (1951); Kulinskaya et al. (2003). Kulinskaya et al. (2011a) and Kulinskaya et al. (2011b) suggest improvements to the null distribution of Q statistic for risk difference and standardized mean difference respectively. So far the approximations for distribution of Q statistic are obtained under null

fixed effect model. The approximations for the non-null Q statistic and approximation to its power under the random effects model have not been obtained yet.

2.5.2 Breslow-Day test statistic

Mantel-Haenzsel estimate for the common odds ratio is developed under the hypothesis that odds ratios are equal across the studies. In order to test the homogeneity of odds ratios between the studies, Breslow and Day (1980) introduced Breslow-Day test statistic. Breslow-Day test is based on fixed margins of 2×2 contingency tables with estimated Mantel-Haenzsel common odds ratio (Breslow and Day, 1980). The Breslow-Day statistic tests the hypothesis that the odds ratios between studies are homogeneous.

Assuming fixed sample sizes for rows and columns in table 2.1, X_{i1} follows a non-central hypergeometric distribution with probability density function

$$Pr(X_{i1} = x_{i1} | X_{i1} + X_{i2} = x_i, \psi_i) = \frac{\binom{n_{i1}}{x_{i1}}\binom{n_{i2}}{x_{i2}}\psi^{x_{i1}}}{\sum_{i=\max(0,N-n_{i2})}^{\min(n_{i1},n_{i2})}\binom{n_{i1}}{i}\binom{n_{i2}}{x_i-i}\psi_i^{x_i}},$$

or

$$Pr(X_{i1} = x_{i1} | X_{i1} + X_{i2} = x_i, \theta_i) = \frac{\binom{n_{i1}}{x_{i1}}\binom{n_{i2}}{x_{i2}}\exp(\theta_i x_{i1})}{\sum_{j=\max(0,N-n_{i2})}^{\min(n_{i1},n_{i2})}\binom{n_{i1}}{i}\binom{n_{i2}}{x_i-i}\exp(\theta_i x_i)},$$

where $\max(0, n_i - n_{i2}) \leq x_{i1} \leq \min(n_{i1}, n_{i2})$. When $\psi = 1$, the distribution reduces to hypergeometric distribution. The mean of non-central hypergeometric distribution $E(X_{i1}|\hat{\psi}_{MH})$ can be obtained by the quadratic equation

$$\frac{E(X_{i1}|\hat{\psi}_{MH})[n_i - x_i - n_{i1} + E(X_{i1}|\hat{\psi}_{MH})]}{[x_i - E(X_{i1}|\hat{\psi}_{MH})][n_{i1} - E(X_{i1}|\hat{\psi}_{MH})]} = \hat{\psi}_{MH}, \tag{2.5.1}$$

whereas the asymptotic variance $\mathrm{Var}(X_{i1}|\hat{\psi}_{MH})$ is given by

$$\mathrm{Var}(X_{i1}|\hat{\psi}_{MH}) = \Big[\frac{1}{E(X_{i1}|\hat{\psi}_{MH})} + \frac{1}{(x_i - E(X_{i1}|\hat{\psi}_{MH}))} \tag{2.5.2}$$

$$+\frac{1}{(n_{1i} - E(X_{1i}|\hat{\psi}_{MH}))} + \frac{1}{(n_i - x_i - n_{1i} + E(X_{1i}|\hat{\psi}_{MH}))}\Big]^{-1}$$

where $x_i = X_{i1} + X_{i2}$. The values of X_{i1} should be close to $E(X_{i1}|\hat{\psi}_{MH})$ (the expected value for X_{i1} given $\hat{\psi}_{MH}$).

The Breslow-Day test statistic based on assumption of non-central hypergeometric distribution for events in treatment arm under the hypothesis of homogeneity of odds ratio $\hat{\psi}_i = \hat{\psi}_{MH}$, which is a conditional distribution for the X_{i1} given that columns n_{i1}, n_{i2} and rows x_i, $n_i - x_i$ are fixed is

$$BD = \sum_{i=1}^{K} \frac{(X_{i1} - E(X_{i1}|\hat{\psi}_{MH}))^2}{\mathrm{Var}(X_{i1}|\hat{\psi}_{MH})} \tag{2.5.3}$$

where $E(X_{i1}|\hat{\psi}_{MH})$ and $\mathrm{Var}(X_{i1}|\hat{\psi}_{MH})$ denote the expected number and the asymptotic variance of cases, respectively, given the Mantel-Haenzsel fitted odds ratio $\hat{\psi}_{MH}$ under the assumption of homogeneity. The exact Breslow-Day test BD has χ^2_{K-1} distribution with $K-1$ degrees of freedom. Breslow-Day test is asymptotically equivalent to the logistic regression method (Leonard and Duffy, 2002). Tarone (1985) proposed the correction factor for approximation of Breslow-Day test statistic

$$T = \sum_{i=1}^{K} \frac{(X_{i1} - E(X_{i1}|\hat{\psi}_{MH}))^2}{\mathrm{Var}(X_{i1}|\hat{\psi}_{MH})} - \frac{[\sum_{i=1}^{K} X_{i1} - \sum_{i=1}^{K} E(X_{i1}|\hat{\psi}_{MH})]^2}{\sum_{i=1}^{K} \mathrm{Var}(X_{i1}|\hat{\psi}_{MH})},$$

arguing that the distribution of Breslow-Day test is stochastically larger than χ^2_{K-1} under the hypothesis of homogeneity of odds ratios $H_0 : \psi_i = \psi$. The Tarones Breslow-Day test T should also follow χ^2_{K-1} distribution with $K-1$ degrees of freedom. Adjustments to the variance of a non-central hypergeometric distribution (2.5.2) were proposed by Levin (1984)

$$\mathrm{Var}(X_{i1}|\hat{\psi}_{MH}) = \frac{n_i}{n_i - 1}\Big[\frac{1}{E(X_{i1}|\hat{\psi}_{MH})} + \frac{1}{(x_i - E(X_{i1}|\hat{\psi}_{MH}))} + .$$

$$+\frac{1}{(n_{i1} - E(X_{i1}|\hat{\psi}_{MH}))} + \frac{1}{(n_i - x_i - n_{i1} + E(X_{i1}|\hat{\psi}_{MH}))}\Big]^{-1}.$$

Alternative to Breslow-Day test is a conditional likelihood score test for homogeneity of odds ratios. This test has a similar form to (2.5.3) apart from the conditional maximum likelihood estimate of odds ratio used instead of Mantel-Haenzsel odds ratio. Alternative homogeneity tests were suggested by Liang and Self (1985) for the case when the binomial data are sparse and the number of tables is large.

2.5.3 Comparison of Q and Breslow-Day test statistics

Paul and Donner (1989) compared nine tests for testing the homogeneity of odds ratios. Later, Paul and Donner (1992) studied the performance of these tests for the small size studies. The Q statistic is recommended for balanced studies by Paul and Donner (1989), however when the studies are unbalanced the power of the Q statistic decreases.

The power of the Q statistic was investigated by Hedges and Pigott (2001) and Valentine et al. (2010) and also discussed by Biggerstaff and Jackson (2008), Hardy and Thompson (1998) and Jackson (2006). According to Takkouche et al. (1999) and Viechtbauer (2007), the Q statistic outperforms other tests for the presence of heterogeneity. However, Hardy and Thompson (1998) states that Q statistic has a low power. Hardy and Thompson (1998) studies how the power of the heterogeneity test depends on the number of studies K.

Normand (1999) recommends relying on Q statistic for choosing between fixed effect and random effects model. Higgins and Thompson (2002) proposed quantification of heterogeneity by so called I^2 measure which is an increasing function of Q, i.e $I^2 = (Q - (K-1))/Q$. If $\tau^2 = 0$ and assuming known weights, $E(Q) = K - 1$ and $I^2 \approx 0$. This is also approximately true for large sample sizes. If $\tau^2 \neq 0$, then Q statistic increases with the total sample size $N = \sum_{i=1}^{K} n_i$ and $I^2 \to 1$.

Bagheri et al. (2011) compares the empirical power and type I error of the Breslow-Day test statistic, Q statistic and likelihood ratio test for testing the homogeneity of odds ratios among K studies for unequal sizes of the arms within and between studies. For large equal sample sizes Breslow-Day test statistic performs better than Q statistic and likelihood ratio test. In terms of power the Breslow-Day test has shown the lowest reduction in the balanced case compared to Q statistic and likelihood ratio test. The conclusion of Bagheri et al. (2011) is that Breslow-Day test is the most appealing in terms of the type I error and power. According to Jones et al. (1989) and O'Gorman et al. (1990), Breslow-Day test has lower power than tests proposed by Liang and Self (1985), when the number of studies K is increasing and sample size is fixed in sparse data. Thus, for $K \to \infty$ and fixed sample sizes n_{ij}, the Breslow-Day

test is invalid for testing the homogeneity of effects. However, in the simulations by Jones et al. (1989), the Breslow-Day test performed the best for the large and intermediate stratum setting with increasing n_j and fixed K. Jones et al. (1989) and O'Gorman et al. (1990) only recommend Breslow-Day test for non-sparse data. When the data are sparse, Liang and Self (1985) examines five tests for homogeneity of odds ratios by using Monte Carlo experiments. Gavaghan et al. (2000) compares five tests for homogeneity of dichotomous outcome measures on the log-odds scale and the risk difference scale. Out of five tests for homogeneity, in terms of power, Gavaghan et al. (2000) recommends Breslow-Day test based on Mantel-Haenzsel odds ratio. Other alternatives for testing the homogeneity of odds ratios include Tarone's approximate score test (Tarone, 1985) and the likelihood ratio test (Hardy and Thompson, 1998).

Some improvements have been suggested for Q test by Lipsitz et al. (1998) for sparse data and Kulinskaya et al. (2011a) under the null hypothesis for risk differences, Takkouche et al. (1999) and Kulinskaya and Dollinger (2015) for odds ratio. Hardy and Thompson (1996) state that Q test has a low power. The power of the Q statistic is discussed by Hedges and Pigott (2001), Valentine et al. (2010), Biggerstaff and Jackson (2008), Hardy and Thompson (1998) and Jackson (2006). Paul and Donner (1989) recommends Tarone's approximate score test, based on the Mantel-Haenszel estimator of the common odds ratio. Some of these tests for example three-way interaction test in logistic regression has a low power (Thompson, 1994). Bagheri et al. (2011) compares Dersimonian-Laird, Breslow-Day and likelihood ratio test in a mixed logistic model. They show that in terms of power Breslow-Day test outperforms DerSimonian-Laird and likelihood ratio test statistic. Also among the asymptotic tests Breslow-Day test was recommended by Reis et al. (1999) due to simplicity of calculation. Kulinskaya and Dollinger (2015) recommends Breslow-Day test for its superiocity in comparison to Q statistic.

These tests do not perform equally well for different types of effect measures. These tests answer the question whether or not the observed estimates for effect measure are heterogeneous between the studies. It might be more important to quantify the heterogeneity and explore

what are the sources of the heterogeneity.

2.6 Standard Random Effects Model

The random effects model is an alternative to fixed effect model. It assumes that K studies are randomly chosen from a population of studies. The random effects model was introduced by DerSimonian and Laird (1986). Random effects model inflates standard errors and confidence intervals for the overall effect measure in comparison to FEM. The models differ by the type of variation taken into account to estimate the combined effect measure. Random effects model introduces additional variance component for variation in effect measures between studies. Hence, the homogeneity assumption is replaced with its heterogeneity counterpart. The heterogeneity of effect measures is explained by the fact that now the population effect measures vary from study to study and are assumed to be random variables from some distribution. The heterogeneity between studies can be taken into account and quantified by additional variance component of random effects. Quantifying and explaining the heterogeneity plays a significant role in systematic reviews. A general random effect model has a form

$$\hat{\theta}_i \sim F(\theta_i, \sigma_i^2) \qquad \text{and} \qquad \theta_i \sim G(\theta, \tau^2). \tag{2.6.1}$$

The F and G distributions are commonly assumed to be normal. However, other combinations of distributions are permissible. If the within study distribution for each data set is assumed to be normal, then this assumption has to be justified by a choice of an appropriate effect measure and size of data in each study. The between study distribution is commonly assumed to be normal.

In some circumstances an assumption of normality for within study distribution might be problematic. For example, if the majority of studies are dominated by sparse events, some of the standard errors will be extremely variable or even become undefined (Stijnen et al., 2010). Stijnen et al. (2010) proposed the models with corresponding exact likelihood instead

of within study normal approximations. It is important to emphasize that F is a conditional distribution which depends on parameter of interest θ_i and usually does not condition on the within-studies variance σ_i^2. Variance σ_i^2 might or might not depend on the parameter θ_i. Assuming normal distributions for F and G, the resulting marginal random effects model with unconditional distribution for estimated effect measure $\hat{\theta}_i$ in the $i-th$ study is

$$\hat{\theta}_i \sim N(\theta, \sigma_i^2 + \tau^2), \qquad \text{for} \qquad i = 1, ..., K, \tag{2.6.2}$$

where θ is the weighted overall effect measure, σ_i^2 represents the variance of error term or the within-study variance of an effect measure $\hat{\theta}_i$ from each study and τ^2 is an unknown variance of random effect which describes between study variability and the heterogeneity of effect measures. Equivalently, the model can be defined in the form similar to fixed effect model (2.4.5)

$$\hat{\theta}_i = \theta_i + \epsilon_i,$$

with θ_i replacing the common θ due to assumption of heterogeneous effect measures instead of homogeneous ones in the fixed effect model. Redefining the true effects as $\theta_i = \theta + b_i$, the random effects model is

$$\hat{\theta}_i = \theta + b_i + \epsilon_i \tag{2.6.3}$$

with

$$b_i \sim N(0, \tau^2) \qquad \text{and} \qquad \epsilon_i \sim N(0, \sigma_i^2)$$

where b_i is a random effect independent from ϵ_i that introduces heterogeneity quantified by additional variance component τ^2. Also b_i represents an error term by which each individual effect measure θ_i differs from the common effect measure θ. In case of $\tau^2 = 0$, the model reduces to fixed effect model (2.4.5).

Usually in meta-analysis σ_i^2 is assumed to be known and includes the sampling size normalization. However if σ_i^2 is unknown, the unbiased estimator for σ_i^2 is treated as a true value ignoring any associated sampling errors. The problems with using the estimate for σ_i^2 might occur in small studies, where the sampling error of the estimate for σ_i^2 might be large. Other

issues are related to correlation between the estimate for effect measure and its variance, the use of continuity corrections or studies with few events where normal approximation is invalid (see Stijnen et al. (2010) for details). Another problem is that the assumption of normal distribution for log-odds-ratio might not be true for small studies. In our case, for the estimated log-odds-ratio, the within study variance is given by (2.4.2). Substitution of estimates for within-study variances can be problematic in studies with small sample sizes, because of low accuracy for estimators of variances. The low accuracy in estimators for σ_i^2 results in either positive or negative bias in $\hat{\sigma}_i^2$ and in overall effect measure $\hat{\theta}_{RE}$. Thus, the overall effect measure will be wrongly centred from the true effect measure with higher or lower bias respectively. Böhning and Sarol (2000) and Brockwell and Gordon (2001) discussed the danger of the assumption of known variance σ_i^2.

The main difficulty in random effects model lies in estimating the unknown between-studies variance τ^2. When τ^2 is estimated by $\hat{\tau}^2$, the overall estimate of an effect measure can be estimated by the weighted mean

$$\hat{\theta}_{RE} = \frac{\sum_{i=1}^{K} \hat{w}_i(\hat{\tau}^2)\hat{\theta}_i}{\sum_{i=1}^{K} \hat{w}_i} \qquad \text{with weights} \qquad \hat{w}_i = \frac{1}{\hat{\sigma}_i^2 + \hat{\tau}^2}. \tag{2.6.4}$$

The variance of $\hat{\theta}_{RE}$ is estimated by $\mathrm{Var}(\hat{\theta}_{RE}) = 1/\sum_{i=1}^{K} w_i(\hat{\tau}^2)$. This variance is obtained under assumption of known between-study variance τ^2, within-study variances σ_i^2 and hence weights of individual studies. The approximate large sample confidence interval for $\hat{\theta}_{RE}$ is obtained as $\hat{\theta}_{RE} \pm zW^{-1/2}$ with $W = \sum_{i=1}^{K}(\hat{\sigma}_i^2 + \hat{\tau}^2)^{-1}$. A confidence interval for the pooled overall effect measure can be obtained assuming to have its approximate normality. This is true for either approximately normal effects or large number of studies K. In order to account for uncertainty in τ^2, Higgins et al. (2009) suggests a t-distribution instead of normal for combined effect measure $\hat{\theta}_{RE}$. The distribution of the Wald statistic $T = \hat{\theta}_{RE}/\hat{SE}(\hat{\theta}_{RE})$ is studied by Raghunathan and Yoichi (1993), Berkey et al. (1995), Follmann and Proschan (1999) and Hartung and Knapp (2001). The likelihood-based approaches to account for uncertainty in

$\hat{\tau}^2$ are available (Hardy and Thompson, 1996; Vangel and Rukhin, 1999). Sidik and Jonkman (2002) proposed an alternative confidence interval for overall effect measure $\hat{\theta}_{RE}$ based on t distribution.

Different estimators for τ^2 are discussed in this thesis. The accuracy of every estimator for τ^2 depends on a method used for estimation and on how large number of studies K is. As we mentioned previously, the value of τ^2 shows the degree of heterogeneity between studies. We only assume the randomness of effect measures, i.e random effects model. $\tau^2 = 0$ implies that there exist no heterogeneity and the effect measures are homogeneous. In that case, the random effects model reduces to fixed effect model.

In some cases, the estimators for τ^2 may take on negative values given that positive total variance $\sigma_i^2 + \tau^2 > 0$. However, the common approach is the truncation at zero since the τ^2 defines the variance, generally it cannot take negative values, (Rukhin, 2013).

Fixed effect model underestimates the standard error of an overall effect by ignoring the variation between studies. In random effects model when τ^2 on n_i tend to infinity, variances σ_i^2 becomes small in comparison to τ^2 and the estimate for weighted average is

$$\hat{\theta}_{RE} = \lim \frac{\sum_{i=1}^{K}(\sigma_i^2 + \tau^2)^{-1}\hat{\theta}_i}{\sum_{i=1}^{K}(\sigma_i^2 + \tau^2)^{-1}} = \frac{1}{K}\sum_{i=1}^{K}\hat{\theta}_i, \quad \text{when} \quad \tau^2 \to \infty \quad \text{or} \quad \min(n_i) \to \infty$$

i.e. the weighted least squares estimator results in a simple average, (Demidenko, 2004). Due to bias in $\hat{\theta}_{RE}$, Shuster (2010) recommends the unweighted estimator for the mean instead of its weighted counterpart.

In the additive random effects model (2.6.2), τ^2 indicates the degree of heterogeneity. Thompson and Sharp (1999) proposed the multiplicative random effects model in the form

$$\hat{\theta}_i \sim N(\theta, \phi\sigma_i^2) \tag{2.6.5}$$

where ϕ is a multiplicative random effects parameter to incorporate heterogeneity through overdispersion. Parameter ϕ allows deflation and inflation in the variance of $\hat{\theta}_i$. The new version of random effects model based on overdispersion is proposed by Kulinskaya and Olkin

(2014). In this paper, the multiplicative parameter ϕ is defined as $\phi_i = 1 + a(n_i)\gamma$ where $a(n_i)$ are functions linearly dependent on sample sizes n_i. The key idea of this model is an interpretation of nuisance parameter γ as an intra-cluster correlation or transformation of it. In the same paper, for estimated two sample effect measure, the multiplicative random effects model is

$$\hat{\theta}_i \sim N(\theta, \frac{v_i(R_i)}{n_i}(1 + a_i\gamma)) \quad \text{for} \quad \gamma > \frac{-1}{\max(a_i)}$$

where $R_i = n_{i1}/n_{i2}$ is an allocation ratio of treatment to control group sizes and $v_i(R_i)$ is a precision measure. More details are available in the paper by Kulinskaya and Olkin (2014). Higgins et al. (2009) discusses the recent issues in standard two stage random effects model.

2.7 Point estimators for τ^2

Between study variance or heterogeneity variance τ^2 plays an important role in random effects model. Once τ^2 is known, the pooled effect measure can be estimated by weighted average of study specific estimates (2.6.4). Secondly, the value for τ^2 directly indicates the degree of heterogeneity in effect measures between studies. In meta-analysis, the value for τ^2 is never known. The standard practice is to substitute one of the estimators for τ^2 ignoring its variability. Henmi and Copas (2010) discusses the effect of this on confidence intervals for publication bias. Previously, several estimators for τ^2 have been proposed. The list of estimators for τ^2 include estimators derived by Hunter and Schmidt (1990), Hedges (1983), DerSimonian and Laird (1986), Mandel and Paule (1970) for interlaboratory studies, also likelihood-based estimators which require numerical maximization such as maximum likelihood estimator and restricted maximum likelihood estimator (Demidenko, 2004). One more class of estimators are Bayesian estimators. From the previous list, we briefly discuss only unbiased and efficient estimators for τ^2 with their moments, mean square errors and confidence intervals. Viechtbauer (2005) compared theoretically and through simulations estimators for τ^2 proposed by Hunter and Schmidt (1990), Hedges (1983), DerSimonian and Laird (1986), and also maximum likelihood and the restricted maximum likelihood estimators for the unstandardized and

the standardized mean difference as an effect measure.

2.7.1 Hedges Estimator

The unweighted method of moments estimator of τ^2 or an analysis of variance estimator for τ^2 derived by Hedges (1983) is

$$\hat{\tau}^2_{HE} = \frac{\sum_{i=1}^{K}(\hat{\theta}_i - \bar{\theta}_{uw})^2}{K-1} - \frac{1}{K}\sum_{i=1}^{K}\hat{\sigma}^2_i,$$

where $\bar{\theta}_{uw}$ is an unweighted average of the estimated effect measures $\hat{\theta}_i$ for $i = 1, .., K$. This estimator is also known as minimum norm quadratic unbiased estimator (MINQUE Rao (2009), Demidenko (2004))

The main advantage of estimator $\hat{\tau}^2_{HE}$ for τ^2 is its unbiasness for both known σ^2_i or with substituted unbiased estimates of σ^2_i instead. The estimator $\hat{\tau}^2_{HE}$ might yield negative values in which case it has to be truncated to zero. The estimator $\hat{\tau}^2_{HE}$ is unbiased before truncation assuming that the sampling variances $\hat{\sigma}^2_i$ are known (Viechtbauer (2005); Veroniki et al. (2015)). The estimator $\hat{\tau}^2_{HE}$ is consistent when $K \to \infty$ since the variance $\hat{\tau}^2_{HE}$ is of order (1/K). The sampling variance of $\hat{\tau}^2_{HE}$ is given by

$$\mathrm{Var}(\hat{\tau}^2_{HE}) = \frac{2}{(K-1)^2}\left[(1-\frac{2}{K})\sum_{i=1}^{K}(\sigma^2_i+\tau^2)^2 + \frac{(\sum_{i=1}^{K}(\sigma^2_i+\tau^2))^2}{K^2}\right],$$

(Friedman, 2000; Viechtbauer, 2007).

2.7.2 DerSimonian and Laird estimator

Non-iterative method, which gives an unbiased estimate of τ^2 for known σ^2_i was proposed by DerSimonian and Laird (1986) and later studied by Whitehead and Whitehead (1991). This is the weighted method of moments estimator for τ^2. This method is based on Cochran's Q statistic

$$Q = \sum_{i=1}^{K} w_i(\theta_i - \theta_{FE})^2 \sim \chi^2_{K-1}$$

where w_i are inverse variance weights and θ_{FE} is the weighted average of effect measures with $w_i = \sigma_i^{-2}$ from fixed effect model, see Section (2.5.1) for details. The Der-Simonian and Laird estimator for τ^2 is

$$\hat{\tau}^2_{DL} = [Q - K + 1] \Big/ \Big[\sum_{i=1}^{K} w_i - \frac{\sum_{i=1}^{K} w_i^2}{\sum_{i=1}^{K} w_i} \Big]. \tag{2.7.1}$$

The DerSimonian and Laird estimator is obtained by equating an estimate of expected value of Q to its observed value. The DerSimonian and Laird estimator is unbiased and consistent for $K \to \infty$ under assumption that σ_i^2 are known. It is possible to obtain negative values for $\hat{\tau}^2_{DL}$, in that case it has to be truncated $\hat{\tau}^2_{DL} = \max(0, \hat{\tau}^2_{DL})$. Due to truncation, the estimator $\hat{\tau}^2_{DL}$ obtains positive bias. According to Biggerstaff et al. (1997) and Viechtbauer (2005) the sampling variance of $\hat{\tau}^2_{DL}$ is approximated to order $O(1/K)$ by

$$\mathrm{Var}(\hat{\tau}^2_{DL}) = \frac{2}{\big[\sum_{i=1}^{K} w_i - \frac{\sum_{i=1}^{K} w_i^2}{\sum_{i=1}^{K} w_i} \big]} \Big[\sum_{i=1}^{K} w_i^2 (\sigma_i^2 + \tau^2)^2 - 2 \frac{\sum_{i=1}^{K} w_i^3 (\sigma_i^2 + \tau^2)^2}{\sum_{i=1}^{K} w_i} + \frac{\big(\sum_{i=1}^{K} w_i^2 (\sigma_i^2 + \tau^2) \big)^2}{\big(\sum_{i=1}^{K} w_i \big)^2} \Big].$$

Friedman (2000) derived alternative form of the variance $\mathrm{Var}(\hat{\tau}^2_{DL})$. Even though $\hat{\tau}^2_{DL}$ is unbiased and consistent for known σ_i^2, it is common practice to substitute unbiased estimators for σ_i^2 when the true values of within-study variance are unknown. Böhning et al. (2002) discusses the issues of using estimated study-specific variances instead of theoretical values of counterparts for σ_i^2 which can result in bias in estimating $\hat{\tau}^2_{DL}$.

In general, the DerSimonian and Laird method underestimates the true value for between-study variance τ^2, (Böhning et al., 2002; DerSimonian and Laird, 1986; DerSimonian and Kacker, 2007). Also, Brockwell and Gordon (2001) shows some deficiencies of $\hat{\tau}^2_{DL}$ in their simulations for small values of K, such as $K < 20$. The large mean-squared error of τ^2_{DL} is obtained in the simulations by Malzahn et al. (2000), Jackson et al. (2010) and from theory in Rukhin (2013).

2.7.3 The Mandel-Paule estimator

The random effects model has also become popular in interlaboratory studies, where the outcome $\hat{\theta}_i$ is an estimate of interlaboratory effect measure or so called consensus value with estimated sample variance $\hat{\sigma}_i^2$. In order to obtain the weighted average $\hat{\theta}_{RE}$ of consensus values, Mandel and Paule (1970) introduced the Mandel Paule algorithm for estimating the parameter for between study variability in analysis of interlaboratoty studies by solving iteratively the equation

$$Q(\tau^2) = \sum_{i=1}^{K} \frac{(\theta_i - \hat{\theta}_{RE})^2}{\sigma_i^2 + \tau_{MP}^2} = K - 1. \tag{2.7.2}$$

The Mandel-Paule method for estimating τ^2 produces a moment type estimator. Function $Q(\tau^2)$ for τ^2 is the Cochran's Q statistic with weights including τ^2 under alternative hypothesis about the presence of heterogeneity. It is a convex monotonically decreasing function of $\tau^2 > 0$, (Rukhin, 2009). The difference between Mantel-Paule and Der-Simonian and Laird method is that Mandel-Paule algorithm uses weights $w_i = (\tau_{MP} + \hat{\sigma}_i^2)^{-1}$ instead of $w_i = (\hat{\sigma}_i^2)^{-1}$. Rukhin and Vangel (1998) and Rukhin et al. (2000) shows that (2.7.2) has at most one positive solution for τ^2 and estimated τ_{MP}^2 can be interpreted as an approximation to REML estimator and as a generalized Bayes estimator (Morris, 1983). The Mandel-Paule estimator for τ^2 was proposed for use in meta-analysis by Rukhin (2003) and DerSimonian and Kacker (2007). Rukhin (2003) compares theoretical properties of estimates for τ^2 by Mandel-Paule method and DerSimonian and Laird method. Rukhin (2003) suggested the modified version of Mandel-Paule algorithm, with K instead $K - 1$ in (2.7.2). For $K = 2$, the Mandel-Paule estimator of the between studies variance coincides with Der-Simonian and Laird estimator and Hedges estimator as

$$\hat{\tau}_{MP}^2 = \hat{\tau}_{DL}^2 = \hat{\tau}_{HE}^2 = \frac{1}{2}\max[0, (\theta_1 - \theta_2)^2 - \hat{\sigma}_1^2 - \hat{\sigma}_2^2].$$

2.7.4 Maximum likelihood estimator of θ and τ^2

The likelihood-based inference is possible under fixed effect and random effects models, since we have distributional assumption for estimates of effect measures $\hat{\theta}_j$ between studies. For the

normal G and F in (2.6.1), we discuss the method for estimation of θ and τ^2 by maximising the log-likelihood function. The maximum likelihood-based estimators for random effects model in meta-analysis were proposed by Hardy and Thompson (1996), Harville (1977) and Raudenbush and Bryk (1985). The maximum likelihood estimators require iterative numerical solution. For the model (2.6.1), the joint likelihood function of θ and τ^2 is

$$L(\theta, \tau^2) \sim \prod_{i=1}^{K} \int_{-\infty}^{\infty} F_i(\cdot|\theta_i) G(\theta_i|\theta, \tau^2) d\theta_i \tag{2.7.3}$$

where $F_i(\cdot|\theta_i)$ is a conditional likelihood function with a true unobserved effect θ_i. The choice of distribution should be performed with care since the distribution might concern different variables of the study. $G(\theta_i|\theta, \tau^2)$ is a function of true unobserved effect measure θ_i for normal density with parameters θ and τ^2. This likelihood function is valid for likelihood-based inference about θ and τ^2. Examples of the joint likelihood $F_i(\cdot|\theta_i)$ in (2.7.3) are normal distribution or non-central hypergeometric distribution in case of combining the logarithms of odds ratios. The normal distribution is applied to the estimators of the log-odds-ratio $\hat{\theta}_i$. The non-central hypergeometric distribution is the distribution of the number of events in treatment arm given the fixed margins. The model with a non-central hypergeometric for $F_i(\cdot|\theta_i)$ and normal distribution for $G(\theta_i|\theta, \tau^2)$ is discussed in Chapter 5.

Under assumption of normal distribution for F and G (2.6.1), the marginal distribution of $\hat{\theta}_i$ is normal with mean θ and variance $\sigma_i^2 + \tau^2$. The marginal density function for $\hat{\theta}_i$ is

$$f(\hat{\theta}_i, \theta, \sigma_i^2 + \tau^2) = (2\pi(\sigma_i^2 + \tau^2))^{-\frac{1}{2}} \exp(-\frac{1}{2}\frac{(\hat{\theta}_i - \theta)^2}{\sigma_i^2 + \tau^2}).$$

The product of marginal likelihoods for K studies is

$$L = \prod_{i=1}^{K} f(\hat{\theta}_i, \theta, \sigma_i^2 + \tau^2).$$

Ignoring the constant term $-\frac{1}{2}\sum_{i=1}^{K} \log(2\pi)$, the log likelihood function for standard additive random effects model with a mix of two normal distributions is

$$l(\theta, \tau^2) = -\frac{1}{2}[\sum_{i=1}^{K}(\log(\sigma_i^2 + \tau^2) + \frac{(\hat{\theta}_i - \theta)^2}{\sigma_i^2 + \tau^2})]. \tag{2.7.4}$$

The score functions for θ and τ^2 are

$$U(\theta) = \frac{dl(\theta,\tau^2)}{d\theta} = \sum_{i=1}^{K} \frac{(\hat{\theta}_i - \theta)}{\sigma_i^2 + \tau^2} = 0 \quad U(\tau^2) = \frac{dl(\theta,\tau^2)}{d\tau^2} = \frac{1}{2}\sum_{i=1}^{K}\left[\frac{1}{\sigma_i^2+\tau^2} - \frac{(\hat{\theta}_i-\theta)^2}{(\sigma_i^2+\tau^2)^2}\right] = 0.$$

The maximum likelihood estimators for θ and τ^2 are

$$\hat{\theta}_{ML} = \frac{\sum_{i=1}^{K}(\sigma_i^2+\tau_{ML}^2)^{-1}\hat{\theta}_i}{\sum_{i=1}^{K}(\sigma_i^2+\tau_{ML}^2)^{-1}} \qquad \text{and} \qquad \hat{\tau}_{ML}^2 = \frac{\sum_{i=1}^{K} w_i^2[(\hat{\theta}_i - \hat{\theta}_{ML})^2 - \sigma_i^2]}{\sum_{i=1}^{K} w_i^2} \quad , \tag{2.7.5}$$

with $w_i = [\hat{\tau}_{ML}^2 + \sigma_i^2]^{-1}$. From the formula above, the maximum likelihood estimates for τ^2 and θ can be found iteratively. Firstly, τ^2 is treated as fixed and the value of θ maximising the log-likelihood is calculated. Next, θ is treated as fixed and the value of τ^2 maximising the log-likelihood is calculated. The iterations can be started from the method of moments estimator for τ^2 or setting $\tau^2 = 0$ (Erez et al., 1996). The iterations continue till convergence of parameters. Second derivatives for θ and τ^2 are

$$\frac{\partial^2 l(\theta,\tau^2)}{\partial\theta^2} = -\sum_{i=1}^{K}\frac{1}{\sigma_i^2+\tau^2}, \qquad \frac{\partial^2 l(\theta,\tau^2)}{\partial\theta\partial\tau^2} = -\sum_{i=1}^{K}\frac{\hat{\theta}_i-\theta}{(\sigma_i^2+\tau^2)^2},$$

$$\frac{\partial^2 l(\theta,\tau^2)}{\partial(\tau^2)^2} = -\frac{1}{2}\sum_{i=1}^{K}\left[\frac{1}{(\sigma_i^2+\tau^2)^2} - \frac{2(\hat{\theta}_i-\theta)^2}{(\sigma_i^2+\tau^2)^3}\right]$$

Using the obtained derivatives, the Hessian matrix for the log likelihood function is

$$\mathbf{H}(\eta) = \begin{pmatrix} \frac{\partial^2 l}{\partial\theta^2} & \frac{\partial^2 l}{\partial\theta\partial\tau^2} \\ \frac{\partial^2 l}{\partial\tau^2\partial\theta} & \frac{\partial^2 l}{\partial(\tau^2)^2} \end{pmatrix} = -\begin{pmatrix} \sum_{i=1}^{K}\frac{1}{\sigma_i^2+\tau^2} & \sum_{i=1}^{K}\frac{\hat{\theta}_i-\theta}{(\sigma_i^2+\tau^2)^2} \\ \sum_{i=1}^{K}\frac{\hat{\theta}_i-\theta}{(\sigma_i^2+\tau^2)^2} & \frac{1}{2}\sum_{i=1}^{K}[\frac{2(\hat{\theta}_i-\theta)^2}{(\sigma_i^2+\tau^2)^3} - \frac{1}{(\sigma_i^2+\tau^2)^2}] \end{pmatrix}.$$

The two most common methods for maximising the log-likelihood iteratively are Newton-Raphson and Fisher scoring algorithms. The general Newton-Raphson algorithm for $\eta = (\theta, \tau^2)$ and score function $U(\eta) = (U(\theta), U(\tau^2))$ is

$$\eta_{i+1} = \eta_j + [H(\eta)^{-1}U(\eta)]_i$$

where $H(\eta)$ is the observed Hessian information matrix. The Newton-Raphson algorithm might fail in maximising the log likelihood function for $\theta_i \sim N(\theta, \sigma_i^2 + \tau^2)$, when the starting

point is far from maximum. In order to obtain the convergence of the maximum likelihood estimators for $\hat{\theta}_{ML}$ and $\hat{\tau}^2_{ML}$ using the Newton-Raphson algorithm, the determinant of matrix $H(\eta)$ should remain positive, hence matrix $H(\eta)$ should remain as positive definite matrix. Taking the expectation of Hessian matrix and multiplying by -1, the Fisher information matrix is written as

$$\mathbf{I}(\eta) = -\mathbf{E}(\mathbf{H}(\eta)) = \begin{pmatrix} \sum_{i=1}^{K} \frac{1}{\sigma_i^2+\tau^2} & 0 \\ 0 & \frac{1}{2}\sum_{i=1}^{K} \frac{1}{(\sigma_i^2+\tau^2)^2} \end{pmatrix}.$$

The elements of information matrix are always positive, hence the matrix itself is positive-definite. Thus, the Fisher scoring algorithm is more appropriate to use in this case than Newton-Raphson algorithm. Inverting the information matrix, the asymptotic covariance matrix is

$$\mathbf{I}(\eta)^{-\mathbf{1}} = \begin{pmatrix} (\sum_{i=1}^{K} \frac{1}{\sigma_i^2+\tau^2})^{-1} & 0 \\ 0 & 2(\sum_{i=1}^{K} \frac{1}{(\sigma_i^2+\tau^2)^2})^{-1} \end{pmatrix}.$$

$(2,2)_{th}$ term of the matrix $I(\eta)^{-1}$ provides a large-sample approximation for the variance of the estimated heterogeneity parameter.

The general Fisher score algorithm is a modified Newton-Raphson algorithm with expected information matrix

$$\eta_{i+1} = \eta_i + [I(\eta)^{-1}U(\eta)]_i.$$

The individual Fisher's scoring algorithms for θ_{ML} and τ^2 are

$$\hat{\theta}_{s+1} = \hat{\theta}_s + (\sum_{i=1}^{K} \frac{1}{\hat{\tau}^2+\sigma_i^2})^{-1} \sum_{i=1}^{K} \frac{\theta_i - \theta_s}{\hat{\tau}^2+\sigma_i^2}$$

and

$$\hat{\tau}^2_{s+1} = \hat{\tau}^2_s + (\sum_{i=1}^{K} \frac{1}{(\hat{\tau}^2+\sigma_i^2)^2})^{-1} \sum_{i=1}^{K} [\frac{\theta_i - \theta_s}{(\hat{\tau}^2+\sigma_i^2)^2} - \frac{1}{\hat{\tau}^2+\sigma_i^2}],$$

with iteration index s (Demidenko, 2004). The algorithm starts with $\hat{\tau}^2_0 = 0$. Demidenko (2004) formulates in his book (Demidenko, 2004, page 253) a condition for the maximum

likelihood estimate for τ^2 being positive as

$$\sum_{i=1}^{K} \frac{(\theta_i - \hat{\theta}_{ML})^2}{\sigma_i^4} > \sum_{i=1}^{K} \frac{1}{\sigma_i^2}.$$

Assuming that the within-study variance of each study remains constant between the studies i.e $\sigma_i^2 = \sigma^2$, the closed form solution for the MLE of the between-study variance is

$$\hat{\tau}^2 = \frac{1}{K} \sum_{i=1}^{K} (\theta_j - \hat{\theta}_{ML})^2 - \sigma^2.$$

Maximum likelihood estimator $\hat{\tau}^2_{ML}$ has a downward bias (Viechtbauer, 2005), which affects the hypothesis testing and confidence intervals of the effect measure $\hat{\theta}_{ML}$. By taking the inverse of the Fisher information matrix and assuming normality for $\hat{\tau}^2_{ML}$, we can find the Wald-type confidence interval for τ^2. The asymptotic variance of maximum likelihood-based estimator τ^2_{ML} is

$$\text{Var}(\hat{\tau}^2_{ML}) = 2(\sum_{i=1}^{K} w_i^2)^{-1} \qquad \text{for} \qquad w_i = \frac{1}{\hat{\tau}^2_{ML} + \sigma_i^2}.$$

2.7.5 Restricted maximum likelihood estimator of θ and τ^2

The restricted or residual maximum likelihood (REML) is a modification of standard likelihood using the generalized least square residuals. The REML is called restricted or residual due to maximization of marginal log-likelihood function for the residuals from a (generalised) least squares fit of the model. The REML is a function of variance components only. The main advantage of using the REML is that it corrects for the downward bias of maximum likelihood estimates for these components.

The restricted maximum likelihood for the model (2.6.2) is

$$l(\theta, \tau^2) = -\frac{1}{2} \sum_{i=1}^{K} [\log(\sigma_i^2 + \tau^2) + \frac{(\theta_i - \theta)^2}{\sigma_i^2 + \tau^2} + \log(\sum_{i=1}^{K} (\sigma_i^2 + \tau^2)^{-1})] \tag{2.7.6}$$

The difference between standard maximum likelihood and the restricted maximum likelihood is the additional term $\log(\sum_{i=1}^{K} (\sigma_i^2 + \tau^2)^{-1})$. The restricted maximum likelihood estimator for

τ^2 is

$$\hat{\tau}^2_{REML} = \frac{\sum_{i=1}^{K} w_i^2[(\theta_i - \hat{\theta}_{REML})^2 - \sigma_i^2]}{\sum_{i=1}^{K} w_i^2} + \frac{1}{\sum_{i=1}^{K} w_i}. \tag{2.7.7}$$

Similarly to maximum likelihood estimator, $\hat{\tau}^2_{REML}$ can be estimated iteratively with the starting point of either moment based estimator or $\hat{\tau}^2_0 = 0$. The condition for the restricted maximum likelihood estimate of τ^2 to stay positive is

$$\sum_{i=1}^{K} \sigma_i^{-4}(\theta_i - \hat{\theta}_{REML})^2 > \sum_{i=1}^{K} \sigma_i^{-2} - \frac{\sum_{i=1}^{K} \sigma_i^{-4}}{\sum_{i=1}^{K} \sigma_i^{-2}},$$

see (Demidenko, 2004, page 255). The restricted maximum likelihood estimate for a case of constant within studies variance $\sigma_i^2 = \sigma^2$ is

$$\hat{\tau}^2 = \frac{1}{K-1}\sum_{i=1}^{K}(\theta_i - \hat{\theta}_{REML})^2 - \sigma^2. \tag{2.7.8}$$

The variance of $\hat{\tau}^2_{REML}$, obtained from inverted Fisher information matrix is

$$\mathrm{Var}(\hat{\tau}^2_{REML}) = 2[\sum_{i=1}^{K} w_i^2 - 2\frac{\sum_{i=1}^{K} w_i^3}{\sum_{i=1}^{K} w_i} + \frac{(\sum_{i=1}^{K} w_i^2)^2}{(\sum_{i=1}^{K} w_i)^2}].$$

with $w_i = \frac{1}{\hat{\tau}^2_{REML}+\sigma_i^2}$. The approximation for restricted maximum likelihood estimator τ^2_{REML} is suggested by Morris (1983) in the Bayes setting.

The empirical evidence indicates that the restricted maximum likelihood estimator $\hat{\tau}^2_{REML}$ is approximately unbiased as opposed to negatively biased maximum likelihood estimator $\hat{\tau}^2_{ML}$. The approximate unbiasedness of $\hat{\tau}^2_{REML}$ is shown by Viechtbauer (2005). On the other hand, REML estimator is less efficient than regular maximum likelihood estimator $\hat{\tau}^2_{ML}$, because it has greater sampling variance, Viechtbauer (2005). However, among all unbiased estimators for τ^2, Viechtbauer (2005) recommends to use $\hat{\tau}^2_{REML}$, because of its balance between approximate unbiasedness and efficiency.

2.7.6 Sidik-Johnkman estimator of τ^2

Sidik and Jonkman (2005) proposed a simple variance estimator

$$\hat{\tau}^2_{SJ} = \frac{\sum_{i=1}^{K} \hat{v}_i(\hat{\theta}_i - \hat{\theta}_{RE})^2}{K-1} \tag{2.7.9}$$

where $\hat{v}_i = r_i + 1$, $r_i = \hat{\sigma}^2_i/\hat{\tau}^2_0$, and $\hat{\tau}^2_0$ is the initial naive estimator of between-study variance τ^2 defined by

$$\hat{\tau}^2_0 = \frac{\sum_{i=1}^{K}(\hat{\theta}_i - \hat{\theta}_{uw})^2}{K}$$

where $\hat{\theta}_{uw}$ is the unweighted mean of effect measures across K studies. Sidik and Jonkman (2005) compared the performance of $\hat{\tau}^2_{SJ}$ and $\hat{\tau}^2_{DL}$ by simulation study. For moderate to large values of heterogeneity, Sidik and Jonkman (2005) conclude that $\hat{\tau}^2_{SJ}$ is less biased than $\hat{\tau}^2_{DL}$. However, for small values of $\hat{\tau}^2_{DL}$, $\hat{\tau}^2_{DL}$ performs better than $\hat{\tau}^2_{SJ}$. $\hat{\tau}^2_{SJ}$ is very simple estimator of τ^2. Another advantage of $\hat{\tau}^2_{SJ}$ is that it always produces non-negative value for τ^2. Thus, no truncation at zero is required.

2.8 Confidence intervals for τ^2

It is important to provide information about the uncertainty about the τ^2. Therefore in this section, we provide confidence intervals for $\hat{\tau}^2$.

2.8.1 Wald-type confidence intervals

Using asymptotic normality property of ML and REML estimators, 95 % maximum likelihood and restricted maximum likelihood Wald type confidence intervals for $\hat{\tau}^2$ are

$$\hat{\tau}_{ML} \pm 1.96\sqrt{\text{Var}(\hat{\tau}_{ML})}$$

and

$$\hat{\tau}_{REML} \pm 1.96\sqrt{\text{Var}(\hat{\tau}_{REML})}$$

2.8.2 Q profile confidence interval

Viechtbauer (2007) proposed a method for constructing confidence interval for the heterogeneity parameter τ^2 based on inverting the Q test statistic. Under random effects model, the generalized Q-statistic for testing the heterogeneity is

$$Q(\tau^2) = \sum_{i=1}^{K} \frac{(\theta_i - \hat{\theta}_{RE})^2}{\tau^2 + \sigma_i^2} \sim \chi^2_{K-1}$$

where $\hat{\theta}_{RE} = \sum_{i=1}^{K} w_i\theta_i / \sum_{i=1}^{K} w_i$ is the weighted average of effect measures and $w_i = \frac{1}{\tau^2+\sigma_i^2}$ are the weights. Since

$$P(\chi^2_{K-1;0.025} \leq Q(\tau^2) \leq \chi^2_{K-1;0.975}) = 0.95,$$

the 95 % Q profile confidence intervals for τ^2 can be found from lower and upper quantiles of χ^2_{K-1} distribution

$$Q(\tau^2_L) = \chi^2_{K-1;0.975} \qquad Q(\tau^2_U) = \chi^2_{K-1;0.025} \tag{2.8.1}$$

The upper and lower bounds for τ^2 can be calculated iteratively for increasing values of τ^2.

2.8.3 Profile likelihood confidence interval

The profile likelihood confidence interval is proposed by Hardy and Thompson (1996). The confidence interval based on $\hat{\tau}^2_{REML}$ can be estimated from the likelihood (2.7.6). A 95 per cent confidence interval for τ^2 is given by set of values which satisfy

$$l_R(\hat{\tau}^2) > l_R(\hat{\tau}^2_{REML}) - \frac{1}{2}C_{0.95}(\chi^2_1) \tag{2.8.2}$$

where $C_{0.95}(\chi^2_1)$ is the 0.95 quantile of the χ^2_1 distribution and l_R is the restricted likelihood ratio test statistic. The distribution

$$-2\log(\frac{l_R(\tau^2)}{l_R(\tau^2_{REML})}) \to \chi^2_1 \qquad \text{for} \qquad K \to \infty$$

where $l_R(\tau^2)$ is the restricted maximum likelihood function calculated at the τ^2 and τ^2_{REML}. The profile likelihood confidence interval might not be centred at τ^2_{REML} due to absence of symmetry.

According to Hardy and Thompson (1998), the advantage of profile likelihood confidence interval is that it takes into account the absence of known parameters and the uncertainty of estimator for τ^2. Sørensen (2008) showed the correspondence between the distribution of likelihood ratio test and τ^2. For the confidence interval with higher order asymptotic properties see Sharma and Mathew (2011). An interval centred at the fixed effect model combined effect and based on the conditional distribution of Q statistic is suggested by Henmi and Copas (2010). Interval suggested by Henmi and Copas (2010) performs well when publication bias is present.

2.8.4 Biggerstaff-Tweedie confidence interval for τ^2

According to Biggerstaff et al. (1997), the expected value and variance of Q statistic are given by

$$E(Q) = (K-1) + (S_1 + \frac{S_2}{S_1})\tau^2$$

and

$$\mathrm{Var}(Q) = 2(K-1) + 4(S_1 + \frac{S_2}{S_1})\tau^2 + 2(S_2 - 2\frac{S_3}{S_2} + \frac{S_2^2}{S_1^2})\tau^4$$

where $S_t = \sum_{i=1}^{K} w_i^t$. Hence, distribution of the Q statistic can be approximated by a gamma distribution with the shape and scale parameters

$$\gamma(\tau^2) = \frac{(E(Q))^2}{\mathrm{Var}(Q)} \qquad \text{and} \qquad \phi(\tau^2) = \frac{\mathrm{Var}(Q)}{E(Q)},$$

respectively. The lower and upper bounds for τ^2 can be found iteratively from

$$\int_{Q/\phi(\tau^2)}^{\infty} f(x|\gamma(\tau^2))dx = 0.025 \qquad \text{and} \qquad \int_{0}^{Q/\phi(\tau^2)} f(x|\gamma(\tau^2))dx = 0.025,$$

where $f(x|\gamma(\tau^2))$ is the density function for a gamma distribution with shape parameters $\gamma(\tau^2)$ and scale parameter 1. The obtained lower and upper bounds for τ^2 are constrained to non-negative values.

2.8.5 Sidik-Johnkman confidence interval

Using the estimator 2.7.9 for τ^2 derived by Sidik and Jonkman (2005) and the assumption that $(K-1)\hat{\tau}^2_{SJ}/\hat{\tau}^2 \sim \chi^2_{K-1}$, the suggested 95 % confidence interval for τ^2 is

$$\left(\frac{(K-1)\hat{\tau}^2_{SJ}}{\chi^2_{K-1;0.975}}, \frac{(K-1)\hat{\tau}^2_{SJ}}{\chi^2_{K-1;0.025}}\right)$$

Compared to moment based estimator of τ^2, the estimator $\hat{\tau}^2_{SJ}$ is always positive. This implies that the lower and upper bounds for τ^2 are also positive

2.8.6 Parametric and non-parametric bootstrap confidence interval

Viechtbauer (2007) described two methods of obtaining the confidence interval for τ^2 based on parametric and non-parametric bootstrap. In parametric bootstrap method, K values of θ_i for each iteration $b = 1, \ldots, B$ are generated from $N(\hat{\theta}_{RE}, \sigma^2 + \hat{\tau}^2)$. Estimated values of $\hat{\tau}^2_b$ and $\hat{\theta}_{RE}$ can be obtained from any method providing a non-negative consistent estimator. Repeating the bootstrap process B times, we obtain B values of $\hat{\tau}^2_b$. A 95 % parametric bootstrap confidence interval is obtained by the $2.5th$ and $97.5th$ empirical percentiles of the bootstrapped τ^2_b values.

A non-parametric bootstrap confidence interval for τ^2_b is obtained in a similar way. Firstly, B estimates $\hat{\tau}^2_b$ are obtained by re-sampling the K values of θ_i's and corresponding σ^2_i's. Then, a 95 % parametric bootstrap confidence interval is given by the $2.5th$ and $97.5th$ empirical percentiles of these bootstrapped $\hat{\tau}^2_b$ values.

Any negative values for $\hat{\tau}^2_b$ can be left unchanged or truncated to zero.

2.9 Summary

In this chapter, the methods for meta-analysis of binary data have been reviewed. These methods assume either homogeneity or heterogeneity of effects across studies. For homogeneous effects, the methods were reviewed under fixed effect model. In particular, two main methods are the inverse-variance and the Mantel-Haenzsel method. For the heterogeneous effects, the

random effects model is an attractive alternative to fixed effect model. In random effects model, the between-study variance needs to be estimated. The methods of point and interval estimation of the between-study variance were discussed. For the point estimators of τ^2, Der-Simonian and Laird moment-based method, Restricted maximum likelihood and Mandel-Paule method produce the most reasonable results, (Viechtbauer, 2005). Der-Simonian and Laird method is non-iterative as opposed to parametric estimators derived from likelihood-based methods. DerSimonian and Kacker (2007) and Rukhin (2003) introduced to meta-analysis the Mandel-Paule method which produces a reasonable approximation to REML estimator (Rukhin et al., 2000). Friedman (2000) discusses Hedges and DerSimonian-Laird estimator, with suggestion that for small heterogeneity τ^2_{DL} is more efficient than τ^2_{HE} while τ^2_{HE} is more efficient for large variations between studies. Other point estimators of between-study variance include the estimators by Malzahn et al. (2000), Hartung and Makambi (2003), DerSimonian and Kacker (2007) and the class of estimators proposed by Rukhin (2013).

A number of other estimators for τ^2 are biased and are not discussed here because of the presence of estimation bias, for example, Hunter-Schmidt estimator $\hat{\tau}^2_{HS}$ (Schmidt and Hunter, 2014) is negatively biased, maximum likelihood estimator $\hat{\tau}^2_{ML}$ underestimates τ^2 with negatively biased variance (Viechtbauer, 2005). Bias may lead to wrong inference about the heterogeneity of effect measures and their central tendency. Any bias in estimation of τ^2 affects to the accuracy of pooled effect measure and its sampling variance. If the sampling variance of an estimate for τ^2 is positively or negatively biased, then it will result in over or underestimated overall effect measure $\hat{\theta}_{RE}$ and its sampling variance. On the other hand, when the unbiased estimate for τ^2 is substituted into the equation for estimation of overall effect measure $\hat{\theta}_{RE}$ and its variance, this results in a negatively biased estimate for the variance of $\hat{\theta}_{RE}$. Hence, in order to obtain a reasonable estimate for τ^2, the bias, efficiency and mean-squared error of $\hat{\tau}^2$ have to be assessed simultaneously in order to guarantee optimality for the estimate of τ^2. There exist no uniformly optimal estimator of τ^2 (Kulinskaya et al., 2014). We assess the performance of standard and new proposed estimators of between-study

variance in Chapter 6.

The reviewed methods for interval estimation of between-study variance include Wald-type confidence interval, inverted Q-profile confidence interval, Profile-likelihood confidence intervals, Biggerstaff-Tweedie and Sidik-Jonkman confidence intervals. It is an important task to be able to construct the confidence intervals for the heterogeneity parameter τ^2. By constructing the confidence interval, we can identify the accuracy of our estimator. Through analysis and simulation studies of different methods for interval estimation of τ^2, Viechtbauer (2007) suggests Q-profile and Profile-likelihood confidence intervals. Sidik and Jonkman (2002) suggested to replace of normal quantiles by t_{K-1} quantiles to account for variability in $\hat{\tau}^2$. The most recent review of the methods for estimation of between-study variance and its uncertainty is summarised in Veroniki et al. (2015).

After estimation of between-study variance, the estimator of overall effect measure is obtained usually by the inverse-variance method. The confidence interval for overall effect measure is given by Wald statistic under an assumption of normal distribution. For the Wald statistic based on the estimator $\hat{\theta}_{DL}$, Higgins et al. (2009) suggested an approximation by t distribution with degrees of freedom from $K-4$ to $K-1$.

The asymptotic properties of $\hat{\theta}_{RE}$ and τ^2 can be derived for three scenarios in a meta-analysis. The first scenario is when the number of studies K is increasing and sample sizes are fixed. The second scenario is when the number of studies K is fixed and sample sizes are increasing. The last case is when the number of studies and sample sizes are increasing simultaneously. For each asymptotic scenario, the inference in meta-analysis is different, since the approximations of exact distributions for the same statistic will be different (Kulinskaya et al., 2014). In general, for asymptotic results to hold, for studies with large K, large sample sizes are obligatory.

One of most important problem in the meta-analysis is the sparse data. We have discussed the use of continuity corrections in case of rare events. However, the continuity corrections

themselves might introduce a bias which results in wrong inference in a meta-analysis. Different ways of inferences for sparse data with either continuity correction or avoidance of them in meta-analysis have been discussed by many authors including Sankey et al. (1996), Zhou et al. (1999), Sweeting et al. (2004), Shuster et al. (2007), Friedrich et al. (2007), Bradburn et al. (2007), Rücker et al. (2009), Cai et al. (2010), Bhaumik et al. (2012), Böhning and Mylona (2015), Kuss (2015). Kuss (2015) summarized all the most recent non-standard methods for meta-analysis of binary data without using the continuity correction.

The continuity correction $c = 1/2$ still stays the most efficient correction in contingency tables. For arcsine, $c = 3/8$ is the most used continuity correction. For odds ratios, the Mantel-Haenzsel method seems to solve the problem, since no continuity corrections are required in this method. The Mantel-Haenzsel method under an assumption of fixed effect model is an attractive alternative to an inverse-variance method. The Mantel-Haenzsel method estimators are usually less biased than inverse-variance method estimators in the case of small sample sizes and/or small risks. Breslow (1981) studies the standard fixed effect methods for combining odds ratio. Breslow (1981) reports that for sparse data, the Mantel-Haenzsel-based estimator of odds ratio has the good efficiency in comparison to inverse-variance approach under fixed effect model. For Mantel-Haenzsel and DerSimonian-Laird method, the addition of standard continuity correction $c = 1/2$ was compared to a method of combining results without continuity correction by Sankey et al. (1996). The conclusion from simulations of comparison of Mantel-Haenszel odds ratio and Dersimonian and Laird odds ratio was that Mantel-Haenzsel odds ratio performs better in the presence of little heterogeneity without an addition of continuity correction, whereas for moderate and large heterogeneity the addition of continuity correction improves the coverage rates. Unfortunately, Mantel-Haenzsel method is applicable only in fixed effect model. The Mantel-Haenzsel method combines odds ratios themselves by inverse variance concept instead of using transformations into log scale. Transformation of random variables into different scale might result in transformation biases. The

issues of transformations biases are discussed in Chapter 3. Transformation of random variables is wide-spread in the meta-analysis. In randomised controlled trials, the effect is always given as a function of prevalences.

Chapter 3

Transformation bias

3.1 Introduction

The main focus of this chapter is the bias that arises as a result of transformations of random variables in random or mixed effects models and the deleterious effects of these biases on inference in meta-analyses. At the start, the theoretical derivation of the transformation bias is provided for the general transformation of single random variable. For a single sample, the influential paper by Cox (1983) investigated transformations in some detail.

Suppose that X_n is an unbiased estimator based on a sample of size n for some real parameter θ and furthermore that we are interested in the estimator $f(X_n)$ of the transformed parameter $\eta = f(\theta)$ for a nonlinear transformation $f(\cdot)$. The estimator $f(X_n)$ will then exhibit a finite-sample bias, but it retains consistency. If an unsuspected random effect is introduced, however, the estimator loses its consistency, because the bias is enlarged by overdispersion.

If the overdispersion is small and undetectable in the data, it may still severely affect the inference on transformed effects in a meta-analysis. Combining studies using meta-analytic methods are increasingly popular in biomedical applications. In epidemiology, studies are often based on routinely collected administrative unit level data, such as prevalence or incidence of a disease or condition in a population. Cluster randomised trials are motivated by the convenience of group-level treatment allocation. Meta-analyses aim to combine evidence from the existing studies. Requisite statistical methods are essentially the same; they are based on random or mixed effects models, and are well established, especially so in meta-analysis.

We illustrate our findings with the comparatively simple example of overdispersed binomial data, where overdispersion arises as a result of an intra-cluster correlation (ICC) ρ between Bernoulli random variables in cluster-randomised trials or within studies in meta-analyses. The most general assumptions about dependent Bernoulli variables are discussed in Chapter 2. Our general findings are based on the examples of biases from arcsine and logit (log-odds) transformations in single studies and in meta-analysis concentrating on the small values of the ICC $\rho < 0.1$. These transformations are very popular in analysis of binomial proportions (Kulinskaya et al., 2008, Ch.18) and they are also used in the other popular effect measures for binary data, such as the log-odds ratios or the differences of arcsine-transformed proportions (Hedges and Olkin, 1985; Rücker et al., 2009). The small value of intra-cluster correlation is chosen, since small ρ's commonly appear in bio-medical applications, where the number of clusters K is moderate to large. Clustering is mostly due to the same healthcare provider (health practitioner, general practice, clinic, etc.). Gulliford et al. (2005) analysed the data on 188 ICCs obtained from the General Practice Research Database (GPRD) for variation of outcomes and performance between United Kingdom general practices and 136 results from a Health Technology Assessment (HTA) review for a range of outcomes in community and health services settings. In the GPRD, the median prevalence p was 13.1% (interquartile range IQR 3.5 to 28.4%) and median ICC was 0.051 (IQR 0.011 to 0.094%). In the HTA review, the median prevalence was 6.5% (IQR 0.4% to 20.7%) and median ICC 0.006 (IQR 0.0003 to 0.036). Similarly, the paper by Littenberg and MacLean (2006) calculated the ICC for 62 binary variables measured as part of the Vermont Diabetes Information System, a cluster-randomized study of adults with diabetes from 73 primary care practices in Vermont, USA and surrounding areas. The median ICC was 0.022; IQR (0.006, 0.040). Prevalence of some comorbidities and complications and certain aspects of quality of life varied much more across patients with only small correlation within practices (ICC$<$ 0.001). Eldridge et al. (2004) provided a systematic review of 152 publised and 47 unpublished cluster randomized trials in primary health care, published between 1997 and 2000. The median number of clusters in

the published trials was 29, and in the unpublished trials it was 32. Our findings apply both to the standard additive random effects model (REM) of meta-analysis (Hedges and Olkin, 1985), and to the multiplicative version of REM (Kulinskaya and Olkin, 2014).

The structure of this chapter is as follows. The transformation bias is introduced in Section 3.2. Section 3.3.1 explores the consequences of the transformation bias when combining results in meta-analysis, and Section 3.5 applies our methodology to two examples of meta-analyses of prevalence of a disease or a condition. Final comments and discussion are given in Section 3.6. This Chapter represents the novel work of this thesis.

3.2 Theoretical derivation of transformation bias

Consider a real-valued statistic X, an estimator of a real parameter based on a sample of size n. Let $g_\tau(x)$ denote the density of X, where $\tau \geq 0$ is an overdispersion parameter. When $\tau = 0$, we have the null or "fixed effect" model with density $g(x) = g_0(x)$, and for $\tau > 0$ we have a "random effects" model (REM). Denote the expected value and the central moments of X by $x_\tau = \mathrm{E}_\tau(X)$ and $\mu_j(\tau) = \mathrm{E}_\tau((X - x_\tau)^j),\ j > 1$, respectively. We assume that all moments $\mu_j(\tau),\ j > 1$ of X exist for values of τ close to zero, and that the variance is of order $O(1/n)$ and the higher moments are $o(1/n)$. When we state "the variance is of order $O(1/n)$", this is one of the assumptions that we make about a real valued statistic X . We are mostly interested in statistics related to binary data. For example, in meta-analysis of binary data, p_{i1} and p_{i2} are the means of Bernoulli variables X_{i1k} and X_{i2k} and therefore variance is of order $O(1/n)$. Any effect measure such as log-odds ratio which is the transformation of p_{i1} and p_{i2} has the variance which has order $O(1/n)$. The same is true for relative risk and risk difference. This setting is similar to that of Cox (1983). Let $f(X)$ be a transformation of X such the derivatives of all orders exist. Think of X as an estimator of x_0 and of τ as an effect that is not part of the model used by the statistician. The expected value of the transformed

variable is

$$\mathrm{E}_\tau(f(X)) = \int f(x)g_\tau(x)dx = f(x_0) + \sum_{j=1}^{\infty} \frac{1}{j!}\frac{d^j f(x)}{dx^j}\Big|_{x_0} \mathrm{E}_\tau((X-x_0)^j)\,. \tag{3.2.1}$$

The first two terms of this series collect all the terms up to order $O(1/n)$ at the model $\tau = 0$

$$\mathrm{E}_\tau(f(X)) - f(x_0) = \frac{df(x)}{dx}\Big|_{x_0} \mathrm{E}_\tau(X-x_0) + \frac{1}{2}\frac{d^2 f(x)}{dx^2}\Big|_{x_0} \mathrm{E}_\tau(X-x_0)^2 + \text{ remainder }\,. \tag{3.2.2}$$

The left-hand side is the bias of $f(X)$ as an estimator of $f(x_0)$. The first term on the right-hand side measures the influence of the bias, and the second term the one of the mean squared error of X introduced by the "random effects" model. The formula shows that $f(X)$ is an unbiased estimator of $f(x_0)$ to order $O(1/n)$ only if X remains unbiased even under the REM, i.e. $x_\tau = x_0$, and if furthermore the transformation is linear.

The bias to order $O(1/n)$ can be calculated directly if the first two moments of X under the density $g_\tau(x)$ are known:

$$f'(x)\Big|_{x_0}(x_\tau - x_0) + \frac{1}{2}f''(x)\Big|_{x_0}(\mu_2(\tau) + (x_\tau - x_0)^2)\,.$$

Further, setting $\tau = 0$ shows that for a nonlinear transformation, $f(X)$ is in general not an unbiased estimator of $f(x_0)$. Now consider a similar expansion for the variance. Using (3.2.1) for $f^2(X)$, and subtracting the relevant order terms from $[\mathrm{E}_\tau[f(X)]^2]$, we obtain

$$\begin{aligned}\mathrm{Var}_\tau(f(X)) \;&= \mathrm{Var}_0(f(X)) + [f'(x)|_{x_0}]^2[\mu_2(\tau) - \mu_2(0)] \\ &\quad +2[f'(x)|_{x_0}][f(x_0) - \mathrm{E}_0(f(X))][x_\tau - x_0] \\ &\quad +\tfrac{1}{2}[f''(x)|_{x_0}][f(x_0) - \mathrm{E}_0(f(X))][\mu_2(\tau) - \mu_2(0)] \\ &\quad -[f''(x)|_{x_0}][x_\tau - x_0]^2 + \cdots\,.\end{aligned} \tag{3.2.3}$$

If τ omit itself of order $O(1/\sqrt{n})$, only the terms in the first line are of order $O(1/n)$, and the rest can be neglected.

Example 1: Transformation of a normal random variable

Consider a random variable X from a normal distribution with density $g_\tau(x) = \varphi(x_0, \sigma^2/n + \tau^2)$,

where $\varphi(\mu, \sigma^2)$ is the normal density with mean μ and variance σ^2. Here $x_\tau = x_0$, so X is unbiased for x_0 even under the REM, and the variance is $\sigma^2/n + \tau^2$. The first term in equation (3.2.2) is zero for $x_0 = \mathrm{E}_0(X)$, and the variance of X is $\sigma^2 + \tau^2$. Therefore,

$$\mathrm{E}_\tau(f(X)) = f(x_0) + \frac{\sigma^2 + \tau^2}{2} \frac{d^2 f(x)}{dx^2}|_{x_0} + o(1/n)$$

or

$$\mathrm{E}_\tau(f(X)) = f(x_0) + \frac{1}{2}(\mathrm{Var}_0(X) + \tau) \frac{d^2 f(x)}{dx^2}|_{x_0} + o(1/n). \tag{3.2.4}$$

In case of fixed effects, usually the second order term is neglected. However, under the assumption of random effects, the second order term consists of the random effect variance τ^2 in addition to $\mathrm{Var}_0(X)$, which is assumed to be non-zero. Hence, the overdispersion is of size τ and in a non-linear transformation it causes an added bias of size $0.5\tau^2 f''(x)|_{x_0}$. This does not tend to 0 with $n \to \infty$.

Example 2: Standard random effects model for log-odds

In meta-analysis, the following standard additive REM for the empirical log-odds is used routinely: $\hat{\theta} = X \sim N(\theta, [np(1-p)]^{-1} + \tau^2)$, where p is the probability of an event of interest. We are interested in going from the logit scale θ to the probability scale p. This transformation is

$$p = f(\theta) = [1 + \exp(-\theta)]^{-1},$$

and let its estimate $\hat{p}$ be the same function of $\hat{\theta}$. Therefore, $\hat{p}$ is biased in the standard REM model:

$$E_\tau(\hat{p}) - p = \frac{1}{2} f''(\theta)\{[np(1-p)]^{-1} + \tau^2\} = \frac{1}{2} p(1-p)(1-2p) \left[\frac{1}{np(1-p)} + \tau^2\right],$$

since $f'(\theta) = f(\theta) - f(\theta)^2$, and

$$f''(\theta) = \frac{\exp(-\theta)(\exp(-\theta) - 1)}{(1 + \exp(-\theta))^3} = p(1-p)(1-2p).$$

$\hat{p}$ is biased unless $p = 1/2$. For $p = 0.1$, this bias is $0.4/n + 0.036\tau^2$; and for $p = 0.2$, it is $0.3/n + 0.048\tau^2$. This is not a large bias but it matters in meta-analysis, as we shall see later. Note again the dramatic effect of the REM parameter τ. The similar expansion of $\hat{p}$ under the logistic-normal model is discussed in Hinde and Demétrio (1998). However, Hinde and Demétrio (1998) does not account for the second order terms which introduces a transformation bias.

Example 3: Overdispersed binomial model for log-odds

Consider the sample log-odds in any overdispersed binomial model, i.e. a model for a number of successes X out of n dependent Bernoulli variables. We denote the overdispersion parameter in this model by ρ instead of τ, as this corresponds to correlation between each pair of underlying Bernoulli variables. The transformation of interest is $f(p) = \log(p/(1-p))$. The derivatives are $f'(p) = [p(1-p)]^{-1}$ and $f''(p) = (2p-1)/[p(1-p)]^2$. When the log-odds transformation is applied to $\hat{p} = X/n$, the equation (3.2.2) is used with the first two moments of an overdispersed binomial distribution $\mu_\rho(\hat{p}) = p$ and $\mathrm{Var}_\rho(\hat{p}) = n^{-1}p(1-p)[1 + (n-1)\rho]$ to obtain

$$\begin{aligned} \mathrm{E}_\rho(\log(\tfrac{\hat{p}}{1-\hat{p}})) = \mathrm{E}_\rho(\log(\tfrac{X}{n-X})) \quad &= \log(\tfrac{p}{1-p}) + \tfrac{(2p-1)}{2p^2(1-p)^2}\mathrm{Var}_\rho(\hat{p}) + \cdots \\ &= \log(\tfrac{p}{1-p}) - \tfrac{(1-2p)(1+(n-1)\rho)}{2np(1-p)} + \cdots \end{aligned} .$$

Therefore, the sample log-odds has a bias term linear in ρ and of order $O(\rho)$ in the overdispersed binomial model. For $p = 0.1$, this bias is $-4.4(4)[1 + (n-1)\rho]/n$; and for $p = 0.2$, this bias is $-1.875[1 + (n-1)\rho]/n$.

The standard continuity correction for log-odds is to add $1/2$ to numerator and denominator, i.e. equivalently to use $\tilde{p} = (X + 1/2)/(n + 1)$ when estimating log-odds. As is well known, this continuity correction takes care of the $1/n$ bias term of the log-odds under the fixed effects, i.e. for $\rho = 0$. When using this correction, the bias term $f'(p)\mathrm{E}(\hat{p} - p) = [p(1-p)]^{-1}(1-2p)/[2(n+1)] + O(n^{-2})$ is added in the equation (3.2.2), and the variance is multiplied by $[n/(n+1)]^2$. We are left with

$$\mathrm{E}_\rho(\log(\frac{\tilde{p}}{1-\tilde{p}})) = \mathrm{E}_\rho(\log(\frac{X + 1/2}{n - X + 1/2})) =$$

$$\log(\frac{p}{1-p}) + \frac{(1-2p)}{2p(1-p)(n+1)} - \frac{(1-2p)n(1+(n-1)\rho)}{2p(1-p)(n+1)^2} + \cdots .$$

To assess the precision of these two-moment approximations to bias with and without the continuity correction, we performed 10000 simulations for $p = 0.1$ at each value of $\rho = 0(0.01)0.1$ for various n values from 10 to 1000, generating overdispersed Binomial variables from Beta-Binomial distribution, GC model by Emrich and Piedmonte (1991) and from the model by Lunn and Davies (1998). The results are given in Figure 3.1. The values of p and ρ were assumed known in these simulations. It can be seen that the approximation is not too bad for small values of ρ in the case of Beta-Binomial and GC models, but works much worse for the Lunn-Davies model. Thus the knowledge of just two moments of a distribution does not provide sufficient information on the magnitude of bias.

3.2.1 Variance-stabilizing transformations in over-dispersed families

Variance-stabilizing transformations (v.s.t.'s) are used when the variance (under the fixed effect model) is a function of the mean: $\mathrm{Var}_0(X) = h(\mathrm{E}_0(X))$. The aim of a v.s.t. is to achieve $\mathrm{Var}_0(f(X)) \approx 1$. To be a v.s.t. when $\tau = 0$, a transformation $f(X)$ needs to satisfy

$$[f'(\mathrm{E}_0(X)]^2 = [\mathrm{Var}_0(X)]^{-1},$$

see Kulinskaya et al. (2008) for details and examples. Substituting this expression in equation 3.2.3, we obtain that, up to terms of smaller order,

$$\mathrm{Var}_\tau[f(X)] = \frac{\mathrm{Var}_\tau(X)}{\mathrm{Var}_0(X)}. \tag{3.2.5}$$

It follows that for an additive REM, where under $g_\tau(x)$ the variance $\mathrm{Var}_\tau(X) = \mathrm{Var}_0(X) + \tau^2$, the null v.s.t. $f(x)$ does not stabilise the variance for $\tau \neq 0$. On the other hand, for any over-dispersed family with $\mathrm{Var}_\tau(X) = \mathrm{Var}_0(X)\phi(\tau)$, the null v.s.t. $f(x)$ would achieve variance stabilization with $\mathrm{Var}_\tau[f(X)] = \phi(\tau)$.

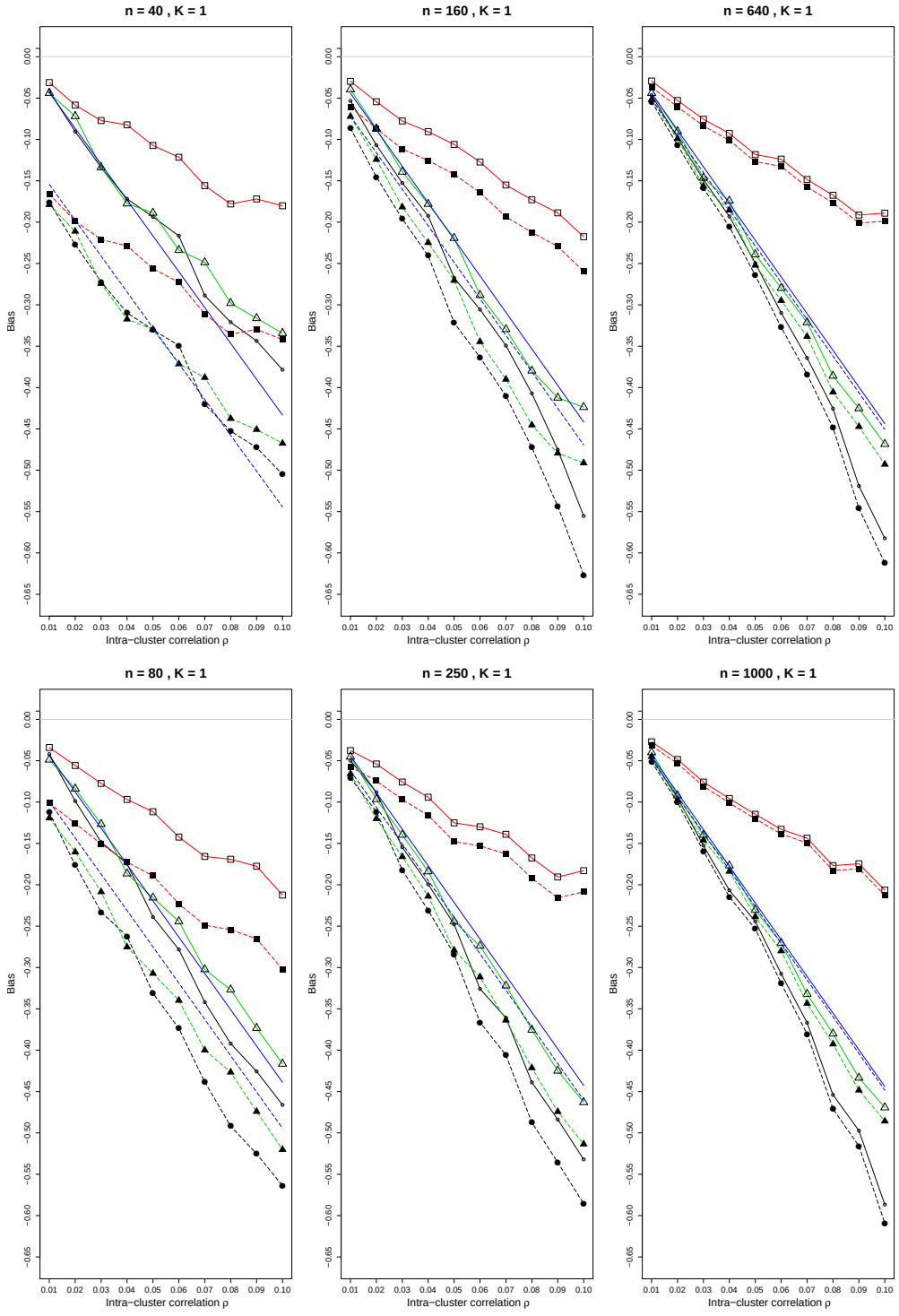


Figure 3.1: Bias on log-odds scale in overdispersed binomial model for $p = 0.1$ ($\log(p/(1-p)) = 2.20$) and $0 \leq \rho \leq 0.1$. 10000 simulations for each value of ρ from the beta-binomial distribution (black, circles); from the Lunn and Davies model Lunn and Davies (1998) model (red, squares); from the Gaussian copula Emrich and Piedmonte (1991) (green, triangles), with and without the Gart Gart et al. (1985) continuity correction (solid and dashed lines, respectively), and the first order linear bias term given by the first two terms of equation (3.2.1) using known values of p and ρ (blue). Light grey line at zero.

Example 4: arcsine transformation for over-dispersed binomial

As an example, consider the estimated probability of success $\hat{p} = X/n$ based on the sum of n dependent Bernoulli variables X. Then $\mathrm{Var}_\rho(X/n) = n^{-1}p(1-p)[1+(n-1)\rho]$. The arcsine transformation is routinely used to variance-stabilize binomial variables. (Kulinskaya et al., 2008, Ch. 18, p.139) recommend using the Anscombe (1948) transformation $2\mathrm{arcsin}(\sqrt{\tilde{p}})$ for $\tilde{p} = (X+3/8)/(n+3/4)$. Then $f(\tilde{p}) = 2\mathrm{arcsin}(\sqrt{\tilde{p}})$ has the mean $\mathrm{E}_0(f(\tilde{p})) = f(p) = 2\mathrm{arcsin}(\sqrt{p})$. Consider, first, the arcsine transform without continuity correction. The derivatives are $f'(p) = [p(1-p)]^{-1/2}$ and $f''(p) = -(1/2)[p(1-p)]^{-3/2}(1-2p)$. The bias under overdispersion $\rho > 0$ is

$$\mathrm{E}_\tau[2\mathrm{arcsin}(\sqrt{\hat{p}})] - 2\mathrm{arcsin}(\sqrt{p}) = -\frac{1}{4}\frac{(1-2p)}{\sqrt{p(1-p)}}\frac{[1+(n-1)\rho]}{n}, \tag{3.2.6}$$

and the variance is $n^{-1}[1+(n-1)\rho]$. With Anscombe's continuity correction, we need to add the first order bias term in the above formula. Using $\tilde{p} = (n\hat{p}+3/8)/(n+3/4)$, the bias is

$$\mathrm{E}_\tau[2\mathrm{arcsin}(\tilde{p})] - 2\mathrm{arcsin}(p) = \frac{3(1-2p)}{2\sqrt{p(1-p)}(4n+3)} - \frac{(1-2p)}{\sqrt{p(1-p)}}\frac{4n[1+(n-1)\rho]}{(4n+3)^2}. \tag{3.2.7}$$

For $p = 0.1$, the bias is $-(2/3)[1+(n-1)\rho]/n$ and the additional bias from the overdispersion is $-(2/3)\rho$; and for $p = 0.2$, the bias is $-0.375[1+(n-1)\rho]/n$ with an additional bias of -0.375ρ.

To assess the bias of the arcsine transform and the precision of our two-moment approximation to the bias for $p = 0.1$, we performed 10000 simulations for $p = 0.1$ at each value of $\rho = 0(0.01)0.1$ for various n values from 10 to 1000, generating overdispersed Binomial variables from the Beta-Binomial distribution, from the model by Lunn and Davies (1998) and from the GC model, Emrich and Piedmonte (1991). The results are given in Figure 3.2. The linear bias term was plotted for known values of p and ρ. Overall, the bias of the arcsin transformation is rather small. The approximation provides correct slope but not the intercept of a linear trend for smaller values of n. For larger n, the approximation is very good for the Beta-Binomial and for the GC, but not for the Lunn-Davies model unless $\rho \leq .01$. For larger ρ, the bias of the arcsin transform in the Lunn-Davies model is clearly not linear. The

Anscombe's continuity correction reduces bias for all values of n, though it does not matter much for larger n. In this case the Lunn-Davies model results in a somewhat smaller bias than the Beta-Binomial and the GC models.

We also studied the coverage of the confidence intervals for p based on the normal approximation with the variance $[1+(n-1)\rho]$ for known ρ, to the arcsine transformation of the $\hat{p}$ for the three models. The results are given in Figure (3.3).

Overall the coverage in the Beta-binomial and the GC models in case of the continuity correction is pretty good. It becomes increasingly conservative with increasing ρ. The coverage deteriorates for larger sample sizes in the Lunn-Davies model. This is due to its asymptotic non-normality, as was discussed in Section (2.2.2).

3.3 Transformation bias in meta-analysis

3.3.1 Small biases in meta-analysis

In meta-analysis, a relation between average sample size n and the number of studies K is important for the quality of the inference for the combined effect. In our context, if a statistic X estimating some parameter μ has a bias of order $1/n$, the mean (weighted or not) of K such statistics has the bias of the same order, but its variance is of order $1/K$. So keeping n fixed and increasing K results in a diminishing coverage of μ as the narrow confidence intervals are centered on a biased estimator. This observation was originally made in Kulinskaya et al. (2014). In the current setting, a minor bias from a transformation used under REM may result in sub-standard coverage of the combined effect, as is demonstrated for the arcsine transform in Section 3.3.2. Therefore in meta-analysis we cannot afford even small biases in a case of a large number of studies K.

Denote by $n_i, i = 1, \cdots, K$ the sample sizes and by $\bar{n}$ the "average" sample size of the K studies, and let the total sample size be $N = \bar{n}K$. Denote by X_i a summary statistic from the study i, and denote its expectation and variance by μ and σ_i^2. Let the bias of the combined weighted mean $\bar{X}$ be $c/\bar{n}$, so the combined mean is centered at $\mu + c/\bar{n}$. If the inverse

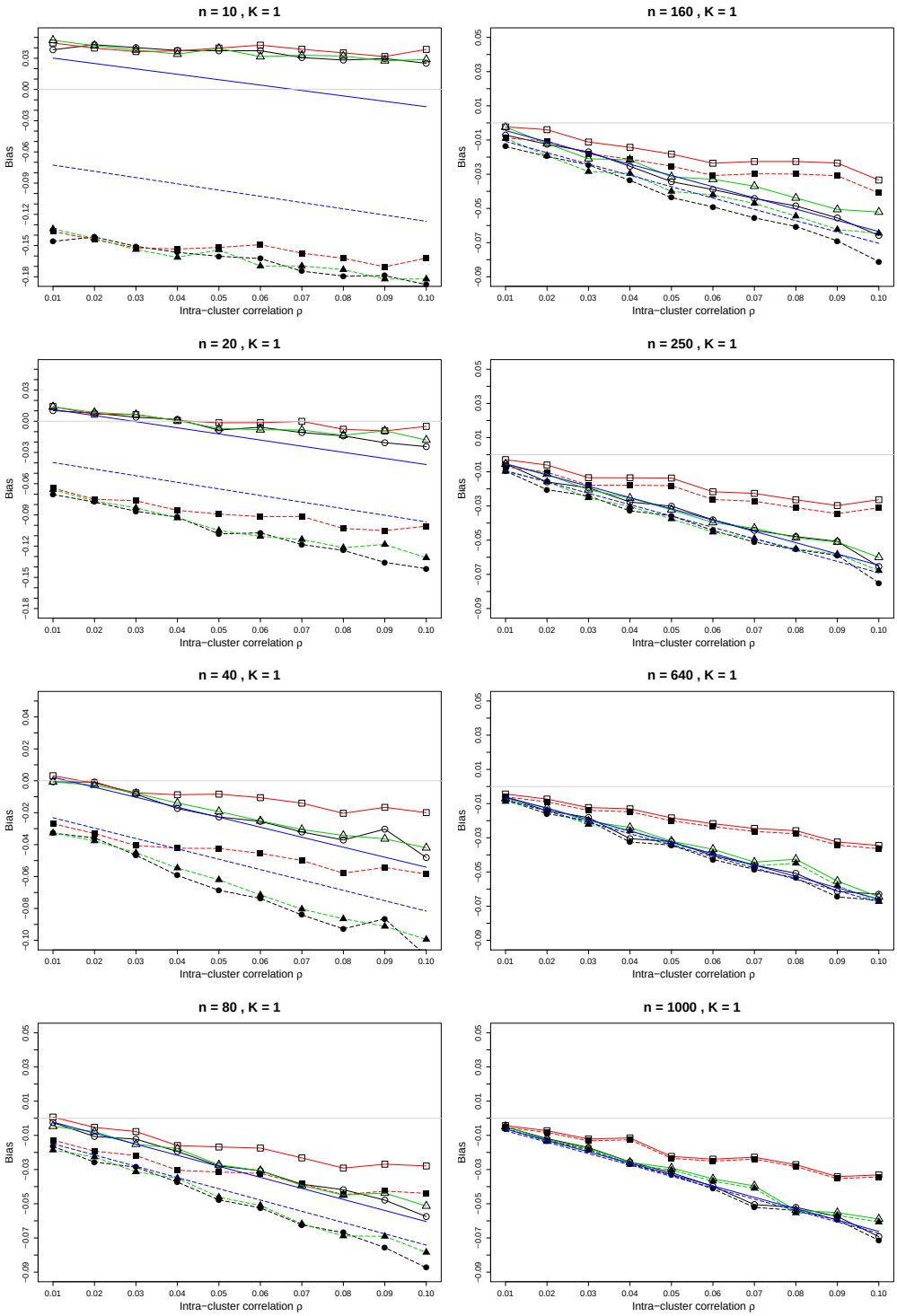


Figure 3.2: Bias on the arcsine scale of the arcsine transformation in overdispersed binomial model for $p = 0.1$ and $0 \leq \rho \leq 0.1$. 10000 simulations for each value of ρ from the beta-binomial distribution (black), from the Lunn and Davies (1998) model (red) and from the Gaussian copula Emrich and Piedmonte (1991) (green) with and without the continuity correction (solid and dashed lines, respectively). Also the first order bias terms given by the first two terms of equation (3.2.1) and plotted for known p and ρ (solid or dashed blue lines). Light grey line at zero.

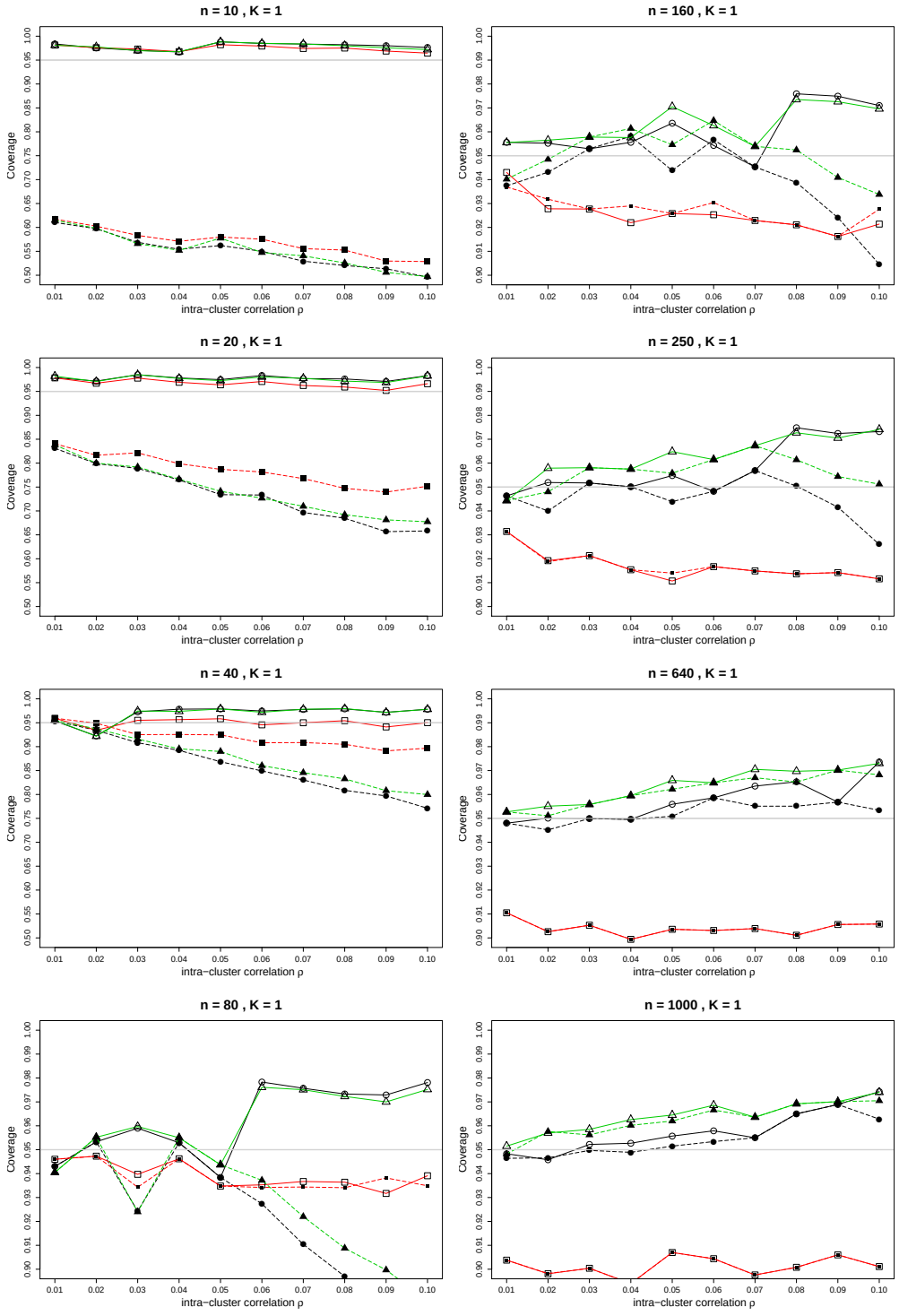


Figure 3.3: Coverage (for a known value of ρ) at the nominal 95% level of the true value of p using the acrsine transformation in overdispersed binomial model for $p = 0.1$ and $0 \leq \rho \leq 0.1$. 10000 simulations for each value of ρ from the beta-binomial distribution (black lines), from the Lunn and Davies (1998) model (red lines) and from the Gaussian copula Emrich and Piedmonte (1991) (green) with and without the continuity correction (solid and dashed lines, respectively). Light grey line at 0.95.

variance weights w_i are used to obtain the combined mean, its variance is approximated by $[\sum w_i]^{-1} = [K\bar{w}]^{-1} = O(1/K)$, where $\bar{w}$ is the average weight. Denote by $\sigma^2 = \bar{w}^{-1} = O(1/\bar{n})$ the "average" variance of the within-studies summary statistics X_i. The half-width of the confidence interval (CI) for the combined mean of K studies is $z_{1-\alpha/2}\sigma/\sqrt{K}$. For the CI for the combined effect to reliably cover the mean μ, the requirement is

$$z_{1-\alpha/2}\sigma/\sqrt{K} >> |c|/\bar{n}.$$

This may not be satisfied when the number of studies K is too large, or the sample size $\bar{n}$ is too small. To achieve a good coverage of the combined effect, given biased estimates from the individual trials whose bias is of rough order $1/\bar{n} = K/N$, the following relationship between the number of studies K and the overall sample size N should hold:

$$K = O(N^{2/3-\gamma}) \text{ for } \gamma > 0.$$

This means that the sample sizes of the individual studies in meta-analysis cannot be too small in relation to the number of studies. In practice, as the constant c is not known, particular caution is required in the case of a large number of comparatively small studies. A similar complication arises, for instance, when combining penalised GLM regressions (which are intentionally somewhat biased) on subsamples of a big dataset. The even stronger restriction $K < N^{1/5}$ is require in that case in order for the combined result to be equivalent to the regression on the full dataset, Chen and Xie (2014).

3.3.2 Arcsine transformation

We have studied by simulation the bias and the coverage of the parameter $2\text{arcsin}(p)$ when the data is generated from an overdispersed binomial distribution with the correlation coefficient $\rho \leq 0.1$, the estimated probabilities $\hat{p}_i$, $i = 1, \cdots, K$ from individual studies are arcsine-transformed and the meta-analysis is performed on the variance-stabilized scale. We varied sample sizes n from 10 to 1000 and the number of studies K from 10 to 80. We assumed known probability p and in these simulations. Inverse-variance weights on the variance-stabilized scale

$w_i = n_i/[1 + (n_i - 1)\rho]$ with known ρ were used in meta-analysis.

A representative selection of these simulation results when $p = 0.1$ is given in Figures 3.4 and 3.5 for the bias and the coverage (with the known ρ in weights), respectively, of the combined mean of the arcsine-transformed estimated probabilities from K studies. Results for $p = 0.2$ and $p = 0.4$ are given in A.1 - A.4 in the Appendix A. The coverage of the combined mean when the parameter ρ is estimated is explored in Section 3.4.1. The bias in meta-analysis is exactly the same as in one study. Fact that the confidence intervals were inflated by known value os the standard deviation $(1 + (n - 1)\rho)^{1/2}$ on the variance-stabilizing scale. This may differ from actual variance inflation, becoming unacceptable from $n = 80$ for $\rho \geq .01$. This happens because the confidence intervals are $[\sin(\arcsin(\sqrt{\hat{p}}) \pm 1.96/(2\sqrt{n}))]^2$, and their width is quickly reducing with n, whereas the mean-square-error of arcsine-based estimate $\hat{p}$ is comparatively large due to bias.

Bias in Figure 3.4 does not much depend on the number of studies and is very similar to that for one study, see Figure 3.2. The coverage given in Figure 3.5 is considerably worse than that for one study, given in Figure 3.3. See Section 3.3.1 for explanation of these findings.

It can be seen that the continuity correction substantially reduces the bias and improves the coverage. For each sample size, the coverage deteriorates when the number of studies K increases. The reason for this is that the bias, though small, becomes non-negligible for large K, as discussed in Section 3.3.1. Interestingly, for large n (starting from $n = 80$), there is a substantial difference in coverage between the beta-binomial and GC models, on one hand, and the Lunn-Davies model on the other hand. The coverage in the Lunn-Davies model is close to nominal, whereas in the previous two models, the coverage deteriorates quite dramatically. For $p = 0.2$ and $p = 0.4$ the above overall patterns also apply but on a milder scale, especially for $p = 0.4$, see A.1 - A.4 in the Appendix.

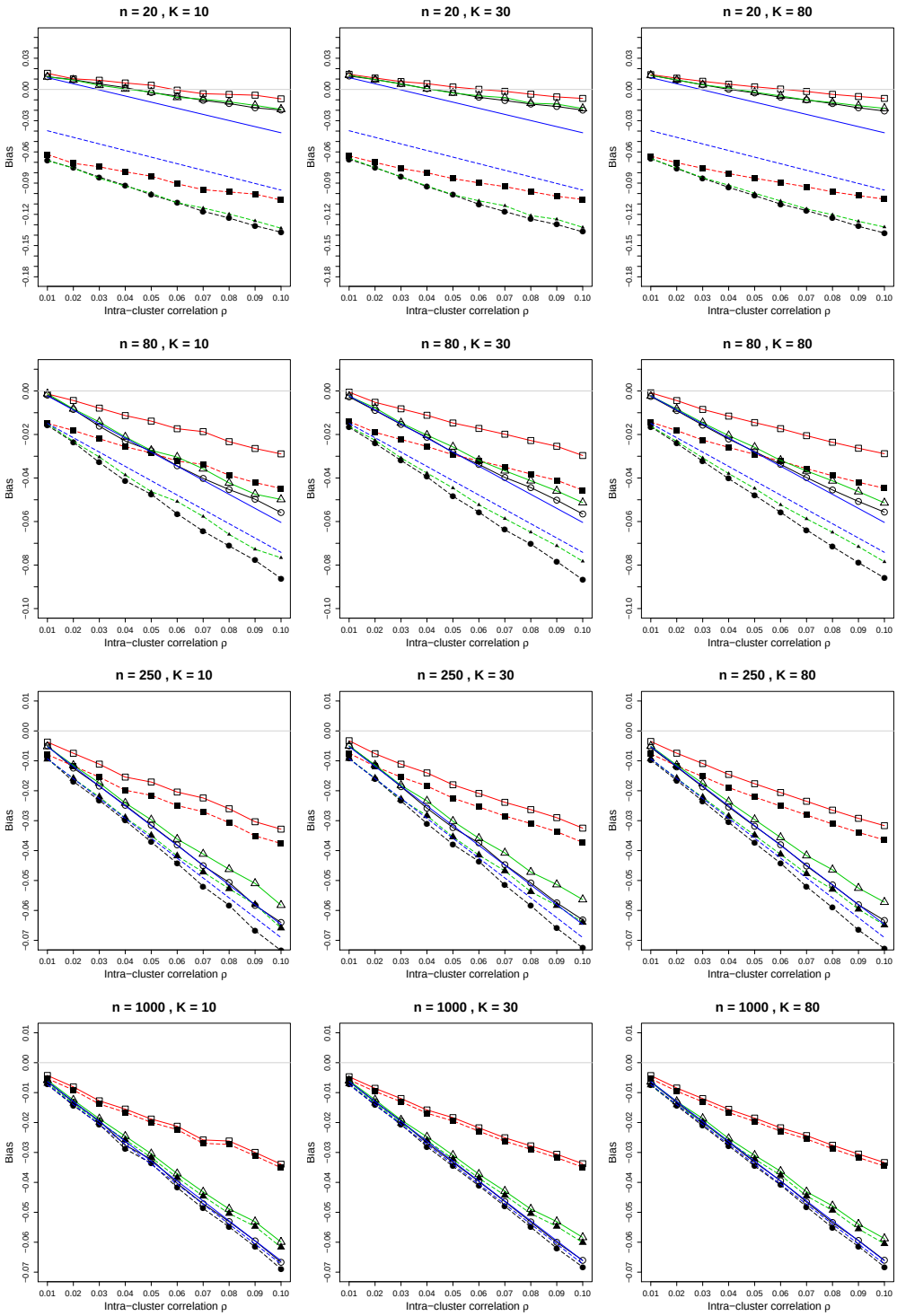


Figure 3.4: Bias on the arcsine scale of the meta-analysis of arcsine transformations from K studies in overdispersed binomial model for $p = 0.1$ and $0 \leq \rho \leq 0.1$. 10000 simulations for each value of ρ from the beta-binomial distribution (black), from the Lunn and Davies (1998) model (red) and from the GC model by Emrich and Piedmonte (1991) (green) with and without the continuity correction (solid and dashed lines, respectively). Also the first order bias terms given by the first two terms of equation (3.2.1) and plotted for known p and ρ (solid or dashed blue lines). Light grey line at zero.

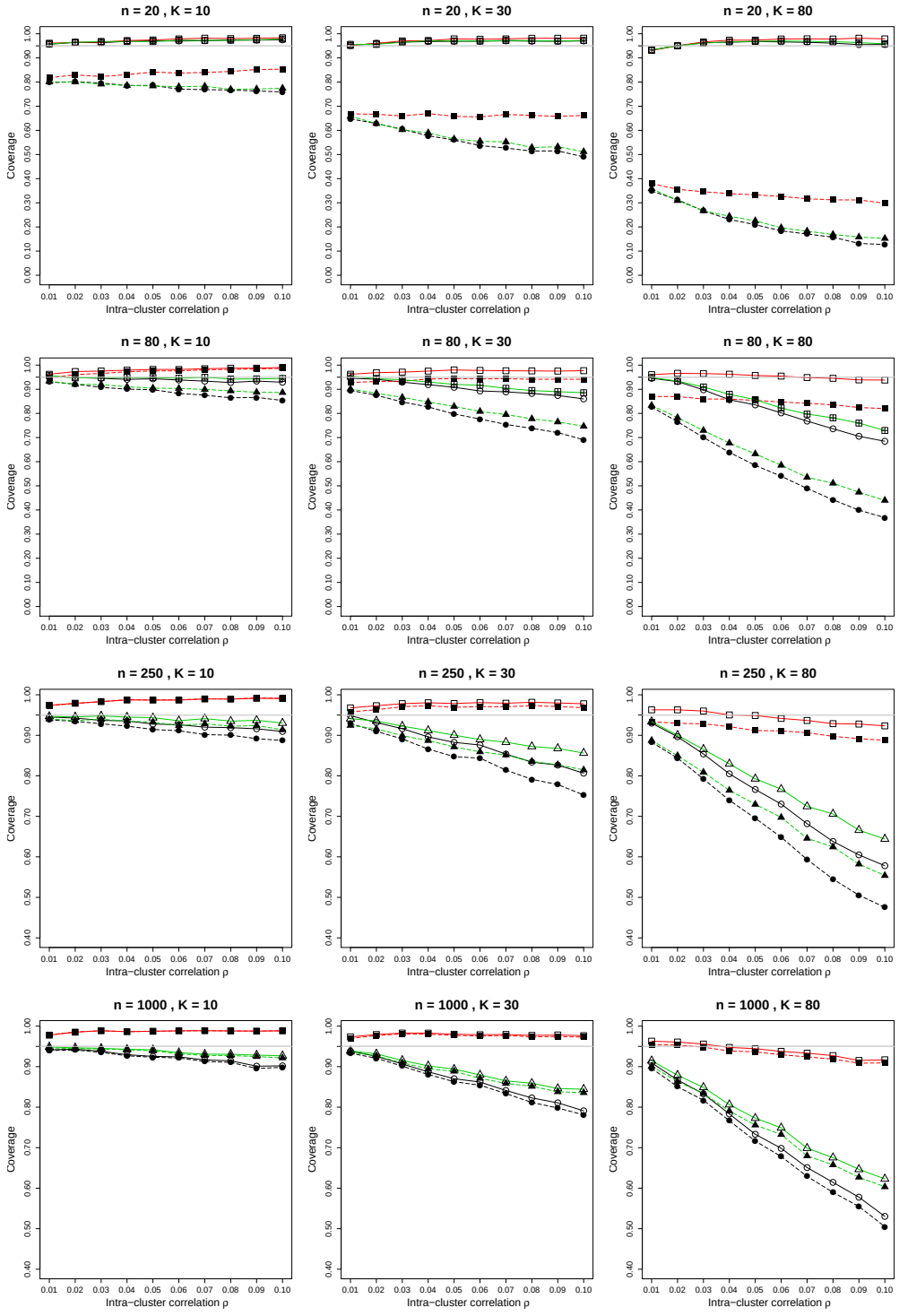


Figure 3.5: Coverage (for a known value of ρ) at the nominal 95% level of the true value of p using the meta-analysis of acrsine transformation from K studies in overdispersed binomial model for $p = 0.1$ and $0 \leq \rho \leq 0.1$. 10000 simulations for each value of ρ from the beta-binomial distribution (black lines), from the Lunn and Davies (1998) model (red lines) and from the GC model by Emrich and Piedmonte (1991) (green) with and without the continuity correction (solid and dashed lines, respectively). Light grey line at 0.95.

3.3.3 Log-Odds transformation

In meta-analysis of the log odds $\log(p/(1-p))$ from an overdispersed binomial distribution with correlation ρ, the weight of an estimated log-odds is given by the inverse estimated variance $w = [\sigma^2]^{-1} = [(1 + (n-1)\rho)/np(1-p)]^{-1}$. In contrast to arcsine transformation, the weights of log odds depend on the unknown probabilities. Estimation of the probabilities affects the bias of the log-odds, and even its sign.

Trikalinos et al. (2013) studied by simulation the log, logit and arcsine transformations for overdispersed binomial data. They rightly point out that "All these functions are concave for proportions between 0, 0.50, and therefore introduce a negative bias: The mean in the transformed scale will be smaller than the transformation of the mean in the proportion scale." However, this theoretical finding is reversed when the probabilities p are estimated. Figures 3.6 and 3.7 show the bias and coverage of log odds in the meta-analysis of K studies using known probabilities p and intra-cluster correlation ρ in the weights. Here the first order bias term given by the first two terms of equation (3.2.1) approximates the bias of the log-odds transformation reasonably well. Compare these results to those in Figures 3.8 and 3.9 showing the bias and coverage when the weights include estimated probabilities $\hat{p}$ and a known value of ρ. The coverage in the meta-analysis of log odds is pretty dismal in both settings. However, the sign of the bias changes from negative to positive with substitution of $\hat{p}$ in the weights. New terms taking the random weights into account are required to estimate the bias. It is therefore considerably more difficult to provide bias correction for meta-analysis of log-odds, and we do not pursue this further.

3.4 Theory of bias-correction

In the previous section, we have shown that there exist a bias from the transformation of a random variable in case of a random effects model. The random effect can be either additive or multiplicative. In this section, we correct the bias of $O(1/n)$ adding the second order bias correction term to the given transformations. For general transformation $f(X)$, the new bias

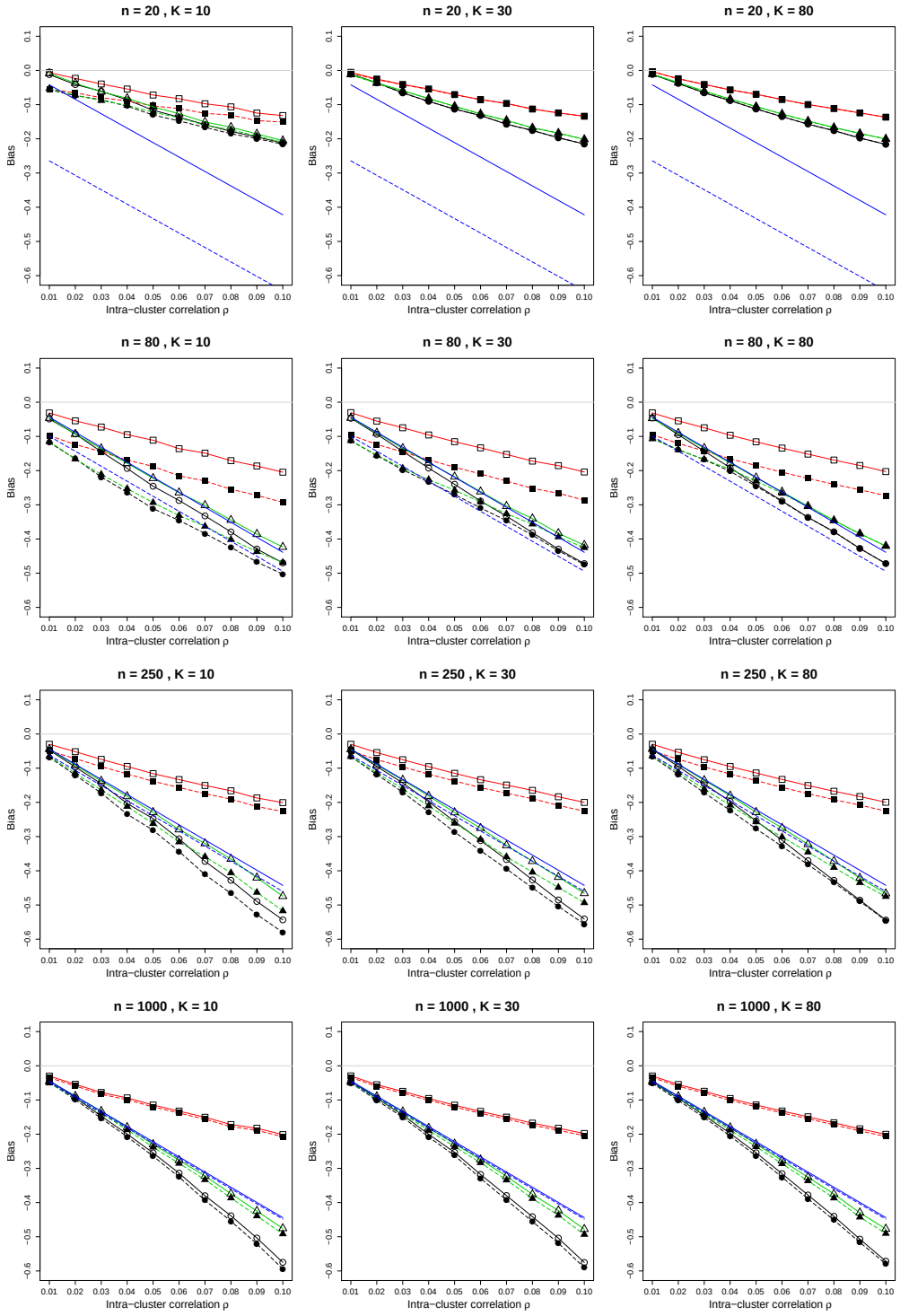


Figure 3.6: Bias on log-odds scale in the meta-analysis of log-odds from K studies in overdispersed binomial model for $p = 0.1$ ($\log(p/(1-p)) = 2.20$) and $0 \leq \rho \leq 0.1$ with known p and ρ in the weights. Simulations (10000 for each values of ρ) from the beta-binomial distribution (black); from the Lunn and Davies (1998) model (red); from the GC model by Emrich and Piedmonte (1991) (green); and the first order bias term given by the first two terms of equation (3.2.1) and plotted for known values of p and ρ (blue) with and without the continuity correction (solid and dashed lines, respectively).

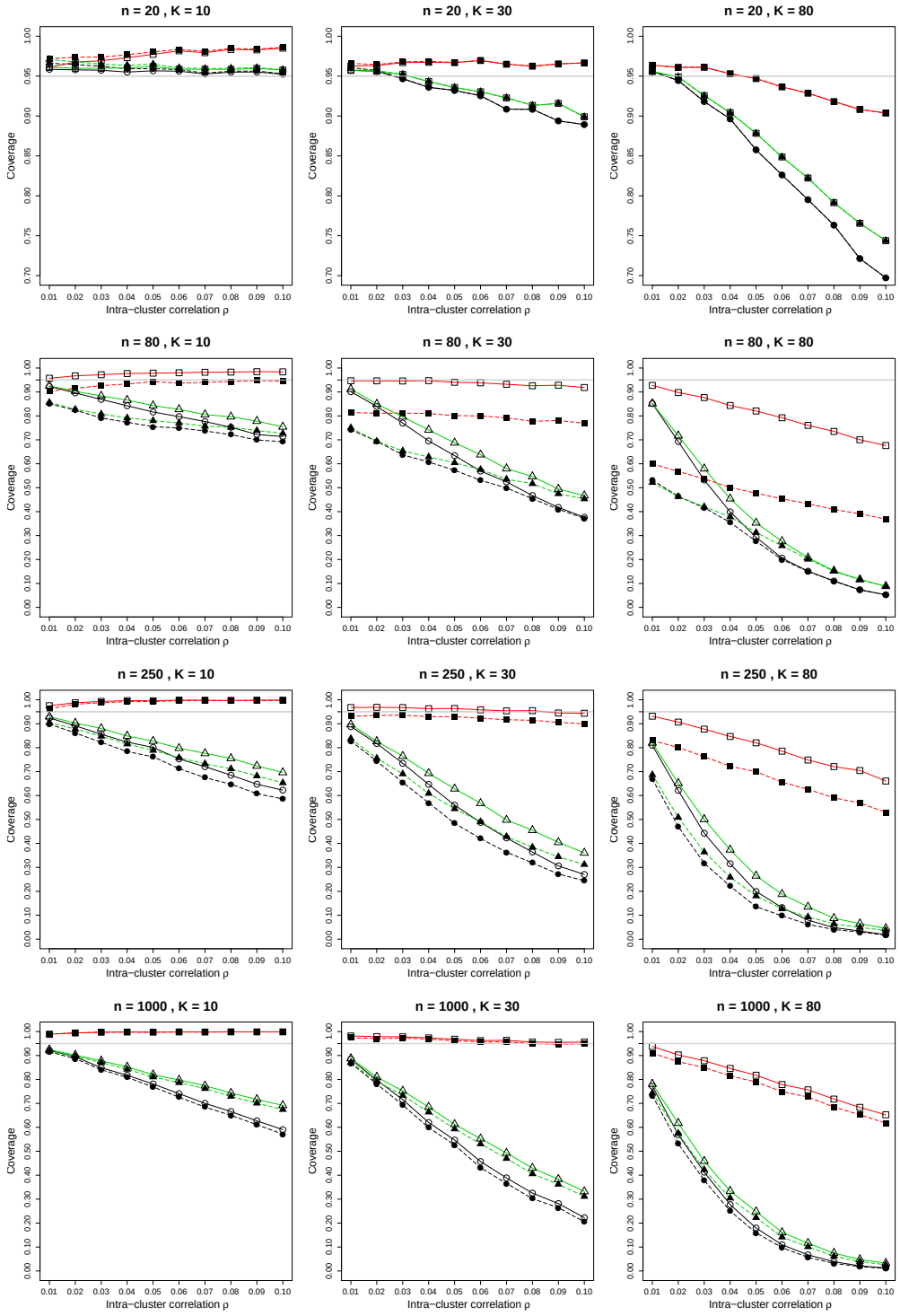


Figure 3.7: Coverage of the combined effect on log-odds scale in the meta-analysis of log-odds from K studies in overdispersed binomial model for $p = 0.1$ ($\log(p/(1-p)) = 2.20$) and $0 \leq \rho \leq 0.1$ using known p and ρ in the weights. 10000 simulations for each value of ρ from the beta-binomial distribution (black); from the Lunn and Davies (1998) model (red) and from the GC model by Emrich and Piedmonte (1991) (green line), with and without the continuity correction (solid and dashed lines, respectively).

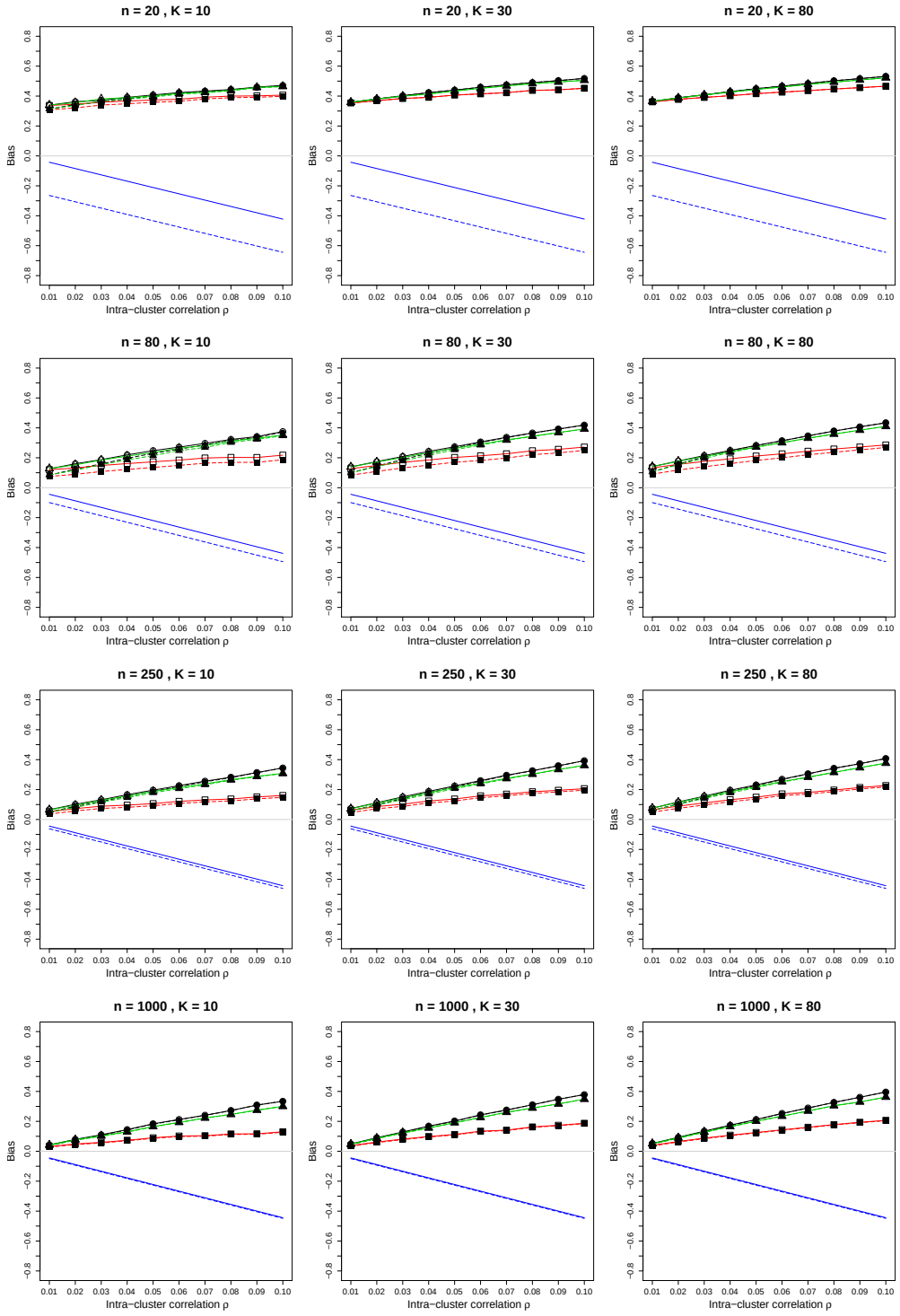


Figure 3.8: Bias on log-odds scale in the meta-analysis of log-odds from K studies in overdispersed binomial model for $p = 0.1$ ($\log(p/(1-p)) = 2.20$) and $0 \leq \rho \leq 0.1$ using estimated p and known ρ in the weights. 10000 simulations for each value of ρ from the beta-binomial distribution (black); from the Lunn and Davies (1998) model (red); from the GC model by Emrich and Piedmonte (1991) (green) and the first order bias term given by the first two terms of equation (3.2.1) and plotted for known values of p and ρ (blue), with and without the continuity correction (solid and dashed lines, respectively).

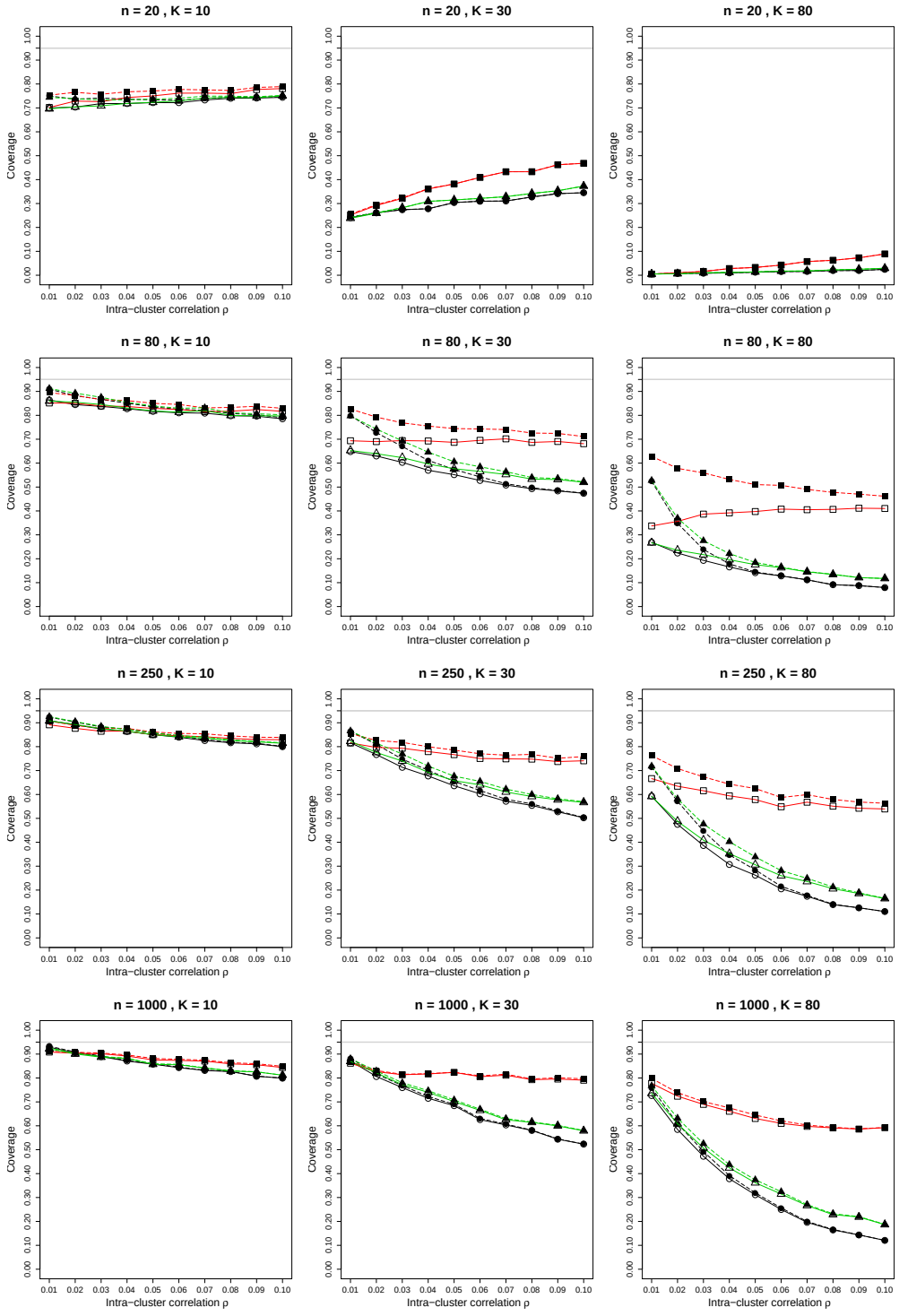


Figure 3.9: Coverage of the combined effect on log-odds scale in the meta-analysis of log-odds from K studies in overdispersed binomial model for $p = 0.1$ ($\log(p/(1-p) = 2.20$) and $0 \leq \rho \leq 0.1$ using estimated p and known ρ in the weights. 10000 simulations for each value of ρ from the beta-binomial distribution (black); from the Lunn and Davies (1998) model (red) and from the GC model by Emrich and Piedmonte (1991) (green), with and without the continuity correction (solid and dashed lines, respectively).

corrected estimator is

$$f(X) + \frac{1}{2}\frac{d^2 f(x)}{dx^2}|_{x_0}\mathrm{E}_\tau(X - x_0)^2.$$

In case of using the continuity correction in estimated probabilities, the bias correction should include the first order bias term so that

$$f(X) + \frac{df(x)}{dx}|_{x_0}\mathrm{E}_\tau(X - x_0) + \frac{1}{2}\frac{d^2 f(x)}{dx^2}|_{x_0}\mathrm{E}_\tau(X - x_0)^2.$$

For example, bias corrected estimator for arcsine transformation without continuity correction is

$$2\mathrm{arcsin}(\sqrt{\hat{p}}) + \frac{1}{4}\frac{(1-2p)}{\sqrt{p(1-p)}}\frac{[1+(n-1)\rho]}{n}.$$

With continuity correction 3/8 in $\tilde{p}$ adding the first order term to the bias correction, the corrected estimator of $2\mathrm{arcsin}(\sqrt{p})$ is

$$2\mathrm{arcsin}(\tilde{p}) - \frac{3(1-2p)}{2\sqrt{p(1-p)}(4n+3)} + \frac{(1-2p)}{\sqrt{p(1-p)}}\frac{4n[1+(n-1)\rho]}{(4n+3)^2}$$

The bias correction depends on probability p and intra-cluster correlation ρ. Both of these parameters are unknown and have to be estimated. Therefore the bias correction itself may become biased as it includes non-linear function of estimated probabilities.

3.4.1 Bias correction for arcsine transformation

In this Section we aim to correct the biases in the arcsine transformation with and without the Anscombe (1948) continuity correction $\tilde{p} = (n\hat{p} + 3/8)/(n + 3/4)$ by taking out the first order bias terms given by equations (3.2.6) and (3.2.7). The bias terms depend on $(1-2p)/\sqrt{p(1-p)}$ and ρ. The bias correction using known values of p and ρ substantially improves both bias and coverage, see A.5 - A.7 in Appendix A. Substituting an estimate $\hat{p}$ in the expressions for bias results in an additional bias in the expected value of $(1-2\hat{p})/\sqrt{\hat{p}(1-\hat{p})}$. This bias is minimised when the Gart et al. (1985) continuity correction $\hat{p} = (X + 0.5)/(n+1)$ is used in the bias term.

For the intra-cluster correlation coefficient ρ, different estimators were reviewed by Ridout

et al. (1999). Among all the estimators, the analysis of variance (AOV) estimator and an estimator based on a weighted average of Pearson correlation coefficients between pairs of observations within each group, denoted by $\hat{\rho}_{PPR}$ by Ridout et al. (1999) perform the best in terms of bias. Our own simulations show that the AOV estimator, $\hat{\rho}_{AOV}$, defined in Appendix A, is superior to the PPR estimator, see 3.12.

The analysis of variance (AOV) estimator for intra-cluster correlation is

$$\hat{\rho}_{AOV} = \frac{MS_b - MS_w}{MS_b + (n_0 - 1)MS_w},$$

where MS_w and MS_b are the within and between group mean squares for a one-way analysis of variance applied to Bernoulli r.v.'s, and where

$$n_0 = \frac{1}{K-1}[N - \sum_{i=1}^{K} \frac{n_i^2}{N}], \qquad \text{with} \qquad N = \sum_{i=1}^{K} n_i.$$

For binary outcomes, the within and between group mean squares are

$$MS_b = \frac{1}{K-1}[\sum_{i=1}^{K} \frac{X_i^2}{n_i} - \frac{1}{N}(\sum_{i=1}^{K} X_i)^2]$$

and

$$MS_w = \frac{1}{N-K}[\sum_{i=1}^{K} X_i - \sum_{i=1}^{K} \frac{X_i^2}{n_i}],$$

respectively.

We have studied by simulation the changes to bias and coverage of $\arcsin(p)$ when the bias-correction based on the estimated first-order bias term is applied to arcsine transformation and meta-analysis is performed on the variance-stabilized scale. A representative selection of these simulation results when $p = 0.1$ is given in Figures 3.10 and 3.11 for the bias and the coverage, respectively. The intra-class correlation was estimated by $\hat{\rho}_{AOV}$, and the standard $1/2$ continuity correction was applied to $\hat{p}$ in the bias term. We varied sample sizes n from 10 to 1000 and the number of studies K from 10 to 80. In meta-analysis, inverse variance weights on the variance-stabilized scale were used: $w_i = n_i/[1 + (n_i - 1)\hat{\rho}_{AOV}]$.

Comparing Figure 3.10 to Figure 3.4, we see that the bias correction reduced the bias, especially for small values of the intra-class correlation $\rho < 0.06$. For larger values of ρ, the bias

correction results in a positive bias, compared to negative bias without the correction. Bias increases for large values of ρ.

Comparing Figure 3.11 to Figure 3.5, it is clear that the bias correction improves the coverage in the beta-binomial and the GC models for $\rho < 0.06$ and $K \geq 30$. The coverage deteriorates for larger values of $\hat{\rho}$. For $K = 10$ the bias correction results in coverage at about 90% at nominal 95% level. The reason is the inferior estimation of ρ for small values of K. For the Lunn-Davies model, coverage is considerably lower than in the beta-binomial and the GC models.

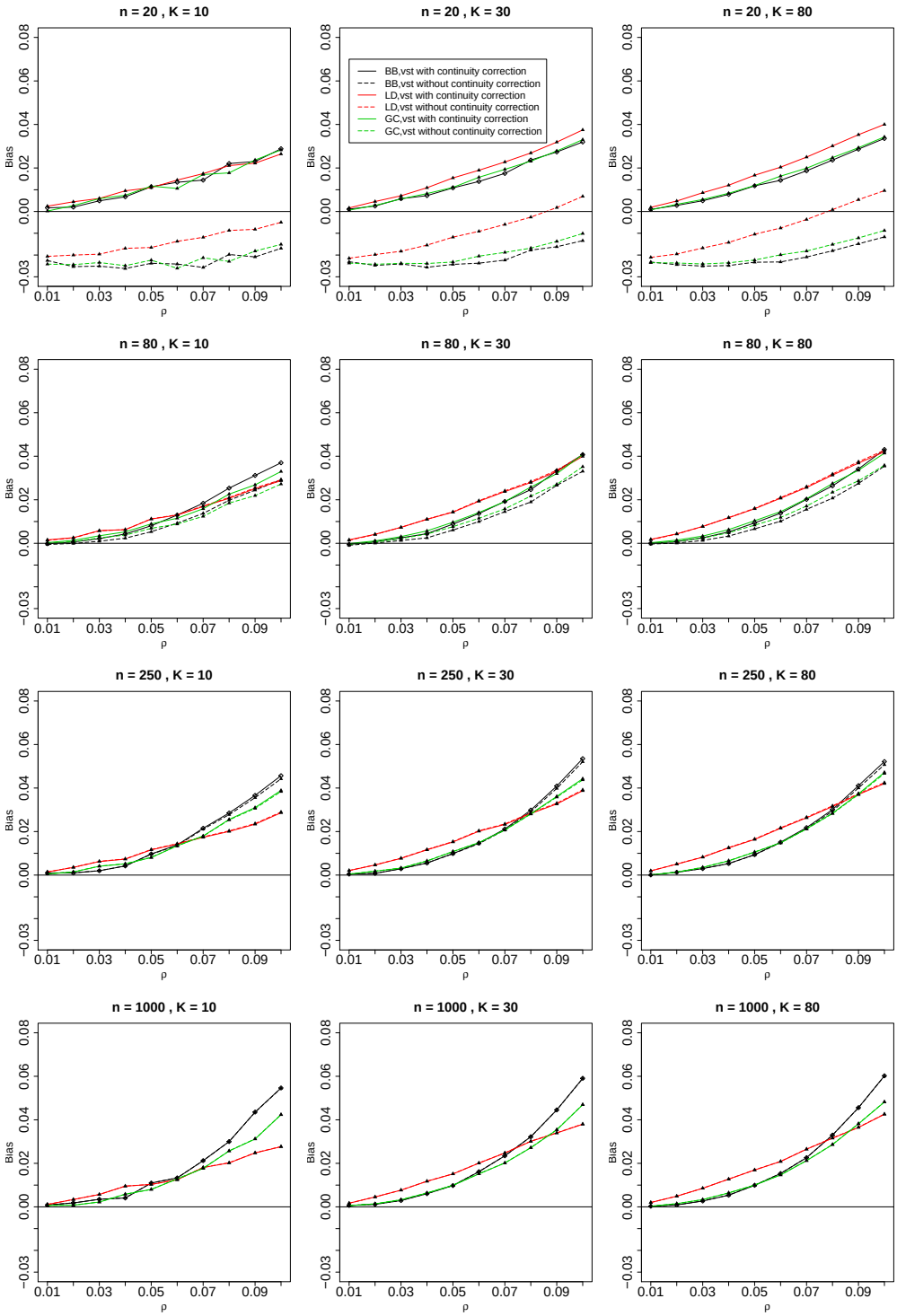


Figure 3.10: Bias on the arcsine scale in the meta-analysis of bias-corrected arcsine transformations from K studies in overdispersed binomial model for $p = 0.1$ and $0 \leq \rho \leq 0.1$ with estimated probabilities $\hat{p}_j$ and $\hat{\rho}_{AOV}$ in the bias correction terms. 10000 simulations for each value of ρ from the beta-binomial distribution (black), from the Lunn and Davies (1998) model (red) and from the GC model by Emrich and Piedmonte (1991) (green) with and without the continuity correction (solid and dashed lines, respectively). Light grey line at zero.

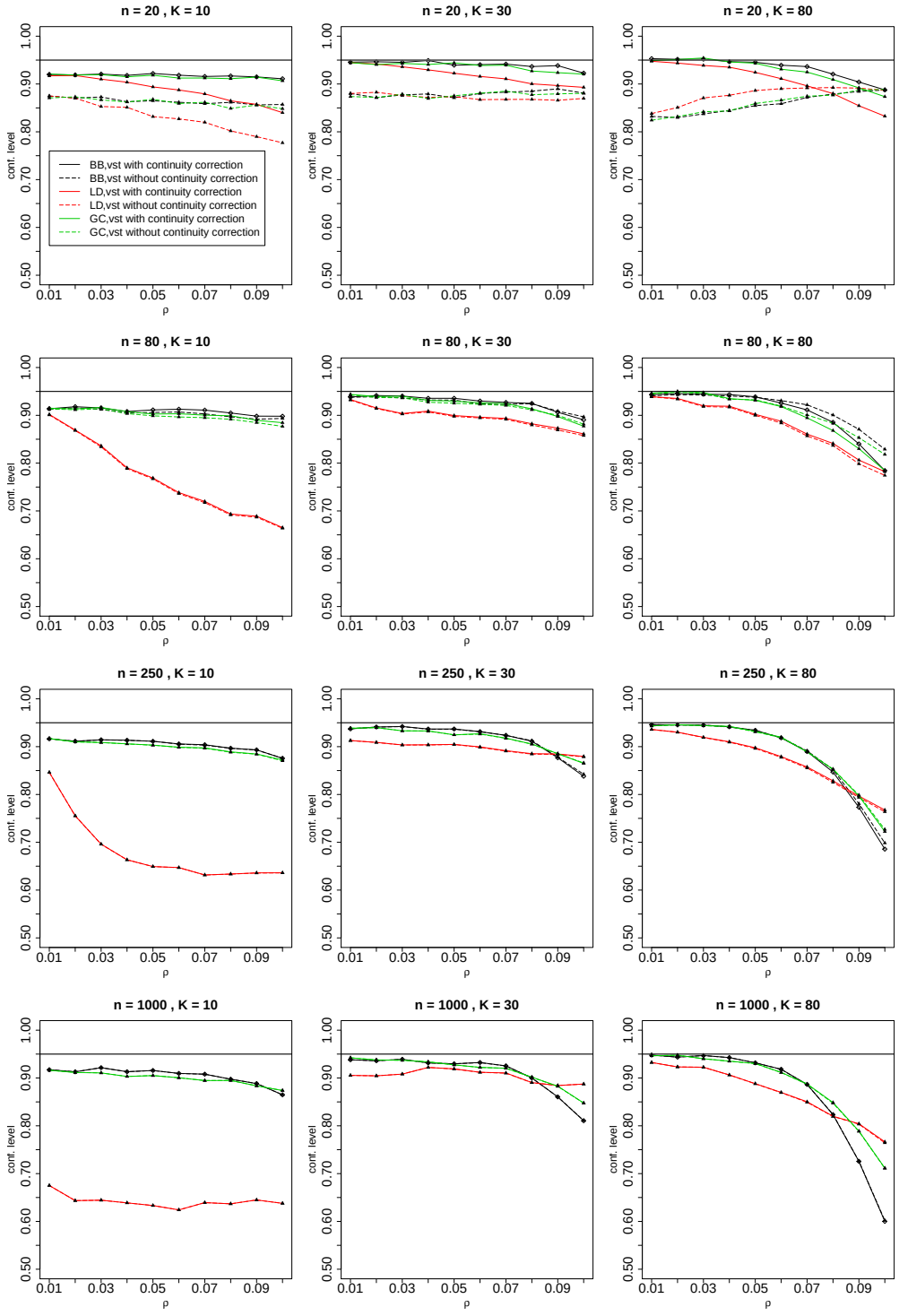


Figure 3.11: Coverage at the nominal 95% level of the true value of p in the meta-analysis of bias-corrected acrsine transformations from K studies in overdispersed binomial model for $p = 0.1$ and $0 \leq \rho \leq 0.1$ with estimated probabilities $\hat{p}_j$ and $\hat{\rho}_{AOV}$ in the bias correction terms. 10000 simulations for each values of ρ from the beta-binomial distribution (black lines), from the Lunn and Davies (1998) model (red lines) and from the GC model by Emrich and Piedmonte (1991) (green) with and without the continuity correction (solid and dashed lines, respectively). Light grey line at 0.95.

3.5 Examples

An important application of our methodology is to meta-analyses of prevalence of a disease or a condition. In this section, we consider the severity of transformation bias and the usefullness of our correction to this bias on two examples of meta-analyses of prevalence. The first example is that of syndromal depression in chronic kidney disease (Palmer et al. (2013)), and the second is the prevalence of HIV infection in homeless people (Beijer et al. (2012)). For both examples we obtained the results using the standard meta-analytic methods for the arcsine-transformed prevalences, and also our bias-correcting methods. These varying techniques result in somewhat different estimates of prevalence. To evaluate which method is likely to provide a correct inference, we have performed three simulation studies for each example, using the three methods for generation of overdispersed binomial outcomes, the GC, the BB and the LD methods. In all simulations we used the sample mean prevalence $\bar{p}$ and the estimated correlation ρ_{AOV} as the true values, and simulated 1000 new meta-analytic data-sets with the same number of studies and their sample sizes as in the original meta-analyses. For each simulation, we estimated the combined prevalence using the arcsine transformation with and without the Anscombe (1948) continuity correction, and also with and without our bias correction.

3.5.1 Prevalence of syndromal depression for paients on dialysis

A meta-analysis of forty-one studies by Palmer et al. (2013) evaluated the prevalence of syndromal depression in chronic kidney disease (CKD). We consider the subset of 28 studies with $N = 2855$ patients in total undergoing dialysis for CKD. According to Palmer et al. (2013), the dialysis stage has the highest rate of depressive symptoms. These data are provided in Table 3.1. The Table 3.1 also includes estimated prevalences, their arcsine transformations and corresponding variances. The sample sizes in these 28 studies are unbalanced and the range of the estimated prevalences is (0.0808, 0.5484). The results of meta-analysis of these data by various meta-analytic techniques are summarized in Table 3.2. The Cochran's Q statistic

Study	**No of events**	**No of participants**	$\hat{p}$	**Var**$(\hat{p})$	2**arcsin**$(\sqrt{\hat{p}})$	**Var**(2**arcsin**$(\sqrt{\hat{p}})$)
Craven et al., 1987	8	99	0.0808	0.0008	0.5765	0.0101
Moura et al., 2006	21	244	0.0861	0.0003	0.5955	0.0041
Cohen et al., 2002	2	22	0.0909	0.0038	0.6126	0.0455
Preljevic et al., 2011	3	25	0.12	0.0042	0.7075	0.04
Jouet et al., 1994	5	40	0.125	0.0027	0.7227	0.025
Huang et al.,1995	15	107	0.1402	0.0011	0.7675	0.0093
Alsuwaida et al., 2006	4	26	0.1538	0.005	0.8061	0.0385
Preljevic et al., 2011	13	84	0.1548	0.0016	0.8086	0.0119
Chan et al., 2011	23	141	0.1631	0.001	0.8315	0.0071
Lowry et al., 1980	15	83	0.1807	0.0018	0.8782	0.012
Eltayeb et al., 2010	55	300	0.1833	0.0005	0.8849	0.0033
Wuerth et al., 2005	70	380	0.1842	0.0004	0.8872	0.0026
Birmele et al., 2012	53	238	0.2227	0.0007	0.9829	0.0042
Chilcot et al., 2008	9	40	0.225	0.0044	0.9884	0.025
Chen et al., 2010	47	200	0.235	0.0009	1.0122	0.005
Soykan et al., 2004	12	50	0.24	0.0036	1.0239	0.02
Hinrichsen et al., 1989	30	124	0.2419	0.0015	1.0285	0.0081
Kalender et al., 2007	11	42	0.2619	0.0046	1.0745	0.0238
Hedayati et al., 2006	26	98	0.2653	0.002	1.0822	0.0102
Drayer et al., 2006	17	62	0.2742	0.0032	1.1022	0.0161
Cukor et al., 2008	20	70	0.2857	0.0029	1.1279	0.0143
Kweon et al., 2011	15	50	0.3	0.0042	1.1593	0.02
Taskapan et al., 2003	9	30	0.3	0.007	1.1593	0.0333
Loosman et al., 2010	21	62	0.3387	0.0036	1.2423	0.0161
Hong et al., 2006	22	64	0.3438	0.0035	1.253	0.0156
Cruz et al., 2010	25	70	0.3571	0.0033	1.281	0.0143
Ceyhun et al., 2010	22	42	0.5238	0.0059	1.6184	0.0238
Koo et al., 2003	34	62	0.5484	0.004	1.6677	0.0161

Table 3.1: Data for Example 1: Prevalence of syndromal depression diagnosed by clinical interview with chronic kidney disease at the stage of dialysis , Palmer et al. (2013)

value is $Q = 142.5$, at $K - 1 = 27$ degrees of freedom, indicating significant heterogeneity. In the standard random effects model for the arcsine-transformed data, the DerSimonian and Laird estimate of the between-studies variance $\hat{\tau}^2_{DL} = 0.043$ and the combined estimate of prevalence of 0.227. The beta-binomial model provides an estimated intra-cluster correlation of $\hat{\rho}_{AOV} = 0.046$, and a very similar combined estimate of prevalence. The Anscombe correction increases this estimate to 0.230 for both models. The proposed bias correction increases it further to 0.236 when used both with or without the Anscombe correction, but the Anscombe correction does not seem to matter when the bias correction is used. The value of $\rho = 0.046$, and the sample mean prevalence of $\bar{p} = 0.2367$ were used in further simulations, summarised in Table 3.3. In tables 3.2 and 3.3, beta-binomial model assumes a beta-binomial distribution for number of events across K studies. In this model, intra-cluster correlation is estimated

Table 3.2: Combined estimates of prevalence of syndromal depression and their confidence intervals for the data by Palmer et al. (2013)

model	continuity correction	$\hat{p}_w$	$\hat{p}_L$	$\hat{p}_U$
Fixed effects model(FE)	None	0.2060	0.1914	0.2211
	3/8	0.2081	0.1934	0.2232
Random effects model(REM)	None	0.2267	0.1904	0.2652
	3/8	0.2299	0.1937	0.2682
Beta-binomial model (BB)	None	0.2269	0.1900	0.2661
without bias correction	3/8	0.2302	0.1931	0.2696
Beta-binomial model (BB)	None	0.2357	0.1982	0.2753
with bias correction	3/8	0.2356	0.1982	0.2752

by analysis of variance method. After estimation of intra-cluster correlation, this estimator $\hat{\rho}$ is substituted into the variances of arcsine transformations $(1 + \rho(n_i - 1))/n_i$ and arcsine transformations are combined using the inverse-variance method.

Table 3.3: Quality of estimation of prevalence in meta-analyses using the arcsine transformation and estimated or theoretical value of ρ in weights evaluated from 1000 simulated meta-analyses of 28 studies with the value of $\rho = 0.046$, and the prevalence of $p = 0.23$ with sample sizes from Palmer et al. (2013)

Generation	Continuity	Bias	Estimated ρ				Known ρ			
method	correction	Correction	$2\arcsin(\sqrt{\hat{p}})$	Bias of vst	$\hat{p}$	Coverage	$2\arcsin(\sqrt{\hat{p}})$	Bias of vst	$\hat{p}$	Coverage
Beta-binomial	None	NO	0.9967	-0.0194	0.2285	0.9020	0.9974	-0.0187	0.2288	0.9170
	3/8	NO	1.0051	-0.0110	0.2320	0.9200	1.0059	-0.0102	0.2323	0.9370
	None	YES	1.0157	-0.0004	0.2365	0.9400	1.0195	0.0034	0.2381	0.9630
	3/8	YES	1.0157	-0.0004	0.2365	0.9410	1.0196	0.0034	0.2381	0.9630
Lunn Davies	None	NO	0.9983	-0.0178	0.2291	0.8850	1.0013	-0.0148	0.2304	0.9540
	3/8	NO	1.0061	-0.0100	0.2324	0.9030	1.0092	-0.0069	0.2337	0.9590
	None	YES	1.0180	0.0019	0.2375	0.9250	1.0184	0.0023	0.2376	0.9790
	3/8	YES	1.0179	0.0018	0.2374	0.9260	1.0182	0.0021	0.2375	0.9790
Gaussian Copula	None	NO	0.9974	-0.0187	0.2287	0.9100	0.9980	-0.0181	0.2290	0.9170
	3/8	NO	1.0057	-0.0104	0.2322	0.9210	1.0063	-0.0098	0.2325	0.9400
	None	YES	1.0148	-0.0013	0.2361	0.9310	1.0158	-0.0003	0.2365	0.9690
	3/8	YES	1.0148	-0.0013	0.2361	0.9320	1.0158	-0.0003	0.2365	0.9710

Overall, the bias of the arcsine transformation is reduced by bias correction, and the coverage is noticeably improved. Known ρ results in somewhat higher, and the estimated ρ in somewhat lower than nominal coverage, but the differences are within 2 percentage points in both cases when the bias correction is used. Once more, the Anscombe's correction does not seem to be needed when the bias correction is used.

3.5.2 Prevalence of HIV in homeless people

A meta-analysis of the data on $N = 10,886$ participants in sixteen studies by Beijer et al. (2012) evaluated prevalence of HIV infection in homeless people. These data are provided in Table 3.4. The main feature of these data is low prevalences across the studies, varying from 0 to 0.13. The results of meta-analysis by various meta-analytic techniques after the arcsine transformation are summarized in Table 3.5.

Country	Study	Study size	$\hat{p}$	Var($\hat{p}$)	2arcsin($\sqrt{\hat{p}}$)	Var(2arcsin($\sqrt{\hat{p}}$))
USA	Zolopa et. al., 1994	1005	0.1	0.0000896	0.6435011	0.0009950
USA	Paris et. al., 1996	331	0.11	0.0002958	0.6761305	0.0030211
USA	Magura et. al., 2000	90	0.13	0.0012567	0.7377260	0.0111111
USA	Hahn et. al., 2004	639	0.01	0.0000155	0.2003348	0.0015649
USA	Robertson et. At., 2004	1958	0.12	0.0000539	0.7074832	0.0005107
France	Brouqui et al., 2005	848	0	0.0000000	0.0000000	0.0011792
USA	Grimpley et. al.,2006	285	0.01	0.0000347	0.2003348	0.0035088
Brazil	Brito et al, 2007	267	0.02	0.0000734	0.2837941	0.0037453
India	Talukdar et. Al, 2007	493	0.05	0.0000963	0.4510268	0.0020284
Sweden	Burstrom et, al, 2007	123	0.08	0.0005984	0.5735131	0.0081301
Sweden	Beijer, 2007	1757	0.02	0.0000112	0.2837941	0.0005692
USA	Forney et. Al. 2007	161	0.05	0.0002950	0.4510268	0.0062112
Iran	Vahdani et al 2009	2002	0.07	0.0000325	0.5355267	0.0004995
France	Laporte et al 2010	402	0.01	0.0000246	0.2003348	0.0024876
France	Colson et al 2011	220	0.01	0.0000450	0.2003348	0.0045455
USA	Wenzel et al 2011	305	0.08	0.0002413	0.5735131	0.0032787

Table 3.4: Data for Example 2: estimated prevalence of HIV infection in homeless people, Beijer et al. (2012)

Cochran's Q statistic value is 536.40 at $K - 1 = 15$ degrees of freedom, indicating significant heterogeneity. The standard random effects model for the arcsine-transformed data provides the DerSimonian and Laird estimate of between study variance $\hat{\tau}_{DL} = 0.054$ and combined estimate of prevalence of 0.043. The beta-binomial model provides the estimated intra-cluster correlation of $\hat{\rho}_{AOV} = 0.037$, and the same combined prevalence value. The beta-binomial model assumes beta-binomial distributions for number of events across K studies, where $\hat{\rho}_{AOV}$ is analysis of variance estimator of ρ. The Anscombe corrections increase these estimates to 0.045 and 0.044, respectively, for the two models. The proposed bias correction increases estimated prevalence to 0.059 when used both with or without the Anscombe correction.

The value of $\rho = 0.037$, and the sample mean prevalence of $\bar{p} = 0.054$ were used in further simulations, summarised in Table 3.6. Overall, the negative bias of the arcsine transforma-

Table 3.5: Combined estimates of prevalence of HIV in homeless people and their confidence intervals for the data by Beijer et al. (2012)

	Continuity correction	$\hat{p}_w$	$\hat{p}_L$	$\hat{p}_U$
Fixed effects model (FEM)	None	0.0482	0.0442	0.0523
	3/8	0.0495	0.0455	0.0536
Random effects model (REM)	None	0.0429	0.0223	0.0697
	3/8	0.0445	0.0241	0.0708
Beta-binomial model (BB)	None	0.0427	0.0252	0.0645
without bias correction	3/8	0.0444	0.0265	0.0666
Beta-binomial model (BB)	None	0.0587	0.0379	0.0836
with bias correction	3/8	0.0590	0.0382	0.0841

Table 3.6: Quality of estimation of prevalence in meta-analyses using the arcsine transformation and estimated or theoretical value of ρ in weights evaluated from 1000 simulated meta-analyses of 16 studies with the value of $\rho = 0.037$, and the prevalence of $p = 0.054$ with sample sizes from Beijer et al. (2012)

Generation	Continuity	Bias	Estimated ρ				Known ρ			
method	correction	Correction	$2\arcsin(\sqrt{\hat{p}})$	Bias of vst	$\hat{p}$	Coverage	$2\arcsin(\sqrt{\hat{p}})$	Bias of vst	$\hat{p}$	Coverage
Beta-binomial	None	NO	0.4303	-0.0404	0.0456	0.8000	0.4264	-0.0443	0.0448	0.8600
	3/8	NO	0.4381	-0.0326	0.0472	0.8300	0.4346	-0.0361	0.0465	0.8990
	None	YES	0.4823	0.0116	0.0570	0.9070	0.4855	0.0148	0.0578	0.9630
	3/8	YES	0.4836	0.0129	0.0573	0.9110	0.4867	0.0160	0.0581	0.9630
Lunn Davies	None	NO	0.4491	-0.0216	0.0496	0.5790	0.4533	-0.0174	0.0505	0.9890
	3/8	NO	0.4532	-0.0175	0.0505	0.5830	0.4587	-0.0120	0.0517	0.9860
	None	YES	0.4790	0.0083	0.0563	0.5170	0.4945	0.0238	0.0599	0.9680
	3/8	YES	0.4790	0.0083	0.0563	0.5170	0.4945	0.0238	0.0599	0.9670
Gaussian Copula	None	NO	0.4317	-0.0390	0.0459	0.8050	0.4380	-0.0327	0.0472	0.9140
	3/8	NO	0.4387	-0.0320	0.0473	0.8450	0.4454	-0.0253	0.0488	0.9410
	None	YES	0.4820	0.0113	0.0570	0.9050	0.4859	0.0152	0.0579	0.9640
	3/8	YES	0.4829	0.0122	0.0572	0.9070	0.4868	0.0161	0.0581	0.9630

tion is reduced and becomes positive due to bias correction, and the coverage is noticeably improved. Known ρ results in somewhat higher, and the estimated ρ in considerably lower than nominal coverage, reaching 91% at 95% nominal level for the BB and the GC generated data as compared to 96-97% for all generation mechanisms when ρ is known and the bias correction is used. Unfortunately, for the LD generation the estimation of ρ by ρ_{AOV} clearly does not work, resulting in abysmal coverage with or without the bias correction. Once more, the Anscombe correction does not seem to be of much benefit when the bias correction is

used. To summarise, low prevalence is considerably more challenging to estimate correctly. The perils of routine use of transformations are very clear in this example, and the proposed bias correction is of much benefit.

Bias in the estimation of intra-class correlation ρ by ρ_{AOV} and ρ_{PPR} is plotted in Figure 3.12.

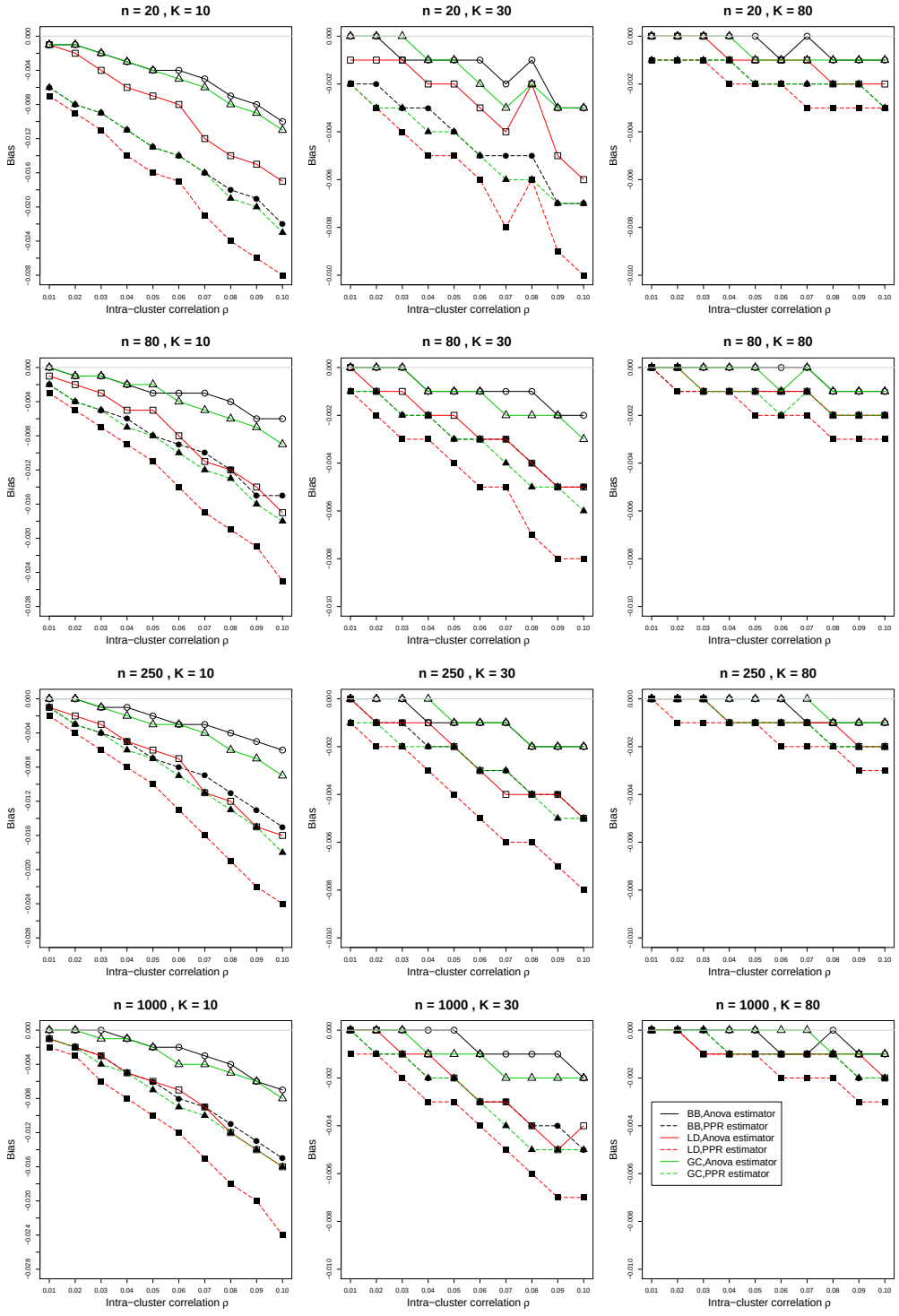


Figure 3.12: Bias of ρ from K studies in overdispersed binomial model for $p = 0.1$. Simulations (10000 for each values of ρ) from the beta-binomial distribution (black); from the Lunn and Davies (1998) model (red); from the Gaussian Copula model by Emrich and Piedmonte (1991) (green); the bias term plotted from $\hat{\rho}_{AOV}$ and $\hat{\rho}_{PPR}$ estimators (solid and dashed lines, respectively).

3.6 Summary

We have investigated bias arising in the estimation of transformed probabilities under the assumptions of random or mixed effects models, and its deleterious effects on inference in a meta-analysis. We demonstrated and quantified these effects in the examples of arcsine and log-odds transformations for overdispersed binomial data. In the standard additive REM of meta-analysis, the random effect is modeled as the between-study variance component τ^2. In the overdispersion model (Kulinskaya and Olkin, 2014)), the overdispersion parameter can be interpreted as the intra-cluster correlation coefficient. Both models can be described in the common framework of overdispersion.

Let $Y_1, \ldots, Y_n$ be identically and independently distributed variables with cumulative density function $F(Y_i, \mu)$. Alternatively, assume an overdispersed model with $Y_i \sim F(Y_i, \mu)$ and $\mu \sim G(\mu, \tau^2/n)$. Cox (1983) compared ML estimates $\hat{\mu}$ and $\hat{\mu}^+$ for the original and the overdispersed model, respectively, under contiguous alternatives $g_\tau(x)$. He found that $\hat{\mu}^+ - \hat{\mu}$ is proportionate to τ unless the parametrization is chosen to eliminate the bias of order n^{-1} in μ. The model specification is important in this context: "if the log linear model specifies a Poisson distribution for Y_i with $\log \mathrm{E}(Y_i) = x_i^T\beta$, the overdispersed model should have $E(Y_i) = \exp(x_i^T\beta)$, with $\mathrm{Var}(Y_i) > \mathrm{E}(Y_i)$. An overdispersed model in which Y_i is considered to have a Poisson distribution with $\log \mathrm{E}(Y_i) = x_i^T\beta + \epsilon_i$, where ϵ_i in turn is a random variable of expectation zero, would, however, lead to the inconsistencies..." (Cox, 1983, p. 273).

In the same vein, we have demonstrated in Section 3.2 that for close alternatives to the fixed effect model, any nonlinear transformation of an overdispersed random variable has a bias that is linear in τ. We have seen in simulations, both for log-odds and arcsine transformation, that the reduction in bias of a transformation under the fixed effect model reduces bias under the REM. We have used the Gart et al. (1985) and Anscombe (1948) continuity corrections to this end. Unfortunately, this, in general, is not sufficient to correct the bias under the REM. Additionally, the continuity correction is more complicated in regression setting.

Gart et al. (1985) discuss the bias reduction for the logit. Let the empirical logit be $L_a(X) =$

$\log(X+a)/(n-X+a)$."It is not possible to recommend a universal correction a for $L_a(x)$ in weighted linear regression; sometimes $a = 1/2$ is best, other times $1/4,\ 0,\ 1/2$, or intervening values are appropriate. The estimation of its variance also presents problems of bias and correlation." (Gart et al., 1985, p. 187).

We have seen that depending on the way the REM is defined, the primary statistic may be unbiased but any transformation of this statistic is biased unless the transformation is a linear function of the primary statistic. Thus, the linear models are bias-free, but the Generalised Linear Mixed Models (GLMMs) are not, because they involve a transformation. Other popular classes of transformations which are affected are the variance stabilizing and the normalizing transformations. This may have important implications in data analysis, where these kinds of transformations are routinely performed. In section 3.3.1, we demonstrated how large an effect of these small biases may be in the context of meta-analysis, and explained the reasons for these findings.

Model misspecification bias in meta-analysis of rates and proportions is discussed in (Trikalinos et al., 2013, p.81). The authors studied by simulation the log, log-odds and arcsine transformations of the estimated probability p under beta-binomial and binomial-uniform (i.e. discrete uniform) distributions. They noted very small bias of the arcsine transformation as compared to log-odds and log transformations, and recommended the use of inference based on the arcsine transformation without Anscombe continuity correction in meta-analysis. Our results show that this recommendation cannot be accepted without reservations. They also noticed that for both the log-odds and arcsine transformations "coverage appears to become worse with increasing K, and more so for scenarios where heterogeneity is large," but failed to explain this pattern.

Our simulations confirm that the bias of log-odds and arcsine transformations are linear in ρ for small values of intra-cluster correlation. These biases do not depend on the sample sizes or the number of studies in meta-analysis and result in abysmal coverage of the combined effect for large K. As a remedy, we proposed a plug-in bias correction for the arcsine transformation

in meta-analysis. For a large number of studies $K \geq 30$, and for well-behaved overdispersed binomial distributions such as the beta-binomial or the GC model, this correction improves the coverage and reduces the bias for $\rho < 0.06$. For $\rho > 0.06$, the coverage still deteriorates. For the log-odds transformation of proportions, it is more difficult to provide a similar bias correction due to the dependence between the probabilities of the outcome and the weights. Random effects models are often written without any details on how the overdispersion is generated. However, we demonstrated that knowing just two moments of a distribution is not sufficient. When meta-analysis includes just a few studies, the mechanism of randomness is difficult to ascertain. In such cases, our examples show that it will be nearly impossible to get a realistic bias correction for large ICC. How to safeguard against misspecification of the REM and which method to use in a meta-analysis is an open question. If the REM is specified on the original scale X, the transformed effect measure $f(X)$ is biased. It appears to be safer to specify the REM on the transformed scale when the inference on this scale is preferable. These considerations may apply in the context of a meta-analysis, where the REM is rather artificial to start with, and therefore there is some freedom on how to define it. Such a freedom is not ordinarily present in the analysis of real data, where the correct model is paramount.

Chapter 4

Multiplicative random effects model for binary data

4.1 Introduction

In chapter 2, we discussed the standard models of meta-analysis: the fixed effect model (FEM) and the random effects model (REM) for log-odds ratios (see Chapter 2 for details). The former assumes that the LORs $\theta_i,\ i = 1, \cdots, K$, do not differ across the studies, i.e. $\theta_i \equiv \theta$; the latter assumes that the LORs $\hat{\theta}_i$ themselves are a random sample from, usually, normal distribution, $\theta_i \sim N(\theta_{RE}, \tau^2)$ with the between-studies variance τ^2. Further, for large sample sizes, estimated LORs are approximately normally distributed, $\hat{\theta}_i \sim N(\theta_i, \sigma_i^2)$. Therefore, the REM considers that $\hat{\theta}_i \sim N(\theta_{RE}, \sigma_i^2 + \tau^2)$, and the FEM follows for $\tau^2 = 0$. Importantly, the variances σ_i^2 are of order $O(1/n_i)$ for sample sizes $n_i,\ i = 1, \cdots, K$ of the studies. Standard inference concerns the combined effect $\hat{\theta}$, estimated as the weighted mean of the individual effects from (2.6.4) with weights equal to inverse estimated variances, $w_i = \hat{\sigma}_i^{-2}$ in FEM, and $w_i = (\hat{\sigma}_i^2 + \hat{\tau}^2)^{-1}$ in REM. The distribution of the combined effect $\hat{\theta}_{RE}$ is customarily approximated by a normal distribution, $N(\theta_{RE}, (\sum_{i=1}^{K} w_i)^{-1})$. Estimated within-studies variances $\hat{\sigma}_i^2$ are often assumed to be known. Establishing an effect of treatment corresponds to testing the null hypothesis $\theta_{RE} = 0$, and a confidence interval calculation in REM requires an estimate of the between-studies variance τ^2, which is also of interest for quantifying heterogeneity.

The shortcomings of the inverse-variances method, as described above, in meta-analysis in

general, and in its application to the LORs are well known. They include the bias in estimation of the combined effect, underestimation of its variance, and poor coverage of the obtained confidence intervals, especially for sparse data and/or small sample sizes, see Kulinskaya et al. (2014) for discussion and further references. Under FEM, a considerably better way to combine odds ratios is the Mantel and Haenszel (1959) method. Unfortunately, there is no analogue to this method under the REM. Van Houwelingen et al. (1993) proposed the random-effects conditional logistic model with the natural generalization of the Mantel-Haenszel method to random-effects models. However, the assumption of our model and the model by Van Houwelingen et al. (1993) are slightly different. We assume a pair of beta-binomial distributions across K studies. Van Houwelingen et al. (1993) assume standard binomially distributed events within each study and normally distributed random effects between K studies (see Van Houwelingen et al. (1993) for details).

Further, the most popular method of estimating the between-studies variance τ^2 is the DerSimonian and Laird (1986) method based on the approximate chi-square moments of the Cochran's Q statistic, Cochran (1937), and this method is not satisfactory both in general, and in application to the heterogeneity estimation of LORs, see Hoaglin (2016); Kulinskaya and Dollinger (2015, 2016). Kulinskaya and Dollinger (2015) recommend the use of the Breslow-Day test (Breslow and Day, 1980) for testing the heterogeneity of ORs, and also provide a new gamma-based approximation to distribution of Q.

Alternative approaches to REM include the use of fixed weights (Shuster, 2010; Shuster and Walker, 2016) and the overdispersion model (ODM) introduced by Kulinskaya and Olkin (2014). The ODM allows the interpretation of overdispersion through intra-cluster correlation ρ or its transformation.

In this chapter we further develop the ODM for odds ratios, using a pair of independent beta-binomial distributions to describe the variability in both arms. This model includes binomial distributions for positive responses in both arms, conditional on the probabilities, and allows beta-distributed variation of the probabilities across the studies. For the log odds ratios from

a pair of beta-binomial distributions, the normal approximation has been suggested by Zelen and Parker (1986) and Ashby et al. (1993). To obtain the combined effect, we study the standard inverse-variance method and a version of the Mantel-Haenszel method adjusted for clustering, based on the work by Donner et al. (2001) and Chen (2012). Both methods require estimation of the intra-cluster correlation ρ. We study several methods of estimating ρ, including two new methods, one based on the profiling of the Breslow-Day test, and another based on the gamma approximation to the distribution of Q by Kulinskaya and Dollinger (2015).

The structure of this chapter is as follows. The proposed beta-binomial model for meta-analysis of odds ratios and the Mantel-Haenzsel-inspired estimation of the combined odds ratio are introduced in Section 4.2 and 4.3. Five methods of estimation of an overdispersion parameter, ρ, including a new method based on the BD test, are given in 4.4. An example is provided in Section 4.5. A large simulation study is described in Section 4.6. Discussion and conclusions are in Section 4.7. This Chapter represents the novel work of this thesis.

4.2 Odds ratio under beta-binomial model

In the paper by Kulinskaya and Olkin (2014), the idea of proposed overdispersion model comes from taking into account an intra-cluster correlation between observations of normal and binomial data in each study. Similarly, by accounting for the intra-cluster correlation between observations in 2×2 contingency tables, we propose a version of random effects model for the Mantel-Haezsel odds ratio. The standard Mantel-Haenzsel method for odds ratios given by (2.4.6) is applied under assumption of the fixed effect model. In fixed effect model, X_{i1} and X_{i2}, $i = 1, \ldots, K$ are assumed to be independent binomial random variables. Each of the binomial random variables is the sum of Bernoulli variables such that $X_{ij} = \sum_{k=1}^{n_{ij}} X_{ijk}$ for j=1,2. Fixed effect model assumes that odds ratios ψ_i or logarithms of odds ratio $\log(\psi_i)$ are homogeneous across K 2×2 contingency tables. In practice, studies might differ due to environmental or experimental factors and may produce contradictory results. The random effects model which includes an assumption of heterogeneity between studies is more realistic.

Applying the standard random effects model for logarithms of odds ratios, a heterogeneity of studies or an inflation in the variance of overall effect measure is explained by variation of log odds ratios between studies. This variation is usually quantified by an extra between-study variance τ^2 in the additive model (2.6.2). In terms of overdispersion, the additive random effects model (2.6.2) is overdispersed whenever $\tau^2 \neq 0$ and multiplicative model (2.6.5) is overdispersed for $\phi > 1$. When, the opposite is true ($\tau^2 = 0$ or $\phi = 1$), the models reduce to fixed effect model. In general the random effects models for meta-analysis are overdispersed relative to fixed effect model.

In this Section we propose a new version of multiplicative random effects model (2.6.2) which can be used with the Mantel-Haenzsel method for combining K odds ratios. In this model, inflation in the variance is also explained by overdispersion. The overdispersion in the random effect model can be explained by the intra-cluster correlation between Bernoulli observations. Usual fixed effect model assumes no dependence between Bernoulli observations. In our case, this dependence between Bernoulli observations is considered for meta-analysis of contingency tables. Dependence may occur for example due to repeated measurements on the same patient as noted by Zhang and Boos (1997). Overdispersion on the data from 2×2 tables can also be referred to as heterogeneity of log odds ratio $\hat{\theta}_i$ according to Liu and Pierce (1993).

In order to introduce a new version of random effects model, consider the model where the distribution of group level sums are

$$X_{i1} \sim Bin(n_{i1}, p_{i1}) \qquad \text{and} \qquad X_{i2} \sim Bin(n_{i2}, p_{i2})$$

with probabilities under a common beta distribution

$$p_{i1} \sim Beta(\alpha_{i1}, \beta_{i1}) \qquad \text{and} \qquad p_{i2} \sim Beta(\alpha_{i2}, \beta_{i2}) \tag{4.2.1}$$

with parameters α_{ij}, β_{ij} for $j = 1, 2$ such that $\alpha_{ij} > 0$, $\beta_{ij} < 1$. The beta distribution is chosen due to conjugacy requirement. In the Bayesian language, beta distribution is a prior distribution for p_{ij} that ensures that the posterior distribution for X_{ij} belongs to beta-distribution.

In the standard fixed effect model, the probabilities p_{ij} are assumed fixed. In our case, we let probabilities p_{i1} and p_{i2} vary across the studies according to independent beta distributions. The variation of p_{i1} and p_{i2} corresponds to heterogeneity in the meta-analysis, since it affects to variation of logarithms of odds ratios. From above, marginally we have a pair of beta-binomial distributions

$$X_{i1} \sim BetaBinom(n_{i1}, \alpha_{i1}, \beta_{i1}) \qquad \text{and} \qquad X_{i2} \sim BetaBinom(n_{i2}, \alpha_{i2}, \beta_{i2})$$

instead of usual pair of binomial distributions for X_{i1} and X_{i2}. The variance (2.2.7) for observed number of positive responses X_{ij} shows that we also have a multiplicative linear factor $1 + (n_{ij} - 1)\rho_{ij}$ which allows the deflation and inflation in the variance of X_{ij} and therefore in the variance for log odds ratio discussed further. The parameters of beta-binomial distribution α_{ij} and β_{ij} are expressed in terms of a single parameter $\rho_{ij} = 1/(1 + \alpha_{ij} + \beta_{ij})$. Beta-binomial model is the true random effects model as it is a mixture model. Beta-binomial model allows variance inflation for X_{i1} and X_{i2} relative to binomial distribution. In meta-analysis we have variance inflation relative to fixed effect model. Thus, we have an over-dispersed random effects model with pair of beta-binomial distributions relative to fixed effect model with pair of binomial distributions. The presence of over-dispersion in binomial outcomes hence in log odds ratios can be explained by heavier tails of a distribution with clustering in comparison to the distribution without clustering (Crowder, 1979). In synthesising multiple studies, beta-binomial distribution is widely used in combining overdispersed data for event rates, see, for example Young-Xu and Chan (2008).

For a pair of beta-binomial distributions, the odds ratio (2.4.1) has the same form as before. However, assuming a common intra-cluster correlation $\rho_{ij} = \rho$ across groups within K studies, the variance of individual log transformed odds ratio has to be adjusted for intra-cluster correlation ρ and is given by

$$\text{Var}(\log(\hat{\psi}_i)) = \frac{1 + (n_{i1} - 1)\rho}{n_{i1}p_{i1}(1 - p_{i1})} + \frac{1 + (n_{i2} - 1)\rho}{n_{i2}p_{i2}(1 - p_{i2})} \tag{4.2.2}$$

with restrictions for ρ

$$\rho > \max_{1\leq i\leq K}\left\{-\frac{1}{n_{i1}-1},\ \ -\frac{1}{n_{i2}-1}\right\}.$$

A standard delta method can be used for derivation. For $\rho = 0$ this model is the standard fixed effect model for binomially distributed data. By the correspondence between the variance of logarithms of odds ratio and the variance of odds ratio, the variance of a logarithms of odds ratio is also given by (4.2.2) with restrictions mentioned above. We are mostly interested in the variance of LOR, because the distribution of estimated LORs is approximately normal. Assuming a normal distribution for estimated log-odds-ratio $\hat{\theta}_i = \log(\hat{\psi}_i)\ i = 1, ...K$, the fixed effect model with usual binomial distributions in treatment and control arms is

$$\hat{\theta}_i \sim N(\theta, \sigma_i^2)$$

where with substitution of (2.4.2)

$$\sigma_i^2 = \mathrm{Var}(\hat{\theta}_i) = \frac{1}{n_{i1}p_{i1}(1-p_{i1})} + \frac{1}{n_{i2}p_{i2}(1-p_{i2})}.$$

Re-parametrizing the variance of log odds ratio

$$\mathrm{Var}(\hat{\theta}_i) = \frac{1}{n_{i1}}\left[\frac{1}{p_{i1}} + \frac{1}{1-p_{i1}}\right] + \frac{1}{n_{i2}}\left[\frac{1}{p_{i2}} + \frac{1}{1-p_{i2}}\right]$$

or

$$\mathrm{Var}(\hat{\theta}_i) = \frac{1}{n_{i2}}\frac{n_{i2}}{n_{i1}}\left[\frac{1}{p_{i1}} + \frac{1}{1-p_{i1}}\right] + \frac{1}{n_{i2}}\left[\frac{1}{p_{i2}} + \frac{1}{1-p_{i2}}\right]$$

taking the common factor n_{i2}^{-1} out of the brackets, we obtain

$$\mathrm{Var}(\hat{\theta}_i) = \frac{1}{n_{i2}}\left[\frac{n_{i2}}{n_{i1}}\left(\frac{1}{p_{i1}} + \frac{1}{1-p_{i1}}\right) + \left(\frac{1}{p_{i2}} + \frac{1}{1-p_{i2}}\right)\right]$$

or

$$\mathrm{Var}(\hat{\theta}_i) = \frac{1}{n_{i1}+n_{i2}}\frac{n_{i1}+n_{i2}}{n_{i2}}\left[\frac{n_{i2}}{n_{i1}}\left(\frac{1}{p_{i1}} + \frac{1}{1-p_{i1}}\right) + \left(\frac{1}{p_{i2}} + \frac{1}{1-p_{i2}}\right)\right].$$

Denote $n_i = n_{i1} + n_{i2}$ and $R_i = n_{i1}/n_{i2}$, the allocation ratio, the variance can be rewritten as a function of n_i and R_i as

$$\mathrm{Var}(\hat{\theta}_i) = \frac{v_i(R_i)}{n_i} \approx {n_i}^{-1}(R_i+1)(R_i^{-1}(p_{i1}^{-1} + (1-p_{i1})^{-1}) + p_{i2}^{-1} + (1-p_{i2})^{-1}).$$

Now the fixed effect model for pair of binomials is

$$\hat{\theta}_i \sim N(\theta, \frac{v_i(R_i)}{n_i}).$$

$\mathrm{Var}(\hat{\theta}_i) = v_i(R_i)/n_i$ is the alternative form for variance (2.4.2) with the overall sample size $n_i = n_{i1} + n_{i2}$ and $R_i = n_{i1}/n_{i2}$ as an allocation ratio, $i = 1, .., K$. This results agrees with the result in the paper by Kulinskaya and Olkin (2014).

Opening the brackets in (4.2.2), variance of LOR for overdispersed binomial data is

$$\mathrm{Var}(\hat{\theta}_i) = \frac{v_i(R_i)}{n_i} + \left[\frac{(n_{i1}-1)}{n_{i1}p_{i1}(1-p_{i1})} + \frac{(n_{i2}-1)}{n_{i2}p_{i2j}(1-p_{i2})}\right]\rho$$

or

$$\mathrm{Var}(\hat{\theta}_i) = \frac{1}{n_{i1}p_{i1}(1-p_{i1})} + \frac{1}{n_{i2}p_{i2}(1-p_{i2})} + \big[\frac{n_{i1}-1}{n_{i1}p_{i1}(1-p_{i1})} + \frac{n_{i2}-1}{n_{i2}p_{i2}(1-p_{i2})}\big]\rho.$$

This variance is clearly inflated in comparison with the standard variance (2.4.2). Inflation term is of order $O(1)$ and increases with ICC, it also may be large for probabilities in each arm close to 0 or 1.

Defining

$$\tau_i = \left[\frac{(n_{i1}-1)}{n_{i1}p_{i1}(1-p_{i1})} + \frac{(n_{i2}-1)}{n_{i2}p_{i2}(1-p_{i2})}\right]\rho, \tag{4.2.3}$$

it is clear that when the variance component τ_i does not depend on i, the beta-binomial model results in the same two first moments of the LORs $\hat{\theta}_i$ as the standard REM, for appropriate choice of ρ and τ^2. This holds when the probabilities in the treatment and control arms do not differ $p_{ij} \equiv p_i$ and the sample sizes are all equal, or, at least approximately, when the sample sizes are all large, i.e. when $(n_{ij}-1)/n_{ij} \approx 1$.

The reformulated variance under beta-binomial model can be written as

$$\mathrm{Var}(\hat{\theta}_i) = \frac{1}{n_{i1}}(\frac{1}{p_{i1}} + \frac{1}{(1-p_{i1})}) + \frac{1}{n_{i2}}(\frac{1}{p_{i2}} + \frac{1}{(1-p_{i2})}) +$$

$$+\big[\frac{(n_{i1}-1)(p_{i1}^{-1} + (1-p_{i1})^{-1})}{n_{i1}} + \frac{(n_{i2}-1)(p_{i2}^{-1} + (1-p_{i2})^{-1})}{n_{i2}}\big]\rho$$

The first term is the standard within-study variance, which can be re-written exactly the same as before for fixed effects model. Hence defining the first term in alternative form and taking out of the brackets the factor $1/n_{i1}$, we have

$$\mathrm{Var}(\hat{\theta}_i) = \frac{1}{n_{i2}}(\frac{n_{i2}}{n_{i1}}(\frac{1}{p_{i1}} + \frac{1}{1-p_{i1}}) + (\frac{1}{p_{i2}} + \frac{1}{1-p_{i2}}))+$$

$$+\frac{1}{n_{i2}}\big[\frac{(n_{i1}-1)n_{i2}(p_{i1}^{-1} + (1-p_{i1})^{-1})}{n_{i1}} + (n_{i2}-1)(p_{i2}^{-1} + (1-p_{i2})^{-1})\big]\rho$$

which is the same as

$$\mathrm{Var}(\hat{\theta}_i) = \frac{1}{n_{i2}}(\frac{n_{i2}}{n_{i1}}(\frac{1}{p_{i1}} + \frac{1}{1-p_{i1}}) + (\frac{1}{p_{i2}} + \frac{1}{1-p_{i2}}))(1+a_i\rho)$$

for

$$a_i = \frac{(n_{i1}-1)R_i^{-1}(p_{i1}^{-1} + (1-p_{i1})^{-1}) + (n_{i2}-1)(p_{i2}^{-1} + (1-p_{i2})^{-1})}{R_i^{-1}(p_{i1}^{-1} + (1-p_{i1})^{-1}) + (p_{i2}^{-1} + (1-p_{i2})^{-1})} \tag{4.2.4}$$

Thus, the variance of logarithmic of odds ratio for a pair of beta-binomials is

$$\mathrm{Var}(\hat{\theta}_i) = \frac{1}{n_{i2}}(\frac{n_{i2}}{n_{i1}}(\frac{1}{p_{i1}} + \frac{1}{1-p_{i1}}) + (\frac{1}{p_{i2}} + \frac{1}{1-p_{i2}}))(1+a_i\rho)$$

or introducing the total sample size term $n_j = n_{1j} + n_{2j}$, the variance is

$$\mathrm{Var}(\hat{\theta}_i) = \frac{1}{n_{i1}+n_{i2}}\frac{n_{i1}+n_{i2}}{n_{i2}}(\frac{n_{i2}}{n_{i1}}(\frac{1}{p_{i1}} + \frac{1}{1-p_{i1}}) + (\frac{1}{p_{i2}} + \frac{1}{1-p_{i2}}))(1+a_i\rho)$$

which can be expressed as a function of n_i and R_i

$$\frac{v_i(R_i)}{n_i}\phi_i = {n_i}^{-1}(R_i+1)(R_i^{-1}(p_{i1}^{-1} + (1-p_{i1})^{-1}) + (p_{i2}^{-1} + (1-p_{i2})^{-1}))(1+a_i\rho).$$

The overdispersed random effects model with a pair of beta-binomial distributions is

$$\hat{\theta}_i \sim N(\theta, \frac{v_i(R_i)}{n_i}\phi_i), \tag{4.2.5}$$

where $\phi_i = (1+a_i\rho)$ and a_i is given by (4.2.4). Alternatively, the term a_i is

$$a_i = a(n_i, R_i, p_{i1}, p_{i2}) = n_i v_i(R_i)^{-1}\left[\frac{1}{p_{i1}(1-p_{i1})} + \frac{1}{p_{i2}(1-p_{i2})}\right] - 1.$$

Thus, a_i is a linear function of n_i and has the same order as n_i. Reparametrising a_i as a function of the control arm probability p_{i2} and the odds ratio ψ_i, a_i can be written as

$$a_i = \frac{n_i R_i[(1-p_{i2}(1-\psi_i))^2+\psi_i]}{(R_i+1)[(1-p_{i2}(1-\psi_i))^2+R_i\psi_i]} - 1.$$

For balanced studies $R_i = 1$, and a_i simplifies to $a_i = n_i/2 - 1$.

Alternative model is to consider overdispersion only in the treatment arm:

$$X_{i1} \sim BetaBinom(n_{i1}, \alpha_{i1}, \beta_{i1}) \qquad X_{i2} \sim Bin(n_{i1}, p_{i1})$$

This is perhaps closer to the standard random effects model which usually has a random effect only in the treatment arm (Thompson and Sharp, 1999)). In this case, the variance for odds ratio is

$$\mathrm{Var}(\hat{\psi}_i) = \frac{1+(n_{i1}-1)\rho}{n_{i1}p_{i1}(1-p_{i1})} + \frac{1}{n_{i2}p_{i2}(1-p_{i2})}.$$

which is still inflated, in comparison to the FEM variance, by the term $[(n_{i1}-1)/(n_{i1}p_{i1}(1-p_{i1}))]\rho$. Subsequent methods are easily adapted to this version of the ODM, and we do not pursue this model further. Later, Section 6.2.3 of chapter 6 studies the correspondence of heterogeneity parameters between standard and novel overdispersed random effects models.

4.3 Adjusted Mantel-Haenzsel method for combining odds ratios

Applying the Standard Mantel-Haenzsel method to clustered binary data produces wrong Type I error and downward bias for the p-values associated with χ^2_{MH} according to Darlington and Donner (2007). χ^2_{MH} is the Mantel-Haenzsel test statistic for testing if odds ratio $\psi = 1$. In the paper by Donner and Klar (2002) and Darlington and Donner (2007) an adjusted version of the Mantel-Haenszel test and related estimator of the odds-ratio (OR) appropriate for the meta-analysis of cluster-randomised trials are introduced. Darlington and Donner (2007) compared several methods including the Adjusted Mantel-Haenszel method for combining

the binary data from contingency tables by Monte-Carlo simulations. Other methods for combining clustered binary data include: a ratio procedure based on the idea of design effect suggested by Scott and Holt (1982), the general inverse variance approach provided by the Cochrane Collaboration and the Woolf procedure suggested by Woolf et al. (1955).

In the paper by Darlington and Donner (2007) about cluster-randomised trials, each arm j of trial i contains m_{ij} clusters of size n_{ij}, and there is an intra-class correlation ρ_i common for all clusters in the trial i. This can be adapted to a case of one cluster in each arm, which is equivalent to the over-dispersion based random effects model (or ODM) introduced in Kulinskaya and Olkin (2014). The resulting statistic is as follows. The correction factor for the Mantel-Haenszel odds ratio is

$$C_{ij} = 1 + (n_{ij} - 1)\hat{\rho}_i \text{ for } j = 1, 2;\ i = 1, \cdots, K, \tag{4.3.1}$$

where $\hat{\rho}_i$ is the estimated intra-class correlation ρ_i. The correction factors are referred to design effects by Scott and Holt (1982). For the Mantel-Haenzsel odds ratio, defining the adjusted weights

$$W_{iC} = \big[\frac{C_{i1}}{n_{i1}} + \frac{C_{i2}}{n_{i2}}\big]^{-1}(1 - \hat{p}_{i1})\hat{p}_{i2}.$$

the corrected Mantel-Haenzsel odds ratio for intra-cluster correlation is

$$\hat{\psi}_{CMH} = \frac{\sum_{i=1}^{K} W_{iC}\hat{\psi}_i}{\sum_{i=1}^{K} W_{iC}} \qquad \text{for} \qquad \hat{\psi}_i = \frac{(1 - \hat{p}_{i2})\hat{p}_{i1}}{(1 - \hat{p}_{i1})\hat{p}_{i2}}. \tag{4.3.2}$$

This is the Mantel-Haenzsel odds ratio which can be used in the multiplicative random effects model. When there is no intra-cluster correlation ($\rho_i = 0$), then $C_{ij} = 1$ and the expression (4.3.2) reduces to standard Mantel-Haenzsel odds ratio (2.4.6) for fixed effects model in meta-analysis.

When $\rho \to -1/\max(a_i)$, then

$$\hat{\psi}_{CMH} \to \frac{\sum_{n_{i1}=n_{i2}=\max(n_i)} n_{i1}n_{i2}p_{i1}(1 - p_{i2})/(n_{i1} + n_{i2})}{\sum_{n_{i1}=n_{i2}=\max(n_i)} n_{i1}n_{i2}p_{i2}(1 - p_{i1})/(n_{i1} + n_{i2})}, \tag{4.3.3}$$

which is the standard Mantel-Haenzsel odds ratio without corrections (4.3.1). In (4.3.3), $n_{i1} = n_{i2} = \max(n_i)$ is the maximum values sample sizes across K studies. The proof is

provided in B.2 of Appendix B.

To obtain the asymptotic variance of $\hat{\psi}_{CMH}$, we adjusted for overdispersion the asymptotic variance of Mantel-Haenzsel odds ratio derived by Robins et al. (1986) and Phillips and Holland (1987):

$$\mathrm{Var}(\hat{\psi}_{CMH}) = \frac{\sum_{i=1}^{K} R_i P_i}{2R^2} + \frac{\sum_{i=1}^{K}(P_i S_i + Q_i R_i)}{2RS} + \frac{\sum_{i=1}^{K} S_i Q_i}{2S^2},$$

where

$$P_i = \frac{C_{i1}(n_{i2} - X_{i2}) + C_{i2}X_{i1}}{C_{i1}n_{i2} + C_{i2}n_{i1}}, \quad R_i = \frac{(n_{i2} - X_{i2})X_{i1}}{C_{i1}n_{i2} + C_{i2}n_{i1}},$$
$$Q_i = \frac{C_{i2}(n_{i1} - X_{i1}) + C_{i1}X_{i2}}{C_{i1}n_{i2} + C_{i2}n_{i1}}, \quad S_i = \frac{(n_{i1} - X_{i1})X_{i2}}{C_{i1}n_{i2} + C_{i2}n_{i1}},$$

and $R = \sum_{i=1}^{K} R_i$ and $S = \sum_{i=1}^{K} S_i$.

The variance for logarithms of Mantel-Haenzsel odds ratio can be obtained by correspondence between ψ and $\log(\psi)$. The confidence interval for adjusted logarithms of common odds ratio $\log(\hat{\psi}_{MH})$ can be obtained as

$$\log(\hat{\psi}_{MH}) \pm Z_{\alpha/2}[\mathrm{Var}(\log(\psi_{MH}))^{1/2}]$$

where $Z_{\alpha/2}$ is the $(1-\alpha)100$ two sided critical value of the standard normal distribution. The confidence interval for ψ is obtained by inverting the interval above from logarithmic to odds ratio scale.

4.4 Estimation of ρ

To be able to evaluate the corrected MH odds-ratio (4.3.2) and an estimate of LOR from the inverse-variance method in ODM (4.2.5), an estimate of the intra-cluster correlation ρ is required. We consider two modifications of established methods, namely a moment estimator based on Cochran's Q statistic similar to DerSimonian and Laird (1986) estimator of τ^2, and a restricted maximum likelihood (REML) estimator. We also consider related confidence intervals: an interval based on profiling the Q statistic as in Viechtbauer (2007) and a REML-based interval. Both approaches were proposed in Kulinskaya and Olkin (2014) but

were not explored by simulation. As an alternative, we propose to invert the Breslow and Day (1980) (BD) test for both point and interval estimation. The point estimation is based on an adaptation of the Mandel and Paule (1970) method, and the interval estimation is achieved through profiling the modified BD test. We also propose a point and interval estimator similar to Mandel-Paule point estimator and Q-profile interval estimator of ρ based on the approximation of Q statistic by gamma distribution (Kulinskaya and Dollinger, 2015).

4.4.1 Q-statistic based estimation of ρ

Cochran's Q statistic is $Q = \sum_{i=1}^{K} w_i(\hat{\theta}_i - \bar{\theta}_w)^2$, for the inverse variance weights $w_i = \sigma_i^{-2}$ and $\bar{\theta}_w = \sum_{i=1}^{K} w_i\hat{\theta}_j / \sum_{i=1}^{K} w_i$. Under the null hypothesis of no over- or underdispersion $\rho = 0$, the Q-statistic is approximately chi-square distributed with $K-1$ degrees of freedom, so that $\mathrm{E}(Q) = K - 1$. Under the ODM,

$$\mathrm{E}(Q) = \mathrm{E}\left[\sum_{i=1}^{K} w_i(\hat{\theta}_i - \bar{\theta}_w)^2\right] \quad \rightarrow \quad \mathrm{E}(Q) = \sum_{i=1}^{K} \mathrm{E}\left[w_i(\hat{\theta}_i - \bar{\theta}_w)^2\right]$$

Assuming w_i and $(\hat{\theta}_i - \bar{\theta}_w)^2$ are independent and weights w_{ij} are known

$$\mathrm{E}(Q) = \sum_{i=1}^{K} \mathrm{E}(w_i)\mathrm{E}(\hat{\theta}_i - \bar{\theta}_w)^2 = \sum_{i=1}^{K} w_i\mathrm{E}(\hat{\theta}_i - \bar{\theta}_w)^2$$

with

$$\mathrm{E}(\hat{\theta}_i - \bar{\theta}_w)^2 = \mathrm{Var}(\hat{\theta}_i) - 2\mathrm{Cov}(\hat{\theta}_i, \bar{\theta}_w) + \mathrm{Var}(\bar{\theta}_w)$$

or

$$\mathrm{E}(\hat{\theta}_i - \bar{\theta}_w)^2 = \frac{1 + a_i\rho}{w_i} - \frac{2(1 + a_i\rho)}{W} + \frac{(1 + \bar{a}_w\rho)}{W}$$

since

$$\mathrm{Cov}(\hat{\theta}_i - \bar{\theta}, \hat{\theta}_j - \bar{\theta}) = \delta_{ij}\frac{1 + a_i\rho}{w_i} - \frac{(1 + a_i\rho)}{W} - \frac{(1 + a_j\rho)}{W} + \frac{(1 + \bar{a}_w\rho)}{W}$$

where δ_{ij} is the Kronecker delta (Kulinskaya and Olkin, 2014). Hence, the expected value of Q statistic is

$$\mathrm{E}(Q) = \sum_{j=1}^{K}(1 + a_i\rho) - \frac{2\sum_{i=1}^{K} w_i(1 + a_i\rho)}{W} + \frac{\sum_{i=1}^{K} w_i(1 + \bar{a}_w\rho)}{W}$$

$$\mathrm{E}(Q) = K - 1 + \sum_{i=1}^{K} a_i\rho - \frac{2\sum_{i=1}^{K} w_i a_i \rho}{W} + \bar{a}_w\rho$$

which results in

$$\mathrm{E}(Q) = K - 1 + (K\bar{a} - \bar{a}_w)\rho, \tag{4.4.1}$$

where $\bar{a} = \sum_{i=1}^{K} a_i/K$, $\bar{a}_w = \sum_{i=1}^{K} w_i a_i/W$, and $W = \sum_{i=1}^{K} w_i$ (Kulinskaya and Olkin, 2014). The estimate of ρ from equation (4.4.1) should satisfy the condition $\hat{\rho} > -1/\max(a_i)$.

The estimator in the spirit of random effects estimates of variance (e.g., DerSimonian and Laird, 1986) proposed by Kulinskaya and Olkin (2014) is

$$\hat{\rho}_M = \max\left(\frac{Q - (K-1)}{K\bar{a} - \bar{a}_w}, -\frac{1}{a_{\max}}\right); \tag{4.4.2}$$

underdispersion is present for $Q < K - 1$. If only positive values of ρ are acceptable, then $\hat{\rho}$ can be truncated at zero.

Related confidence interval is obtained by inverting the Q test. When the correct weights $w_i^* = w_i(\rho) = w_i/(1 + a_i\rho)$ are used, the corrected Q statistic given by $Q^*(\rho) = \sum_{i=1}^{K} w_i^*(\hat{\theta}_i - \bar{\theta}_{w^*})^2$ has approximately the χ^2_{K-1} distribution. The confidence interval is constructed as

$$\left\{\rho > -1/a_{max} : \chi^2_{K-1;\alpha/2} \le Q^*(\rho) \le \chi^2_{K-1;1-\alpha/2}\right\}. \tag{4.4.3}$$

Viechtbauer (2007) shows that the standard random effects model confidence intervals for τ^2 based on this approach, named Q-profile, perform very well, better than the restricted maximum likelihood confidence intervals described next.

4.4.2 Restricted maximum likelihood based estimation of ρ

The restricted likelihood for the normal distribution with mean θ and variances $v_i(1 + a_i\rho)/n_j$ is

$$l_R(\rho, \theta) = -\frac{1}{2}\log(\sum_{i=1}^{K} w_i^*) - \frac{1}{2}\sum_{i=1}^{K} w_i^*(\theta_i - \theta)^2 + \frac{1}{2}\sum_{i=1}^{K}\log(w_i^*), \tag{4.4.4}$$

for inverse-variance weights $w_i^* = w_i/(1 + a_i\rho)$. Following Kulinskaya and Olkin (2014), the REML equation for ρ is

$$(W^*)^{-1}\sum_{i=1}^{K} w_i^*\frac{a_i}{1 + a_i\rho} + \sum_{i=1}^{K} w_i^*(\theta_i - \theta)^2\frac{a_i}{1 + a_i\rho} = \sum_{i=1}^{K}\frac{a_i}{1 + a_i\rho}, \tag{4.4.5}$$

where $W^* = \sum_{i=1}^{K} w_i^*$. The mean $\hat{\theta}_{\text{REML}}$ is obtained as $\hat{\theta}_{\text{REML}} = \sum_{i=1}^{K} w_i^* \hat{\theta}_i / W^*$, and an iterative procedure readily yields a solution.

The REML confidence intervals are given by all values of ρ which satisfy

$$l_R(\rho) \geq l_R(\hat{\rho}_{\text{REML}}) - \chi^2_{1;1-\alpha/2}, \tag{4.4.6}$$

where $\chi^2_{1;1-\alpha}$ is the $(1-\alpha)$ percentage point of chi-square distribution with 1 degree of freedom.

4.4.3 Mandel-Paule estimation of ρ

The Mandel-Paule method of estimation of between-studies variance τ^2 in the standard REM was introduced by Mandel and Paule (1970) and studied subsequently by Rukhin (2003) and DerSimonian and Kacker (2007). This method uses the first moment of the approximate chi-square distribution of the Cochran's Q statistic under homogeneity to find an estimate of τ^2. In ODM, under the alternative hypothesis $\rho \neq 0$, the Q statistic with adjusted weights $w_i^*(\rho) = w_i/(1 + a_i\rho)$ is

$$Q^*(\rho) = \sum_{i=1}^{K} \frac{(\hat{\theta}_i - \bar{\theta}_{w^*(\rho)})^2}{\sigma_i^2(1 + a_i\rho)}.$$

The unique estimate of ρ can be obtained iteratively by Mandel and Paule (1970) method from equation $Q^*(\rho) = K - 1$ given that a solution exist. This method is based on the $Q^*(\rho)$ has an approximately chi-square distribution with $K - 1$ degrees of freedom. Thus the first moment of $Q^*(\rho)$ is $K - 1$.

4.4.4 Corrected Q-statistic based estimation of ρ

According to Kulinskaya and Dollinger (2015), the distribution of Q statistic can be better approximated by a gamma distribution with shape and scale parameters

$$r(\rho) = \frac{\text{E}(Q)^2}{\text{Var}(Q)} \quad \text{and} \quad \lambda(\rho) = \frac{\text{Var}(Q)}{\text{E}(Q)}.$$

The expected value and variance of Q statistic for log odds ratio can be estimated from the relations

$$(K - 1) - \text{E}(Q) = 0.678[(K - 1) - \text{E}_{th}(Q)] \tag{4.4.7}$$

and

$$\mathrm{Var}(Q) = 4.74(K-1) - 12.17\mathrm{E}[Q] + 9.42\mathrm{E}[Q]^2/(K-1), \tag{4.4.8}$$

where $\mathrm{E}_{th}(Q)$ is the theoretical approximation to the mean of Q for log odds ratio (Kulinskaya and Dollinger, 2015). The Mandel-Paule estimate of ρ based on moments (4.4.7) and (4.4.8) of Q statistic is $Q^*(\rho) = E(Q)$ given that a solution exist, where $E(Q)$ is the solution of equation (4.4.7).

The related confidence interval based on gamma approximation of distribution for Q statistic can be obtained from

$$\left\{\rho > -1/a_{max} : \Gamma_{r(\rho),\lambda(\rho);\alpha/2} \leq Q^*(\rho) \leq \Gamma_{r(\rho),\lambda(\rho);1-\alpha/2}\right\}, \tag{4.4.9}$$

where $\Gamma_{r(\rho),\lambda(\rho);\alpha/2}$ and $\Gamma_{r(\rho),\lambda(\rho);1-\alpha/2}$ are the quantiles of gamma distribution with $r(\rho)$ and $\lambda(\rho)$ as shape and scale parameters.

4.4.5 Breslow-Day based estimation of ρ

The chi-square distribution is a poor approximation to the distribution of Q statistic for LORs (Kulinskaya and Dollinger, 2015), and the Breslow-Day (BD) test is an attractive alternative for testing the heterogeneity of ORs. In the Section we propose a new method of estimation of ρ based on modification of the BD test for the overdispersed data.

The Breslow-Day test is based on the statistic

$$X^2_{BD} = \sum_{i=1}^{K} \frac{(X_{i1} - E(X_{i1}|\hat{\psi}_{MH}))^2}{\mathrm{Var}(X_{i1}|\hat{\psi}_{MH})} \sim \chi^2_{K-1}, \tag{4.4.10}$$

where $E(X_{i1}|\hat{\psi}_{MH})$ and $\mathrm{Var}(X_{i1}|\hat{\psi}_{MH})$ denote the expected number and the asymptotic variance, respectively, of the number of cases in the treatment arm under the assumption of homogeneity of odds ratios, given the fitted Mantel-Haenzsel odds ratio $\hat{\psi}_{MH}$. The expected value $E(X_{i1}|\hat{\psi}_{MH})$ in (4.4.10) is obtained from the quadratic equation

$$\frac{E(X_{i1}|\hat{\psi}_{MH})[N_i - x_i - n_{i1} + E(X_{i1}|\hat{\psi}_{MH})]}{[x_i - E(X_{i1}|\hat{\psi}_{MH})][n_{i1} - E(X_{i1}|\hat{\psi}_{MH})]} = \hat{\psi}_{MH}, \tag{4.4.11}$$

where $x_i = X_{i1} + X_{i2}$. Its asymptotic variance $\mathrm{Var}(X_{i1}|\hat{\psi}_{MH})$ is a particular case, for $\rho = 0$, of the variance of X_{i1} under overdispersion given by (Song, 2004):

$$\mathrm{Var}(X_{i1}|\hat{\psi}_{CMH}) = \big[\frac{1}{E(X_{i1}|\hat{\psi}_{CMH})C_{i1}} + \frac{1}{(x_i - E(X_{i1}|\hat{\psi}_{CMH}))C_{i2}} + \frac{1}{(n_{i1} - E(X_{i1}|\hat{\psi}_{MH}))C_{i1}} + \frac{1}{(N_i - x_i - n_{i1} + E(X_{i1}|\hat{\psi}_{MH}))C_{i2}}\big]^{-1}, \tag{4.4.12}$$

where C_{ij} terms are given by (4.3.1). The asymptotic variance given above is not defined when any of the cells of the i-th 2 by 2 table are empty. In these cases, a continuity correction of 0.5 is added to each cell of such a table.

The Breslow-Day statistic $X^2_{BD} = X^2_{BD}(\rho)$ is now a function of ρ and $X^2_{BD}(\hat{\rho})$ has an approximately χ^2_{K-1} distribution under the homogeneity of odds ratios given that the value of ρ is estimated correctly. Equating the BD statistic to its first moment $K - 1$,

$$\sum_{i=1}^{K} \frac{(X_{i1} - E(X_{i1}|\hat{\psi}_{MH}))^2}{\mathrm{Var}(X_{i1}|\hat{\psi}_{MH})} = K - 1, \tag{4.4.13}$$

and solving this estimating equation for $\hat{\rho}$, provide a Mandel-Paule type estimator $\hat{\rho}_{BD}$, which can be used for the calculation of odds ratio $\hat{\psi}_{CMH}$ given by (4.3.2).

The range of values for overdispersion parameter ρ is constrained to an interval

$$(\max\{-1/a_{\max}, -1/\max(n_{ij} - 1)\},\ 1).$$

When $\rho \to -1/(n_{ij} - 1)$, the variance in the denominator of $X^2_{BD}(\rho)$ converges to zero and Breslow-Day statistic tends to infinity. When $X^2_{BD}(\rho = 0) < K - 1$, the solution of equation (4.4.13) always exists. On the other hand, $\rho = 1$ provides the lower limit for Breslow-Day statistic, so if this lower limit of $X^2_{BD}(\rho = 0) > K - 1$, the equation (4.4.13) does not have a solution; in this case we set $\hat{\rho} = 1$.

The confidence interval for ρ can be obtained by profiling the Breslow-Day test, similarly to confidence interval for τ^2 obtained by profiling Cochran's Q under REM by Viechtbauer (2007). The confidence interval with 95 percent coverage probability for the intra-cluster correlation parameter ρ based on the modified Breslow-Day test is given by

$$\left\{1 > \rho > \max\{-\frac{1}{a_{\max}}, -\frac{1}{\max(n_{ij} - 1)}\} :\ \chi^2_{K-1,0.025} \le X^2_{BD}(\rho) \le \chi^2_{K-1,0.975}\right\}. \tag{4.4.14}$$

Table 4.1: Data for effects of diuretics on pre-eclampsia

	e.T	e.C	n.T	n.C	p.T	p.C
1	14	14	131	136	0.10	0.10
2	21	17	385	134	0.05	0.13
3	14	24	57	48	0.24	0.50
4	6	18	38	40	0.15	0.45
5	12	35	1011	760	0.01	0.05
6	138	175	1370	1336	0.10	0.13
7	15	20	506	524	0.03	0.04
8	6	2	108	103	0.05	0.02
9	65	40	153	102	0.42	0.39

4.5 Example: effects of diuretics on pre-eclampsia

A well-known meta-analysis of nine trials which include the total of 6942 patients, evaluated an effect of diuretics on pre-eclampsia (Collins et al., 1985). These data have been studied repeatedly, as for example in Hardy and Thompson (1996), Biggerstaff and Tweedie (1997), Viechtbauer (2007) and Kulinskaya and Olkin (2014). The basic data with odds ratios, and their logs are provided in (Kulinskaya and Olkin, 2014, Table 2a) and are reproduced in table 4.1. These data demonstrate considerable heterogeneity in incidence of pre-eclampsia in both the treatment and the control groups, Kulinskaya and Olkin (2014), suggesting that the BB model may be appropriate. There is also considerable heterogeneity in effect sizes. The overall incidence of pre-eclampsia varies from 0.015 in study 6 to 0.412 in study 9. The odds ratios of effect of diuretics vary from 0.229 in study 4, a study with high incidence of 0.308, to 2.971 in study 8, a study with low incidence of 0.038. The Cochran's Q-statistic value is $Q = 27.265$, and the total sample size $N = 6942$. Estimated values of τ^2 for standard REM, and of ρ assuming the BB model and using various estimating methods are provided in Table 4.2. The Der-Simonian-Laird estimate of the variance component in standard REM is $\tau^2_{DL} = 0.23$, and $\tau^2_{REML} = 0.30$. In beta-binomial model, five methods of estimation provide estimates of $\hat{\rho}$ varying from 0.008 for the moment estimator, to 0.019 for the Breslow-Day based estimator. Confidence interval for ρ is the shortest for REML, and the longest for the BD estimator. These values are directly interpretable as the estimated ICCs and their

Table 4.2: Values and confidence intervals for ρ, for log odds ratios and for odds ratios for diuretics on pre-eclampsia example; FEM is the fixed effect, REM is the random effects, and BB is the beta-binomial model. Heterogeneity parameter estimated is τ^2 in REM, and ρ in BB model. L and U are the lower and upper limits of the respective confidence intervals (CIs).

Model	Method	Hetero geneity	L	U	LOR	L	U	length of CI	OR	L	U
FEM		0.000			-0.398	-0.553	-0.223	0.530	0.672	0.564	0.800
REM	DL	0.230	0.072	2.202	-0.517	-0.916	-0.117	0.799	0.596	0.400	0.889
REM	REML	0.300	0.043	1.475	-0.518	-0.956	-0.080	0.876	0.596	0.384	0.923
BB	M&IV	0.008	0.002	0.095	-0.436	-0.792	-0.080	0.712	0.647	0.453	0.923
	M&MH				-0.427	-0.775	-0.080	0.695	0.652	0.461	0.923
BB	REML&IV	0.010	0.001	0.060	-0.447	-0.835	-0.059	0.776	0.640	0.434	0.942
	REML&MH				-0.431	-0.809	-0.053	0.756	0.650	0.445	0.949
BB	MP&IV	0.017	0.002	0.095	-0.469	-0.920	-0.018	0.902	0.626	0.399	0.982
	MP&MH				-0.459	-0.898	-0.020	0.879	0.632	0.407	0.981
BB	CMP&IV	0.018	0.003	0.094	-0.474	-0.942	-0.007	0.936	0.623	0.390	0.993
	CMP&MH				-0.472	-0.927	-0.016	0.911	0.624	0.396	0.984
BB	BD&IV	0.019	0.003	0.107	-0.475	-0.944	-0.006	0.938	0.622	0.389	0.994
	BD&MH				-0.463	-0.920	-0.021	0.899	0.630	0.399	0.980

confidence limits. To see the effect of these estimates of heterogeneity on the inference about the odds ratio, we compare the corresponding estimates for LOR and OR, and their confidence intervals, in the same table. The odds ratio is the highest (0.672) in the fixed effect model, and, not surprisingly, its confidence interval is the shortest. The OR is the lowest (0.596) in the standard REM, and various estimators based on the inverse-variances or the MH method provide intermediate values of OR, the one based on the MH and method of moments estimator ρ_M providing the highest value of OR, 0.652. For each estimator of ρ, the MH estimation of OR results in a somewhat higher value of OR than the inverse-variances based estimation, with a somewhat shorter confidence interval for OR. The sample sizes are reasonably large in all included studies, and based on the results of simulations reported in Section 4.6, we recommend to use the estimated ICC $\hat{\rho}_{BD} = 0.019$ and corresponding value of the pooled OR $\hat{\psi}_{IV} = 0.622$ with confidence interval $(0.389,\ 0.994)$.

4.6 Simulation study

In this Section we provide a simulation study to access the performance of point and interval estimators of overdispersion parameter ρ and the combined LOR θ in beta-binomial model of meta-analysis. We assess five point estimators of ρ in respect to their bias: the moment method (4.4.2), the Mandel-Paule inspired method ρ_{MP}, the corrected Mandel-Paule estimator based on the gamma approximation to Q distribution ρ_{CMP}, the REML method (4.4.5) and the BD-based method (4.4.13). We also assess four related confidence intervals for ρ (4.4.3), (4.4.9), (4.4.6) and (4.4.14) in respect to their coverage at the 95% confidence level. Additionally, we compare two estimation methods for obtaining point and interval estimators of the combined odds ratio or its log, the inverse-variance method $\hat{\theta}_w = \sum w_i(\rho)\hat{\theta}_i / \sum w_i(\rho)$ and the modified Mantel-Haenszel method (4.3.2). We combine five above-mentioned point estimators of ρ with these two methods of obtaining combined effect $\hat{\theta}$, resulting in ten possible combinations, and we assess these estimators of $\hat{\theta}$ for bias and for coverage.

Typically, small values of ρ, below 0.1, appear in bio-medical applications, Gulliford et al. (2005), Littenberg and MacLean (2006). Overdispersion is mostly due to clustering by the same healthcare provider. However our range of values of ρ up to 0.3 is comparable to τ^2 values of up to 5 in the standard REM for our choice of values of probabilities and LORs provided below. This correspondence between heterogeneity in the additive REM and beta-binomial model is given by equation (4.2.3).

4.6.1 Simulation design

Sizes of the control and treatment groups were taken equal $n_{i1} = n_{i2} = n_i$ and were generated from a normal distribution with mean n and variance $n/4$ rounded to the nearest integer and left truncated at 5. For a given probability p_{i2}, the number of cases in the control group X_{i2} was simulated from a beta-binomial (n_{i2}, p_{i2}, ρ) distribution using the R package *emdbook* (Bolker, 2011). The number of cases in the treatment group X_{i1} was generated from a beta-binomial (n_{i1}, p_{i1}, ρ) distribution with $p_{i1} = p_{i2}\exp(\theta)/(1 - p_{i2} + p_{i2}\exp(\theta))$ for a given

LOR value of θ. When $\rho = 0$, the numbers of events for treatment and control arm X_{ij} were generated from binomial distributions with sample size n_{ij} and probabilities p_{ij}, preserving the above relationship between the probabilities in the treatment and control arms.

The following configurations of parameters were included in the simulations. The number of studies $K = (10, 20, 30, 50, 80)$; average sample sizes in each arm are $n = (10, 20, 40, 80, 160, 250, 640, 1000)$; overdispersion parameter ρ varies between 0 and 0.1 (small to moderate heterogeneity) with steps 0.01, and between 0.1 and 0.3 in steps 0.05 (moderate to large heterogeneity). The values of LOR θ vary from 0 to 3 in steps of 1. The probability in the control group p_{i2} takes values 0.1, 0.2, 0.4. A total of 10000 simulations were produced for each combination.

4.6.2 Simulation results

Figures 4.1 and 4.2 show the bias and coverage of ρ estimated by the five methods mentioned above for different combinations of K and n for the case of $p_{i2} \equiv 0.1$ and $\theta = 0$ and varying values of $0 \leq \rho \leq 0.3$. The bias and coverage of true log odds ratio θ estimated by the inverse-variance (θ_{IV}) for values of $\theta = 0$, 1, 2, are shown in Figures 4.3 - 4.6, respectively. Similar figures for bias and coverage of θ by the modified Mantel-Haenzsel method (ψ_{MH}) are given in Appendix, B.25 – B.28.

Bias and coverage in estimation of intra-cluster correlation ρ

Bias of estimated ICC ρ is negative and it clearly increases in ρ, Figure 4.1 for $p_{i2} = 0.1$, B.1 and B.2 in Appendix for $p_{i2} = 0.2$ and 0.4. For small number of studies K combined with small sample sizes ($n \leq 50$), $\hat{\rho}_{CMP}$ estimation appears to be the best option. However, for larger sample sizes ($n \geq 100$), the BD-based estimator $\hat{\rho}_{BD}$ is the clear winner. Still, its negative bias increases almost linearly with ρ and is acceptable only for $\rho < 0.1$. Coming to coverage of ρ (Figure 4.2 and B.3, B.4 in Appendix), once more, the Breslow-Day based estimation appears to be the safest option apart from the case of very small sample sizes $n \leq 50$, where the gamma-based approximation appears to provide better coverage for $K \geq 10$. Both bias and coverage improve when the probabilities in both arms are farther from the edges. B.5 and

B.6 in the Appendix provide the bias and coverage in estimation of ρ for different values of θ and increasing sample size n, keeping $\rho = 0.1$ fixed. Similar plots of bias and coverage of ρ for $p_{i2} \equiv 0.2$ and $p_{i2} \equiv 0.4$ are given in B.7 - B.10 in Appendix. Breslow-Day based estimator ρ_{BD} remains the best estimator of ρ for all scenarios for $n \geq 100$, though it acquires a small positive bias when $p_{i2} = 0.4$ and $\theta = 3$, the case corresponding to $p_{i1} = 0.93$.

Bias in estimation of odds-ratio ψ

Bias of estimated odds ratio $\hat{\psi}$ was practically the same regardless of a method used for estimation of intra-class correlation ρ. This may be due to similarity of sample sizes across studies in our simulations, as the inflation terms $(1 + (n_i - 1)\rho)$ in the normalized individual weights "almost" cancel. Without loss of generality, we plotted the results for bias of $\hat{\psi}$ obtained when using the moment estimator $\hat{\rho}_M$ in Figure 4.3 for values of log-odds $\theta = 0, 1$ and 2. There is no bias when $\theta = 0$, i.e. when the probabilities of an event in two arms are the same, but the bias clearly increases with increasing values of θ, and/or ρ. Still the bias for the inverse variance weights is within 10% for $\rho \leq 0.1$ or $\theta \leq 2$, which would cover the major part of values of these parameters in practice, as $\theta = 2$ corresponds to the odds ratio of 7.39, and the values of ICC ρ are usually small. An explanation of this bias is provided in Section 4.6.2. Unfortunately, the bias is substantially higher for the modified Mantel-Haenszel method, especially for small number of studies K and large values of ρ and n, and the coverage deteriorates accordingly, see B.25 – B.34 in Appendix), and therefore we do not pursue this estimator further.

Bias of sample log-odds ratio under beta-binomial model

The bias of a number of popular effect measures used for binary data under random effects models was discussed in Chapter 3. For log-odds, it is well known that the sample log odds ratio

$$\hat{\theta} = \log(\frac{\hat{p}_1}{1 - \hat{p}_1}) - \log(\frac{\hat{p}_2}{1 - \hat{p}_2}),$$

where the probabilities of events p_1 and p_2 in treatment and control groups are estimated by sample frequencies $\hat{p}_i = X_i/n_i$, $i = 1, 2$, has a bias of order $1/n$ under the fixed effects model $\rho = 0$. The standard bias correction due to Gart et al. (1985) adds $1/2$ to X_i and to $n_i - X_i$, i.e., uses $\tilde{p}_i = (X_i + 1/2)/(n_i + 1)$ when estimating the log-odds to eliminate the $1/n$ bias term at the null model $\rho = 0$.

Expanding the log odds by Taylor series for a general ρ, and taking expectations, see Chapter 3 for details of derivation,

$$\mathrm{E}_\rho(\log(\frac{\hat{p}}{1-\hat{p}})) = \log(\frac{p}{1-p}) - \frac{(1-2p)(1+(n-1)\rho)}{2np(1-p)} + \cdots,$$

where, importantly, the second term includes a bias of order $O(1)$ when $\rho \neq 0$. Therefore, the bias of the sample log odds ratio $\hat{\theta}$ is

$$bias(\hat{\theta}) = -\frac{(1-2p_1)(1+(n_1-1)\rho)}{2n_1p_1(1-p_1)} + \frac{(1-2p_2)(1+(n_2-1)\rho)}{2n_2p_2(1-p_2)}.$$

When log odds ratio $\theta = 0$, i.e. when the probabilities in both arms are equal, the biases for sample log-odds in each arm cancel out. Thus, the estimate $\hat{\theta}$ is unbiased to order $1/n$. However, when $\theta \neq 0$ and the probabilities in both arms are not equal, the sample odds ratio is biased to order $O(1)$, and this bias is not ameliorated by the continuity correction. For example, when $p_1 = 0.1$ and $p_2 = 0.4$, i.e. $\theta = -1.791$, the main bias term is $(-4.444 + 0.417)\rho$, increasing linearly with the intra-class correlation ρ. B.35 in Appendix illustrates quality of this linear approximation to bias. It works well for small values of ρ, but the bias increases and higher order terms become of more importance for larger values of ρ.

In meta-analysis with fixed weights, it would be possible to correct the resulting bias of the overall effect measure for small values of ρ, but the use of inverse variance weights also affects the bias and makes such a correction much more difficult. Luckily, the resulting bias is not very large, as we have seen in Section 4.6.2. We believe that the origin of the higher bias in the corrected Mantel-Haenszel method is the combination of the transformation bias with the bias in estimation of ρ, and the consequences of these biases are graver.

Coverage of odds-ratio ψ

The method used for estimation of intra-class correlation ρ is of utmost importance for correct estimation of variance, and therefore the coverage of the odds-ratio ψ, presented for $p_{i2} = 0.1$ in Figures 4.4–4.6 for $\theta = 0$, 1 and 2, respectively. Overall, exactly like in the case of bias, the modified Mandel-Paule estimator $\hat{\rho}_{CMP}$ results in the best coverage for small sample sizes up to 50, and the ρ_{BD} provides superior coverage for $n \geq 100$. All other estimators of ρ result in inferior coverage, especially for large values of ρ. However, there are important differences in coverage when using the best estimators of ρ due to differences in true value of the odds ratio. For the small number of studies $K = 5$, the coverage is too low for all values of θ, but it drifts from about 90% to about 87% even when the best estimator of ρ is used. Starting from $K = 10$, the coverage is good for $\theta = 0$, but becomes lower than nominal when θ increases. It is still reasonable, at about 93%, for $\theta = 1$, but reaches 90% or even somewhat lower for $\hat{\rho}_{BD}$ used with large sample sizes $n = 1000$. This is due to the increasing biases in the estimation of ψ combined with the "improved" precision for larger sample sizes. Similar plots of coverage for $p_{i2} = 0.2$ and 0.4 when $\theta = 0$ are given in Appendix (B.17, B.18). B.19 - B.24 in Appendix present the bias and coverage when estimating θ by $\hat{\theta}_{IV}$ for different values of p_{i2} and increasing sample size n, keeping the value of θ fixed. These figures clearly show the biases and reduced coverage of OR due to transformation bias discussed in the previous Section. Coverage achieved when $\hat{\rho}_{BD}$ is used in the weights is superior for moderate to large sample sizes.

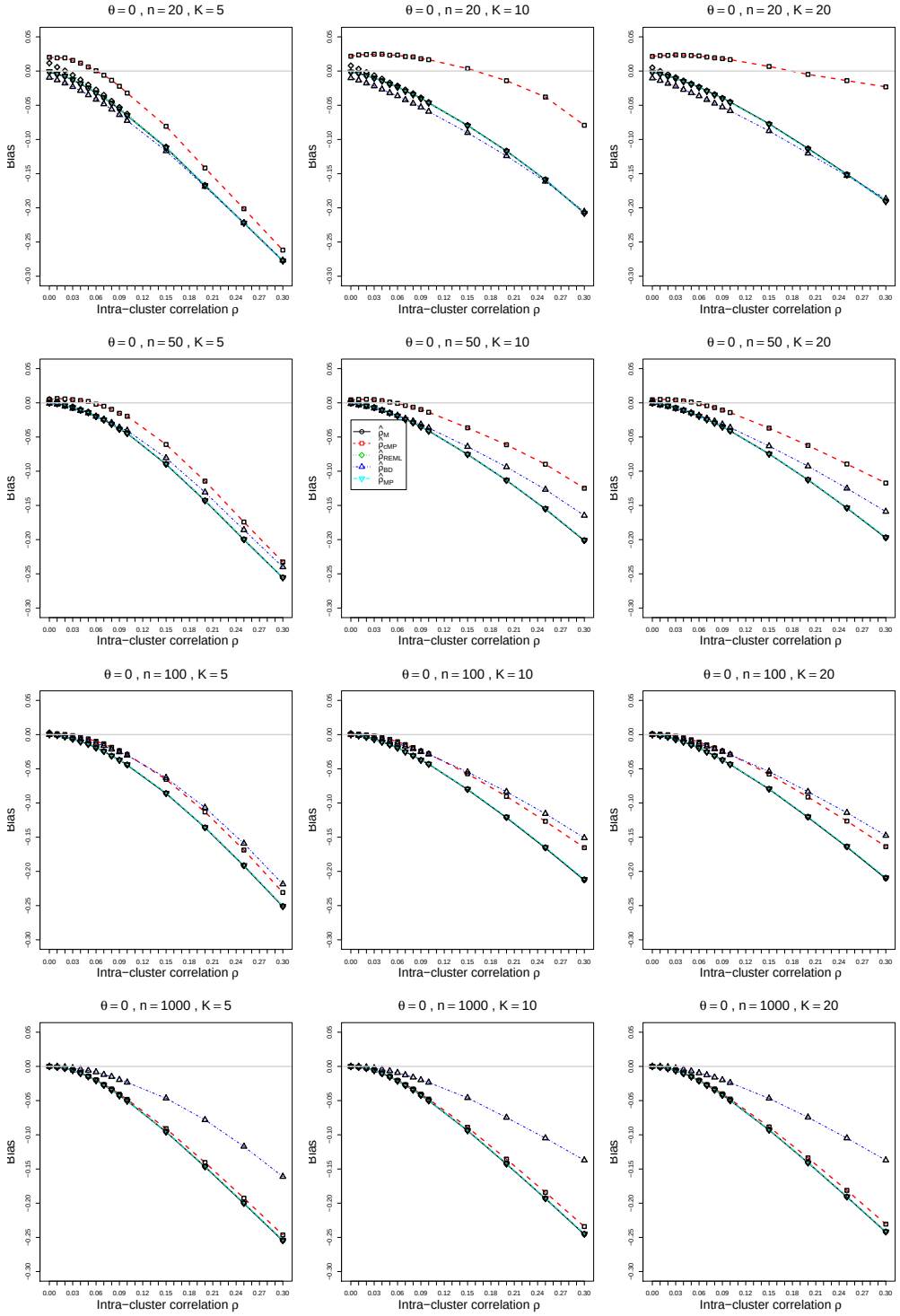


Figure 4.1: Bias of the estimated from K studies intra-cluster correlation ρ in beta-binomial model for $p_{i2} = 0.1$, $\theta = 0$ and $0 \leq \rho \leq 0.3$. Estimation methods: circles – Moment estimator $\hat{\rho}_M$, squares – Corrected Mandel-Paule estimator $\hat{\rho}_{CMP}$), diamonds – $\hat{\rho}_{REML}$), triangles-Breslow-Day estimator based $\hat{\rho}_{BD}$), reverse-triangles – Mandel-Paule estimator $\hat{\rho}_{MP}$). Light grey line at 0.

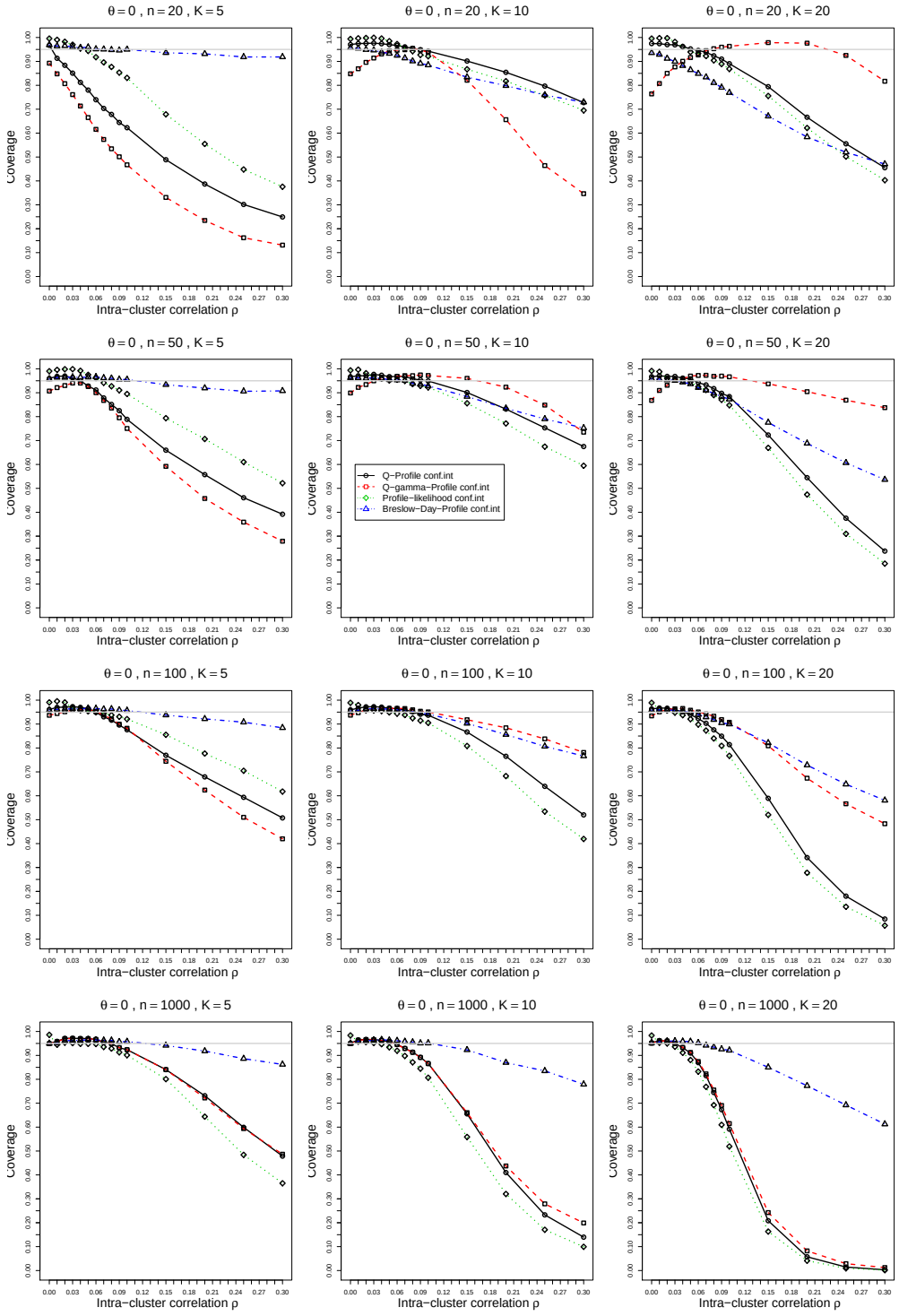


Figure 4.2: Coverage at the nominal confidence level of 0.95 of the intra-cluster correlation ρ estimated from K studies in beta-binomial model for $p_{i2} = 0.1$, $\theta = 0$ and $0 \leq \rho \leq 0.3$. Interval estimation methods: circles – Q-profile confidence interval for ρ based on χ^2 distribution, squares – Q-profile confidence interval for ρ based on $\Gamma_{r(\rho),\lambda(\rho)}$ distribution), diamonds – Profile likelihood confidence intervals, triangles – Breslow-Day-Profile confidence intervals. Light grey line at 0.95.

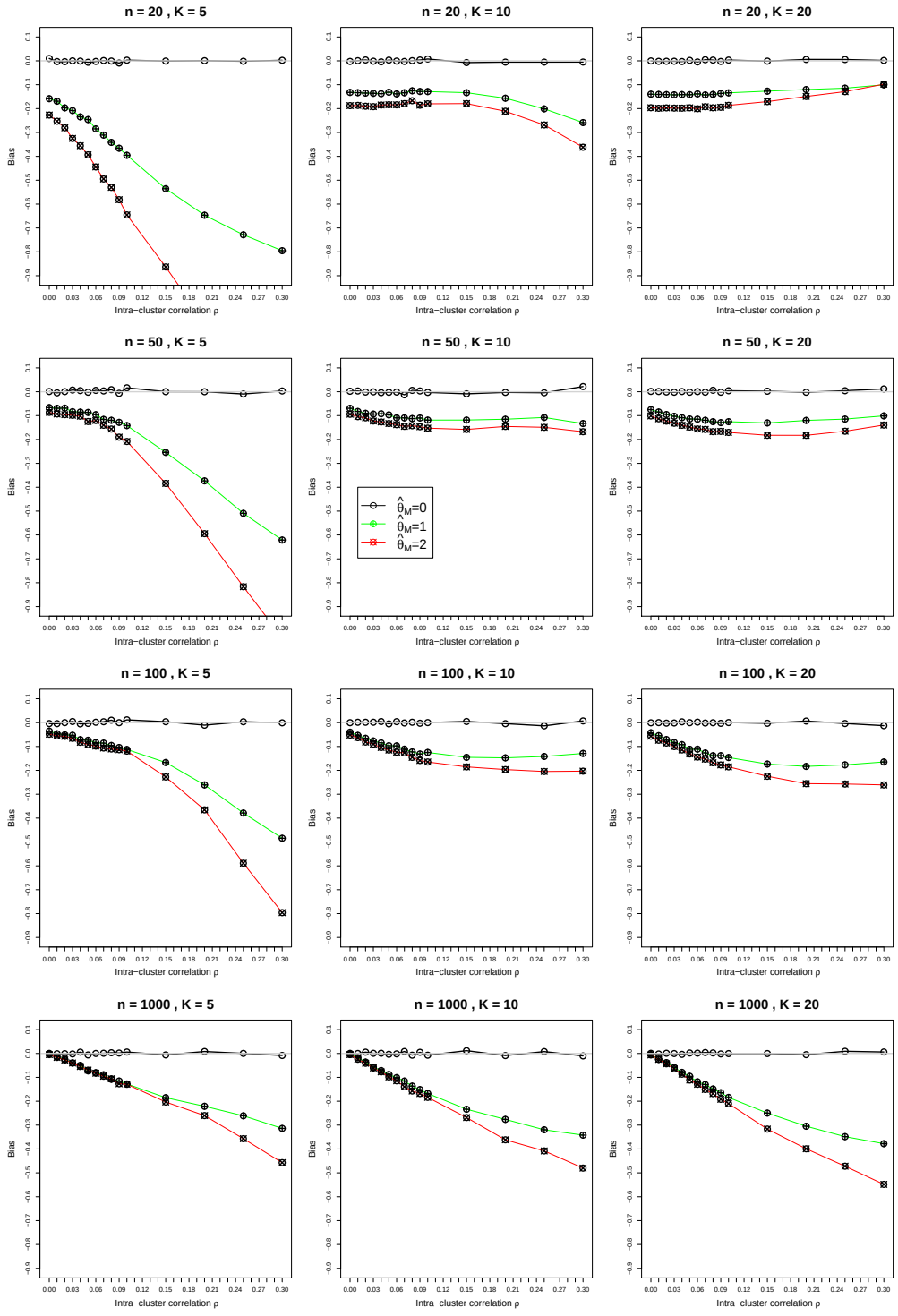


Figure 4.3: Bias of overall odds ratio ψ_{IV} obtained from K studies by the inverse-variance method with the moment estimator $\hat{\rho}_M$ in the weights, for $p_{i2} = 0.1$, and $0 \leq \rho \leq 0.3$. The biases are given for $\theta = 0$ (circles), $\theta = 1$ (circle plus), and $\theta = 2$ (circle cross). Light grey line at 0.

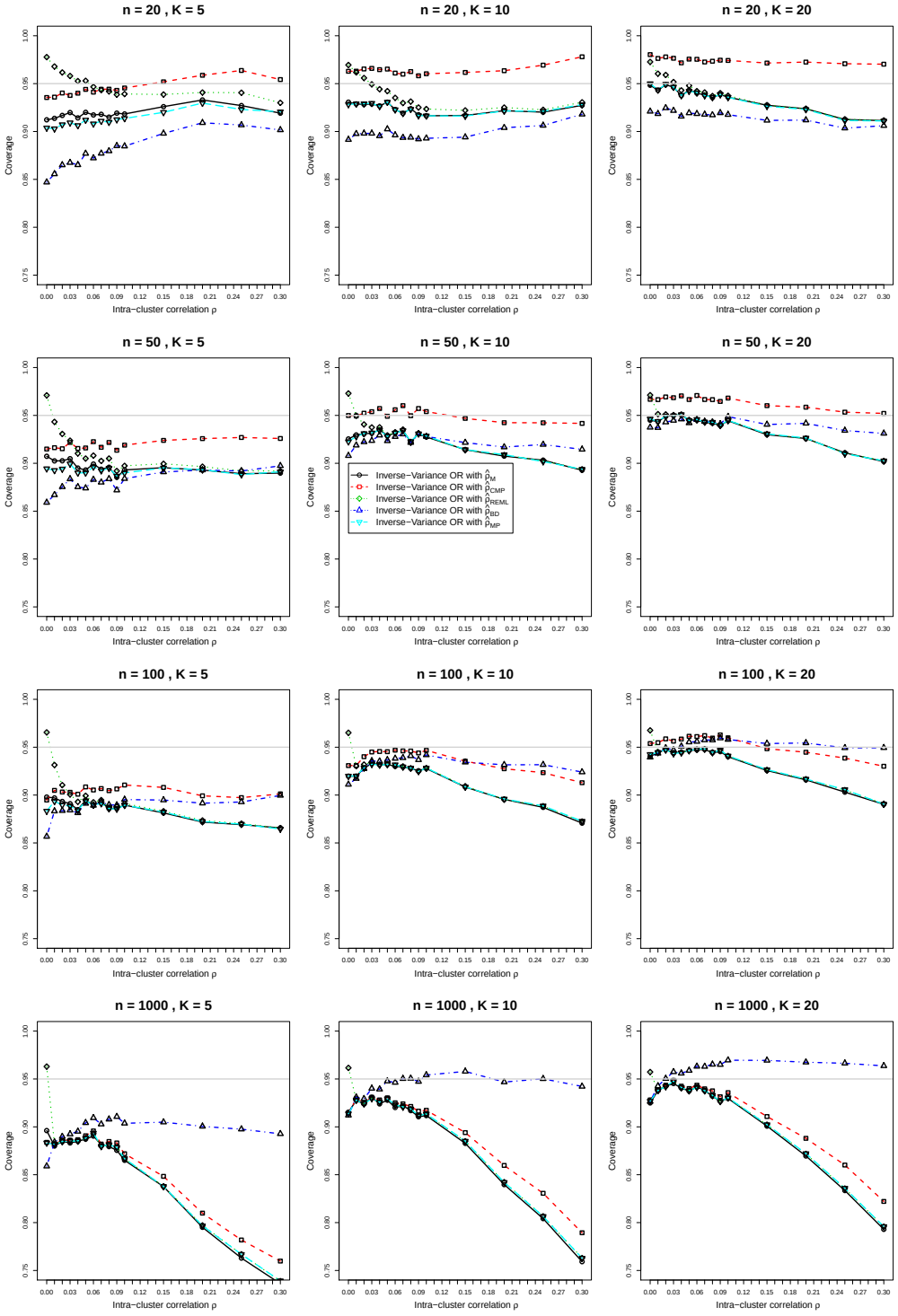


Figure 4.4: Coverage at the nominal confidence level of 0.95 of the overall odds ratio ψ obtained from K studies by the inverse-variance method, for $p_{i2} = 0.1$, $\theta = 0$ and $0 \leq \rho \leq 0.3$. The inverse-variance weights use the following estimators of ρ: circles – $\hat{\rho}_M$, squares – Corrected Mandel-Paule estimator $\hat{\rho}_{CMP}$, diamonds – restricted maximum likelihood estimator $\hat{\rho}_{REML}$, triangles – Breslow-Day estimator $\hat{\rho}_{BD}$ and reverse-triangles (Mandel-Paule estimator $\hat{\rho}_{MP}$). Light grey line at 0.95.

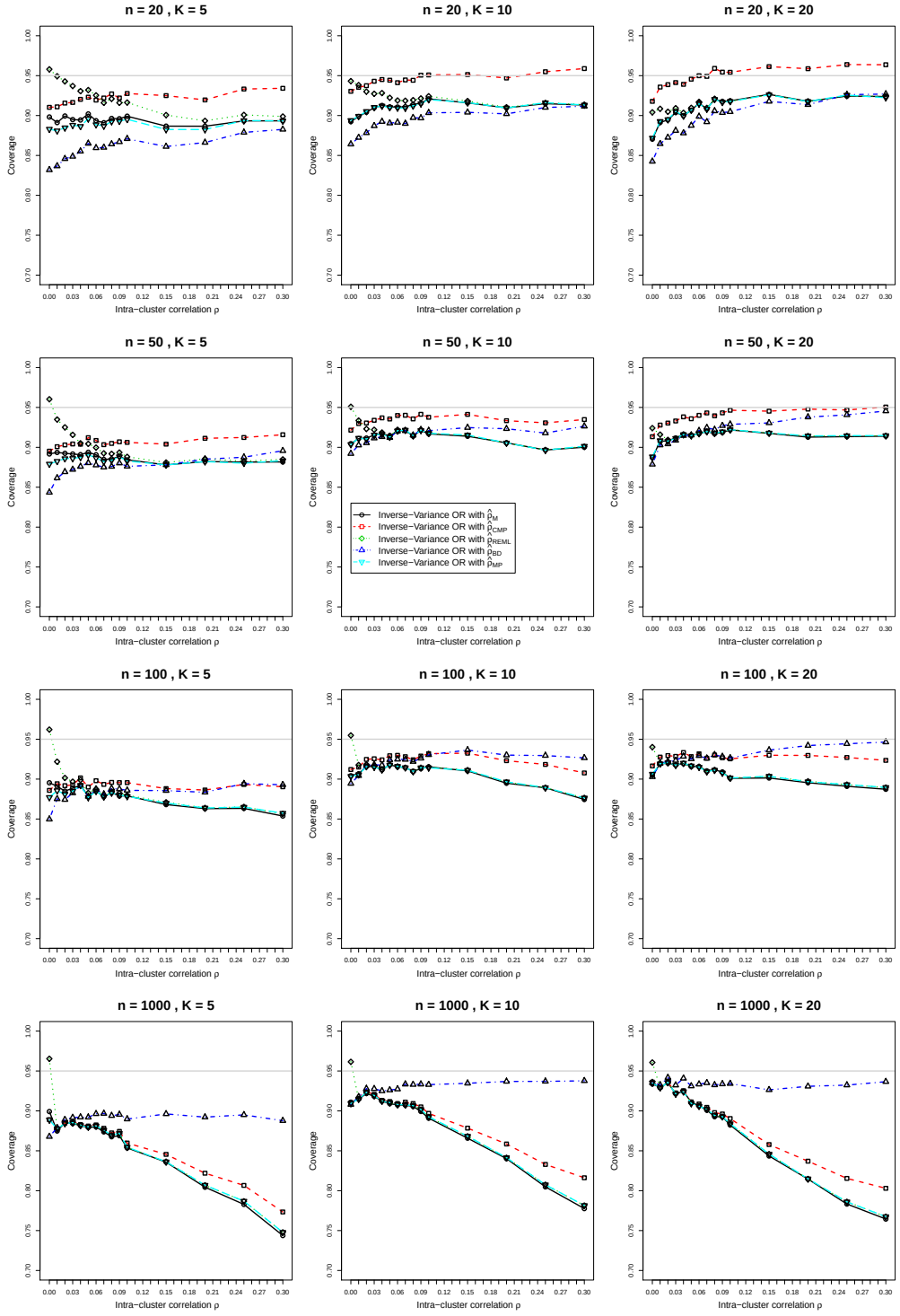


Figure 4.5: Coverage at the nominal confidence level of 0.95 of the overall odds ratio ψ obtained from K studies by the inverse-variance method, for $p_{i2} = 0.1$, $\theta = 1$ and $0 \leq \rho \leq 0.3$. The inverse-variance weights use the following estimators of ρ: circles – $\hat{\rho}_M$, squares – Corrected Mandel-Paule estimator $\hat{\rho}_{CMP}$, diamonds – restricted maximum likelihood estimator $\hat{\rho}_{REML}$, triangles – Breslow-Day estimator $\hat{\rho}_{BD}$ and reverse-triangles (Mandel-Paule estimator $\hat{\rho}_{MP}$). Light grey line at 0.95.

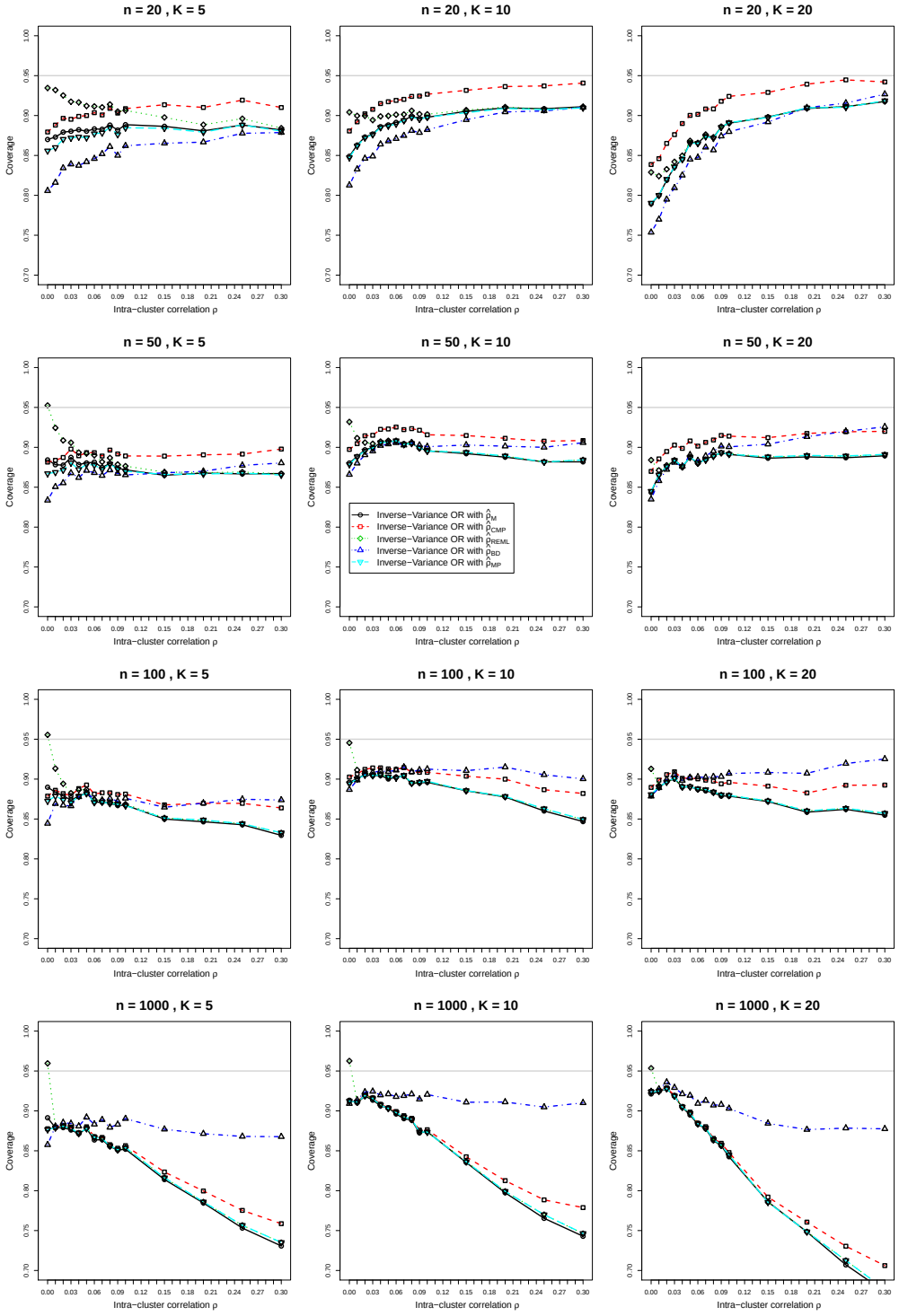


Figure 4.6: Coverage at the nominal confidence level of 0.95 of the overall odds ratio ψ obtained from K studies by the inverse-variance method, for $p_{i2} = 0.1$, $\theta = 2$ and $0 \leq \rho \leq 0.3$. The inverse-variance weights use the following estimators of ρ: circles – $\hat{\rho}_M$, squares – Corrected Mandel-Paule estimator $\hat{\rho}_{CMP}$, diamonds – restricted maximum likelihood estimator $\hat{\rho}_{REML}$, triangles – Breslow-Day estimator $\hat{\rho}_{BD}$ and reverse-triangles (Mandel-Paule estimator $\hat{\rho}_{MP}$). Light grey line at 0.95.

4.7 Summary

In this chapter we developed theory of meta-analysis of odds ratios based on the beta-binomial model. This model is a natural alternative to the standard random effects model based on normality of random effects. Of course, other combinations of distributions are possible for meta-analysis of binomially distributed data. Stijnen et al. (2010) suggest using exact hypergeometric likelihood for individual studies combined with Normally distributed random effect for log-odds. Alanko and Duffy (1996) discuss a family of compounded binomial distributions obtained by using mixing distributions from the generalized inverse gaussian family of distributions, but these distributions had not been used so far in meta-analysis.

We have concentrated on the case of two independent beta-binomial distributions in two arms of each study. We have proposed two new methods of estimation of the intra-cluster correlation ρ in meta-analysis based on this model. Both our methods work considerably better than other, more traditional methods suggested by Kulinskaya and Olkin (2014), and they complement each other by being applicable to meta-analyses of smaller or larger studies. This model is similar to bivariate binomial-normal REM for log odds ratios discussed by (Stijnen et al., 2010, p.3056). The latter model can also incorporate a correlation between the two arms of the same study. However, a similar extension of the beta-binomial model is not straightforward.

For estimation of intra-cluster correlation, another alternative is an analysis of variance estimator $\hat{\rho}_{AOV}$ discussed in Chapter 3. However, $\hat{\rho}_{AOV}$ is also biased and analysis of variance method estimates the intra-cluster correlation in a each arm separately.

A version of a bivariate beta-binomial distribution was proposed by Bibby and Væth (2011), but this distribution has a strictly positive lower bound for correlation between the marginals, so it does not include the case of independent beta-binomial distributions. Moreover, Bibby and Væth (2011) show that "independence cannot be obtained as a limit in the parameters without sacrificing the overdispersion". They also discuss other, previously suggested, versions of a bivariate beta-binomial distribution, and possible extensions aimed at resolving this problem, but none are satisfactory. However, a different version that allows a range of correlation

values, including zero correlation, was applied to meta-analysis in Chu et al. (2012). A new bivariate beta distribution was recently proposed by Olkin and Trikalinos (2015), but so far it has not been used for mixing binomial distributions.

We also briefly considered a model with beta-binomial distribution in treatment arm only. This model is analogous to a version of unconditional random effects logistic regression by Turner et al. (2000). In this model the study specific log odds of the control groups constitute K additional parameters, and this model is not appropriate when $K \to \infty$, Stijnen et al. (2010).

We also proposed a variation of Mantel-Haenszel method for meta-analysis of odds ratios. Unfortunately, in simulations this method was found to be very biased, especially for odds ratios greater than 1. Elimination of this bias will be pursued elsewhere. The traditional inverse variance approach with estimated by one of our methods ICC ρ results in reasonable, though somewhat low coverage for a realistic range of values of odds ratios and intra-class correlations.

Chapter 5

Meta-Analysis via Generalized Linear Mixed-Effects Models

5.1 Introduction

The standard additive random effects model (REM) and multiplicative overdispersed model (ODM) introduced by Kulinskaya and Olkin (2014) are special cases of hierarchical generalized linear mixed effects model. When the outcome of interest is a transformation of some statistic such as a logarithmic transformation of odds, the standard additive random effects model assumes that within-study variability is accounted for through an approximate normal within-study likelihood, i.e $\hat{\theta}_i \sim N(\theta_i, \sigma^2)$. Combining this assumption with a normal approximation for true effects between studies, $\theta_i \sim N(\theta, \tau^2)$, the model results in $\hat{\theta}_i \sim N(\theta, \sigma^2 + \tau^2)$ (see Chapter 2 for details). In our context, the standard REM is modelling the estimated logarithm of ORs $\hat{\theta}_i$. In Chapter 4, we have introduced a multiplicative random effects model (ODM) for ORs. ODM for ORs is modelling the binomial numbers of events X_{1i} and X_{2i} rather than the logarithmic transformation of odds. In ODM, the variability is modelled through a pair of independent beta-binomial distributions.

Both models, the standard REM and the multiplicative ODM, have some potential problems. The standard REM makes a strong assumption about known within-study variances and does not account for the correlation between $\hat{\sigma}_i^2$ and $\hat{\theta}_i$. Also, in REM, the continuity corrections have to be applied in case of sparse data. Both models suffers from the transformation bias

of order $1/n$ and the bias in estimation of random effect variance component. In Chapter 4, we have shown that the standard methods in meta-analysis fail to estimate the random effect component correctly. We also introduced two new methods for estimation of the random effect parameter in ODM. These methods appear less biased than the standard methods. However, we do not have a fully unbiased method for estimation of intra-cluster correlation.

In this chapter, we concentrate on an attractive alternative for meta-analysis of binary data via general class of generalized linear mixed-effects models. Generalized linear mixed-effects models are believed to overcome the problems of standard random effects model in meta-analysis (Stijnen et al., 2010). Particularly, our interest lies in often recommended non-central-hypergeometric normal model for meta-analysis introduced by Van Houwelingen et al. (1993), Liu and Pierce (1993), Sidik and Jonkman (2008) and Stijnen et al. (2010). Non-central-hypergeometric-normal model (NCHGN) is a mix of non-central-hypergeometric and normal distributions applied to the number of events in the treatment arm. Without assumption of normality, the non-central-hypergeometric distribution can be used for inference in a fixed effect model. NCHGN model is a conditional generalized linear mixed-effects model with exact likelihood. The non-central-hypergeometric distribution is the exact distribution for the number of events conditional on marginal totals. In the mix of non-central hypergemometric and normal distributions, the true unobserved conditional odds ratio is obtained by the transformation of the number of successes under non-central hypergeometric distribution. Due to the use of logit transformation, the model might suffer from transformation bias. Previously, no simulations has been performed to investigate the NCHGN model and its approximation by the binomial-normal distribution. This is due to complexity of the models and enormous time requirement for each combination of simulations. We run the simulation study for NCHGN model with two scenarios of generating the data. In the first scenario, the data is simulated with a pair of binomial distributions with a normally distributed log odds across the studies. The second scenario generates the data from a pair of beta-binomial distributions similar to simulations in Chapter 4. We show that the maximum likelihood estimates of overall effects

and between-study variance are biased and the biases are of order $1/n$. Chapter 5 discusses the biases of order $O(1/n)$ similar to those as Chapter 3. However, the models in Chapter 5 differ, they are the models for transformed expected values as opposed for transformed summary statistics in Chapter 3. This Chapter mainly concentrates on the biases in a model based on the non-central hypergeometric distribution mixed by a normal distribution. This Chapter represents the novel work of this thesis.

5.2 Generalized linear mixed effects model

Generalized linear mixed effects model (GLMM) is an extension of a generalized linear model. GLMM includes both fixed and random effects (hence mixed effect model). The inference in GLMM is based on maximum likelihood theory. Usually, the likelihood is obtained as a mixture of two distributions for fixed and random effects. The mixture of distributions might include discreet and continuous distributions.

In meta-analysis of a binary data, the mixture distributions might include: binomial and normal distributions or non-central hypergeometric and normal. For incidence rates, the example of generalized linear mixed effects model is Poisson-normal model. From the listed models, we concentrate our attention on non-central hypergeometric and normal model (NCHGN).

For general case, let the univariate observation in the i_{th} study be y_i, and the vectors of covariates are denoted by x_i and z_i of dimensions p and q for fixed and random effects, for $i = 1, \ldots, K$, respectively. Given a q-dimensional vector b, the generalized linear mixed effects model has general form

$$\eta_i^b(b) = x_i^t\beta + z_i^t b. \tag{5.2.1}$$

The responses y_i are assumed to be independent with mean $\mathrm{E}(y_i|b_i) = \mu_i^b(b_i)$ and variance $\mathrm{Var}(y_i|b_i) = \phi a_i \upsilon(\mu_i^b(b_i))$, where a_i is a known constant and $\upsilon(\cdot)$ is a variance function (Breslow and Clayton, 1993). The conditional mean and variance have a mean-variance relationship and both of them depend on a random effect b_i. Similarly to generalized linear model, the conditional mean is associated with linear predictor through a link function $g(\mu_i(b_i)) = \eta_i(b_i)$.

Inverting the link function, $h = g^{-1}$, and denoting the design matrices with row x_i^T and z_i^T by X and Z, the conditional mean satisfies

$$\mathrm{E}(y|b) = h(X\beta + Zb)$$

where $y = (y_1, \ldots, y_K)$. The random effect b has mean 0 and follows a distribution which is commonly assumed to be multivariate normal with variance-covariance matrix $D = D(\zeta)$. ζ is an unknown vector of variance components. Breslow and Clayton (1993) consider models with binomial, Poisson and hypergeometric specifications for the conditional distribution of y_i with fixed dispersion parameter ϕ at unity in the conditional variance. The parameter ϕ may also be estimated together with other parameters ζ in $D = D(\zeta)$.

In generalized linear mixed-effects models, the parameters are estimated by maximum likelihood theory. However, due to non-linearity of the model and the presence of random effects, the marginal distribution for maximum likelihood approach includes a cumbersome and intractable integration with respect to unobservable random effects. Usually, the integration does not have a closed form, and therefore no analytic solution is possible. Numerical methods such as Gaussian quadrature, adaptive Gaussian quadrature or Gauss-Hermite quadrature have to be applied for evaluation of the integral, approximation of the log-likelihood function, score equations and information matrix (Breslow and Clayton, 1993).

5.3 Likelihood based inference

The inference in generalized linear mixed effects model is based on maximum likelihood methods. By specifying a distribution for the data, log-likelihood function is specified. The maximum likelihood inference is carried out by maximizing the log-likelihood function with respect to unknown parameters.

The likelihood function for the generalized linear mixed effects model with exact and approximate likelihood in meta-analysis is

$$h(X_{1i}|\theta,\tau^2)=\int_{-\infty}^{\infty} g(X_{1i}|\theta_i)f(\theta_i|\theta,\tau^2)d\theta_i, \tag{5.3.1}$$

where $f(\theta_i|\theta,\tau^2)$ is the density of a normal distribution to model the between study variation, and $g(X_{1i}|\theta_i)$ is a conditional or an unconditional distribution of binomial outcomes. For NCHGN, the differentiable log-likelihood function is

$$l=l(X_{1i},\ldots,X_{1K}|\theta,\tau^2)=\log(\prod_{i=1}^{K}h(X_{1i}|\theta,\tau^2))=\sum_{i=1}^{K}\log(h(X_{1i}|\theta,\tau^2)),$$

where the distribution $h(X_{1i}|\theta,\tau^2)$ is the distribution of number of events in treatment arm. The maximum likelihood estimators for θ and τ^2 are the solutions of the score equations $U(\theta)=0$ and $U(\tau^2)=0$, where

$$U(\theta)=\frac{dl}{d\theta} \quad \text{and} \quad U(\tau^2)=\frac{dl}{d\tau^2},$$

provided that the observed information matrix $I(\vartheta)=-\nabla_{\vartheta}\nabla_{\vartheta}^{T}l(\vartheta)$ is positive definite when evaluated at $\hat{\vartheta}$. In maximum likelihood estimation, the standard errors for the parameters θ and τ^2 are obtained from observed information matrix $I(\theta)$, i.e.,

$$\widehat{\mathrm{Var}}(\hat{\vartheta})=[I(\vartheta)]^{-1}.$$

Within-study distributions $g(X_{1i}|\theta)$ such as non-central-hypergeometric or binomial have dependence on sample sizes. Hence, the marginal distribution after integration of unobservable random effects also have dependence on sample sizes. Due to these reasons, the maximum likelihood estimators from (5.3.1) might have a bias of order $O(1/n)$.

5.4 Generalized linear mixed effects model for meta-analysis

In generalized linear mixed-effects model with fixed study effects, the term fixed represents a trial membership. Turner et al. (2000) introduced an unconditional generalized linear mixed-effects model with fixed and random study effects as a multilevel model for meta-analysis in

frequentist setting. The difference between standard additive random effects model and an unconditional generalized linear mixed-effects model is that standard random effects model directly models a measure that reflects the contrast between the two groups (e.g., log odds ratio). The conditional logistic (hypergeometric) model is another approach where we condition out the study effects and deal with the OR directly. Unconditional generalized linear mixed-effects model is the random effects logistic regression model with expected log-odds as an outcome. The parameters in these models can be estimated by maximum likelihood or restricted maximum likelihood methods using the iterative generalized least squares.

In this chapter, we also study the use of the NCHGN model to analyse binary data generated from a mixture of binomial and normal distributions as in standard REM and from a pair of beta-binomial distributions as in ODM. In the former model, the numbers of events in treatment and control arms have a conditionally binomial distribution. The number of events in the treatment arm can be conditioned on total number of events in both arms resulting in a non-central hypergeometric distribution. For a pair of beta-binomial distributions, the resulting conditioned distribution of events in treatment arm no longer follows non-central hypergeometric distribution.

5.4.1 An unconditional generalized linear mixed-effects model with fixed study effects

An unconditional generalized linear mixed-effects model with fixed study effects is a special case of mixed-effects logistic regression model. The model allows to fit the logistic regression model with fixed trial effects and accounting for the heterogeneity across K studies in log odds scale. The model is

$$y_{ij}|\pi_{ij} \sim Binomial(n_{ij}, \pi_{ij}) \quad j = 1, 2; \quad i = 1, \ldots, K,$$

$$\log(\frac{\pi_{ij}}{1-\pi_{ij}}) = \phi_j + (\theta + \nu_j)x_{ij}, \tag{5.4.1}$$

where θ is the overall effect (log odds ratio) and $\nu_i \sim N(0, \tau^2)$. Random effect ν_i is the deviation of the i_{th} study true treatment effect (log-odds ratio) from the average θ. ϕ_i are fixed study effects (log-odds in the control arm). τ^2 is the between-study variance. The indicator $x_{ij} = 0/1$ represents the choice between control and treatment groups. For the control group, $x_{2i} = 0$, the full model (5.4.1) reduces to

$$\log(\frac{\pi_{2i}}{1-\pi_{2i}}) = \phi_i,$$

and for the treatment group, $x_{1i} = 1$, the same model (5.4.1) is

$$\log(\frac{\pi_{1i}}{1-\pi_{1i}}) = \phi_i + \theta + \nu_i$$

where $j = 1, \ldots, K$. Combining both models for control and treatment group, the logistic regression model has a form

$$\log(\frac{\pi_{1i}}{1-\pi_{1i}}) = \log(\frac{\pi_{2i}}{1-\pi_{2i}}) + \theta + \nu_i$$

with $\log(\frac{\pi_{2i}}{1-\pi_{2i}})$ as a fixed study effect parameter that has to be estimated. $\phi_i = \log(\frac{\pi_{2i}}{1-\pi_{2i}})$ can be treated as an intercept. We have

$$\log(\frac{\pi_{1i}}{1-\pi_{1i}}) = \phi_i + \theta + \nu_i.$$

ϕ_i, θ and τ^2 are unknown parameters that have to be estimated. These parameters are estimated iteratively using either marginal quasi-likelihood, penalized quasi-likelihood or first and second order Taylor expansion approximation. In order to remove the bias of between-study variance estimates from penalized quasi-likelihood methods, two step bootstrap procedure can be used (Turner et al., 2000).

5.4.2 An unconditional generalized linear mixed-effects model with random study effects

An unconditional generalized linear mixed-effects model with random study effects is a mixed-effects logistic regression model with random study effects, meaning that random effects corresponding to the study factor are added to the model. The random effects logistic regression

model in log odds scale with random study effects is

$$y_{ij} \sim Binomial(n_{ij}, \pi_{ij}); \quad j = 1, 2, \quad i = 1, \ldots, K,$$

$$\log(\frac{\pi_{ij}}{1-\pi_{ij}}) = \alpha + u_j + (\theta + \nu_j)x_{ij},$$

where θ is the overall effect - log odds ratio and $\nu_i \sim N(0, \tau^2)$, $u_i \sim N(0, \sigma^2)$ and $\text{Cov}(u_i, \nu_i) = \omega\sigma\tau$. In contrast to logistic regression with fixed study effects, the subset of general model for the control group is

$$\log(\frac{\pi_{i2}}{1-\pi_{i2}}) = \alpha + u_i,$$

with random effect $u_i \sim N(0, \sigma^2)$ in control group. In case of treatment group the logistic regression model is

$$\log(\frac{\pi_{i1}}{1-\pi_{i1}}) = \alpha + u_i + \theta + \nu_i,$$

with additional random effect $\nu_i \sim N(0, \tau^2)$ of each study on treatment effect. The heterogeneity between log odds in control group is represented by σ^2 and in treatment group by $\sigma^2 + \tau^2$. This assumption might be inappropriate. In order to avoid this problem a coding of $+1/2$ and $-1/2$ is used for the group dummy for the random effects x_{ij} instead of a coding of 0 and 1 in Turner et al. (2000). Thus, overall heterogeneity is represented by parameters σ^2, τ^2, ω representing variations in control/treatment group and correlation between random study effects respectively. In comparison, standard random effects model assumes that the heterogeneity is usually represented by a single between-study variance τ^2.

α, θ, σ, τ^2 and ω are unknown parameters that have to be estimated. These parameters can be estimated similarly to estimation of parameters in unconditional generalized linear mixed-effects model with fixed study effects (Turner et al., 2000). Hamza et al. (2008) also studied a logistic regression model with a random intercept for meta-analysis of proportions.

5.4.3 A conditional generalized linear mixed-effects model (exact likelihood)

The hypergeometric-normal model was initially proposed for meta-analysis by Van Houwelingen et al. (1993) and Liu and Pierce (1993). Later, Sidik and Jonkman (2008) and Stijnen et al.

(2010) implemented the model into practice. The exact likelihood function of hypergeometric-normal model for each study i is

$$h(x_{1i};\theta,\tau^2)=\int_{-\infty}^{\infty} g(x_{1i}|\theta_i)f(\theta_i|\theta,\tau^2)d\theta_i=$$

$$\int_{-\infty}^{\infty}\binom{n_{1i}}{x_{1i}}\binom{n_{2i}}{x_{2i}}\frac{\exp(x_{1i}\theta_i)}{P(\theta_i)}\frac{1}{\sqrt{2\pi\tau^2}}\exp(-\frac{(\theta_i-\theta)^2}{2\tau^2})d\theta_i.$$

where $g(x_{1i}|\theta_i)$ is the non-central hypergeometric density function for number of events in treatment arm X_{1i} given $X_{1i}+X_{2i}=X_i$, $P(\theta_i)$ is

$$P(\theta_i)=\sum_{i=\max(0,n_i-n_{2i})}^{\min(n_{1i},n_{2i})}\binom{n_{1i}}{i}\binom{n_{2i}}{X_i-i}\exp(X_i\theta_i)$$

is the polynomial in θ_i and true unobservable LOR θ_i. The distribution of true effect measure θ_i is $f(\theta_i|\theta,\tau^2)$, which is the normal probability density function with mean θ and variance τ^2. Density $h(x_{1i}|\theta,\tau^2)$ is the marginal probability function with integrated out unobserved study specific effect. For $g()$ normal and $f()$ non-central-hypergeometric in (5.3.1), the model is referred to as hypergeometric-normal model, Stijnen et al. (2010). According to Stijnen et al. (2010), this approach should solve issues related to an addition of continuity corrections and the existence of correlations between $\hat{\sigma}_i^2$ and $\hat{\theta}_i$ arising in the standard random effects model. This model belongs to a class of generalized linear mixed models. For our case with log odds ratio for effect measure, the model is known as mixed effects logistic model. Liang and Zeger (1986) have shown that the inference based on the non-central hypergeometric likelihood is sensitive to misspecification of the dependence structure.

Liu and Pierce (1993) discussed an accurate closed form approximation to $h(x_{1i};\theta,\tau^2)$ based on Laplace method, which is popular in Bayesian setting. Also in the same paper by Liu and Pierce (1993), other simple approximations to $h(x_{1i};\theta,\tau^2)$ such as William's method based on a weighted least squares and quasi-likelihood approach to the dispersion models are considered. However, these methods do not take into account the variability of estimators for τ^2. Breslow and Clayton (1993) provide a mixed model for log odds ratio based on the non-central hypergeometric distribution and discuss the methods based on the full likelihood

analysis for generalized linear mixed models such as penalized quasi-likelihood and marginal quasi-likelihood methods.

The log-likelihood of the non-central hypergeometric normal model is

$$l(\theta, \tau^2) = \log(\prod_{i=1}^{K} h(x_{1i}; \theta, \tau^2)) =$$

$$= \sum_{i=1}^{K} \log(\int_{-\infty}^{\infty} \binom{n_{1i}}{x_{1i}} \binom{n_{2i}}{x_{2i}} \frac{\exp(x_{1i}\theta_i)}{P(\theta_i)} \frac{1}{\sqrt{2\pi\tau^2}} \exp(-\frac{(\theta_i - \theta)^2}{2\tau^2}) d\theta_i).$$

The parameters θ and τ^2 can be estimated by either using the EM algorithm (Van Houwelingen et al., 1993), the numerical Newton-Raphson iterative algorithm (Sidik and Jonkman, 2008) or maximizing $l(\theta, \tau^2)$ (Stijnen et al., 2010; Viechtbauer et al., 2010). Liu and Pierce (1993) proposed an approximation for the integrand in mix of non-central hypergeometric and normal densities based on Laplace method. However, the most recent approximations for the marginal likelihood of non-central hypergeometric normal distribution are based on adaptive Gauss-Hermite quadrature. The non-central hypergeometric distribution is based on the binomial distribution in treatment and control arms. When the binomial distribution is invalid, X_{1i} no longer follows non-central hypergeometric distribution (Liang, 1985). The non-central-hypergeometric normal model is supposed to solve problems related to dependence between $\hat{\theta}_i$ and $\hat{\sigma}_i^2$.

5.4.4 A conditional generalized linear mixed-effects model (approximate likelihood)

In case of small total number of events relative to the total group sizes, the non-central hypergeometric distribution can be approximated by a binomial distribution (Stijnen et al., 2010). The model with approximate likelihood for a conditional generalized linear mixed-effects model is

$$X_{1i}|(X_{1i} + X_{2i}) \sim Binomial(X_{1i} + X_{2i}, P_{X_{1i}|(X_{1i}+X_{2i})})$$

with

$$\log\left(\frac{P_{X_{1i}|(X_{1i}+X_{2i})}}{1 - P_{X_{1i}|(X_{1i}+X_{2i})}}\right) = \log\left(\frac{n_{1i}}{n_{2i}}\right) + \theta_i.$$

This model arises because

$$\exp(\hat{\theta}_i) = \frac{X_{1i}(n_{2i} - X_{2i})}{X_{2i}(n_{1i} - X_{1i})} = \frac{\hat{P}_{X_{1i}|(X_{1i}+X_{2i})}}{1 - \hat{P}_{X_{1i}|(X_{1i}+X_{2i})}} \frac{(n_{2i} - X_{2i})}{(n_{1i} - X_{1i})}$$

and assuming that X_{1i} and X_{2i} are small relative to the n_{1i} and n_{2i}

$$\frac{(n_{2i} - X_{2i})}{(n_{1i} - X_{1i})} \approx \frac{n_{2i}}{n_{1i}},$$

which results in approximation

$$\exp(\hat{\theta}_i) = \frac{\hat{P}_{X_{1i}|(X_{1i}+X_{2i})}}{1 - \hat{P}_{X_{1i}|(X_{1i}+X_{2i})}} \frac{(n_{2i} - X_{2i})}{(n_{1i} - X_{1i})} \approx \frac{\hat{P}_{X_{1i}|(X_{1i}+X_{2i})}}{1 - \hat{P}_{X_{1i}|(X_{1i}+X_{2i})}} \frac{n_{2i}}{n_{1i}}$$

and

$$\hat{\theta}_i = \log(\frac{\hat{P}_{X_{1i}|(X_{1i}+X_{2i})}}{1 - \hat{P}_{X_{1i}|(X_{1i}+X_{2i})}}) + \log(\frac{n_{2i}}{n_{1i}}) = \log(\frac{\hat{P}_{X_{1i}|(X_{1i}+X_{2i})}}{1 - \hat{P}_{X_{1i}|(X_{1i}+X_{2i})}}) - \log(\frac{n_{1i}}{n_{2i}}).$$

The parameters of this model can be estimated by maximizing a random intercept logistic regression model with offset $\log(n_{1i}/n_{2i})$. Stijnen et al. (2010) states "Using a conditional generalized linear mixed-effects model is the same as using Breslow's approximation for the likelihood as is done in many conditional logistic regression and Cox regression programs".

5.5 Simulation study

In this section we provide a simulation study to assess the performance of point and interval estimators of overall log-odds ratio θ and between-study variance τ^2 for data generated from standard REM and ODM. The estimators of θ and τ^2 are obtained from non-central-hypergeometric-normal model proposed by Stijnen et al. (2010).

In standard REM, the data is generated as follows:

$$X_{1i} \sim Binom(n_{1i}, p_{1i}) \quad \text{and} \quad X_{2i} \sim Binom(n_{2i}, f(p_{1i}, \theta_i)),$$

with $\theta_i = \log(p_{1i}(1 - p_{2i})/p_{2i}(1 - p_{1i}))$, $\theta_i \sim N(\theta, \tau^2)$ and $f(p_{1i}, \theta) = p_{1i}\exp(\theta_i)/(1 - p_{1i} + \exp(\theta_i)p_{1i}))$. This scenario is similar to the method of data generation in a simulation study by Viechtbauer (2007). For a pair of binomial distributions within each study, the conditional

distribution of the number of events in treatment arm given fixed margins is a non-central-hypergeometric distribution. No continuity corrections are added to number of events.

For the ODM, the simulation scenario is similar to simulation study in Chapter 4. In Chapter 4, we introduced a multiplicative random effects model for overdispesed binary data and modified the standard methods of meta-analysis by inclusion of estimated intra-cluster correlation. In the current chapter, we apply the non-central-hypergeometric-normal model proposed by Stijnen et al. (2010) to the data generated from the ODM model as

$$X_{1i} \sim BetaBinom(n_{1i}, p_{1i}, \rho) \quad \text{and} \quad X_{2i} \sim BetaBinom(n_{2i}, f(p_{1i}, \theta_i), \rho),$$

where ρ is the parameter which describes an intra-cluster correlation. The package emdbook by Bolker (2011) is used for simulating data from the beta-binomial distributions. In ODM, ρ is unknown and have to be estimated. However, the non-central-hypergeometric-normal analysis by Stijnen et al. (2010) estimates between-study variance τ^2, not ρ. Hence, given the correspondence (4.2.3) between τ^2 and ρ in Chapter 4 for ODM and equal sample sizes, the bias of τ^2 can be assessed by obtaining the true values of τ^2 from the relationship (4.2.3) in Chapter 4. Again, no continuity corrections are added to number of events in each arm.

In Chapter 4, we have shown that the standard methods perform badly in estimation of ICC ρ and overall odds ratio θ when the probabilities are low in both arms. In particular, the estimate of θ is biased when the probabilities in the arms are not the same. The non-central-hypergeometric-normal model (Stijnen et al., 2010) may be an attractive alternative for standard additive and overdispersed random effects models. In this section, the analysis based on the non-central hypergeometric normal model is applied to both REM and ODM. The sample sizes are assumed to be the same within each arm and across K studies. The maximum-likelihood estimators of θ and τ^2 are assessed in respect to their bias and coverage at the 95% confidence level.

5.5.1 Fitting the non-central-hypergeometric-normal model in R

R package metafor allows to fit the non-central-hypergeometric-normal (NCHGM) model proposed by Stijnen et al. (2010). The NCHGM is the conditional generalized linear mixed-effects model. 'metafor' package-version 1.9-2 was used in the current simulations.

In R, there are two methods to fit the non-central-hypergeometric distribution. One is "dFNCHypergeo" from BiasedUrn package (Fog and Fog, 2013). The other one is "dnoncenhypergeom" from MCMCpack package (Martin et al., 2016). Both of the methods can be used in rma.glmm function from metafor package. In rma.glmm, "dFNCHypergeo" is the default distribution for fitting the conditional generalized linear mixed-effects model (exact likelihood). The model is specified as

$$rma.glmm(measure = \text{“}OR\text{”}, ai =, bi =, ci =, di =, data =, model = \text{“}CM.EL\text{”}).$$

In case of using the "dnoncenhypergeom" function for non-central-hypergeometric distribution, the model should be specified as

$$rma.glmm(measure = \text{“}OR\text{”}, ai =, bi =, ci =, di =,$$

$$model = \text{“}CM.EL\text{”}, control = list(dnchgcalc = \text{“}dnoncenhypergeom\text{”}))$$

where ai, bi ci, di are the binary data from the table

	Event	No event	Total
Treatment	ai	bi	$ai + bi$
Control	ci	di	$ci + di$
Total	$ai + ci$	$bi + di$	$ai + bi + ci + di$

Both methods should perform similarly, when fitting the conditional generalized linear mixed-effects model (exact likelihood). However, the convergence problems might occur when trying to fit a saturated model. Switching to the other method can help to solve the problem.

In case of low number of events, the binomial-normal approximation is possible for non-central-hypergeometric-normal distribution, since the non-central-hypergeometric distribution

can be well approximated by a binomial distribution. In that case, the model is defined as a conditional generalized linear mixed-effects model (approximate likelihood). This model is specified as

$$rma.glmm(measure = \text{“}OR\text{”}, ai =, bi =, ci =, di =, data =, model = \text{“}CM.AL\text{”}).$$

5.5.2 Configurations

The following configurations of parameters were included in the simulations. The number of studies $K = (5, 10, 30)$; average sample sizes in each arm are $n = (50, 100, 250, 1000)$; the between-study variance for standard random effects model $\tau^2 = (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1)$; the overdispersion parameter ρ varies between 0 and 0.1 (small to moderate heterogeneity) with steps 0.01, and between 0.1 and 0.6 in steps 0.1 (moderate to large heterogeneity). The values of LOR θ vary from 0 to 2 in steps of 1. The probability in control group p_{2i} is taken to be 0.1, since we are mostly interested in sparse data. A total of 10000 repetitions were produced for each combination. However, not all the simulations converge due to problems in trying to fit the saturated model, and the actual number of repetitions may be smaller.

5.5.3 Results for a pair of binomial distributions

The results of the simulations for the behaviour of conditional generalized linear mixed-effects model with exact likelihood in case of a pair of binomial distributions as in the standard REM are shown in the Figure 5.1 and Figure 5.2 for $0 \leq \tau^2 \leq 1$ with true LOR $\theta = 0$ and $p_{2i} = 0.1$. This scenario results in generation of the sparse data in treatment and control arm. Figure 5.1 and Figure 5.2 show the bias of between-study variance τ^2, bias of overall odds ratio θ and coverage at the nominal confidence level of 0.95 of the overall odds ratio θ obtained from two methods of fitting the non-central-hypergeometric-normal model for odds ratio. These two methods are dFNCHypergeo distribution from BiasedUrn package and dnoncenhypergeom distribution from MCMCpack package respectively. Similarly, Figure 5.3 and 5.4 show the results of simulations for true $\theta = 1$ and $p_{2i} = 0.1$ using dFNCHypergeo distribution from

BiasedUrn package (5.3) and using dnoncenhypergeom distribution from MCMCpack package (5.4). When $\theta = 1$ with $p_{2i} = 0.1$, the probability $p_{1i} > 0.1$ in the treatment arm. We also fitted the conditional generalized linear mixed-effects model with approximate likelihood for the similar scenario as above with $\theta = 0, 1$ and $p_{2i} = 0.1$. The results for $\theta = 0$ and $\theta = 1$ with $p_{2i} = 0.1$ are shown in the Figure 5.5 and Figure 5.6 respectively.

From the first row of the Figure 5.1 or Figure 5.2 for $0 \leq \tau^2 \leq 1$, it is clear that estimator for between-study variance τ^2 is subject to downward bias. When $\theta = 0$ and $\tau^2 \neq 0$, the bias varies around 0.25 - 77% for different values of $N = 50, 100, 250, 1000$ with highest bias for $K = 5$, $N = 50$ and $\tau^2 = 0.7$. When $\theta = 0$ and $\tau^2 = 0$, the bias varies between values of $0.002 - 0.0992$. Hence, when $\tau^2 = 0$, still some heterogeneity is estimated. This heterogeneity might be also a result of sampling variability. When $\theta = 1$, the bias of τ^2 varies between 0.5-62 % excluding extremely large values for bias for $K = 5$ and $K = 10$. The extremely large values for the bias might be explained by non-stability of the dFNCHypergeo distribution when the probabilities in both arms are not equal. Non-stability of the dFNCHypergeo distribution might the result of small number of studies ($K = 5$ and $K = 10$) in simulations. The similar scenario with more non-stable observations occur when $\theta = 2$. Thus, using dFNCHypergeo distribution for non-equal probabilities in both arms is not recommended for studies with small K. In this case, the inference based on a conditional generalized linear mixed effects model with exact likelihood might be misleading. The bias of τ^2 is similar when using dnoncenhypergeom distribution from MCMCpack package, see Figure 5.2 for $\theta = 0$ and Figure 5.4 for $\theta = 1$. For $\theta = 0$ with $K = 5$ and $K = 10$, the bias of $\hat{\tau}^2_{CM.EL}$ increases asymptotically with N. We can clearly see the latter from the first row in Figure 5.1 or Figure 5.2. However, the increase in bias of $\hat{\tau}^2_{CM.EL}$ is less visible when $K = 30$. Probably, this can be explained by large number of studies ($K = 30$). When $K = 30$, the bias of $\hat{\tau}^2_{CM.EL}$ also increases with increasing N from $N = 50$ to $N = 1000$. The dnoncenhypergeom distribution for fitting the conditional generalized linear mixed effects model with exact likelihood is also non stable when $\theta = 1$. Thus, when probabilities in both arms are equal, it is difficult to find the order for the bias of

$\hat{\tau}^2_{CM.EL}$. The order might be either $1/N$ or $1/K$. Another possibility is that the order of the bias might be depend on the combination of N and K simultaneously.

The bias of the overall odds ratio $\hat{\theta}_{CM.EL}$ from two methods (dFNCHypergeo and dnoncenhypergeom) are quite similar. The bias for $\hat{\theta}_{CM.EL}$ decreases with increasing N. The bias $\hat{\theta}_{CM.EL}$ is linear to τ^2. The bias $\hat{\theta}_{CM.EL}$ is the smallest when true value of $\theta = 0$ and $p_{2i} = 0.1$ with $N = 1000$. In this case $p_{1i} = 0.1$ and both arms provide sparse data. When $\theta = 1$, the data is sparse only in control arm, since $p_{1i} = p_{2i}\exp(\theta_i)/(1 - p_{2i} + p_{2i}\exp(\theta_i))$ and $\theta \sim N(\theta, \tau^2)$. For instance, $\theta = 1$ result in $p_{1i} = 0.232$. The bias of $\hat{\theta}_{CM.EL}$ is higher when $\theta = 1$ (see Figure 5.3 for dFNCHypergeo and Figure 5.4 for dnoncenhypergeom). The bias of $\hat{\theta}_{CM.EL}$ is of order $1/N$.

The coverage of overall log odds ratio $\hat{\theta}_{CM.EL}$ is liberal. When $K = 5$ and $\theta = 0$, the coverage is about 84 - 88%. When $K = 10$ and $K = 30$, the coverage goes up to 90% and 93% for $N = 1000$ respectively. Hence coverage improves with N. When $\theta = 1$, the coverages are pretty similar to the case when $\theta = 0$. Interestingly, when $N = 1000$, the coverage drops dramatically for all values of θ with increasing τ^2. It seems to be due to the bias in estimation of τ^2. We can clearly see that the bias of τ^2 is not negligible in asymptotics. Increasing K and N do no provide an asymptotic results in conditional generalized linear mixed-effects model (exact likelihood). The conditional generalized linear mixed-effects model (approximate likelihood) performs worse than conditional generalized linear mixed-effects model (exact likelihood). Figure 5.5 and Figure 5.6 show that $\hat{\tau}^2_{CM.AL}$ is more biased than $\hat{\tau}^2_{CM.EL}$ for $\theta = 0$ and $\theta = 1$. The bias of $\hat{\tau}^2_{CM.AL}$ is more than 50%. This bias leads to shorter confidence interval of θ and lower coverage shown in third row of the Figure 5.5 and Figure 5.6. It is not surprising that the binomial-normal approximation fails for large values of τ^2, as there will be many studies where the number of cases in the treatment group is large relative to the sample size (and this is not offset by the large number of studies where the number of cases in the treatment group is very small). The conditional generalized linear mixed-effects model with exact and approximate likelihood always underestimate the between-study variance. The

underestimation of τ^2 affects inference about θ, by overestimating weights and information matrix. The overestimation of information matrix results in wrong hypothesis tests with large Type I error and confidence invervals with coverage lower than the nominal level.

We believe that the non-central hypergeometric normal model suffers from transformation bias in maximum likelihood estimates of overall odds ratio θ and between-study variance τ^2. However, the negative bias of τ^2, positive bias and low coverage of $\hat{\theta}_{CM.EL}$ might also be due to nonstability of implementations dFNCHypergeo and dnoncenhypergeom when fitting the model in R programming language. As we have shown, more problems due to non-stability of fitting the conditional generalized linear mixed effects model with exact likelihood appear when probabilities in both arms are not equal. Some of the findings from simulations of NCHGN model depend on numerical issues and not the properties of the model itself. NCHGN model is a rather difficult model to fit, involving computation of the probability mass function of the non-central hypergeometric distribution, numerical integration thereof, in addition to the optimization required for finding the maximum likelihood estimates estimates. The variance-covariance matrix of the fixed (and random) effects is obtained by numerically approximating the Hessian matrix. All of these computations can go horribly wrong. Therefore, the present results say something about the implementation of the NCHGN model in the metafor package, but whether this holds in general for the model is another issue. New ways of improving the numeric methods for conditional generalized linear mixed-effects model with exact likelihood in metafor package in R are required.

In summary, for sparse data, we would recommend to use non-central hypergeometric normal model only when number of studies K is large and sample sizes N are moderate. When K is small, the estimates of between-study variance and overall effect measure are biased resulting in narrow confidence intervals. The bias of between-study variance reduces with K. However, for large values of N, the confidence interval is narrower than for moderate values of N. Thus, when sample sizes are too large, the non-central hypergeometric normal model is not recommended. In that case, the methods of standard random effects might be a better option

since their asymptotic behaviours are well known in meta-analysis. Also, we would not recommend the binomial-normal approximation to non-central hypergeometric normal model, since it provides more biased estimates of between-study variance resulting in wrong confidence intervals.

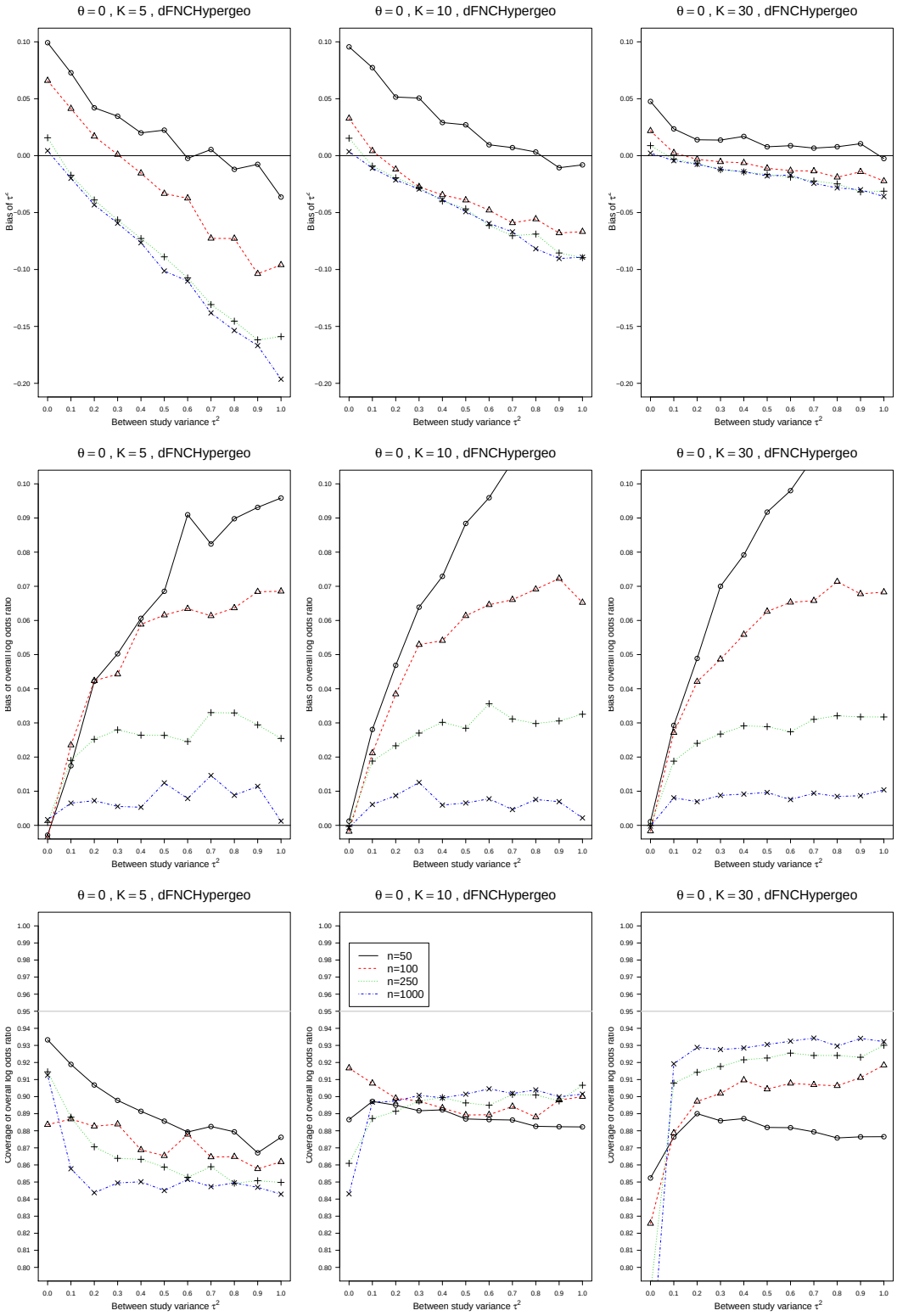


Figure 5.1: Bias of between-study variance τ^2 (first row), bias (second row) and coverage (third row) of overall odds ratio at the nominal 95% level using the non-central hypergeometric-normal in additive random effects model and using dFNCHypergeo for $p_{2i} = 0.1$, $\theta = 0$ and $0 \leq \tau^2 \leq 1$. Light grey line at 0.95.

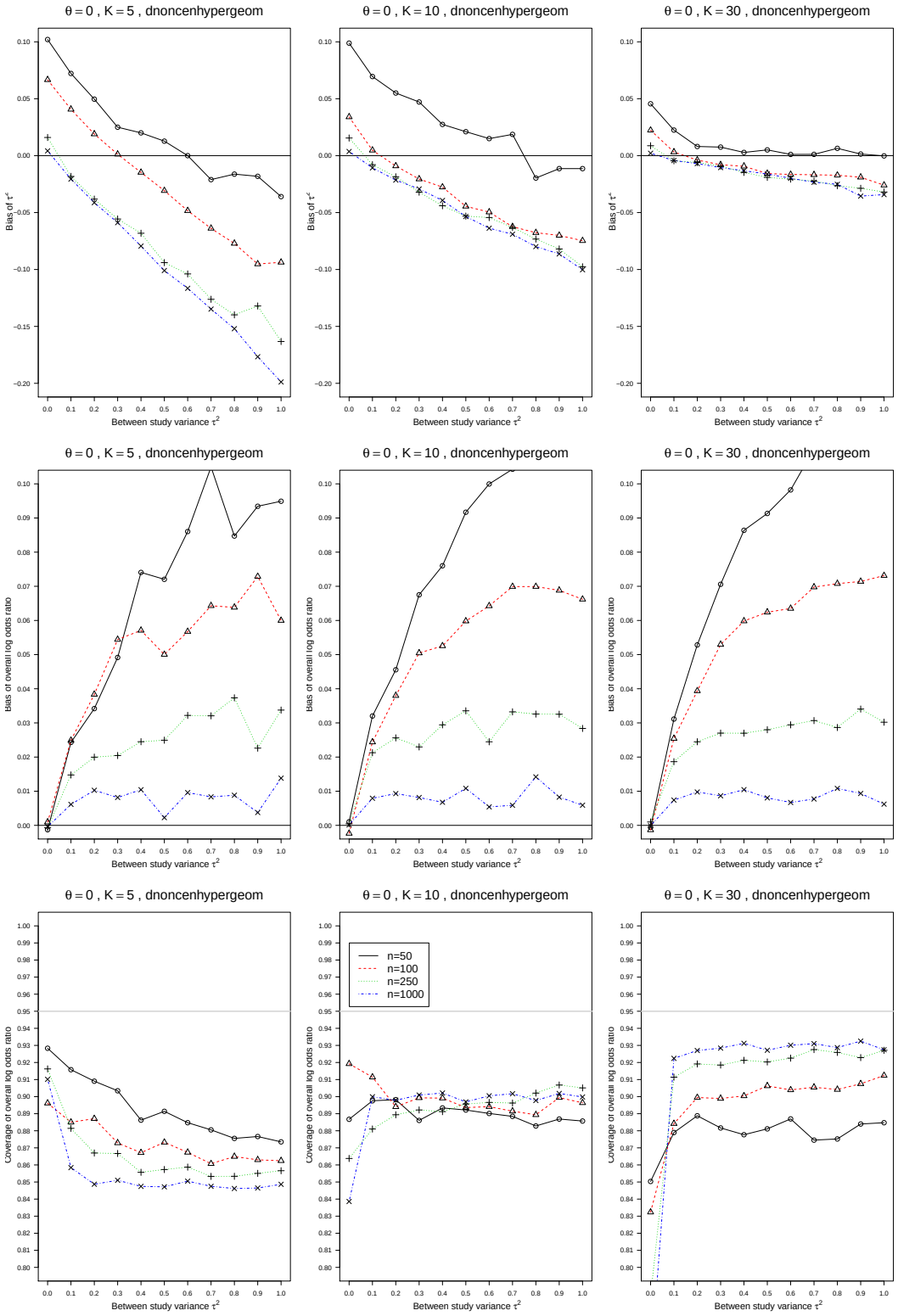


Figure 5.2: Bias of between-study variance τ^2 (first row), bias (second row) and coverage (third row) of overall odds ratio at the nominal 95% level using the non-central hypergeometric-normal in additive random effects model using dnoncenhypergeom for $p_{2i} = 0.1$, $\theta = 0$ and $0 \leq \tau^2 \leq 1$. Light grey line at 0.95.

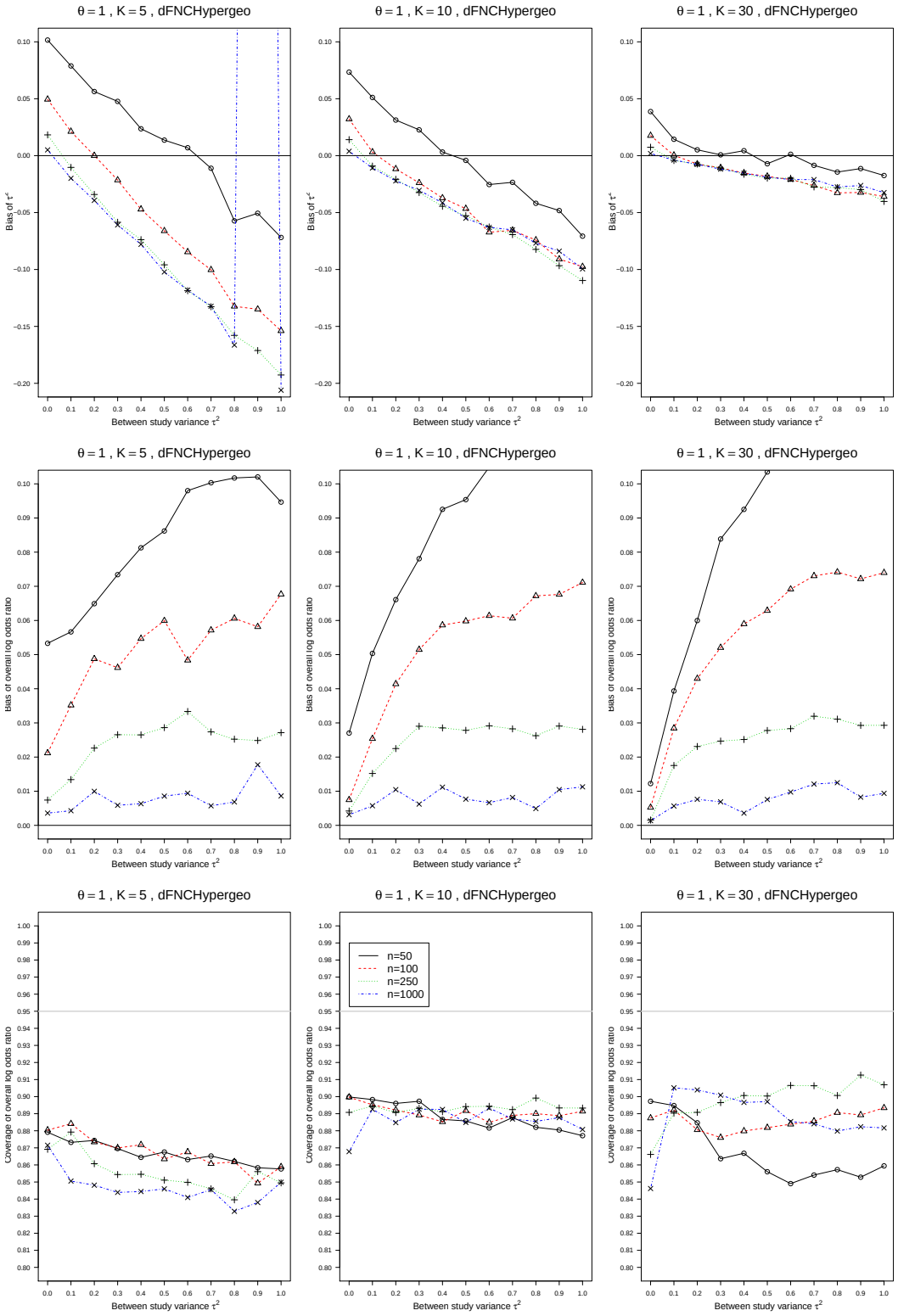


Figure 5.3: Bias of between-study variance τ^2 (first row), bias (second row) and coverage (third row) of overall odds ratio at the nominal 95% level using the non-central hypergeometric-normal in additive random effects model and using dFNCHypergeo for $p_{2i} = 0.1$, $\theta = 1$ and $0 \leq \tau^2 \leq 1$. Light grey line at 0.95.

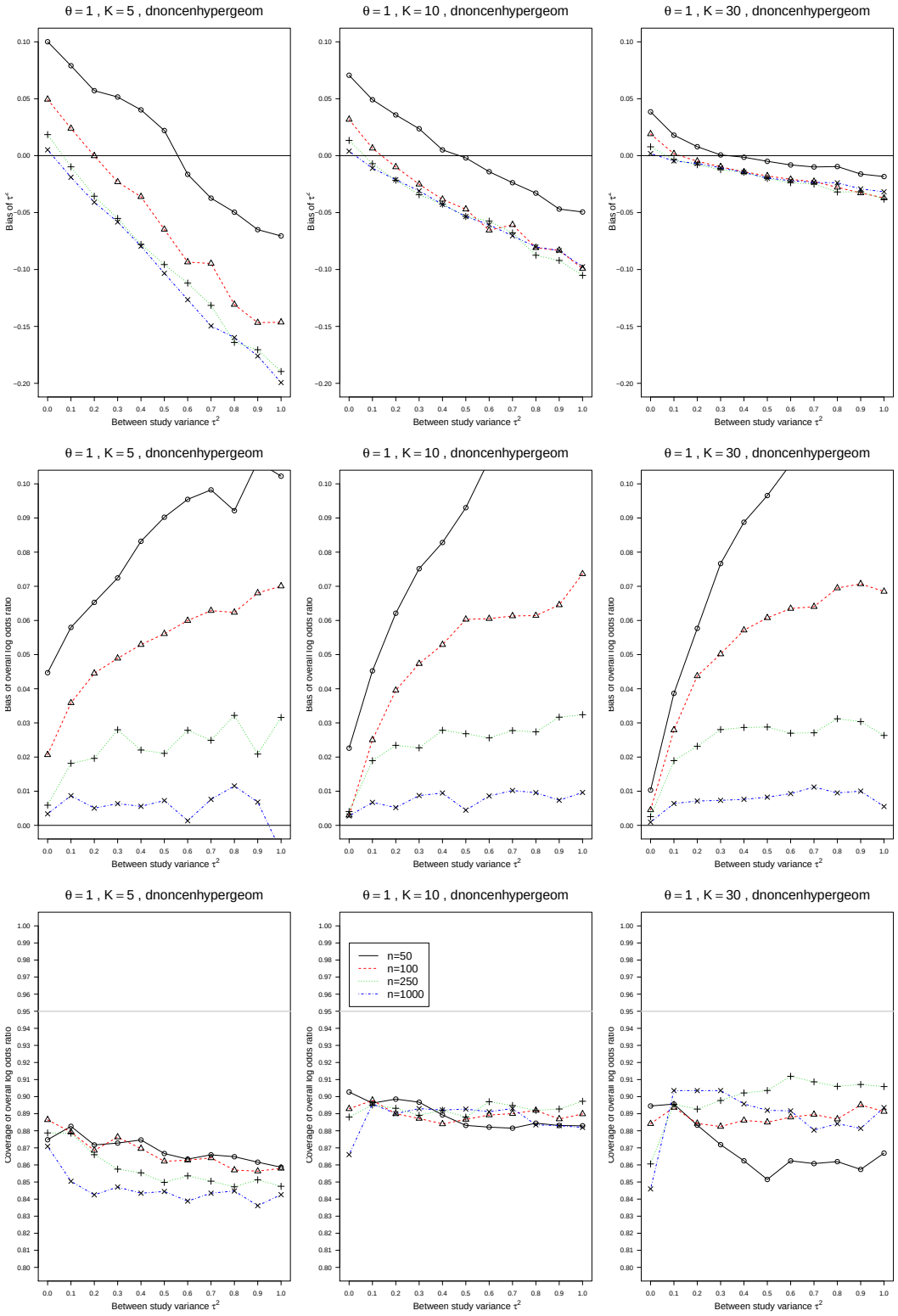


Figure 5.4: Bias of between-study variance τ^2 (first row), bias (second row) and coverage (third row) of overall odds ratio at the nominal 95% level using the non-central hypergeometric-normal in additive random effects model using dnoncenhypergeom for $p_{2i} = 0.1$, $\theta = 1$ and $0 \leq \tau^2 \leq 1$. Light grey line at 0.95.

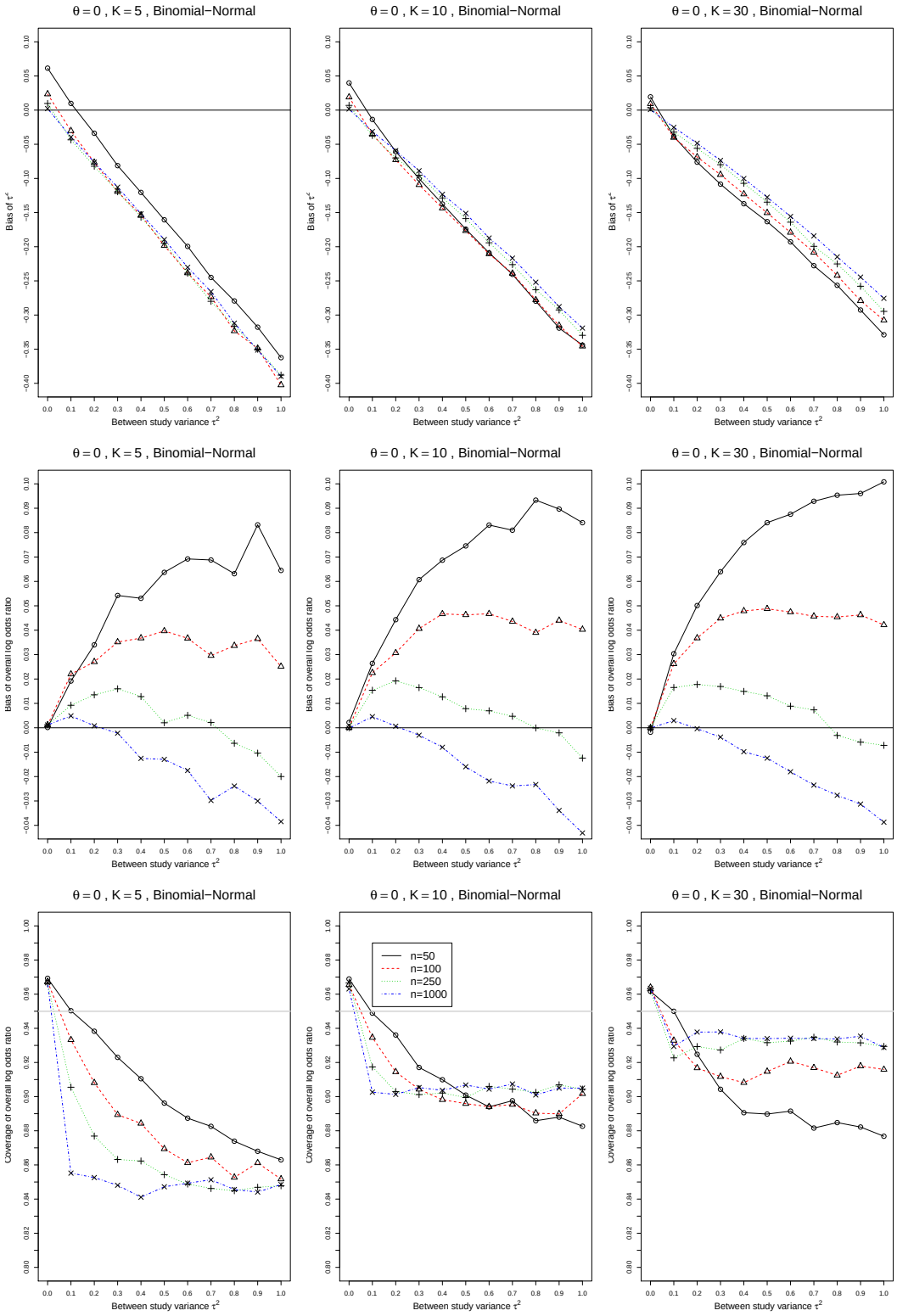


Figure 5.5: Bias of between-study variance τ^2 (first row), bias (second row) and coverage (third row) of overall odds ratio at the nominal 95% level using the binomial normal approximation to non-central hypergeometric-normal in additive random effects model for $p_{2i} = 0.1$, $\theta = 0$ and $0 \leq \tau^2 \leq 1$. Light grey line at 0.95

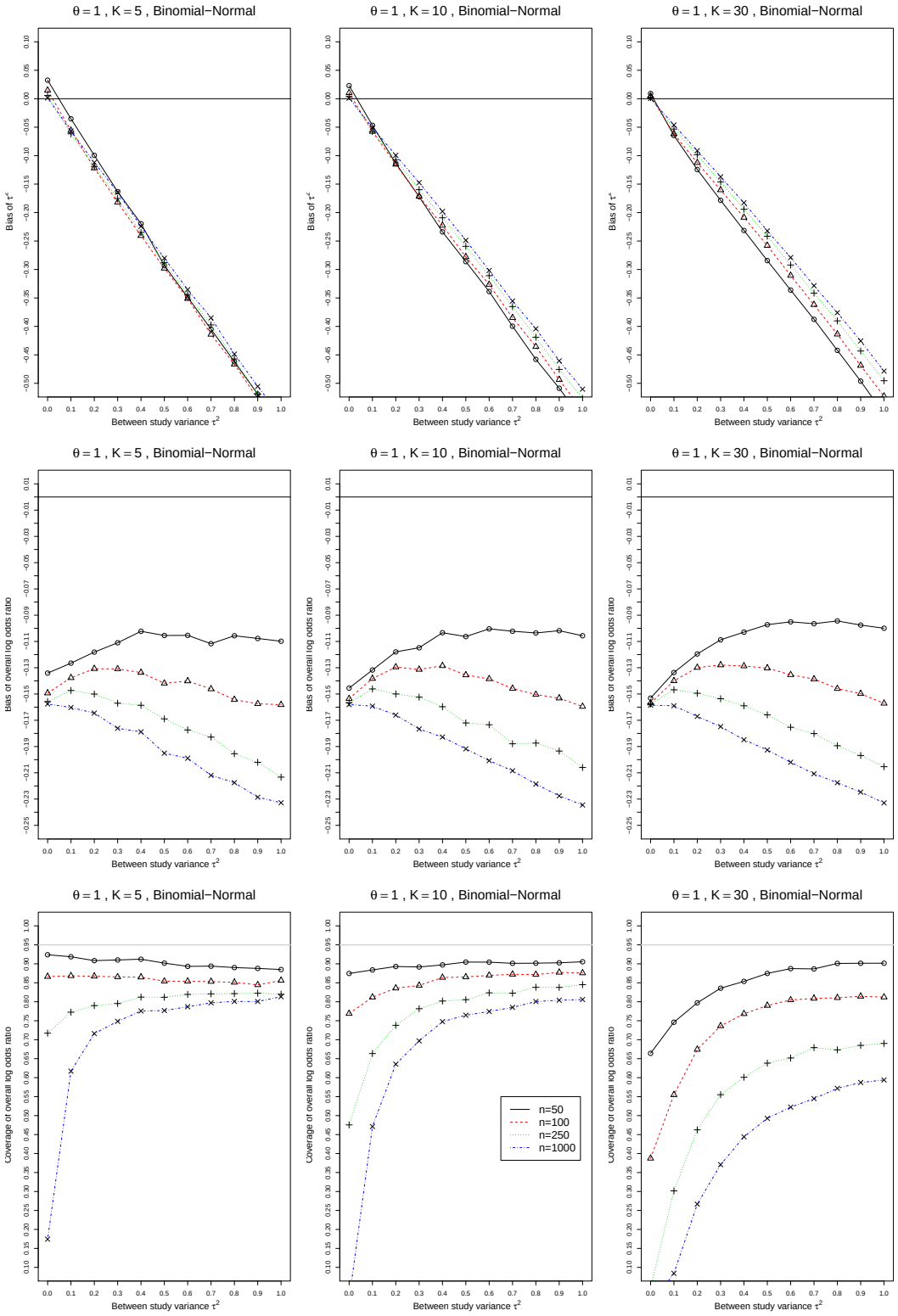


Figure 5.6: Bias of between-study variance τ^2 (first row), bias (second row) and coverage (third row) of overall odds ratio at the nominal 95% level using the binomial normal approximation to non-central hypergeometric-normal in additive random effects model for $p_{2i} = 0.1$, $\theta = 1$ and $0 \leq \tau^2 \leq 1$. Light grey line at 0.95

5.5.4 Results for a pair of beta-binomial distributions

In Chapter 4, we have shown that the standard methods fail in estimation of parameters for overdispersed model (ODM). The non-central hypergeometric normal model is an attractive alternative to standard methods. Figures 5.7, 5.8, 5.9 show the plots of bias of $\hat{\tau}^2$, bias and coverage of $\hat{\theta}_{CM.EL}$ for $\theta = 0, 1, 2$ and $p_{2j} = 0.1$ using the dFNCHypergeo distribution from BiasedUrn package. The results of simulations using the dnoncenhypergeom distribution from MCMCpack package is similar to the results of simulations using the dFNCHypergeo distribution from BiasedUrn package. Due to the close similarity, the results of simulations using the dnoncenhypergeom distribution from MCMCpack package are not reported. Figures 5.10, 5.11 and 5.12 show the results of the simulations when using a conditional approximate likelihood in a generalized linear mixed-effects model.

From the figures 5.7, it is clear that the estimate of the between-study variance $\hat{\tau}^2$ is biased. The bias is positive and it is increasing with τ^2. The bias of τ^2 is smaller for $\theta = 1$ and $\theta = 2$, but it still exists. For $\theta = 1$ and $\theta = 2$, the bias of τ^2also increases with τ^2. The explanation of larger bias of τ^2 when $\theta = 0$ might be that when $\theta = 0$, the value of probabilities are $p_{1i} = 0.1$ and $p_{2i} = 0.1$. Thus, we get sparse data in both arms. The large amount of sparse data might introduce an additional heterogeneity. In contrast, when $\theta = 1$ and $\theta = 2$, the probability of control arm is $p_{2i} = 0.1$, but the probability of treatment arm is $p_{1i} = 0.23$ and $p_{1i} = 0.45$ respectively. Hence, only the treatment arm consists of sparse data with zero-entries in cells. This leads to lower heterogeneity.

The log-odds-ratio $\hat{\theta}_{CM.EL}$ is also biased for $\theta = 0$, $\theta = 1$ and $\theta = 2$. For $\theta = 0$, the bias decreases with increasing K from $K = 5$ to $K = 30$. This decrease is from about 50% to 6%. The bias of log-odds-ratio $\hat{\theta}_{CM.EL}$ is quite pronounced for $\theta = 1$ and $\theta = 2$. The bias of $\hat{\theta}_{CM.EL}$ comes from the combination of transformation bias and bias of $\hat{\tau}^2$. This is the same bias that we discussed in section 4.6.2 of Chapter 4. This bias results in poorer coverage for $\hat{\theta}_{CM.EL}$, when we increase the value of θ from $\theta = 0$ to $\theta = 1$ and to $\theta = 2$ (see Figures 5.7-5.9 for dFNCHypergeo). For small number of studies $K = 5$, Figures 5.7-5.9, the coverage goes

down steadily from 89% to 85% with increasing the sample size from $N = 50$ to $N = 1000$. For moderate number of studies $K = 30$ and $\theta = 0$, the coverage varies between 92% and 93% for different values of N from $N = 50$ to $N = 1000$. When $\theta \neq 0$, the coverage for $K = 5$ and $K = 10$ is around $82\% - 83\%$ for $\theta = 1$ and $75\% - 85\%$ for $\theta = 2$. When the number of studies K increases to $K = 30$, the coverage deteriorates dramatically to $73\% - 76\%$ for $\theta = 1$ and $50\% - 60\%$ for $\theta = 2$. The values of τ^2 in the figures correspond to the values of ρ between 0 and 0.1 for different combinations of p_{2j} and θ.

Some of the Figures (Figure 5.6, Figure 5.8, Figure 5.9, Figure 5.10) show the erratic behaviour of plots when $K = 5$. Figure 5.6, Figure 5.8, Figure 5.9 show results of simulations from non-central-hypergeometric normal model and Figure 5.10 show results of simulations from binomial-normal approximation to non-central-hypergeometric normal model. The erratic behaviour appears when the true value of log-odds ratio θ changes from 0 to 1 and 2. Correspondingly, the probabilities in treatment arms change from $p_{1i} = 0.1$ when $\theta = 0$ to $p_{1i} = 0.23$ for $\theta = 1$ and $p_{1i} = 0.45$ for $\theta = 2$. The reason for erratic behaviour is that when $K = 5$, the non-central-hypergeometric normal model does not estimate τ^2 very well, since variance τ^2 is of order $1/K$. Some scenarios in simulations resulted in huge estimate of τ^2. Hence there is an imbalance between number of events in treatment arm in comparison to number of events in control arm. Suppose we have $K = 5$ and there are 3 or 4 studies with zeros in control arm since $p_{2i} = 0.1$ and large number of events in treatment arm since $p_{1i} = 0.23$ for $\theta = 1$ and $p_{1i} = 0.45$ for $\theta = 2$. Then there is a big imbalance between the numbers of events in treatment arm is comparably large to number of events in control arm. From the results of our simulations it follows that overestimated between-study variance and true effect measure will result in shorter confidence interval $\hat{\theta}_{CM.EL}$. Hence it might lead to reduced coverage. Thus, the non-central-hypergeometric normal model does not perform well in ODM apart from the case when the number of studies is large and $\theta \neq 0$. The explanation of this is that the pair of beta-binomial distributions in ODM model do not result in non-central hypergeometric distribution (Liang, 1985). Originally, the non-central hypergeometric

distribution is derived from a pair of binomially distributed variables for treatment and control arm. Applying the non-central hypergeometric distribution to case of correlated binomial events leads to the failure of assumptions and wrong inference. Liang (1985) has shown that the maximum likelihood estimator of odds ratio based on non-central hypergeometric distribution for fixed effect is biased when the number of studies increases asymptotically. For random effects, we have shown that the maximum likelihood estimators of between-study variance and overall effect measure are also biased. Particularly, the biases are large when the randomness is a result of intra-cluster dependence within each arms in each study. The inference based on non-central hypergeometric distribution is sensitive to misspecification of intra-cluster correlation structure (Hanfelt and Liang, 1998; Liang, 1985). This problem might result in issues in misspecification of the models for meta-analysis of binary data. Thus, inference based on non-central hypergeometric distribution in case of overdispersion in meta-analysis might be misleading. Overall, the results are worse in data generated from ODM than in data generated from REM.

In summary, we would not recommend the non-central hypergeometric-normal model when the binary data is assumed to be correlated and effect measure is far from zero. When, effect measure is far from zero, the main problem in that case is the transformation bias which results in biased estimates of overall effect measure and too narrow confidence intervals. The binomial-normal approximation to non-central hypergeometric-normal model also is not a good option, since it provides poorer confidence intervals than the actual non-central hypergeometric-normal model. Even, when the the effect measure is close to zero, the between-study variance is still biased. This bias does not seem to decrease with K. The bias of overall-effect measure does reduce with increasing K and N. However, the coverage does not still reach nominal 95% significance level.

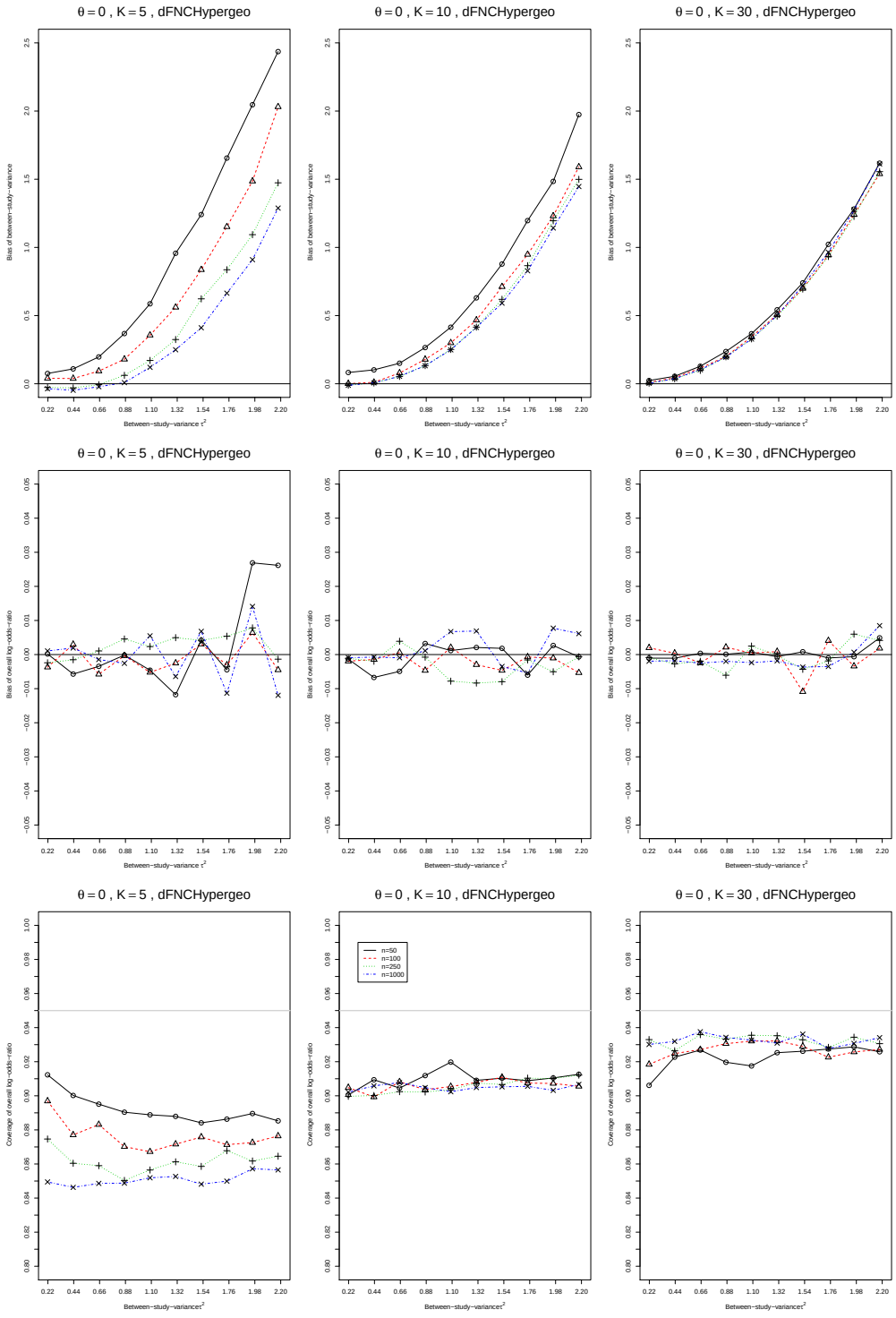


Figure 5.7: Bias of between-study variance τ^2 (first row), bias (second row) and coverage (third row) of overall odds ratio at the nominal 95% level using the non-central hypergeometric-normal in overdispersed random effects model using dFNCHypergeo for $p_{2i} = 0.1$, $\theta = 0$ and $0 \leq \tau^2 \leq 2.2$. Light grey line at 0.95

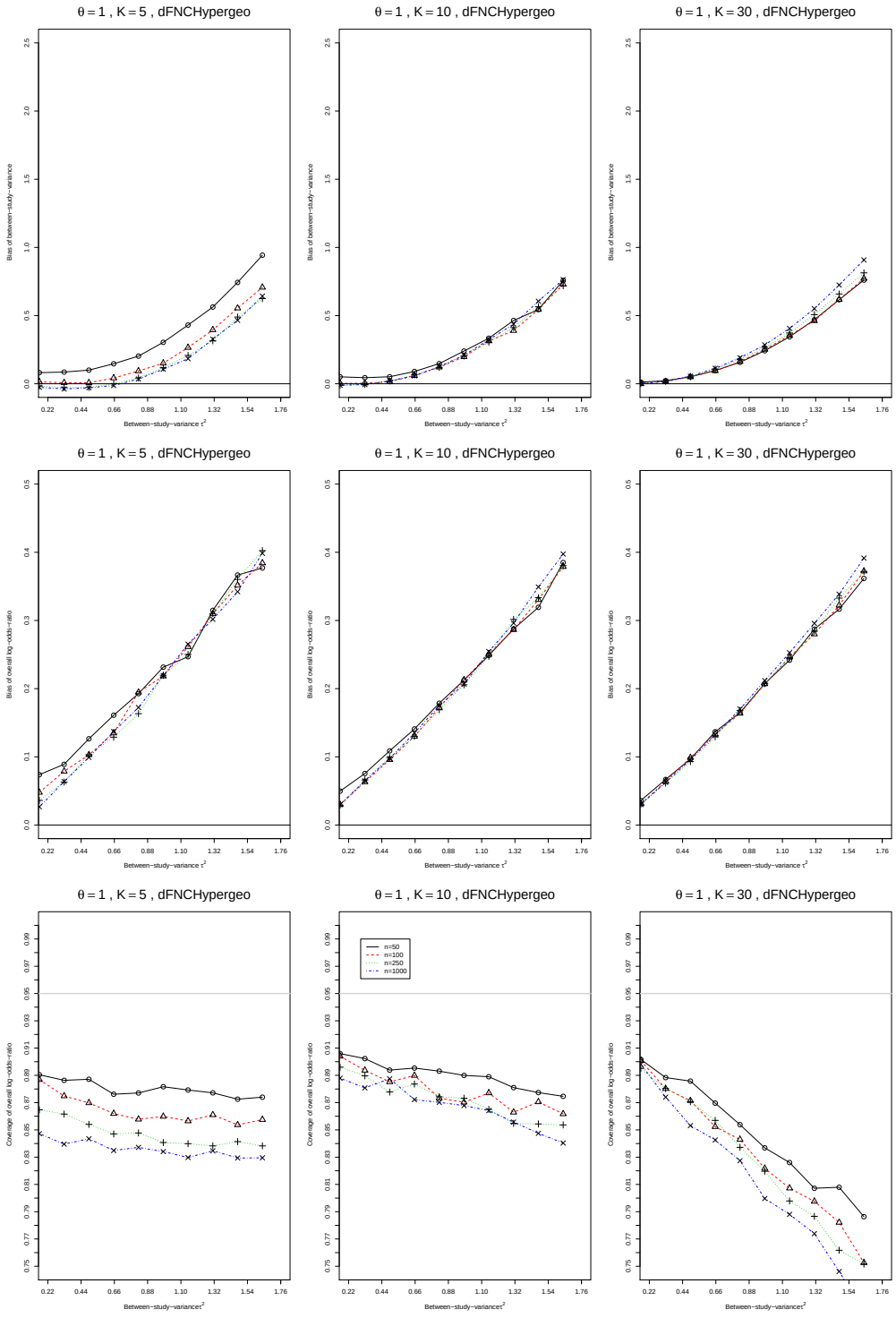


Figure 5.8: Bias of between-study variance τ^2 (first row), bias (second row) and coverage (third row) of overall odds ratio at the nominal 95% level using the non-central hypergeometric-normal in overdispersed random effects model using dFNCHypergeo for $p_{2i} = 0.1$, $\theta = 1$ and $0 \leq \tau^2 \leq 1.55$. Light grey line at 0.95

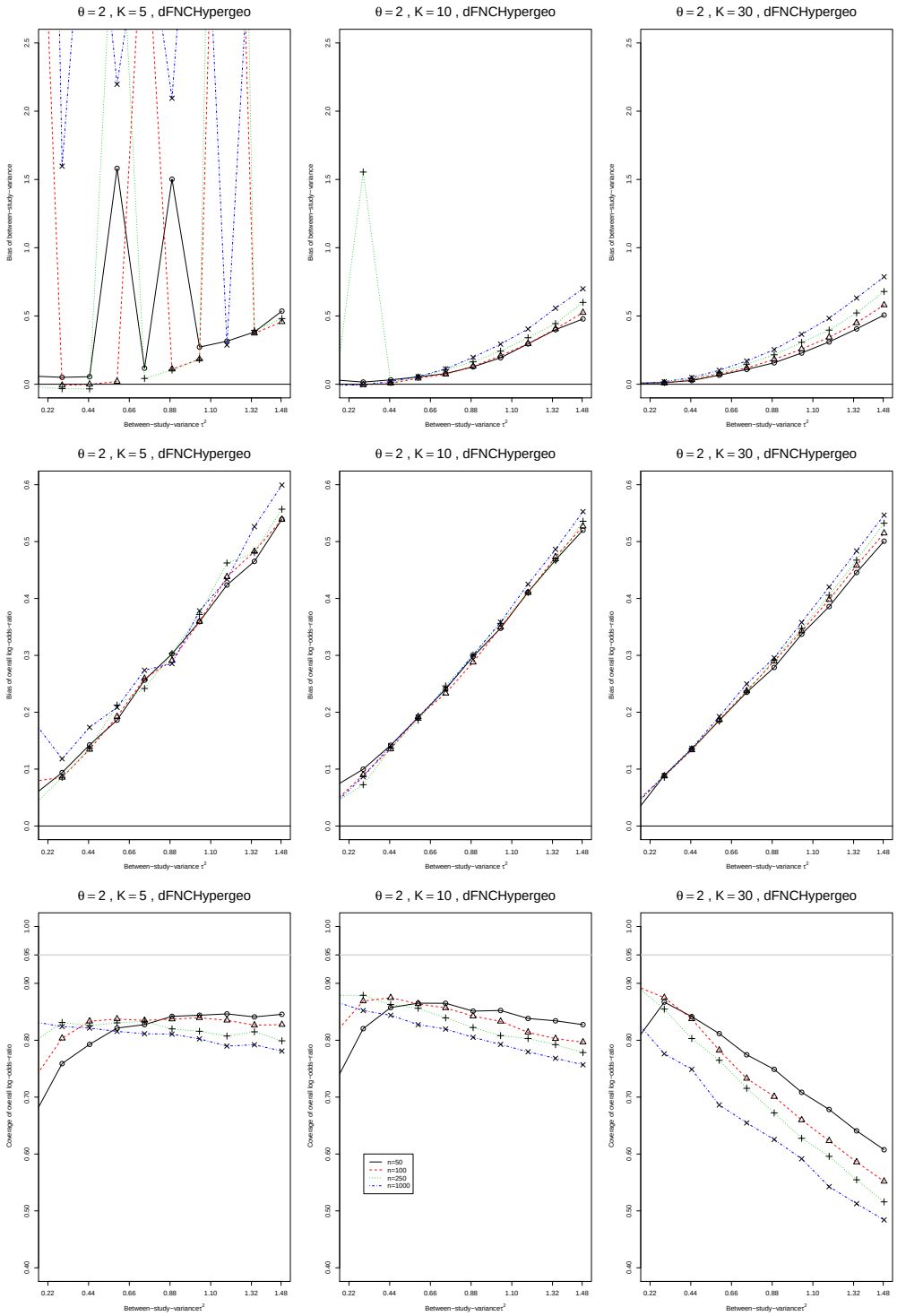


Figure 5.9: Bias of between-study variance τ^2 (first row), bias (second row) and coverage (third row) of overall odds ratio at the nominal 95% level using the non-central hypergeometric-normal in overdispersed random effects model using dFNCHypergeo for $p_{2i} = 0.1$, $\theta = 2$ and $0 \leq \tau^2 \leq 1.48$. Light grey line at 0.95

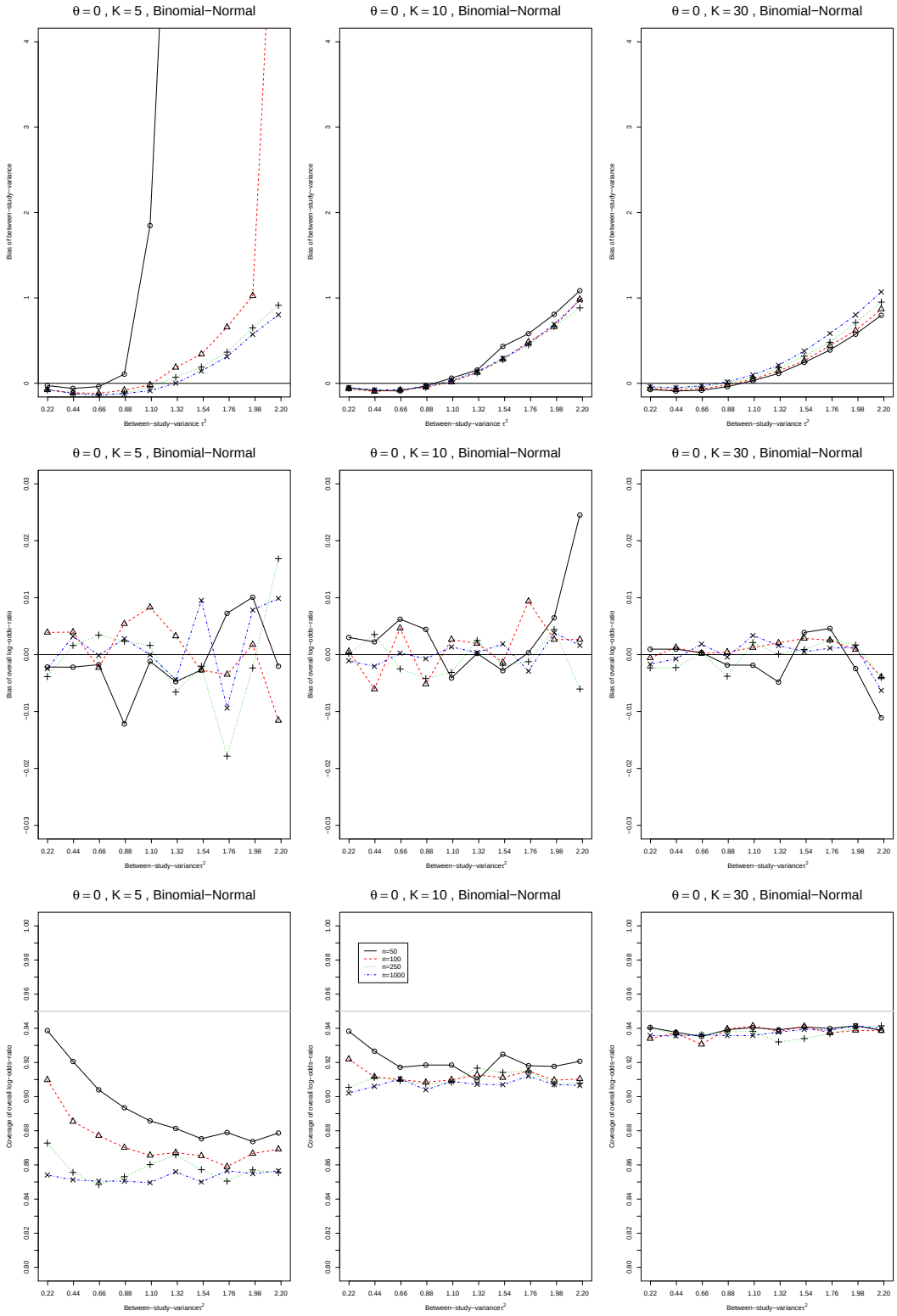


Figure 5.10: Bias of between-study variance (first row), bias (second row) and coverage (third row) of overall odds ratio at the nominal 95% level using the binomial-normal approximation to non-central-hypergemeotric normal model in beta-binomial random effects model for $p_{2i} = 0.1$, $\theta = 0$ and $0 \leq \tau^2 \leq 2.2$. Light grey line at 0.95

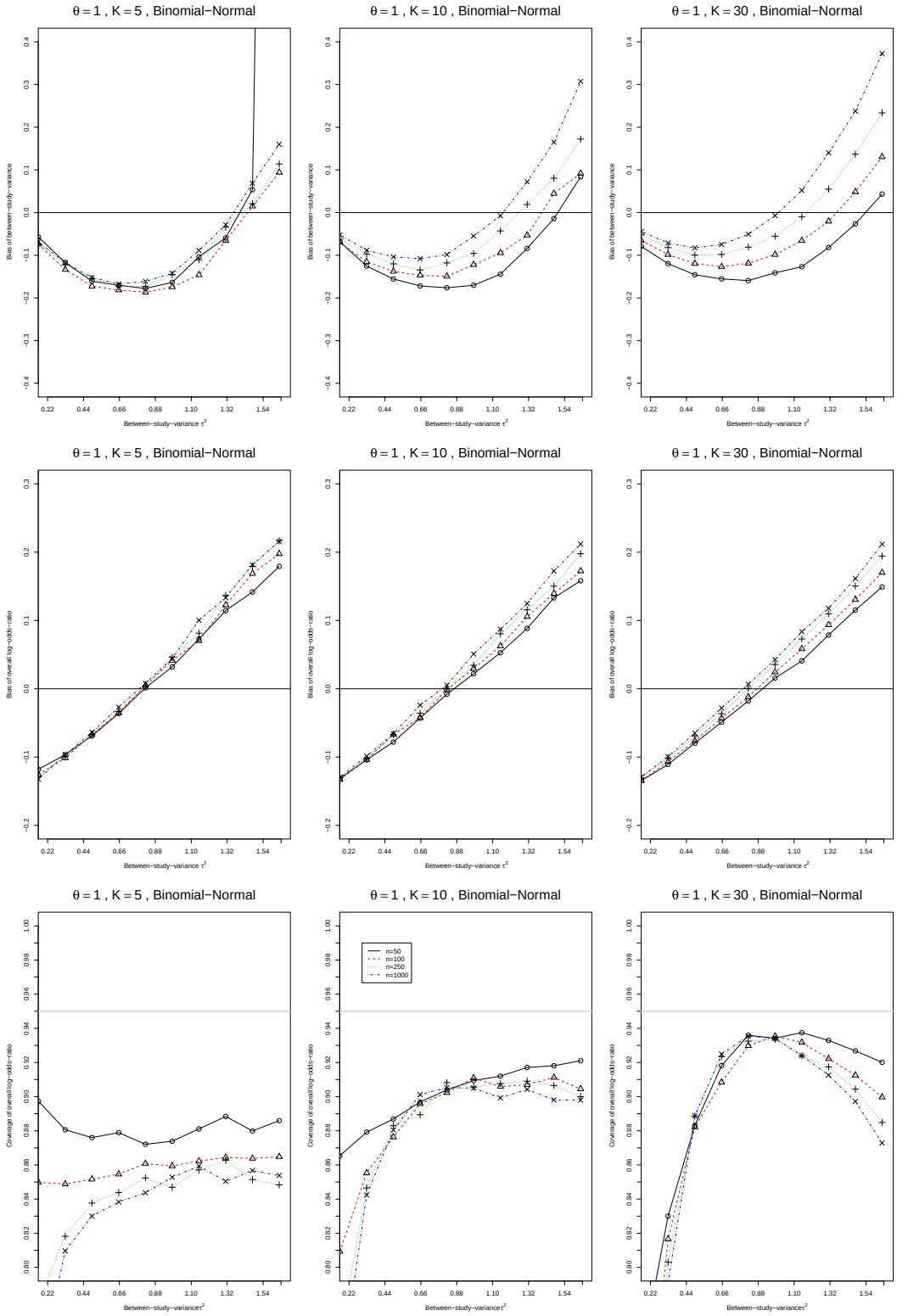


Figure 5.11: Bias of between-study variance (first row), bias (second row) and coverage (third row) of overall odds ratio at the nominal 95% level using the binomial-normal approximation to non-central-hypergemeotric normal model in beta-binomial random effects model for $p_{2i} = 0.1$, $\theta = 1$ and $0 \leq \tau^2 \leq 1.55$. Light grey line at 0.95

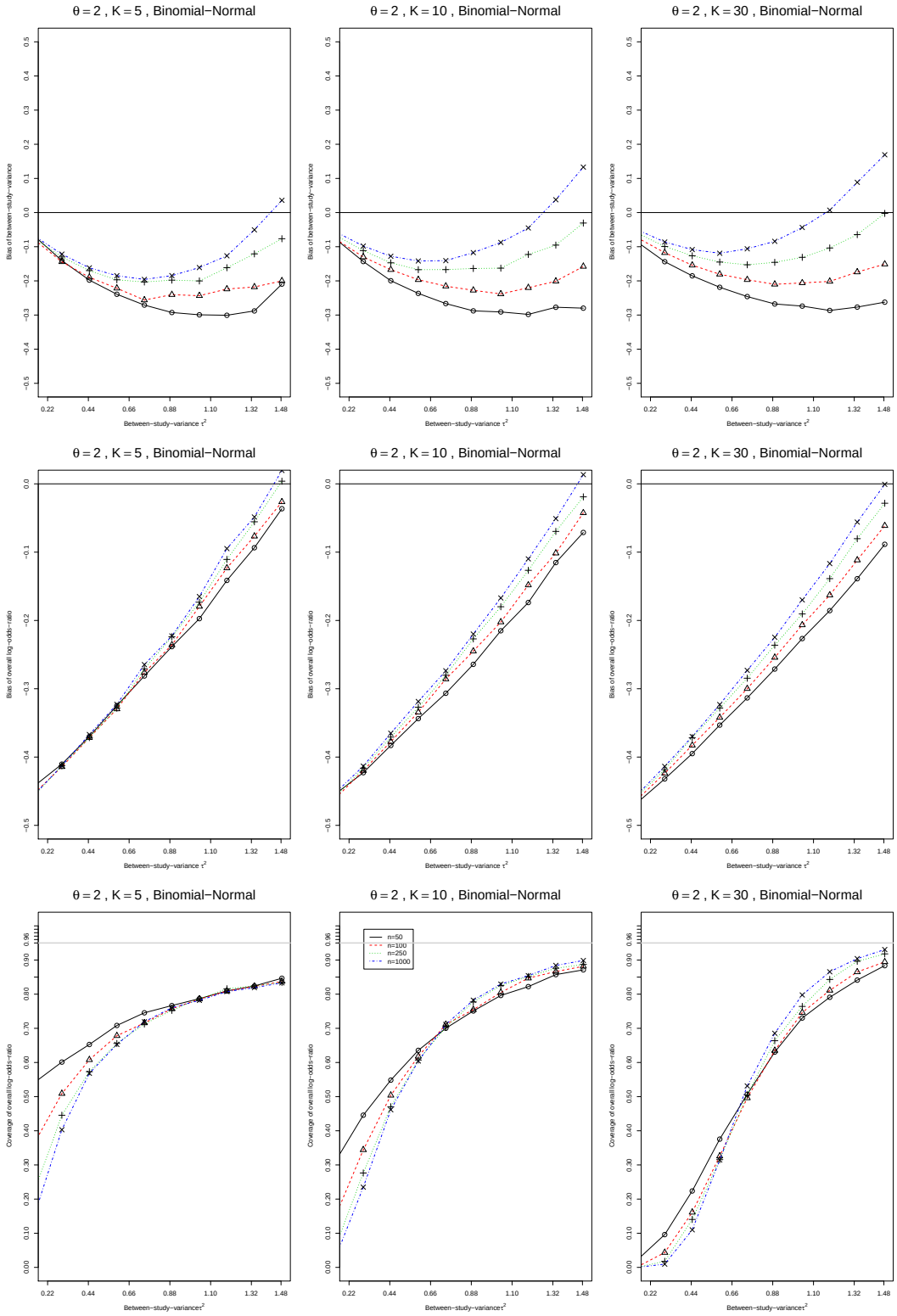


Figure 5.12: Bias of between-study variance (first row), bias (second row) and coverage (third row) of overall odds ratio at the nominal 95% level using the binomial-normal approximation to non-central-hypergemeotric normal model in beta-binomial random effects model $p_{2i} = 0.1$, $\theta = 2$ and $0 \leq \tau^2 \leq 1.48$. Light grey line at 0.95

5.6 Example: effects of diuretics on pre-eclampsia

A meta-analysis of nine-trials on the effect of diuretics on pre-eclampsia (Collins et al., 1985) was studied in section 4.5 of Chapter 4. The same example is re-analysed in this chapter in order to compare the results from standard models with the results from generalized linear mixed effect models. Four classes of generalised models are applied. The first two models are unconditional generalized linear mixed-effects models with fixed and random study effects, Turner et al. (2000). metafor package in R uses a coding of +1/2 and -1/2 for the group dummy for the random effects of unconditional generalized linear mixed-effects models with fixed and random study effects. This is done in order to avoid the problem of lower variance in control group relatively to treatment group when a coding of 0 and 1 used instead. More details can be found in Turner et al. (2000) and Viechtbauer (2015). The second two models are conditional generalized linear mixed-effects models with approximate and exact likelihood (Stijnen et al., 2010). For comparison of the results, we also included the results from standard additive random effects model and multiplicative overdispersed random effects model from Chapter 4. All the results are shown in table 5.1.

The generalized mixed effects models (GLMM) provide similar results for the logarithm of odds ratio (LOR) between -0.513 and -0.516 apart from the model CM.AL where $LOR = -0.434$. The value of estimator for between-study variance, $\hat{\tau}^2_{CM.AL} = 0.165$, in conditional generalized linear mixed-effects model with approximate likelihood is also lower in comparison to unconditional generalized linear mixed-effects model with fixed and random study effects, $\hat{\tau}^2_{UM.FS} = 0.254$ and $\hat{\tau}^2_{UM.FS} = 0.264$, and to conditional generalized linear mixed-effects model with exact likelihood, $\hat{\tau}^2_{CM.EL} = 0.260$. The explanation might be that the binomial approximation to non-central-hypergeometric distribution in conditional generalized linear mixed-effects model is invalid for this data in studies with probabilities $p_{1i} > 0.1$ and $p_{2i} > 0.1$. For example, in studies 3, 4 and 9, the probabilities in both arms are higher than 0.1 (see table 4.1 in Chapter 4). The non-central-hypergeometric normal model produces an estimate of between-study variance $\hat{\tau}^2_{CM.EL} = 0.260$. This estimate is right between the DL and REML

Table 5.1: Estimates and confidence intervals for the ICC ρ, for log odds ratios and for odds ratios diuretics in pre-eclampsia example; GLMM is the generalized linear mixed model, REM is the random effects and BB is the beta-binomial model. Heterogeneity parameters estimated are τ^2 in GLMM, and ρ in BB model. L and U are the lower and upper limits of the respective confidence intervals (CIs).

Model	Method	Hetero geneity	L	U	LOR	L	U	length of CI	OR	L	U
GLMM	UM.FS	0.254			-0.513	-0.923	-0.104	0.819	0.599	0.398	0.901
GLMM	UM.RS	0.264			-0.516	-0.930	-0.102	0.828	0.597	0.395	0.903
GLMM	CM.AL	0.165			-0.434	-0.777	-0.091	0.686	0.648	0.460	0.913
GLMM	CM.EL	0.260	-0.147(0)	0.667	-0.513	-0.927	-0.100	0.827	0.599	0.396	0.905
FEM		0.000			-0.398	-0.573	-0.223	0.530	0.672	0.564	0.800
REM	DL	0.230	0.072	2.202	-0.517	-0.916	-0.117	0.799	0.596	0.400	0.889
REM	REML	0.300	0.043	1.475	-0.518	-0.956	-0.080	0.876	0.596	0.384	0.923
BB	M&IV	0.008	0.002	0.095	-0.436	-0.792	-0.080	0.712	0.647	0.453	0.923
	M&MH				-0.427	-0.775	-0.080	0.695	0.652	0.461	0.923
BB	REML&IV	0.010	0.001	0.060	-0.447	-0.835	-0.059	0.776	0.640	0.434	0.942
	REML&MH				-0.431	-0.809	-0.053	0.756	0.650	0.445	0.949
BB	MP&IV	0.017	0.002	0.095	-0.469	-0.920	-0.018	0.902	0.626	0.399	0.982
	MP&MH				-0.459	-0.898	-0.020	0.879	0.632	0.407	0.981
BB	CMP&IV	0.018	0.003	0.094	-0.474	-0.942	-0.007	0.936	0.623	0.390	0.993
	CMP&MH				-0.472	-0.927	-0.016	0.911	0.624	0.396	0.984
BB	BD&IV	0.019	0.003	0.107	-0.475	-0.944	-0.006	0.938	0.622	0.389	0.994
	BD&MH				-0.463	-0.920	-0.021	0.899	0.630	0.399	0.980

estimator of between-study variance in standard random effects model. The widest confidence intervals among GLMM models of widths 0.827 and 0.828 are from the unconditional generalized linear mixed-effects model with random study effects (UM.RS) and conditional generalized linear mixed-effects model with exact likelihood (CM.EL) respectively. These intervals are still shorter than the interval for the inverse-variance odds ratio obtained from beta-binomial model with ρ_{CMP} and ρ_{BD}. Among all methods, the inverse-variance method with ρ_{BD} provides the widest confidence interval. Among all the estimates of between-study variance Viechtbauer (2005) recommend REML as the most unbiased and efficient estimate of τ^2. The use of REML estimate of τ^2 in standard additive random effects model is well-known. However, Turner et al. (2000) has analysed the current example and showed that $\hat{\tau}^2_{REML}$ is biased downwards.

In order to compare the estimates of τ^2 and θ obtained by various methods, we derive the likely true values of τ^2 and θ using the biases of $\hat{\tau}^2_{CM.EL}$ from simulations of standard REM in table 5.2 and multiplicative ODM in table 5.3. These biases correspond to the case $K = 10$ and true $\tau^2 = 0.1$, $\tau^2 = 0.3$ for REM and $\rho = 0.01$, $\rho = 0.02$ for ODM. We consider this particular case since we have only nine studies and small values of τ^2 in the range of $0.165 - 0.300$ and ρ in the range of $0.008 - 0.019$ in this example.

Table 5.2: Bias of $\hat{\tau}^2_{CM.EL}$ from simulation of REM for $K = 10$ and $\theta = 0$

	Sample size			
τ^2	50	100	250	1000
0.1	0.069289371	0.004323149	-0.006701096	-0.011848884
0.3	0.05059464	-0.02737167	-0.02787429	-0.02933790

Table 5.3: Bias of $\hat{\tau}^2_{CM.EL}$ from simulation of ODM for $K = 10$ and $\theta = 0$

	Sample size				
ρ	τ^2	50	100	250	1000
0.01	0.22	0.072288987	0.000957377	-0.009543923	-0.011582204
0.02	0.44	0.093838612	0.024553661	0.008006127	0.005678536

Thus, the likely true values of between-study variance derived from simulation results of standard REM and multiplicative ODM with $\hat{\tau}^2_{CM.EL} = 0.26$ are shown in the table 5.4 and 5.5 respectively.

Table 5.4: Likely true values of τ^2 derived from simulation of REM for $K = 10$ and $\theta = 0$

	Sample size			
τ^2	50	100	250	1000
0.1	0.1907106	0.2556769	0.2667011	0.2718489
0.3	0.2094054	0.2873717	0.2878743	0.2893379

Table 5.5: Likely true values of τ^2 derived from simulation of ODM for $K = 10$ and $\theta = 0$

		Sample size			
ρ	τ^2	50	100	250	1000
0.01	0.22	0.187711	0.2590426	0.2695439	0.2715822
0.02	0.44	0.1661614	0.2354463	0.2519939	0.2543215

From the table 5.4 and 5.5, the value of true between-study variance is not clear. The results of two models are quire similar for $\tau^2 = 0.1$ in REM and $\rho = 0.01$ in ODM. This might be because $\tau^2 = 0.1$ correspond to $\rho = 0.01$ in relationship 6.2.7. The bias in τ^2 depends on the sample size. The average sample sizes of nine studies are $n_{1i} = 418$ and $n_{2i} = 354$. Thus we would concentrate on the large values of N in simulations such as $N = 250$ and $N = 1000$. For $\theta = 0$ and small amount of heterogeneity, the results of simulations have shown that inference based on a conditional generalized linear mixed-effects model with exact likelihood is not that bad. Thus, from the conditional generalized linear mixed-effects model with exact likelihood, the true value of τ^2 is around $0.2667011 - 0.2718489$ (from data generated with REM) and $0.2695439 - 0.2715822$ (from data generated with ODM). The problems with conditional generalized linear mixed-effects model with exact likelihood might appear in case of moderate to large amount of heterogeneity across the studies. We have shown by simulation study that the NCHGM might be misleading in case of presence of intra-cluster correlation within studies. However, in this particular example the value of intra-cluster correlation is very small. Hence the conditional generalized linear mixed-effects model resulted in similar results in both REM and ODM models. In this particular example, the bias of $\hat{\tau}^2_{CM.EL} = 0.26$ is not that large and

we would still believe this estimator. The likely true values of θ from simulations of REM and ODM model is as following. These biases of θ from simulations of standard REM and multiplicative ODM are shown in tables 5.6 and 5.7.

Table 5.6: Bias of $\hat{\theta}_{CM.EL}$ from simulation of REM for $K = 10$ and $\theta = 0$

	Sample size			
τ^2	50	100	250	1000
0.1	0.028087437	0.021165542	0.018814243	0.006114903
0.3	0.063890995	0.052922304	0.027021773	0.012555648

Table 5.7: Bias of $\hat{\theta}_{CM.EL}$ from simulation of ODM for $K = 10$ and $\theta = 0$

	Sample size				
ρ	τ^2	50	100	250	1000
0.01	0.22	-0.001349553	-0.001946856	-0.00109181	-0.000905712
0.02	0.44	-0.006715632	-0.001437701	-0.0019664942	-0.00075448

Thus, the likely true values of overall odds ratio derived from simulation results of standard REM and multiplicative ODM with $\hat{\tau}^2_{CM.EL} = 0.26$ are shown in the table 5.8 and 5.9 respectively.

Table 5.8: Likely true values of log-odds ratio θ derived from simulation of REM for $K = 10$ and $\theta = 0$

	Sample size			
τ^2	50	100	250	1000
0.1	-0.5410874	-0.5341655	-0.5318142	-0.5191149
0.3	-0.576891	-0.5659223	-0.5400218	-0.5255556

Table 5.9: Likely true values of odds ratio θ derived from simulation of ODM for $K = 10$ and $\theta = 0$

	Sample size				
	τ^2	50	100	250	1000
0.01	0.22	-0.5116504	-0.5110531	-0.5119082	-0.5120943
0.02	0.44	-0.5062844	-0.5115623	-0.5110335	-0.5122455

Again from the table 5.8 and 5.9, similar to the between-study variance, it is not clear the true value of overall odds ratio. The true log-odds-ratio vary between -0.5410874 and -0.5191149 from REM and between -0.5116504 and -0.5120943 from ODM. In Chapter 4, we recommended the estimated ICC $\hat{\rho}_{BD} = 0.019$ and corresponding value of the pooled LOR

$\hat{\theta}_{IV} = -0.475$ with confidence interval $(-0.944,\ -0.006)$. Comparing the results from Chapter 4, our estimator $\hat{\theta}_{IV} = -0.475$ from ODM is close to likely values of true log-odds ratio, however the difference between $\hat{\theta}_{CM.EL} = -0.513$ and $\hat{\theta}_{IV}(\rho) = -0.475$ still exist. Figure 4.1 in Chapter 4 shows that the Breslow-Day method provides less biased estimates of intra-cluster correlation. The absolute bias of ρ_{BD} is much smaller than the bias of $\hat{\tau}^2_{CM.EL}$. Particularly, the bias of ρ_{BD} is minimum for small to moderate values of intra-cluster correlation. Also, Figure 4.3 shows that when for $K = 10$, the bias of overall log-odds ratio almost does not exist when $\theta = 0$ and around $0.05 - 0.1$ when $\theta = 1$. Thus, obtaining the likely true values of odds ratio from ODM, then true value of odds ratio is between $\psi = 0.672$ and $\psi = 0.722$. Even though, in this particular example, a non-central hypergeometric normal model provided similar likely true values between-study variance and overall odds ratio from REM and ODM, there is difference in likely true values of odds ratio from NCHGN and ODM. Which model to believe still remain an open question.

5.7 Summary

In this chapter, we examined by simulations the performance of a conditional generalized linear mixed effects model with exact and approximate likelihood. Both models were applied to data simulated from two different scenarios. The first scenario is a pair of binomially distributed random variables within each study with normally distributed logarithm of odds ratio across studies. This case corresponds to the standard additive random effects model. In random effects model, the overdispersion is introduced through the between-study variance. The second scenario is a pair of beta-binomially distributed random variables within each study. This is a two stage model. The events are assumed to have binomial distributions within each study and the probabilities of events in each arm are assumed to be beta-distributed across K studies. The same model can be obtained by assuming the within study dependence of Bernoulli variables within each arm. The overdispersion is introduced through intra-cluster correction within each arm.

In metafor R package, there exist two methods for fitting the conditional generalized linear mixed effects model with exact likelihood. The first method used by default is using the density function dFNCHypergeo from the BiasedUrn package. The second method is using the density function dnoncenhypergeom from the MCMC package. We examined the stability and performance of dFNCHypergeo or dnoncenhypergeom for estimation of an overall effect measure and between-study variance by simulation study. Both methods perform more or less similarly. Some of our findings depend on numerical issues and not the properties of the conditional generalized linear mixed effects model with exact likelihood. We have also examined the conditional generalized linear mixed effects model with approximate likelihood, which may be used when the probabilities are low due to approximation of non-central-hypergeometric distribution with binomial distribution. This method has shown lower performance than conditional generalized linear mixed effects model with exact likelihood

For beta-binomial model, the conditional generalized linear mixed effects model with exact likelihood provides biased estimates of θ and τ^2 which results in wrong confidence intervals of

θ. Particularly, the coverage deteriorates when the log-odds ratio moves far away from zero. The explanation of the lower coverage and biased estimate of overall effect may be due to the bias of sample log odds ratio under beta-binomial model when the probabilities of two arms are not equal (see Chapter 4).

For standard additive random effects model, the maximum likelihood estimate of τ^2 and θ are also biased. This bias results in lower coverage than the nominal confidence level of 0.95. The bias in non-central-hypergeometric-normal model might be the result of the logit transformation used in non-central-hypergeometric distribution within each study, since the logit transformation has a bias of order 1/n.

From the results of simulations in this current chapter and Chapter 4, when the binary data is sparse it is difficult to propose a universal method for estimating the between-study variance and overall effect measure. There seems to appear a misspecification problems of models. The problem lies in differentiating between the models (standard REM and multiplicative ODM) and their assumptions. In real situation, it is impossible to distinguish if there is a between study heterogeneity across the studies or intra-cluster dependence within each study. The former might be the latter assumption and visa versa. According to Stijnen et al. (2010), generalized linear mixed effects model might provide insights to these problems. Particularly, a conditional generalized linear mixed-effects model with an exact non-central hypergeometric likelihood is suggested as an alternative to standard random effects model in order to avoid the estimated within study variances and continuity corrections in case of sparse data. Nevertheless, we showed that the estimates of parameters from conditional generalized linear mixed-effects model with an exact non-central hypergeometric likelihood are biased, hence the confidence interval of overall effect measure is too narrow. The bias is larger when the binary data is correlated. This correlation is modelled through beta-binomial model. In case of existence of correlation, the inference from a conditional generalized linear mixed-effects model with an exact non-central hypergeometric likelihood is misleading. We believe that the bias in a conditional generalized linear mixed-effects model with an exact non-central hypergeometric

likelihood is the result of transformation. In Chapter 3, we have shown that the transformation bias is present in standard random effects model and overdispersed multiplicative random effects model for log odds. Nemes et al. (2009) shows that logistic regression overestimates the odds ratio due to the bias of order $1/n$ in studies with small and moderate sample sizes. Kosmidis et al. (2017) studies the bias of order $1/n$ in maximum likelihood estimates of overall effect measure and between-study variance under random effects model.

A diagnostic or a method robust to misspecification of dependency structure of Bernoulli variables is required. For differences between beta-binomial and logistic-normal model, Williams (1982) suggests to fit the beta-binomial model and plot standardized residuals against fitted values. If plot indicates that the variance of these residuals decreases markedly as the fitted value approaches zero or one, then standard random effects model may be more appropriate. Another graphical statistic is proposed by Hinde and Demétrio (1998), who consider the plot of half-normal scores against deviance residuals to choose between beta-binomial and logistic normal models. All these and other references for diagnostics are well summarized in Hinde and Demétrio (1998). In regression analysis for binary data, the plot of standardized or deviance residuals against fitted value are particularly useful in checking the model adequacy, detecting outlier or unusual observations. However, the use of this plot to choose between models maybe not the best option. According to Gelman and Hill (2006), it is generally difficult to differentiate between beta-binomial and logistic-normal models. Referencing the paper by Williams (1982), Ganio-Gibbons (1989) suggests that it is only possible to see the difference between models if there are large number of observations with fitted values close to zero or one. Breslow and Clayton (1993) warn that "when the probabilities are small and the data are highly discrete, only limited information is present for estimating the random effects". Hanfelt and Liang (1998) discuss the sparseness of dependent binary data. Hanfelt and Liang (1998) say that for the asymptotic situation when the number of studies is increasing, the maximum likelihood theory leads to inconsistent maximum likelihood estimation for odds ratio regression models. For the dependent binary data, Hanfelt and Liang (1998) do not

recommend the use of likelihood based methods, since these methods are not robust against model specification.

The misspecification of REM is an important issue in meta-analysis of sparse data. How to safeguard against misspecification of REM and which method to use in meta-analysis of sparse data is an open question.

Chapter 6

Comparison of standard and new methods for estimation of random effect component from REM and ODM

6.1 Introduction

In Chapter 5, we have shown that the inference based on conditional generalized linear mixed effects model with exact and approximate likelihood for sparse data might be misleading due to the negative bias in between-study variance and positive bias in overall effect measure. The biases of between-study variance and overall effect measure are not large, but still exist. These biases are the result of transformation from odds ratio to log odds ratio scale, that the true effect measures undergo when they are integrated out across the studies. The bias of between-study variance and overall effect measure are of order $O(1/N)$. These biases are similar to the transformation biases studied in Chapter 3. Similar biases due to transformation might influence the inference in meta-analysis based on methods from standard random effects model (REM).

In Chapter 5, we have also shown the problem of misspecification between two models in meta-analysis. Two main models correspond to standard additive REM and multiplicative ODM. In standard additive REM, we have a pair of binomially distributed random variables within each study with normally distributed logarithm of odds ratio across studies. In multiplicative ODM, binomially-normally distributed random variables from REM are replaced

by a pair of beta-binomially distributed random variables within each study. Two models have different assumptions. In each model, the requisite parameters need to be estimated. In standard REM, the parameters of interest are between-study variance and overall effect measure. In multiplicative ODM, the parameters of interest are intra-cluster correlation and overall effect measure. In reality, even when the raw data is available, it is very difficult to distinguish between these two models. This is a problem of misspecification of random effects models in meta-analysis.

In the current Chapter, we study how the problems of transformation bias and misspecification discussed in Chapter 5 affect methods of estimation of random effect component and overall effect measure in meta-analysis either in standard REM (Chapter 2) or in multiplicative ODM (Chapter 4). In addition to standard methods for estimation of between-study variance and overall effect measure discussed in Chapter 2, we also study novel Profiled Breslow-Day method introduced in Chapter 4, Corrected Mantel-Paule method with gamma approximation from Chapter 4 and a new recently developed method based on penalization of likelihood for estimation of between-study variance (Kosmidis et al. (2017)).

The structure of this Chapter is as follows. Section 6.2 overviews standard REM and multiplicative ODM. In the same section, we also provide the correspondence between standard REM and multiplicative ODM. In section 6.3, we provide the brief overview of different estimators of between-study variance in meta-analysis. Section 6.4 provides a simulation study for standard REM. In section 6.5, we assess the methods for estimation of between-study variance and corresponding inverse-variance effect measures with data simulated from a pair of beta-binomial distributions. The example of meta-analysis from Chapters 4 and 5 is restudied in section 6.6 with estimation of τ^2 from all the possible methods. Section 6.7 summarizes the findings of the current Chapter. This Chapter represents the novel work of this thesis

6.2 Random effects models for meta-analysis of LOR

6.2.1 Standard random effects model

In meta-analysis, the standard random effects model assumes that within and between study variability is accounted for through an approximately normal distribution within and between study effects, i.e

$$\hat{\theta}_i \sim N(\theta_i, \sigma_i^2) \quad \text{and} \quad \theta_i \sim N(\theta, \tau^2), \tag{6.2.1}$$

resulting in a marginal distribution of estimated effect measures $\hat{\theta}_i \sim N(\theta, \hat{\sigma}_i^2 + \tau^2)$. $\hat{\theta}_i$ are the estimates of effect measures, and its within-study variances for each study i are estimated by $\hat{\sigma}_i^2$, $i = 1, \ldots, K$. τ^2 represents an unknown variance for between-study variance. When τ^2 is estimated, the overall estimate of an effect measure can be estimated by the weighted mean

$$\hat{\theta}_{RE} = \frac{\sum_{i=1}^{K} \hat{w}_i \hat{\theta}_i}{\sum_{i=1}^{K} \hat{w}_i}, \tag{6.2.2}$$

where weights $\hat{w}_i = \frac{1}{\hat{\text{Var}}(\hat{\theta}_i)}$ and

$$\hat{\text{Var}}(\hat{\theta}_i) = \frac{1}{n_{i1}\hat{p}_{i1}(1-\hat{p}_{i1})} + \frac{1}{n_{i2}\hat{p}_{i2}(1-\hat{p}_{i2})} + \hat{\tau}^2 \tag{6.2.3}$$

where $\hat{\tau}^2$ is an estimator of τ^2, $\hat{p}_{i1}$ and $\hat{p}_{i2}$ are estimators of probabilities in treatment and control arm.

6.2.2 Overdispersed random effects model

In Chapter 4, we have introduced an overdispersed beta-binomial random effects model (ODM) for ORs. ODM for ORs is modelling the binomial numbers of events X_{1i} and X_{2i} rather than the logarithmic transformation of ORs. In ODM, the variability is modelled through a pair of independent beta-binomial distributions. Assuming normality across K studies, the overdispersed random effects model with a pair of beta-binomial distributions is

$$\hat{\theta}_i \sim N(\theta, \frac{v_i(R_i)}{n_i}\phi_i), \tag{6.2.4}$$

where $\phi_i = (1 + a_i\rho)$ and a_i is given by

$$a_i = a(n_i, R_i, p_{i1}, p_{i2}) = n_i v_i(R_i)^{-1} \left[\frac{1}{p_{i1}(1-p_{i1})} + \frac{1}{p_{i2}(1-p_{i2})}\right] - 1.$$

a_i is a linear function of n_i and has the same order as n_i. Reparametrising a_i as a function of the control arm probability p_{i2} and the odds ratio ψ_i, a_i can be written as

$$a_i = \frac{n_i R_i[(1 - p_{i2}(1-\psi_i))^2 + \psi_i]}{(R_i+1)[(1-p_{i2}(1-\psi_i))^2 + R_i\psi_i]} - 1.$$

For balanced studies $R_i = 1$, and a_i simplifies to $a_i = n_i/2 - 1$.

The inverse-variance method for overdispersed random effects model is

$$\hat{\theta}_{ODM} = \frac{\sum_{i=1}^{K} \hat{w}_i\hat{\theta}_i}{\sum_{i=1}^{K} \hat{w}_i} \qquad \text{with weights} \qquad \hat{w}_i = \frac{1}{\hat{\sigma}_i^2(1 + a_i\rho)}. \tag{6.2.5}$$

where

$$\hat{\sigma}_i^2(1 + a_i\rho) = \mathrm{Var}(\log(\hat{\psi}_i)) = \frac{1 + (n_{i1} - 1)\rho}{n_{i1}p_{i1}(1-p_{i1})} + \frac{1 + (n_{i2} - 1)\rho}{n_{i2}p_{i2}(1-p_{i2})}. \tag{6.2.6}$$

ρ is an unknown parameter for intra-cluster correlation that has to be estimated.

6.2.3 Correspondence between ρ in ODM and τ^2 in REM

We have two versions of random effects model. One is the standard additive REM and another is the multiplicative ODM. The heterogeneity in additive random effects model (6.2.2) is explained by additional variance component of random effects τ^2. In the multiplicative ODM, we explain overdispersion by common intra-cluster correlation ρ. In both cases, we have some additional component in the total variance for $\log(\hat{\psi}_i)$. Comparing the variance (6.2.3) in standard REM (6.2.2) and variance (6.2.6) in multiplicative ODM (6.2.4), there is a monotonic relationship between τ^2 and ρ given by

$$\tau_i^2 = [\frac{(n_{1i} - 1)}{n_{1i}p_{1i}(1-p_{1i})} + \frac{(n_{2i} - 1)}{n_{2i}p_{2i}(1-p_{2i})}]\rho_i. \tag{6.2.7}$$

The relationship between the intra-cluster correlation ρ and the between-study variance τ^2 depends on probabilities and sample sizes of positive response p_{1i} and p_{2i} for treatment and

control arms respectively. Two models, REM and ODM, are largely equivalent when τ^2 or ρ are constant across studies.

Assuming balanced studies, i.e equal sample sizes $n_{1i} = n_{2i} = n_i^*$, the correspondence above becomes

$$\tau_i^2 = \big(\frac{n_i^* - 1}{n_i^*}\big)[\frac{1}{p_{1i}(1-p_{1i})} + \frac{1}{p_{2i}(1-p_{2i})}]\rho_i$$

or in terms of within study variance (6.2.3)

$$\tau_i^2 = \big(n_i^* - 1\big)\sigma_i^2\rho_i$$

where

$$\sigma_i^2 = \frac{1}{n_{1i}p_{1i}(1-p_{1i})} + \frac{1}{n_{2i}p_{2i}(1-p_{2i})}$$

In the case when the probabilities for treatment and control arms are constant, i.e. $p_{1i} = p_1$ and $p_{2i} = p_2$ and $n_{1i} = n_{2i} = n$, then between-study variances τ_i^2 do not depend on i, i.e. $\tau_i^2 = \tau^2$.

6.3 Estimators of between-study variance

There exist numerous estimators of between-study variance in meta-analysis. Some estimators are based on method of moments and some are likelihood based estimators. The literature review of all estimators is given in detail in Chapter 2. In this section, we overview the most popular and new estimators of between-study variance. Two new estimators are based on the idea similar to estimation of intra-cluster correlation parameter in Chapter 4. One more new estimator is the recent development on penalization of likelihood function (Kosmidis et al. (2017)). All these estimators will be used in simulation study in section 6.4.

6.3.1 Der-Simonian and Laird estimator of τ^2

Under random effects model, the Cochran's Q-statistic has approximately chi-square distribution χ^2_{K-1}. The Der-Simonian and Laird estimator of τ^2 is based on Cochran's Q statistic.

The Der-Simonian and Laird is a method of moment estimator introduced by DerSimonian and Laird (1986). It is calculated as

$$\hat{\tau}^2_{DL} = \max\left([Q - K + 1] \Big/ \left[\sum_{i=1}^{K} w_i - \frac{\sum_{i=1}^{K} w_i^2}{\sum_{i=1}^{K} w_i} \right], 0 \right). \tag{6.3.1}$$

6.3.2 Mandel-Paule estimator of τ^2

The Mandel-Paule estimator $\hat{\tau}^2_{MP}$ is another moment based estimator similar to DerSimonian and Laird estimator of between-study variance in meta-analysis. The Mandel-Paule estimator $\hat{\tau}^2_{MP}$ is based on equating Cochran's Q statistic $Q(\tau^2)$ to the first moment of its chi-square distribution χ^2_{K-1} given that a solution exists. It is obtained by solving iteratively the equation

$$Q(\tau^2) = \sum_{i=1}^{K} \frac{(\theta_i - \hat{\theta}_{RE})^2}{\sigma_i^2 + \tau^2} = K - 1. \tag{6.3.2}$$

The Q profiled confidence interval can be estimated from lower and upper quantiles of χ^2_{K-1} distribution

$$Q(\tau^2_L) = \chi^2_{K-1;0.975} \qquad Q(\tau^2_U) = \chi^2_{K-1;0.025} \tag{6.3.3}$$

The upper and lower bounds for τ^2 can be calculated iteratively.

6.3.3 Corrected Q-statistic based estimation of τ^2

As described in (4.4.4) of Chapter 4, the corrected Q-statistic can be used for estimation of between-study variance τ^2 in the standard additive random effects model. Following Kulinskaya and Dollinger (2015), the distribution of Q statistic can be well approximated by a family of gamma distributions with shape and scale parameters

$$r(\tau^2) = \frac{\mathrm{E}(Q)^2}{\mathrm{Var}(Q)} \quad \text{and} \quad \lambda(\tau^2) = \frac{\mathrm{Var}(Q)}{\mathrm{E}(Q)}.$$

The expected value and variance of Q are obtained from the equations

$$(K-1) - \mathrm{E}(Q) = 0.678[(K-1) - \mathrm{E}_{th}(Q)] \tag{6.3.4}$$

and

$$\text{Var}(Q) = 4.74(K-1) - 12.17\text{E}[Q] + 9.42\text{E}[Q]^2/(K-1), \tag{6.3.5}$$

where $\text{E}_{th}(Q)$ is the theoretical approximation to the mean of Q for log odds ratio (Kulinskaya and Dollinger, 2015). Based on gamma approximation of the Q statistics, the Corrected Mandel-Paule estimate of τ^2 is obtained from

$$Q^*(\tau^2) = E(Q) \tag{6.3.6}$$

given that a solution exist, where $E(Q)$ is the solution of equation (6.3.4).

The related confidence interval based on gamma approximation to the distribution of Q statistic can be obtained from

$$\left\{\Gamma_{r(\tau^2),\lambda(\tau^2);\alpha/2} \leq Q^*(\tau^2) \leq \Gamma_{r(\tau^2),\lambda(\tau^2);1-\alpha/2}\right\}, \tag{6.3.7}$$

where $\Gamma_{r(\tau^2),\lambda(\tau^2)}$ is the quantiles of gamma distribution with $r(\tau^2)$ and $\lambda(\tau^2)$ as shape and scale parameters.

6.3.4 Maximum Likelihood estimator of τ^2

Based on assumption that each effect measure $\hat{\theta}_i$ has a marginal normal distribution $N(\theta, \hat{\sigma}_i + \tau^2)$, the maximum likelihood estimator $\hat{\tau}^2_{ML}$ is obtained by maximising the log-likelihood function

$$l(\theta, \tau^2) \approx -\frac{1}{2}\sum_{i=1}^{K}\log(\sigma_i^2 + \tau^2) - \frac{1}{2}\sum_{i=1}^{K}\frac{(\theta_i - \theta)^2}{\sigma_i^2 + \tau^2}, \tag{6.3.8}$$

ignoring constant terms in the log-likelihood. The maximum likehood estimator of τ^2 is

$$\hat{\tau}^2_{ML} = \frac{\sum_{i=1}^{K} w_i^2[(\hat{\theta}_i - \hat{\theta}_{ML})^2 - \sigma_i^2]}{\sum_{i=1}^{K} w_i^2} \tag{6.3.9}$$

Viechtbauer (2005) has also shown that between-study variance estimates based on ML method are biased.

A 95 per cent confidence interval for $\hat{\tau}^2_{ML}$ is given by set of values which satisfy

$$l_R(\hat{\tau}^2) > l_R(\hat{\tau}^2_{ML}) - \frac{1}{2}C_{0.95}(\chi^2_1) \tag{6.3.10}$$

where $C_{0.95}(\chi_1^2)$ is the 0.95 quantile of the χ_1^2 distribution and l_R is the likelihood ratio test statistic. The distribution

$$-2\log(\frac{l_R(\tau^2)}{l_R(\tau^2_{ML})}) \to \chi_1^2 \qquad \text{for} \qquad K \to \infty$$

where $l_R(\tau^2)$ is the maximum likelihood function calculated at the τ^2 and τ^2_{ML}. The profile likelihood confidence interval might not be centred at τ^2_{ML} due to absence of symmetry. The profile likelihood confidence interval is proposed by Hardy and Thompson (1998). The confidence interval based on $\hat{\tau}^2_{ML}$ can be estimated from the likelihood (6.3.8).

6.3.5 Restricted maximum likelihood estimator of τ^2

Based on assumption that each effect measure $\hat{\theta}_i$ has a marginal normal distribution $N(\theta, \hat{\sigma}_i + \tau^2)$, the restricted maximum likelihood estimator $\hat{\tau}^2_{REML}$ is obtained by maximising the log-likelihood function

$$l(\theta, \tau^2) = -\frac{1}{2}\sum_{i=1}^{K}\log(\sigma_i^2 + \tau^2) - \frac{1}{2}\sum_{i=1}^{K}\frac{(\theta_i - \theta)^2}{\sigma_i^2 + \tau^2} - \frac{1}{2}\log(\sum_{i=1}^{K}(\sigma_i^2 + \tau^2)^{-1})$$

ignoring constant terms in the log-likelihood. The restricted maximum likelihood estimator of between-study variance is

$$\hat{\tau}^2_{REML} = \frac{\sum_{i=1}^{K} w_i^2[(\theta_i - \hat{\theta}_{REML})^2 - \sigma_i^2]}{\sum_{i=1}^{K} w_i^2} + \frac{1}{\sum_{i=1}^{K} w_i}. \tag{6.3.11}$$

The restricted maximum likelihood estimator is the the most common method for estimation of between-study variance in meta-analysis. The restricted maximum likelihood is preferred over the method by DerSimonian and Laird (1986) due to its balance between unbiasedness and efficiency (Viechtbauer, 2005).

Similarly to maximum likelihood estimator, a 95 per cent confidence interval for $\hat{\tau}^2_{REML}$ is given by set of values which satisfy

$$l_R(\hat{\tau}^2) > l_R(\hat{\tau}^2_{REML}) - \frac{1}{2}C_{0.95}(\chi_1^2) \tag{6.3.12}$$

where $C_{0.95}(\chi_1^2)$ is the 0.95 quantile of the χ_1^2 distribution and l_R is the restricted likelihood ratio test statistic. The distribution

$$-2\log(\frac{l_R(\tau^2)}{l_R(\tau^2_{ML})}) \to \chi_1^2 \qquad \text{for} \qquad K \to \infty$$

where $l_R(\tau^2)$ is the restricted maximum likelihood function calculated at the τ^2 and τ^2_{ML}.

6.3.6 Penalized likelihood estimator of τ^2

The penalized maximum likelihood estimator of between-study variance proposed by Kosmidis et al. (2017) can be estimated by obtaining the derivative of the penalized log-likelihood function

$$l^*(\theta, \tau^2) = l(\theta, \tau^2) - \frac{1}{2}\log(\sum_{i=1}^{K} \frac{1}{\hat{\sigma}_i^2 + \tau^2}) \tag{6.3.13}$$

where $\sum_{i=1}^{K} \frac{1}{\hat{\sigma}_i^2+\tau^2}$ is the $(1,1)$ block of the information matrix obtained from the initial log-likelihood function $l(\theta, \tau^2)$ given by 6.3.8. Maximizing $l^*(\vartheta)$ results in penalized maximum likelihood estimators $\hat{\theta}_{MPL}$ and $\hat{\tau}^2_{MPL}$. More details about the derivations and theory of the new penalized maximum likelihood estimation in meta-analysis are provided in Kosmidis et al. (2017). Kosmidis et al. (2017) claims that their bias correction reduces the bias of ML for OR. We include the bias corrected estimator of τ^2 proposed by Kosmidis et al. (2017) in our simulation.

6.3.7 Breslow-Day estimation of τ^2

Along with the new method for estimation of random effect parameter ρ in ODM, we propose a new method for estimating the between-study variance in standard additive random effects model (6.2.1). Through the correspondence (6.2.7) between ρ_i and τ^2, we can define

$$\rho_i = [\frac{(n_{1i}-1)}{n_{1i}p_{1i}(1-p_{1i})} + \frac{(n_{2i}-1)}{n_{2i}p_{2i}(1-p_{2i})}]^{-1}\tau^2$$

for the common between-study τ^2. By substituting, ρ_i as a function of τ^2 into the variance in Breslow-Day test through the correction factor (4.3.1), we get

$$\text{Var}(x_{1i}|\hat{\psi}_{MH}, \tau^2) = [\frac{1}{E(X_{1i}; \hat{\psi}_{MH})C_{1i}(\tau^2)} + \frac{1}{(x_i - E(X_{1i}; \hat{\psi}_{MH}))C_{2i}(\tau^2)} +$$

$$\frac{1}{(n_{1i} - E(X_{1i}; \hat{\psi}_{MH}))C_{1i}(\tau^2)} + \frac{1}{(n_i - x_i - n_{1i} + E(X_{1i}; \hat{\psi}_{MH}))C_{2i}(\tau^2)}\Big]^{-1}.$$

Now, though the correspondence (6.2.7), we expressed Breslow-Day statistic as a function of τ. Since the Breslow-Day statistic still follows χ^2_{K-1} with $K-1$ degrees of freedom, equating the $BD(\tau)$ statistic to its first moment $K-1$

$$\sum_{i=1}^{K} \frac{(X_{1i} - E(X_{1i}; \hat{\psi}_{MH}))^2}{\mathrm{Var}(X_{1i}; \hat{\psi}_{MH}, \tau^2)} = K - 1 \tag{6.3.14}$$

results in a new estimator for between-study variance τ^2. The confidence interval for this new estimator $\hat{\tau}^2_{BD}$ can be obtained similarly from the lower and upper percentile of the χ^2 distribution with $K-1$ degrees of freedom as we did before for ρ so that

$$X^2_{BD}(\tau^2_U) = \chi^2_{K-1,0.025} \qquad X^2_{BD}(\tau^2_L) = \chi^2_{K-1,0.975}. \tag{6.3.15}$$

We restrict the estimation of τ^2 to positive values, since between-study variance cannot be negative. $\hat{\tau}^2_{BD}$ can be used in estimation of common effect measure in standard random effects model. The restrictions for $\hat{\tau}^2_{BD}$ will be similar to those for estimation of ρ, apart from the additional restriction that τ^2 cannot be negative.

6.4 Simulation study for OR

In this section, we provide a simulation study to access the performance of point and interval estimators of random effect parameter τ^2 and the combined LOR θ in standard random effects model of meta-analysis. We assess seven point estimators of τ^2 in respect to their bias: the Der-Simonian and Laird estimator (6.3.1), the Mandel-Paule inspired estimator τ^2_{MP} - solution of equation 6.3.2, the corrected Mandel-Paule estimator based on the gamma approximation to Q distribution τ^2_{CMP} - solution of equation 6.3.6, the ML estimator (6.3.9), the REML estimator (6.3.11), the penalized ML estimator by Kosmidis et al. (2017) which maximize likelihood (6.3.13) and the BD-based estimator obtained from (6.3.14). We also assess four related confidence intervals for τ^2 (6.3.3), (6.3.7), (6.3.12) and (6.3.15) in respect to their coverage at the 95% confidence level. The combined odds ratio or its log is obtained by

inverse-variance method $\hat{\theta}_{REM} = \sum w_i(\tau^2)\hat{\theta}_i \sum w_i(\tau^2)$. We assess these estimators of $\hat{\theta}$ for bias and for coverage.

6.4.1 Simulation methods

We simulate the binary data using two different methods. One method is by Viechtbauer (2007) and another method is the same as simulation studies by Abo-Zaid et al. (2013) and Kosmidis et al. (2017). For both methods of simulations, we consider scenarios with p_{2i} values equal to 0.1, 0.2, 0.4 and θ values of 0, 1 and 2. Seven estimators of between-study variance and corresponding effect measure are assessed through a simulation study.

Simulation method by Viechtbauer (2007)

In standard REM for LOR, the data is generated as follows:

$$\theta_i \sim N(\theta, \tau^2) \quad \text{and} \quad p_{1i} = p_{2i}\exp(\theta_i)/(1 - p_{2i} + \exp(\theta_i)) \tag{6.4.1}$$

with

$$X_{1i} \sim Binom(n_{1i}, p_{1i}) \quad \text{and} \quad X_{2i} \sim Binom(n_{2i}, p_{2i}) \tag{6.4.2}$$

where the values of p_{2i} and θ are fixed. The study specific effect measures are estimated as $\hat{\theta}_i = \log(\hat{p}_{1i}(1 - \hat{p}_{2i})/\hat{p}_{2i}(1 - \hat{p}_{1i}))$ and its variance $\hat{\sigma}_i^2$ is estimated by (6.2.3). This scenario is similar to the method of simulation by Viechtbauer (2007).

Simulation method by Kosmidis et al. (2017)

The method of data generation in the simulation study by Abo-Zaid et al. (2013) and Kosmidis et al. (2017) generates the effect measures and their within-study variances using the model similar to an unconditional generalized linear mixed-effects model with random study effects discussed in Chapter 5.

Following the simulation method by Abo-Zaid et al. (2013), Kosmidis et al. (2017) generated K independent sample sizes $n_1, \ldots, n_K$ from uniform distribution with range of $(30, 31, \ldots, 100)$. Instead of using uniform distribution, we have set all K sample sizes $n_1, \ldots, n_K$ to be equal

across K studies. This is done in order to avoid any additional heterogeneity between K studies. Let $n_{iC} \sim Bin(n_i, p = 1/2)$ and $n_{iT} = n_i - n_{iC}$ be the sample sizes in control and treatment arms respectively. Next, two independent random effects u_{i1} and u_{i2} are generated from normal distributions $u_{i1} \sim N(0, 0.1)$ and $u_{i2} \sim N(0, \tau^2)$, where τ^2 is the between study variance. Let X_{ijk} be the Bernoulli variable for outcome k in group j of study i and ν_{ijk} is its logit transformation. The ν_{ijk} values are generated as

$$\nu_{ijk} = \beta_0 + u_{i1} + (\beta_1 + u_{i2})I(j = T), \tag{6.4.3}$$

where $I(j = T)$ is an indicator that takes values 0 for control and 1 for treatment arm, $\beta_0 = \log(p_{i2}/(1 - p_{i2}))$ is log-odds in control group and $\beta_1 = \theta$ is the true log-odds-ratio. In terms of multilevel model structure, the model (6.4.3) is

$$\nu_{iTk} = \beta_0 + u_{i1} + \beta_1 + u_{i2} \quad \text{and} \quad \nu_{iCk} = \beta_0 + u_{i1},$$

so

$$\theta_i = \log(\frac{p_{iT}}{1 - p_{iT}}) - \log(\frac{p_{iC}}{1 - p_{iC}}) = \beta_1 + u_{i2}.$$

Then, using the probabilities $\pi_{ijk} = \exp(\nu_{ijk})/(1 + \exp(\nu_{ijk}))$ from the model (6.4.3), we generated the vector of individual measurements of $X_{ij1}, \ldots, X_{ijn_i}$ in the study i and group j. Next, logistic regression

$$g(X_{ijk}) = \gamma_1 + \theta_i I(j = T),$$

is used to estimate γ_1 and θ_i, where g is the logit link function and θ_i is the individual study specific effect measure. The estimates of within-study variances are based on the evaluation of information matrix at the given estimates $\hat{\theta}_i$. The full description of this data generation method can be found in Abo-Zaid et al. (2013) and Kosmidis et al. (2017).

The function $functionsMPL.R$ is programmed by Kosmidis et al. (2017) and *metaLik* function from R package *metaLik* (Guolo and Varin, 2012) were used together for the method of estimation of between-study variance and overall effect measure proposed by Kosmidis et al. (2017). Kosmidis et al. (2017) claims that maximum penalized likelihood reduces the asymptotic bias of the maximum likelihood estimator of τ^2 and improves the coverage of overall

log odds ratio $\hat{\theta}_{RE}$ in studies with small to moderate sample sizes for all possible cases. The results of our simulation using the method of simulation by Kosmidis et al. (2017) are shown in the C.1.2 in Appendix. We have performed a simulation study for different combination of p_{2i} and θ. We tried to make this simulation study to be comparable to the simulation study using the method by Viechtbauer (2007). Kosmidis et al. (2017) provided simulation studies for two scenarios. One scenario is when the values of probabilities in treatment and control arms are $p_{1i} = 0.40$ and $p_{2i} = 0.219$. Another scenario is when the values of probabilities in treatment and control arms are $p_{1i} = 0.30$ and $p_{2i} = 0.1$. The first case is similar to $p_{2i} = 0.2$ and $\theta = 1$ in our simulations that is done following the method of simulation by Viechtbauer (2007) and Kosmidis et al. (2017). Thus, for comparison of methods for simulations by Kosmidis et al. (2017) and by Viechtbauer (2007), see C.13 for data simulated with 6.4.1 and 6.4.2 similar to Viechtbauer (2007) and C.53 for data simulated through logistic regression similar to Kosmidis et al. (2017) with $p_{2i} = 0.2$, $\theta = 1$ in Appendix.

Configurations

Sizes of the control and treatment groups were taken equal $n_{1i} = n_{2i} = n_i$ and were fixed across K studies. The true values of LOR θ_i across K studies were generated from normal distributions with mean θ_w and variance τ^2. For a given probability p_{2i}, the number of cases in the control group X_{2i} was simulated from a Binomial (n_{2i}, p_{2i}) distribution. The number of cases in the treatment group X_{1i} was generated from a Binomial (n_{1i}, p_{1i}) distribution with $p_{1i} = p_{2i}\exp(\theta_i)/(1 - p_{2i} + p_{2i}\exp(\theta_i))$ for a given LOR values of θ_i.

The following configurations of parameters were included in the simulations. The number of studies $K = (5, 10, 30)$; average sample sizes in each arm are $n = (40, 100, 250, 1000)$; the between-study variance for standard random effects model τ^2 varies between 0 and 1 for small-moderate heterogeneity and between 1 and 10 for moderate-large heterogeneity. The values of LOR θ vary from 0 to 2 in steps of 1. The probability in the control group p_{2i} takes values 0.1, 0.2, 0.4. A total of 10000 repetitions were produced for each combination.

Results of simulation study for bias and coverage of τ^2 in case of small-moderate heterogeneity

Figures 6.1-6.9 shows the results of simulations for five methods mentioned above for small-moderate heterogeneity ($0 \leq \tau^2 \leq 1$) for different combinations of K and n for the case when $p_{2i} \equiv 0.1$ and $\theta = 0, 1$. Due to large positive bias of Profiled-Breslow-Day estimator $\hat{\tau}^2_{BD}$ and negative bias of Der-Simonian and Laird estimator $\hat{\tau}^2_{DL}$, we did not include them in the Figures 6.1-6.9. Also, the bias of Profiled-Breslow-Day estimator $\hat{\tau}^2_{BD}$ is not linear in comparison to other methods. The full results of simulations that include all seven methods are provided in C.1-C.9 in Appendix. The biases of between-study variance estimates $\hat{\tau}^2$ for values of $\theta = 0, 1$ are shown on Figures 6.1 and 6.3. All estimates are negatively biased apart from $\hat{\tau}^2_{CMP}$ which has positive bias when N is small ($N = 40$). The bias of $\hat{\tau}^2_{CMP}$ decreases substantially for $N > 40$ resulting in the least biased estimator of τ^2 in all scenarios with different combinations of p_{2i} and $\theta = 0$. When $\theta \neq 0$, the absolute bias of $\hat{\tau}^2_{MP}$ and $\hat{\tau}^2_{CMP}$ are similar, but bias of $\hat{\tau}^2_{MP}$ is negative and the bias of $\hat{\tau}^2_{CMP}$ is positive. For $\theta = 0$, the second least biased estimator of τ^2 is $\hat{\tau}^2_{MP}$. We can clearly see that the proposed approximation for distribution of Q statistic by gamma distributions improves the estimation of τ^2 in Mandel-Paule method (see section 6.3.3 for details of approximation). Figure 6.1 shows that when data is sparse in both arms ($\theta = 0$), the bias of maximum likelihood estimate $\hat{\tau}^2_{ML}$ varies between 5-53%. Whereas for the same occasion, the bias of penalized maximum likelihood estimate $\hat{\tau}^2_{MPL}$ varies between 0-42%. Overall, penalization of the likelihood reduces the bias of maximum likelihood estimates. The reduction in bias from maximum likelihood estimate to penalized maximum likelihood estimate is between 2-61% for different combinations of K and N. Particularly, the bias reduction is more than 20% for $K = 5$ and $K = 30$. For $K = 10$, the bias is reduced between 9-24%. In Figures 6.1-6.9, we cannot see the Penalized maximum likelihood estimator τ^2_{MPL}, since the bias of Penalized maximum likelihood estimator τ^2_{MPL} and restricted maximum likelihood estimator τ^2_{REML} are identical. Penalized maximum likelihood estimator τ^2_{MPL} is good alternative to restricted maximum likelihood estimator. However, the

bias of both estimators does not disappear completely. For $\theta = 1$ (Figure 6.3) the bias of penalized maximum likelihood estimate varies between 0-23%. When $\theta = 1$, only the control arm has a sparse data - $p_{2i} = 0.1$. Hence, the overall bias of $\hat{\tau}^2_{MPL}$ is lower for $\theta = 1$ than for sparse data in both arms, $\theta = 0$.

Coming to coverage of τ^2 (Figure 6.2 and 6.4) for $p_{2i} = 0.1$ with $\theta = 0$ and 1), the Breslow-Day based estimation, which was the safest option in ODM, results in low coverage for τ^2. Again, similarly to point estimator of τ^2 from the Breslow-Day based method, this might be due to bias of estimated probabilities in 6.2.7 that is used in corrections of Breslow-Day statistic. The bias in probabilities is the result of transformation bias in REM discussed in Chapter 3. Similar deteriorations in coverage of Breslow-Day based interval estimator for τ^2 are also shown for non-sparse data. The Q-profile-based and Profile likelihood confidence intervals perform similarly to simulations by Viechtbauer (2007). The new Q-gamma-profile-based confidence interval does provide similar results to Q-profile-based confidence interval estimation. The Q-profile-based confidence interval estimation is always above the nominal 95% confidence level, whereas Q-gamma-profile-based confidence interval is always somewhere below the nominal 95% confidence level.

Overall, to be on the safe side, we would recommend the corrected-Mandel-Paule estimator $\hat{\tau}_{CMP}$ and corresponding Q-gamma-profile-based confidence interval. This is because $\hat{\tau}_{CMP}$ is the least biased estimator of τ^2 and Q-gamma-profile-based confidence interval perform similarly to Q-profile-based confidence interval.

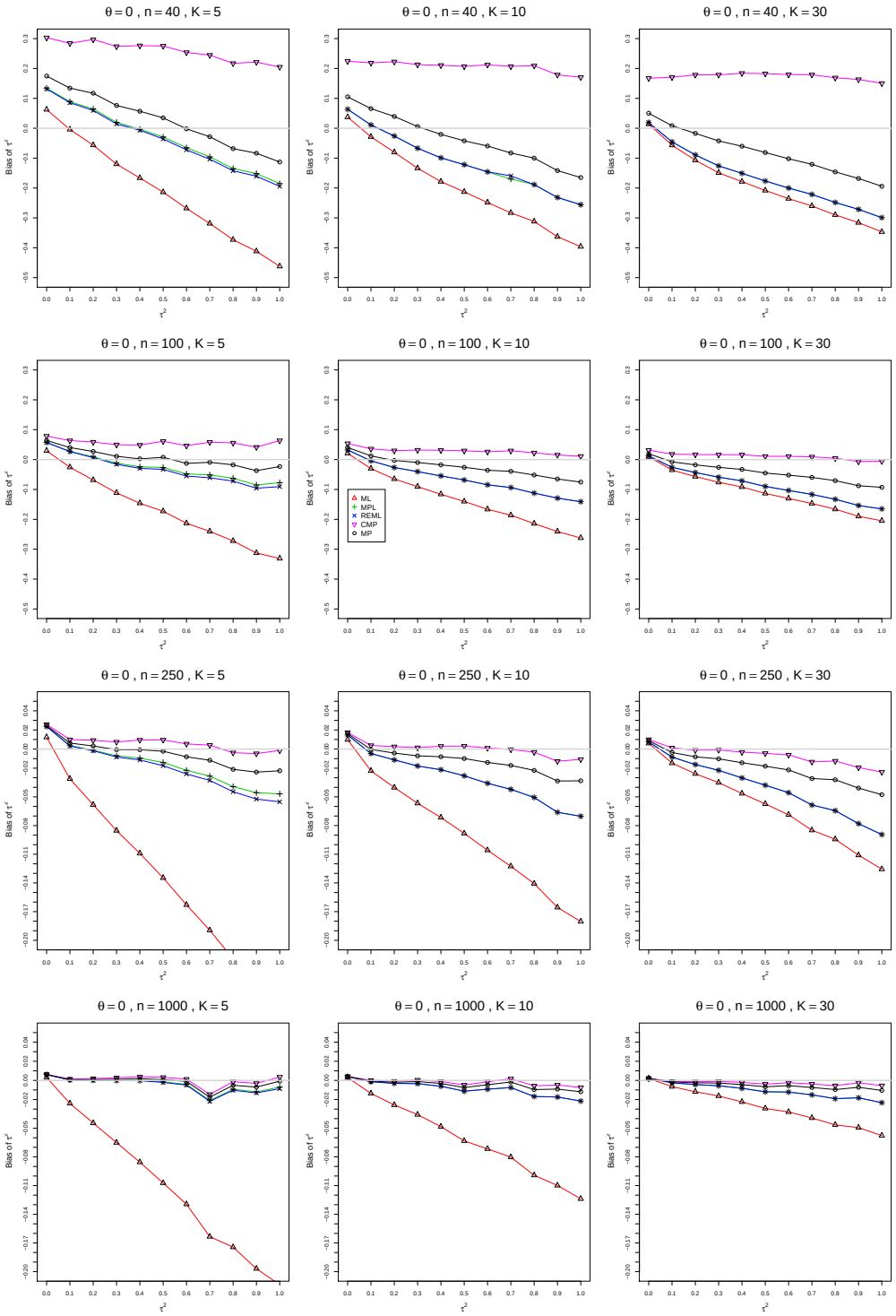


Figure 6.1: Bias of the between-study variance τ^2 obtained from K studies for $p_{2i} = 0.1$, $\theta = 0$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

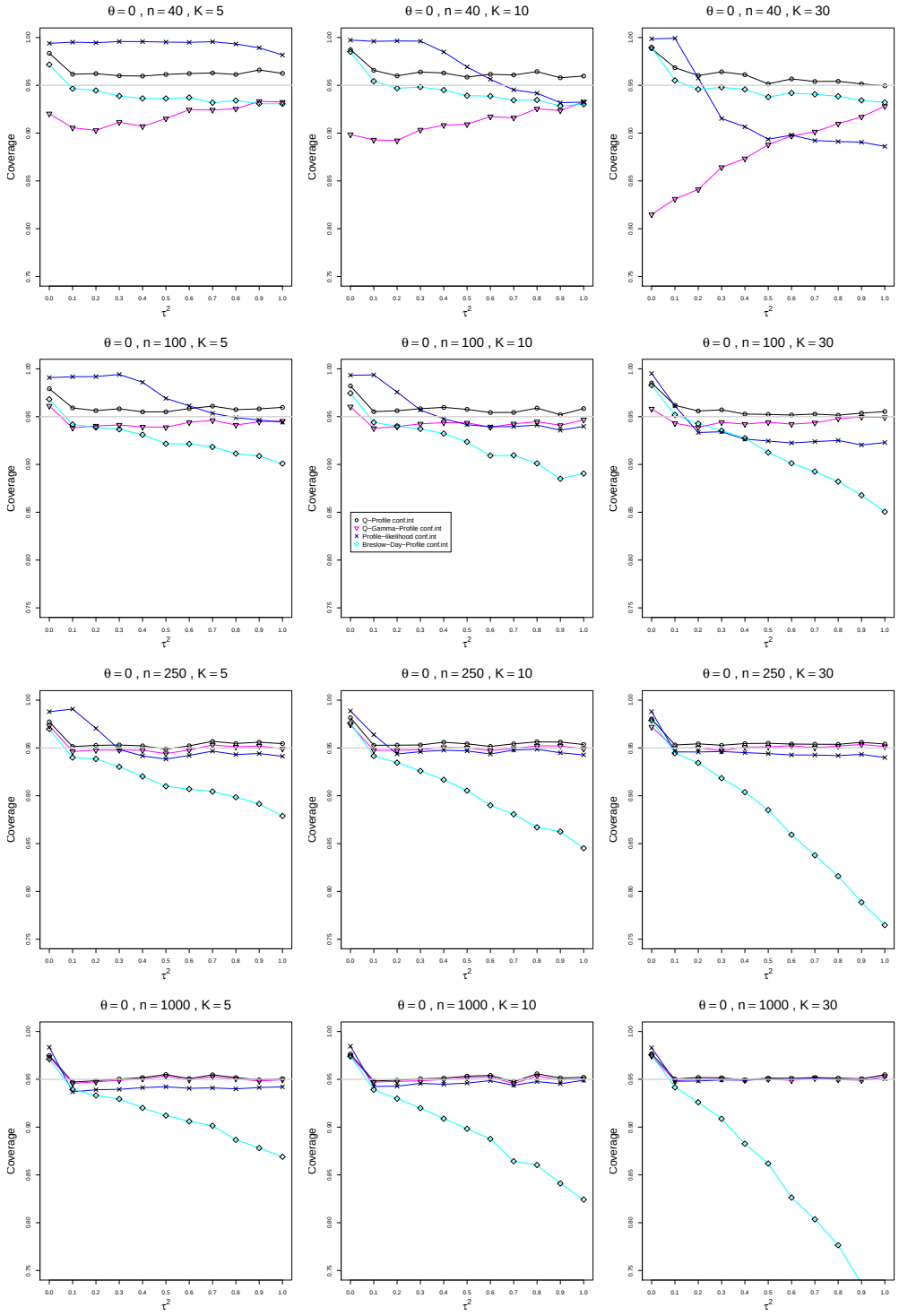


Figure 6.2: Coverage at the nominal confidence level of 0.95 of the between-study variance τ^2 obtained from K studies for $p_{2i} = 0.1$, $\theta = 0$ and $0 \leq \tau^2 \leq 1$. Interval estimation methods: black circles – Q-profile confidence interval for τ^2 based on χ^2 distribution, pink reverse triangles – Q-profile confidence interval for τ^2 based on $\Gamma_{r(\tau^2),\lambda(\tau^2)}$ distribution), dark blue crosses – Profile likelihood confidence intervals, light blue diamonds – Breslow-Day-Profile confidence intervals. Light grey line at 0.95.

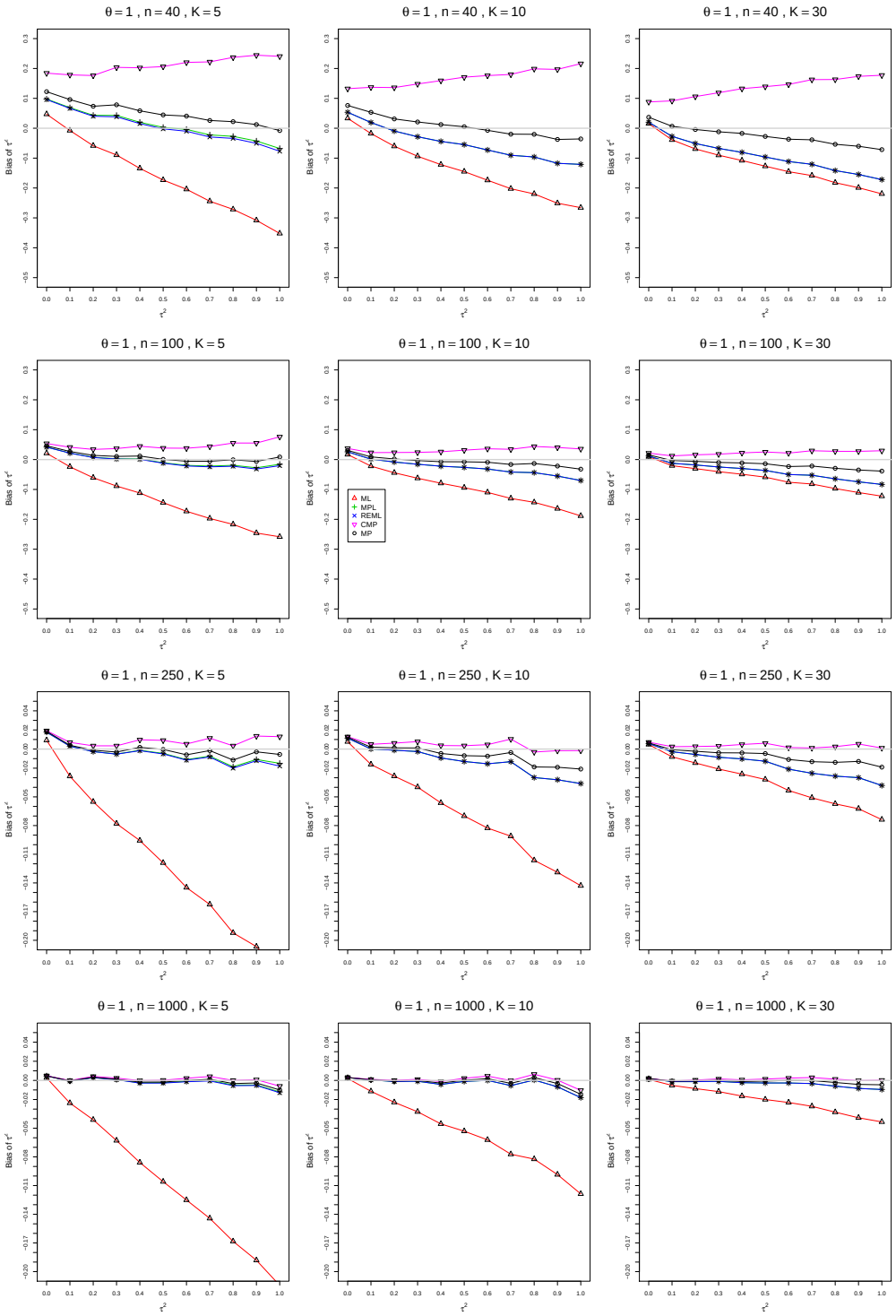


Figure 6.3: Bias of the between-study variance τ^2 obtained from K studies for $p_{2i} = 0.1$, $\theta = 1$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

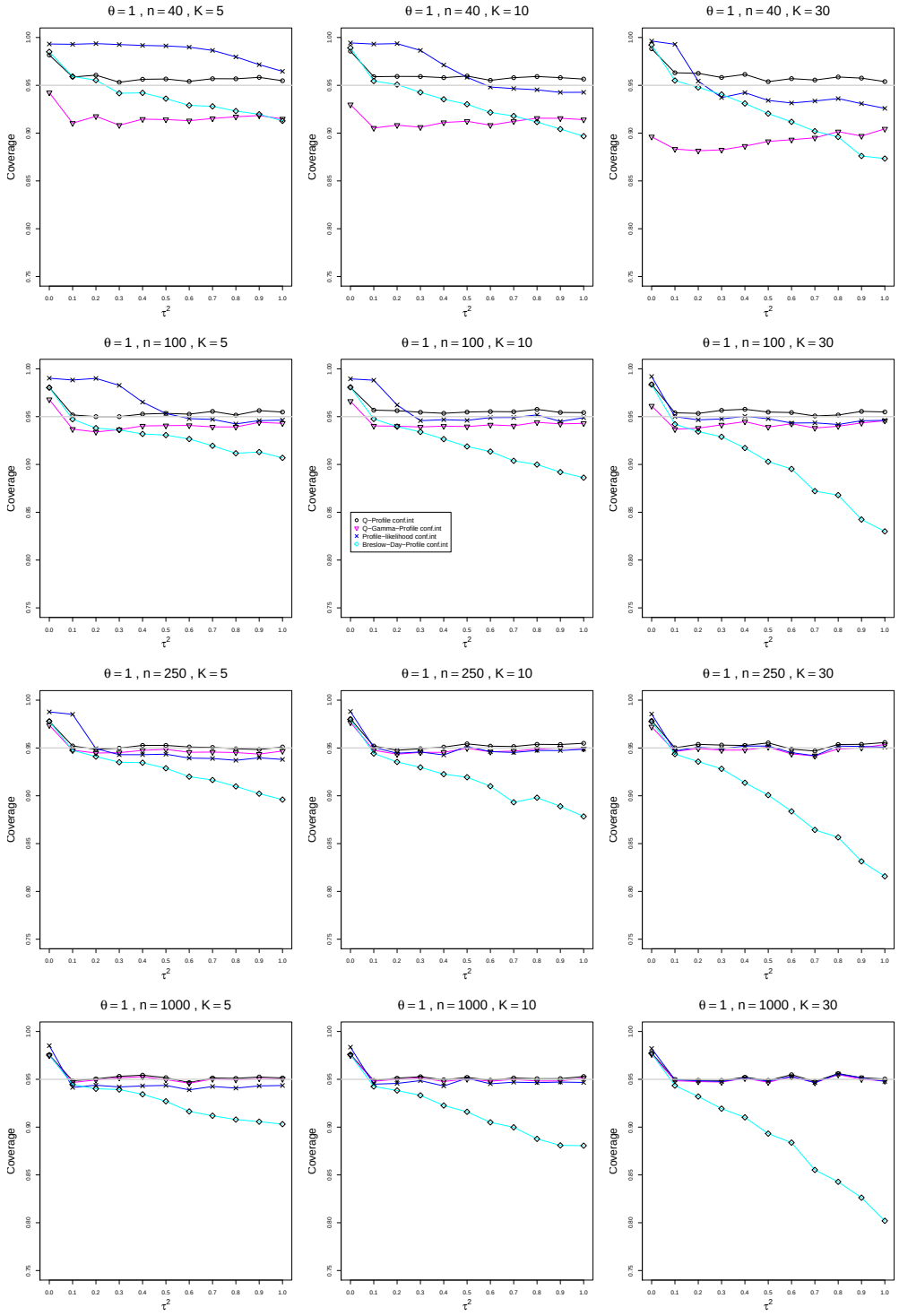


Figure 6.4: Coverage at the nominal confidence level of 0.95 of the between-study variance τ^2 obtained from K studies for $p_{2i} = 0.1$, $\theta = 1$ and $0 \leq \tau^2 \leq 1$. Interval estimation methods: black circles – Q-profile confidence interval for τ^2 based on χ^2 distribution, pink reverse triangles – Q-profile confidence interval for τ^2 based on $\Gamma_{r(\tau^2),\lambda(\tau^2)}$ distribution), dark blue crosses – Profile likelihood confidence intervals, light blue diamonds – Breslow-Day-Profile confidence intervals. Light grey line at 0.95.

Results of simulation study for bias and coverage of θ in case of small-moderate heterogeneity

The improvement in the estimation of τ^2 has a positive impact on bias and coverage of the overall-effect measure. Particularly, the improvement is visible between $\hat{\tau}^2_{MP}$ and $\hat{\tau}^2_{CMP}$ and $\hat{\tau}^2_{ML}$ and $\hat{\tau}^2_{MPL}$ (see Figures 6.6, 6.8 for biases of θ and Figures 6.6 and 6.8 for coverages when $\theta = 0$ and $\theta = 1$). The bias of overall effect measure is positive. The bias is practically the same regardless of the method. The bias decreases with increasing the sample size N. The least biased estimator of overall effect measure θ is $\hat{\theta}_{CMP}$. Figure 6.5 shows the bias of $\hat{\theta}_{CMP}$ for $\theta = 0$, 1 and 2. The bias of $\hat{\theta}_{CMP}$ decreases with $\theta = 0$, 1 and 2 and changes its magnitude. Also, using $\hat{\tau}^2_{CMP}$ in weights provides reasonably good coverages for $\hat{\theta}_{CMP}$ in comparison to all the methods (see Figures 6.7 and 6.9). The coverage of $\hat{\theta}_{BD}$ with $\hat{\tau}^2_{BD}$ in weights becomes conservative with increasing K (see C.3,C.6 and C.9 in Appendix). This is due to the large positive bias of $\hat{\tau}^2_{BD}$. The bias in $\hat{\tau}^2_{BD}$ comes from the biases of estimated probabilities in equation (6.2.7). The Profiled Breslow-Day performs well in estimation of intra-cluster correlation in ODM, but it does not provide a good estimator of between-study variance in standard REM. Thus, $\hat{\theta}_{BD}$ with $\hat{\tau}^2_{BD}$ are not recommended in standard additive random effects model. Overall the coverage of $\hat{\theta}_{MPL}$ is somewhat improved in comparison to the coverage of $\hat{\theta}_{ML}$. Figures 6.7 and 6.9 show the coverage of overall effect measure for values of $\theta = 0, 1$ respectively. Particularly, the improvements are evident for $N > 100$. However, for $N < 100$ the coverages are still low as in standard methods The latter contradicts to the results by Kosmidis et al. (2017) for small-moderate sample sizes. We have chosen the sample sizes N and number of studies K almost similar to simulation results from Chapter 4 and 5.

The results for non-sparse data $p_{2i} = 0.2$ and $p_{2i} = 0.4$ are also provided in the Appendix (see C.10-C.25). When $p_{2i} = 0.2$ and $p_{2i} = 0.4$, the biases are similar to the case when $p_{2i} = 0.1$. The absolute values of the biases are smaller. Overall, the biases of penalized maximum likelihood estimator τ^2_{MPL} varies between 0-40% for $p_{2i} = 0.2$ and $p_{2i} = 0.4$ respectively. Whereas, the bias of maximum likelihood estimator τ^2_{ML} vary between 3-37% for $p_{2i} = 0.2$ and 3-49%

for $p_{2i} = 0.4$. Again, penalized maximum likelihood reduces the bias of maximum likelihood estimator. When $p_{2i} = 0.2$ and $p_{2i} = 0.4$, the coverages are all above 90% apart from the case when $K = 5$. Case $K = 5$ is the scenario when number of studies is small and all methods fail to estimate between-study variance, since the precision of τ^2 has an order $O(1/K)$.

When, $K = 5$ and $p_{2i} = 0.1$, from the C.1, C.4 and C.7 in Appendix, we can clearly see that penalized maximum likelihood is much better than maximum likelihood estimator. However, our sample sizes are big in this situation. Hence, using τ^2_{MPL} leads to small reductions in the bias of τ_{ML} for large biases. The penalized maximum likelihood reduces the bias of maximum likelihood estimates of τ^2. However, some bias does still exist. The bias of overall log odds-ratio is the same for maximum likelihood, restricted maximum likelihood and penalized likelihood method (see C.2, C.5, C.8 for $p_{2i} = 0.1$, C.11, C.14, C.17 for $p_{2i} = 0.2$ and C.20, C.23, C.26 for $p_{2i} = 0.4$ in Appendix) . The DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$ performs either similar to maximum likelihood estimator or worse for large $N > 80$ (see C.1, C.4 and C.7 in Appendix). Particularly $\hat{\tau}^2_{DL}$ has a large bias for moderate-large heterogeneity between studies as we shall see in section 6.4.1. Thus, DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$ is not recommended in an estimation of between-study variance.

In case of sparse data the biases in $\hat{\tau}^2$ might be the combination of transformation bias, bias due to continuity corrections and bias of estimates $\hat{\theta}_i$ that come from the bias of estimated probabilities in $\hat{\theta}_i$. The sparseness in the data itself can introduce an unobserved heterogeneity between studies. All the methods for sparse data have to be applied with care. At the moment, there exist no uniformly unbiased and robust method for estimation of between-study variance and for an overall effect measure. In general, different estimators of τ^2 can be used for different inferential purposes, one to reduce the variance of overall effect measure to minimum and another to obtain a reliable confidence interval (Kulinskaya et al., 2014).

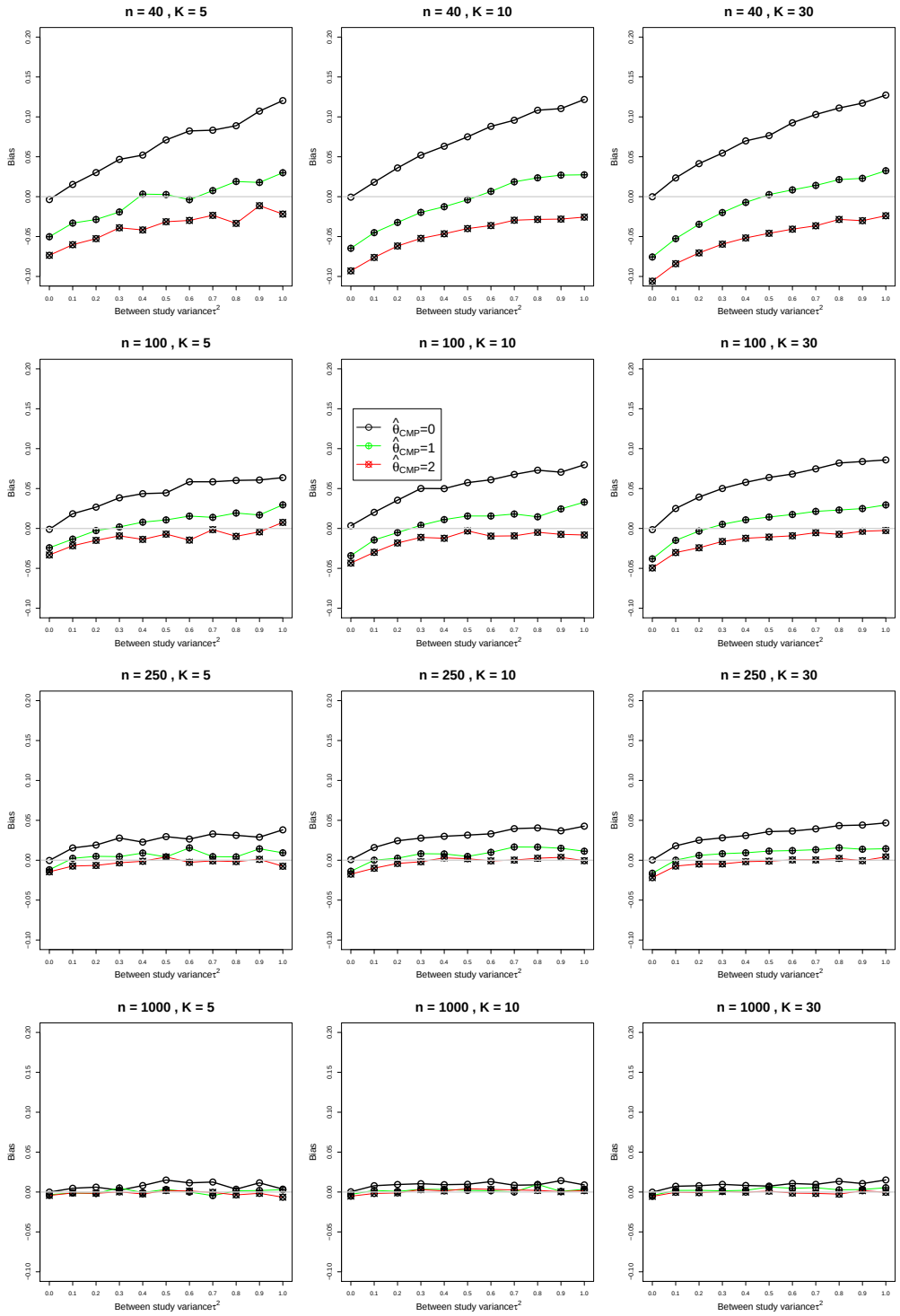


Figure 6.5: Bias of overall odds ratio θ_{IV} obtained from K studies by the inverse-variance method with the moment estimator $\hat{\tau}^2_{CMP}$ in the weights, for $p_{2i} = 0.1$, and $0 \leq \tau^2 \leq 1$. The biases are given for $\theta = 0$ (circles), $\theta = 1$ (circle plus), and $\theta = 2$ (circle cross). Light grey line at 0.

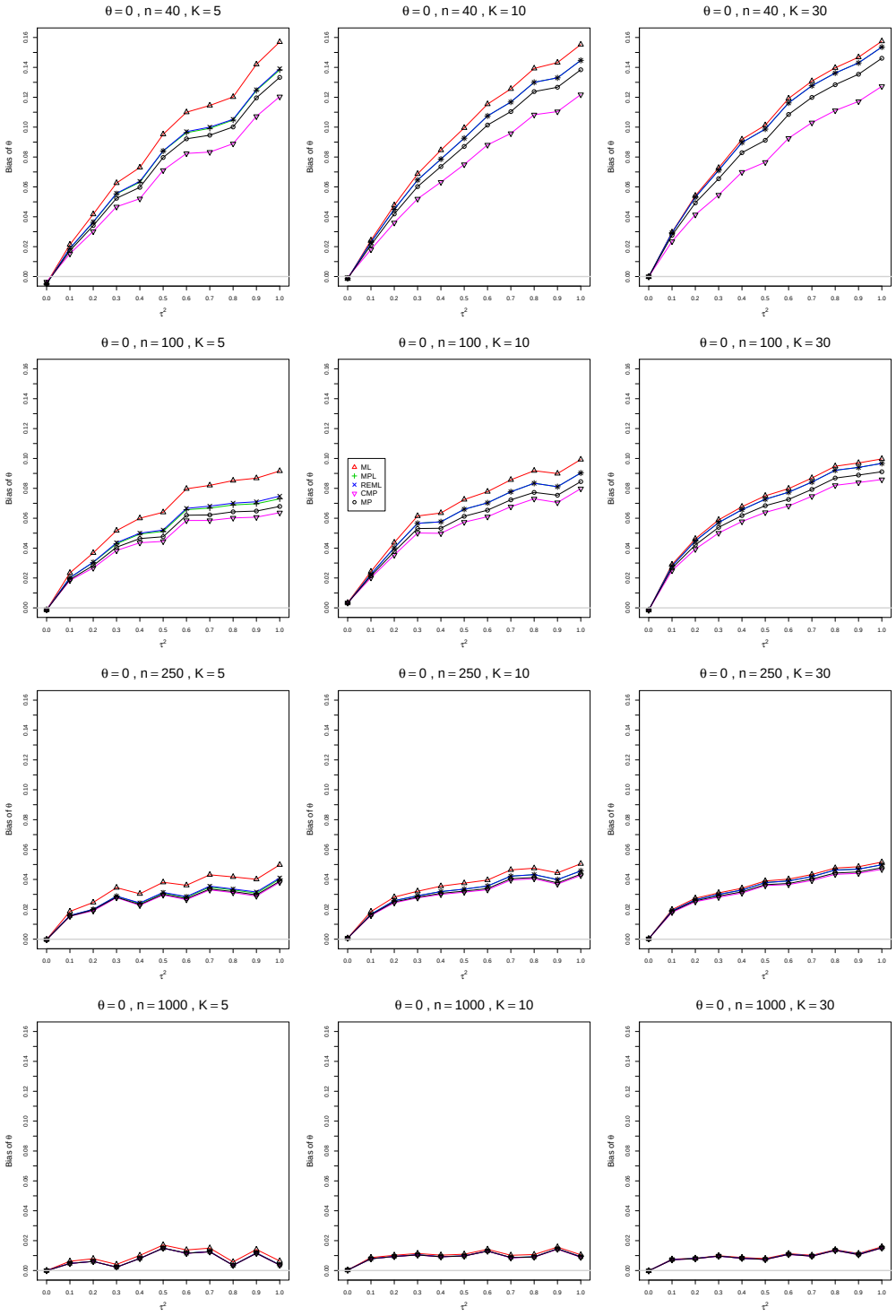


Figure 6.6: Bias of the estimated overall effect measure $\hat{\theta}_{RE}$ obtained from K studies for $p_{2i} = 0.1$, $\theta = 0$ and $0 \leq \tau^2 \leq 1$. The estimators of $\hat{\theta}_{RE}$ include the estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

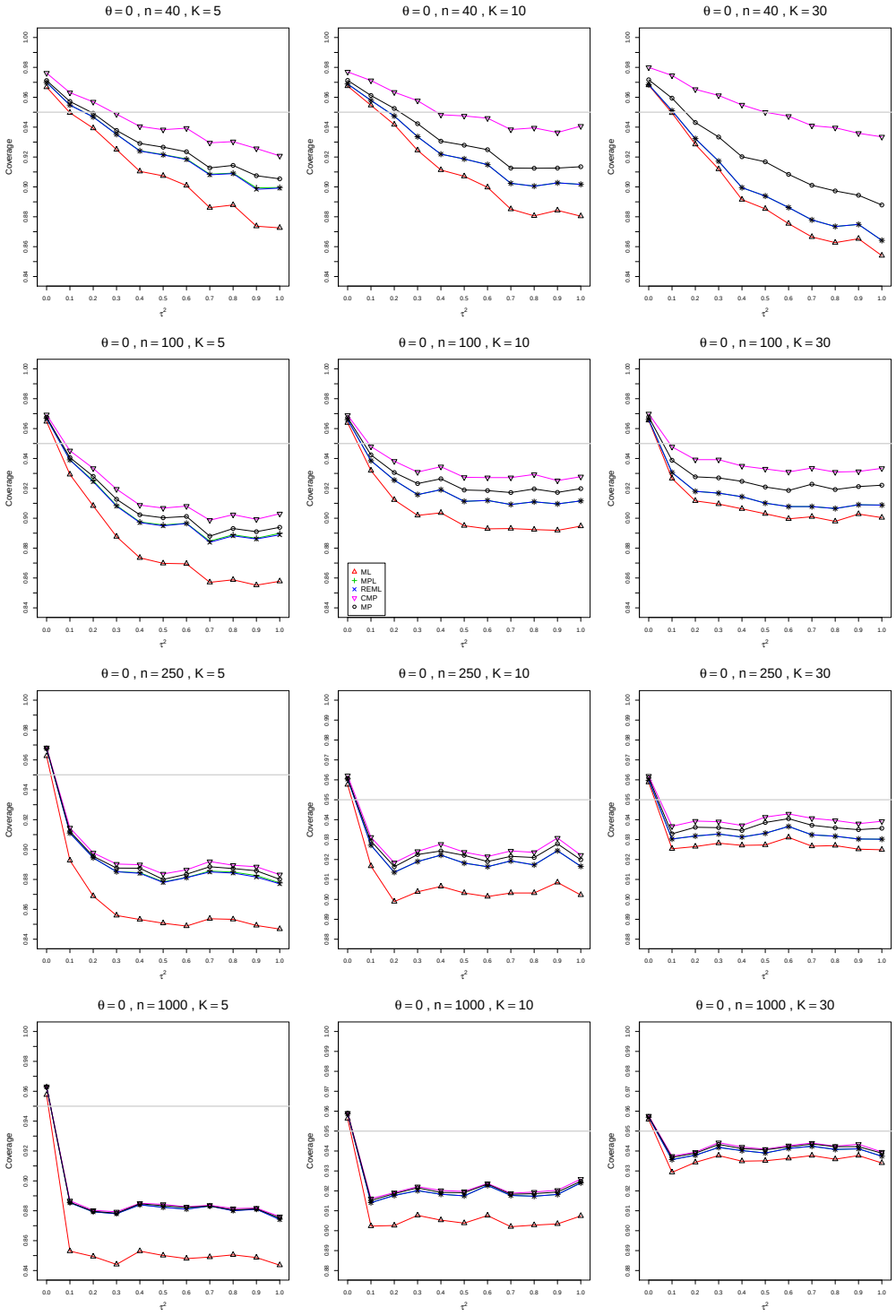


Figure 6.7: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_{RE}$ obtained from K studies by the inverse-variance method, for $p_{2i} = 0.1$, $\theta = 0$ and $0 \leq \tau^2 \leq 1$. The inverse-variance weights use the following estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

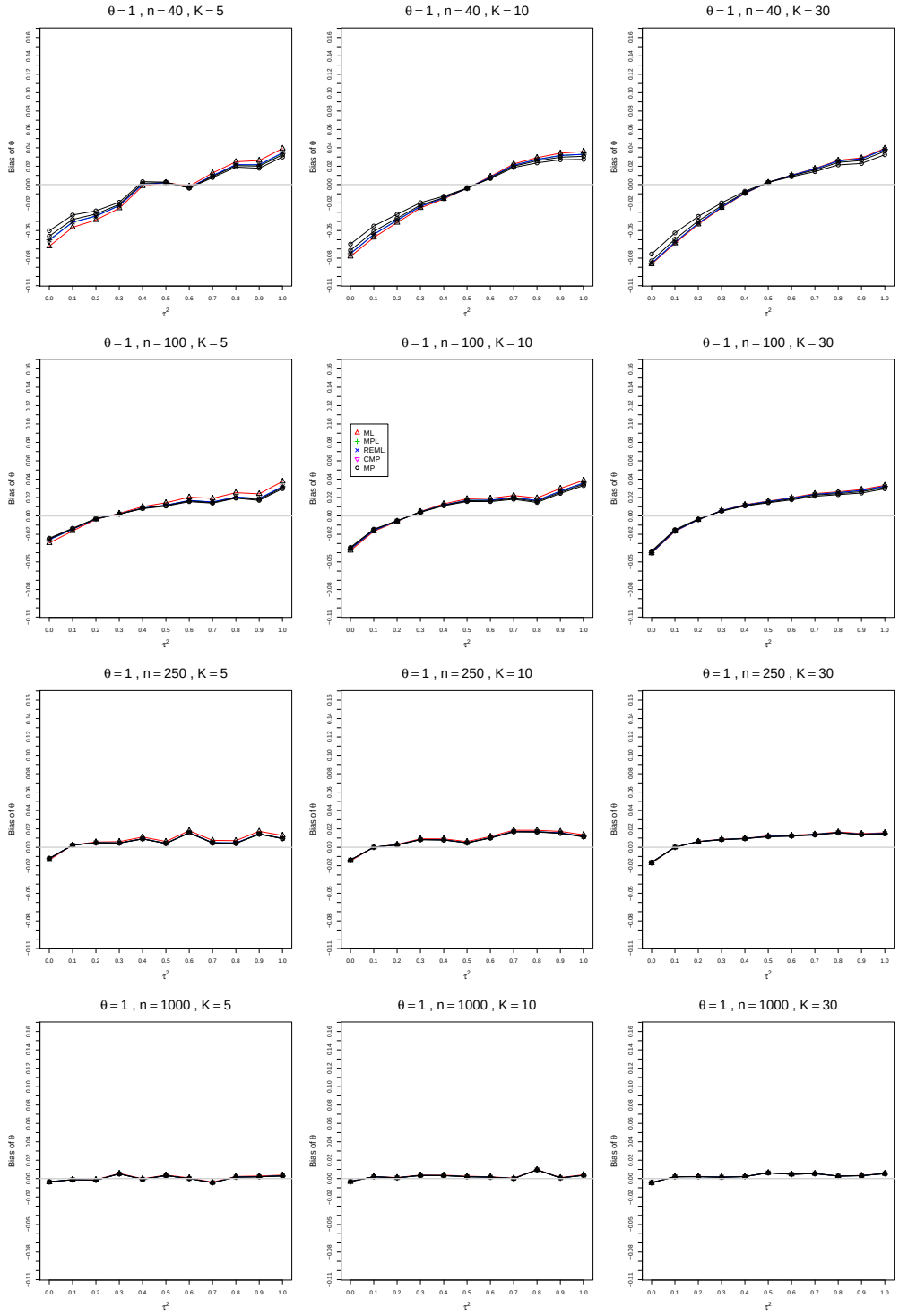


Figure 6.8: Bias of the estimated overall effect measure $\hat{\theta}_{RE}$ obtained from K studies for $p_{2i} = 0.1$, $\theta = 1$ and $0 \leq \tau^2 \leq 1$. The estimators of $\hat{\theta}_{RE}$ include the estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

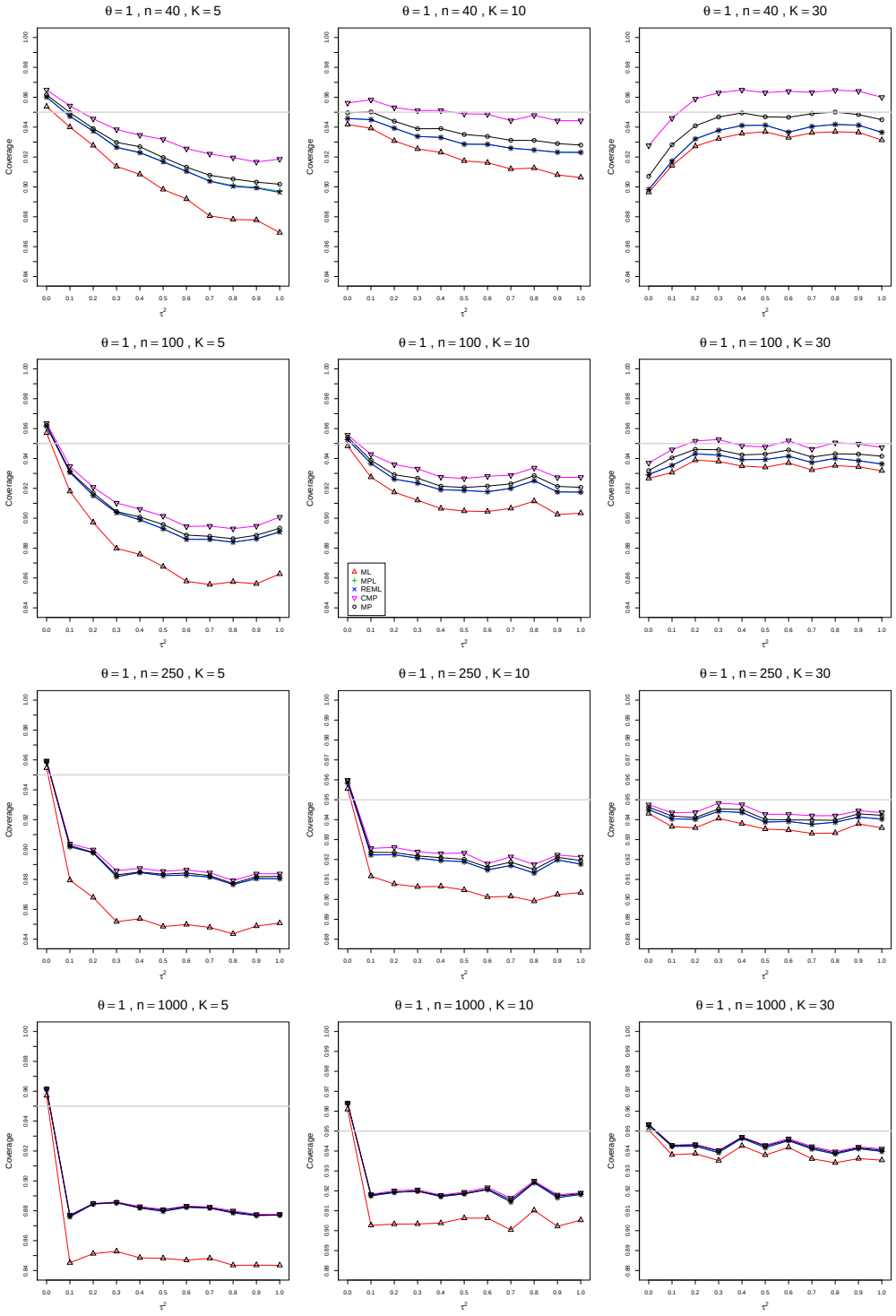


Figure 6.9: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_{RE}$ obtained from K studies by the inverse-variance method, for $p_{2i} = 0.1$, $\theta = 1$ and $0 \leq \tau^2 \leq 1$. The inverse-variance weights use the following estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

Results of simulation study for bias and coverage of τ^2 in case of moderate-large heterogeneity

In addition to simulations with small to moderate heterogeneity, we have performed similar simulations for moderate to large heterogeneity ($0 \leq \tau^2 \leq 10$). Figures 6.10-6.17 show the results of simulations for moderate-large heterogeneity ($0 \leq \tau^2 \leq 10$). Again, we do not include the Profiled Breslow-Day estimator $\hat{\tau}^2_{BD}$ and Der-Simonian and Laird estimator $\hat{\tau}^2_{DL}$ due to the large biases of both estimators. The Profiled Breslow-Day estimator $\hat{\tau}^2_{BD}$ results in the largest positive bias and conservative coverage in $\hat{\theta}_{BD}$ among all methods. This is because of large positive bias of $\hat{\tau}^2_{BD}$. Only $\hat{\tau}^2_{BD}$ has a positive bias, whereas the bias of other estimator $\hat{\tau}^2_{ML}$,$\hat{\tau}^2_{MPL}$, $\hat{\tau}^2_{REML}$, $\hat{\tau}^2_{MP}$, $\hat{\tau}^2_{CMP}$ and $\hat{\tau}^2_{DL}$ are negative.

Figures 6.10 and 6.12 show the biases of between-study variances estimated by five methods similar to results for small-moderate heterogeneity. Correspondingly to the results for small-moderate heterogeneity, $\hat{\tau}^2_{CMP}$ is the first least biased estimator and $\hat{\tau}^2_{MP}$ is the second least biased estimator of between-study variance for all scenarios. Asymptotically with increasing the sample sizes, both estimators $\hat{\tau}^2_{CMP}$ and $\hat{\tau}^2_{MP}$ tend to perform identically. However, for small-moderate sample sizes $\hat{\tau}^2_{CMP}$ is much better than $\hat{\tau}^2_{MP}$ and other estimators based on maximising the likelihood for all the scenarios with different combination of p_{2i} and θ in simulations.

When $p_{2i} = 0.1$, Figures 6.10 and 6.12 show the bias of between-study variance estimators under moderate to large heterogeneity, $\theta = 0$ and $\theta = 1$. Similarly to results for small-moderate heterogeneity, all estimators of τ^2 are biased downwards. The biases for moderate-large heterogeneity ($1 \leq \tau^2 \leq 10$) are the continuation of the biases for small-moderate heterogeneity ($0 \leq \tau^2 \leq 1$).

When data are sparse in both arms ($p_{2i} = 0.1$ and $\theta = 0$) and $1 \leq \tau^2 \leq 10$, the bias of maximum likelihood estimator $\hat{\tau}_{ML}$ ranges between 4-54%. Whereas, the bias of penalized maximum likelihood estimator $\hat{\tau}_{MPL}$ ranges between 0.1-42%. The overall bias reduction for moderate-large heterogeneity is similar to the results of simulations for small-moderate

heterogeneity ($0 \leq \tau^2 \leq 1$). Also, we would expect the similar bias reductions for results with $p_{2i} = 0.1$ and $\theta = 1$ (see Figures 6.12). The bias of $\hat{\tau}_{MPL}$ and $\hat{\tau}_{REML}$ are linearly identical.

Also, similarly for small-moderate heterogeneity, the coverages of τ^2 from four methods are shown in Figures 6.11 and 6.13 for $p_{2i} = 0.1$ and $\theta = 0$ and $\theta = 1$ respectively. Interestingly, for sparse data $p_{2i} = 0.1$ and $\theta = 0$, all the coverages deteriorate with τ^2 when $K = 30$ and $N \leq 250$. The similar reductions in coverages occur for higher probabilities such as $p_{2i} = 0.2$ and $p_{2i} = 0.4$ with $\theta = 0, 1, 2$. For the rest of the combinations of N and K, the coverages from all three methods perform pretty well with Q-profile and Q-gamma-profile based confidence intervals being more conservative than Profile-likelihood confidence interval.

Again, among all the methods, we would recommend $\hat{\tau}_{CMP}$ as the least biased point estimator with corresponding Q-gamma-profile based confidence intervals. The approximation of the Q distribution by gamma distribution with parameters defined in section 6.3.3 works very well. Among the maximum likelihood methods, restricted maximum likelihood and penalized maximum likelihood perform similar hence are recommended to use. However, the bias of maximum likelihood and penalized maximum likelihood estimators still exist and further study of their bias reduction is to be pursued elsewhere.

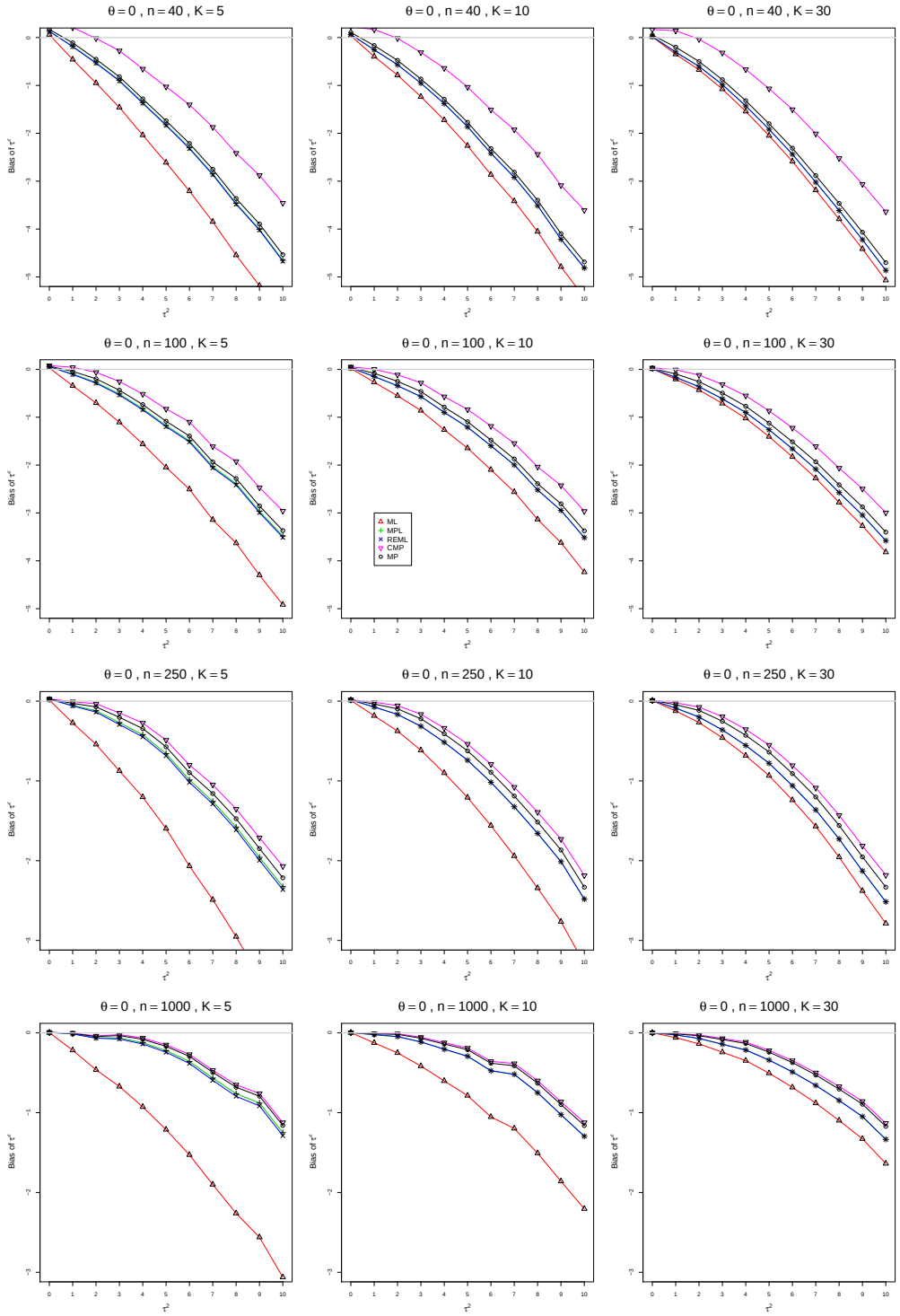


Figure 6.10: Bias of the between-study variance τ^2 obtained from K studies for $p_{2i} = 0.1$, $\theta = 0$ and $0 \leq \tau^2 \leq 10$. The estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

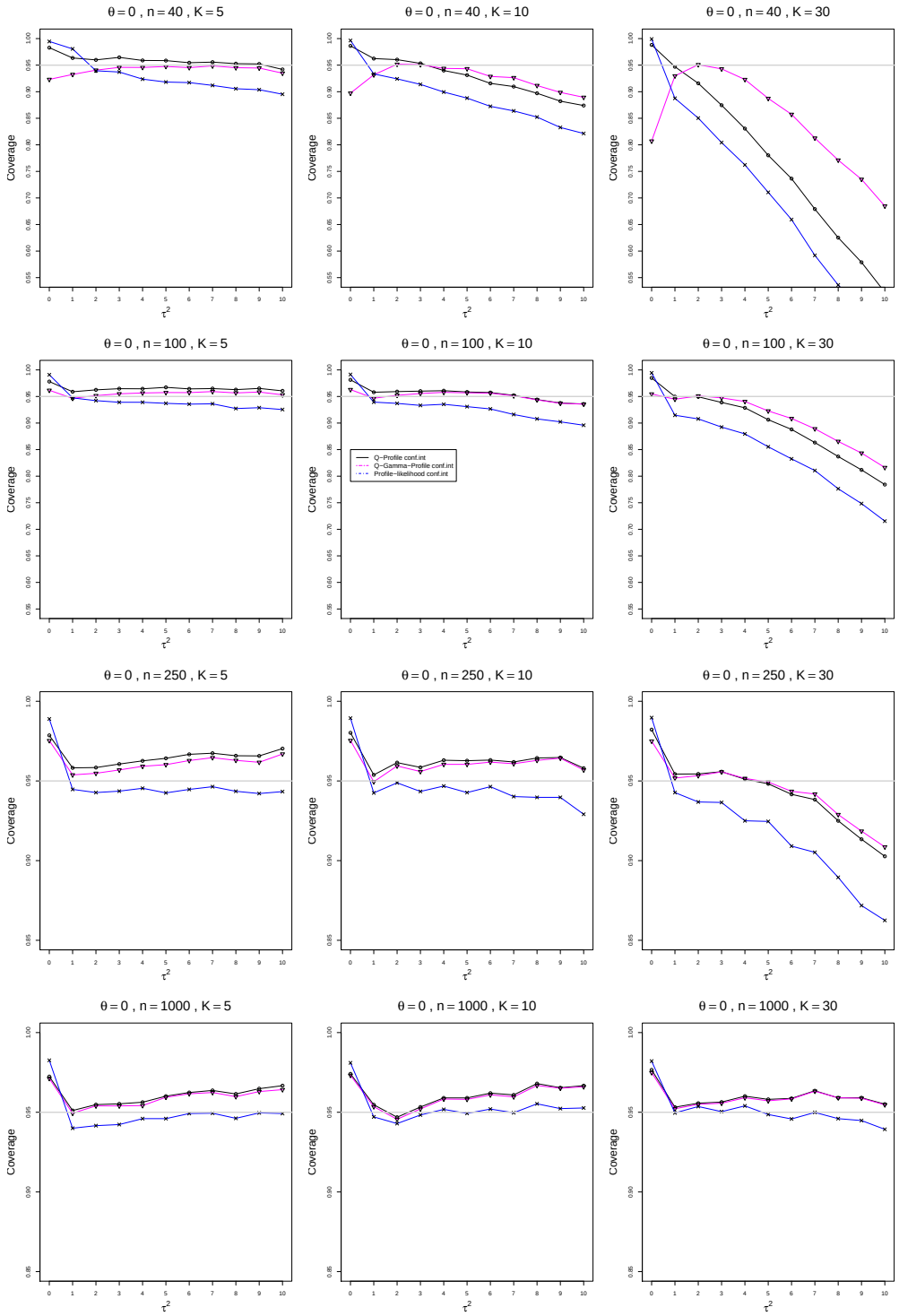


Figure 6.11: Coverage at the nominal confidence level of 0.95 of the between-study variance τ^2 obtained from K studies for $p_{2i} = 0.1$, $\theta = 0$ and $0 \leq \tau^2 \leq 10$. Interval estimation methods: black circles – Q-profile confidence interval for τ^2 based on χ^2 distribution, pink reverse triangles – Q-profile confidence interval for τ^2 based on $\Gamma_{r(\tau^2),\lambda(\tau^2)}$ distribution), dark blue crosses – Profile likelihood confidence intervals. Light grey line at 0.95.

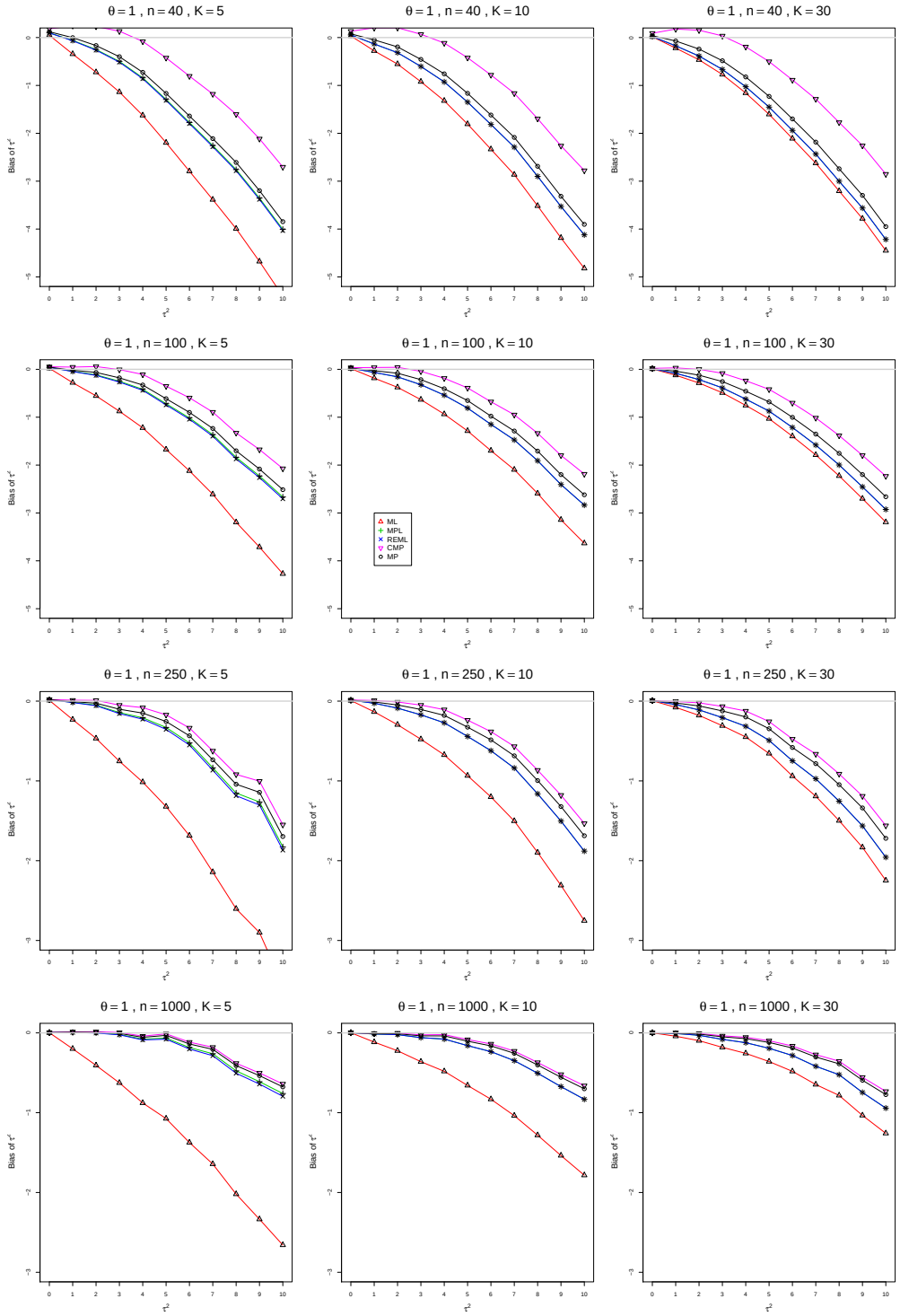


Figure 6.12: Bias of the between-study variance τ^2 obtained from K studies for $p_{2i} = 0.1$, $\theta = 1$ and $0 \leq \tau^2 \leq 10$. The estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

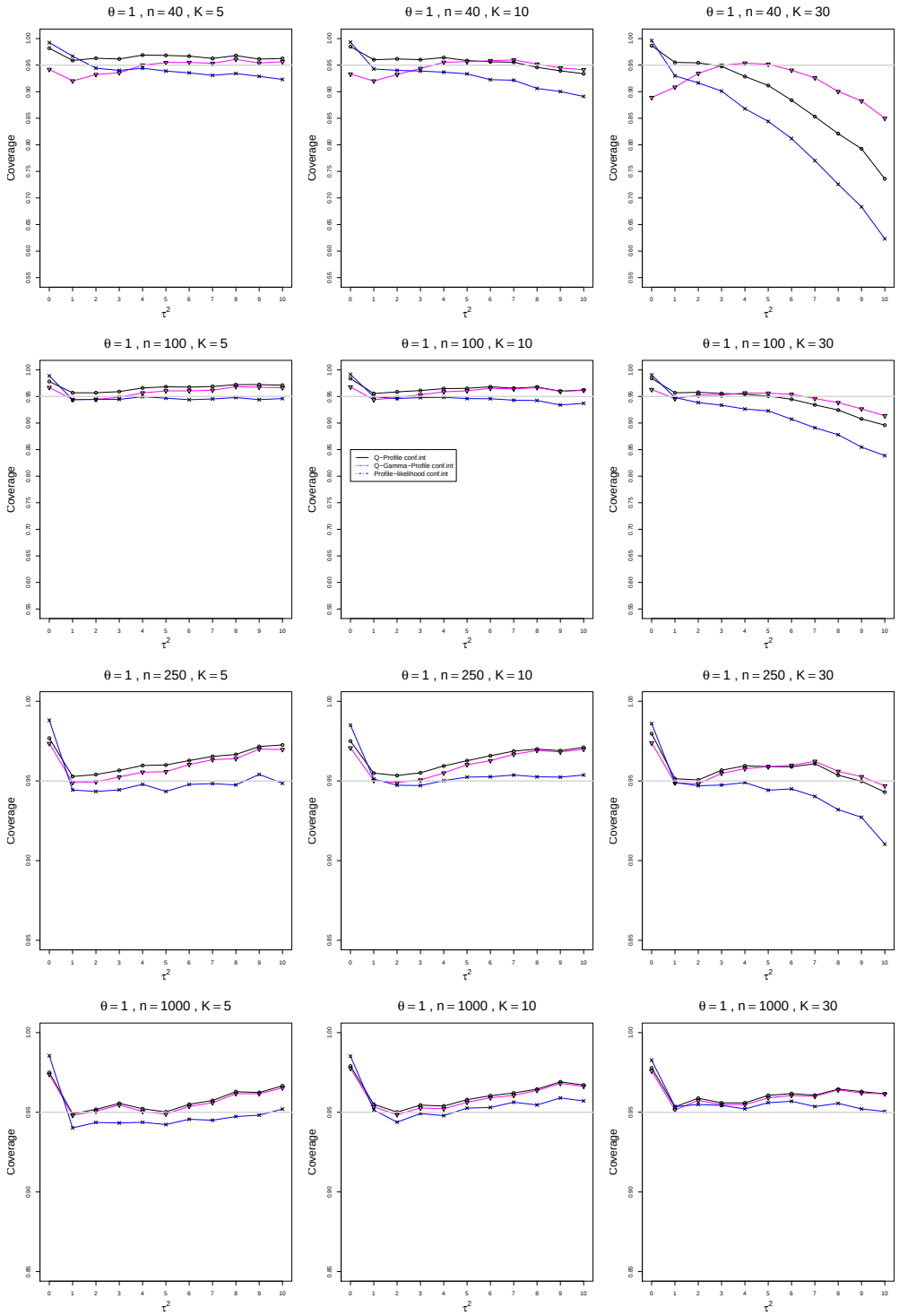


Figure 6.13: Coverage at the nominal confidence level of 0.95 of the between-study variance τ^2 obtained from K studies for $p_{2i} = 0.1$, $\theta = 1$ and $0 \leq \tau^2 \leq 10$. Interval estimation methods: black circles – Q-profile confidence interval for τ^2 based on χ^2 distribution, pink reverse triangles – Q-profile confidence interval for τ^2 based on $\Gamma_{r(\tau^2),\lambda(\tau^2)}$ distribution), dark blue crosses – Profile likelihood confidence intervals. Light grey line at 0.95.

Results of simulation study for bias and coverage of θ in case of moderate-large heterogeneity

Figure 6.14 and Figures 6.15, 6.16, 6.17 show the bias and coverage of overall log-odds-ratio θ of K studies obtained from five methods of estimating of τ^2, respectively. All the figures for bias and coverage of overall log-odds-ratio correspond to the simulated data from the scenarios with $\theta = 0$, $\theta = 1$ and $\theta = 2$.

For moderate-large heterogeneity, $0 \leq \tau^2 \leq 10$, the bias of estimated log-odds-ratio $\hat{\theta}$ was practically the same regardless of a method used for estimation of between-study variance τ^2 (Figure 6.14 shows the bias of overall-log-odds-ratio with $\hat{\tau}^2_{CMP}$ in weights). The bias of estimated log-odds-ratio is pretty much the same in these scenarios. This is probably due to the fact that large values of τ^2 lead to the estimate of weighted inverse variable log-odds ratio becoming just an unweighted average, regardless of the estimator of τ^2 used. The bias of overall log-odds-ratio reduces with increasing θ from $\theta = 0$ to $\theta = 2$.

The coverages of overall log-odds-ratio for sparse data ($p_{2i} = 0.1$ and $\theta = 0$) from all methods are shown in Figure 6.15. The best method that provides the coverage close to nominal 0.95 significance level is new Profiled-Q-gamma-based method. Similarly to simulations for small-moderate heterogeneity the coverages from all methods deteriorate when $N \leq 100$ and $K = 30$. The coverages of maximum likelihood estimator θ_{ML} are below 90% for $N \leq 100$ and above 90% for $N \geq 100$ apart from the case of small number of studies $K = 5$. For small number of studies $K = 5$, the bias of $\hat{\tau}^2_{ML}$ is larger than for $K = 10$ and $K = 30$. Due to this large negative bias, the coverages are lower for small and large values of N. The coverage of θ_{MPL} are somewhat better than coverage of θ_{ML}. Particularly, the difference is noticeable for $K = 5$. This can be explained by less biased estimate of between-study variance θ_{MPL} for $K = 5$ (see Figures 6.10 and 6.15 for bias of $\hat{\tau}^2_{MPL}$ and coverage of $\hat{\theta}_{MPL}$ respectively). When $K = 10$ and $K = 30$, the improvements in coverages of $\hat{\theta}_{MPL}$ relative to $\hat{\theta}_{ML}$ are smaller than for the case of $K = 5$, but stills exists. The similar improvements occur for larger probabilities $p_{2i} = 0.2$ and $p_{2i} = 0.4$. When both arms have equal probabilities (either

$p_{1i} = p_{2i} = 0.2$ and $p_{1i} = p_{2i} = 0.4$), the bias of $\hat{\tau}^2_{ML}$ varies between 4-50% and 4-56% for $p_{2i} = 0.2$ and $p_{2i} = 0.4$ respectively. Whereas, the bias of $\hat{\tau}^2_{MPL}$ ranges between 0-37% and 0-45%. Again, we can clearly see the reduction in the bias of $\hat{\tau}^2_{MPL}$ in comparison to $\hat{\tau}^2_{ML}$. The bias reduction is between 2-25% and 2-24% for $p_{2i} = 0.2$ and $p_{2i} = 0.4$ respectively. We would expect the same proportion of bias reduction similar to small-moderate heterogeneity for non-equal probabilities and non-sparse data, i.e $p_{2i} = 0.2$ and $p_{2i} = 0.4$ with $\theta = 1, 2$.

The overall proportion of bias reduction for small-moderate and moderate-large heterogeneity in between-study variance from $\hat{\tau}^2_{MPL}$ to $\hat{\tau}^2_{ML}$ are quite similar. Again, similarly to small to moderate heterogeneity, the bias of overall log-odds-ratio for moderate large heterogeneity is the same for maximum likelihood and penalized likelihood method. Penalized maximum likelihood does perform better than standard ML in estimation of τ^2 and (marginally) coverage of the confidence interval for θ for all sample sizes N. However, this method still does not provide nominal 95% confidence level, apart from the case when studies have large sample sizes $N = 1000$. Thus, further understanding of the penalized maximum likelihood for log odds ratio and other effect measures from binary data is required. Particularly, when the sample sizes are not large. Higher order terms might matter. Further expansion of the score functions might be required. Much better method for confidence interval of overall effect measure is the new Profiled Q gamma based method. The main reason of good performance of Profiled Q gamma based method can be explained by having a least biased point estimator $\hat{\tau}^2_{CMP}$.

The DerSimonian and Laird estimator $\hat{\theta}_{DL}$ does not perform well at all in comparison to $\hat{\theta}_{ML}$, $\hat{\theta}_{MPL}$, $\hat{\theta}_{REML}$, $\hat{\theta}_{MP}$ and $\hat{\theta}_{CMP}$ for moderate-large heterogeneity. Thus, $\hat{\theta}_{DL}$ as well as $\hat{\theta}_{BD}$ were not include in the Figures. Among all estimators, the least biased estimator of between-study variance $\hat{\tau}^2_{CMP}$ provide a superior coverages for $\hat{\theta}_{CMP}$. Thus, $\hat{\tau}^2_{CMP}$ and $\hat{\theta}_{CMP}$ are recommended to use in standard additive random effects model. From the likelihood based estimators $\hat{\tau}^2_{MPL}$ and $\hat{\tau}^2_{REML}$ perform similarly. Thus, $\hat{\tau}^2_{MPL}$ is a feasible alternative to $\hat{\theta}_{REML}$. However, neither $\hat{\tau}^2_{REML}$ nor $\hat{\theta}_{REML}$ outperform $\hat{\tau}^2_{CMP}$.

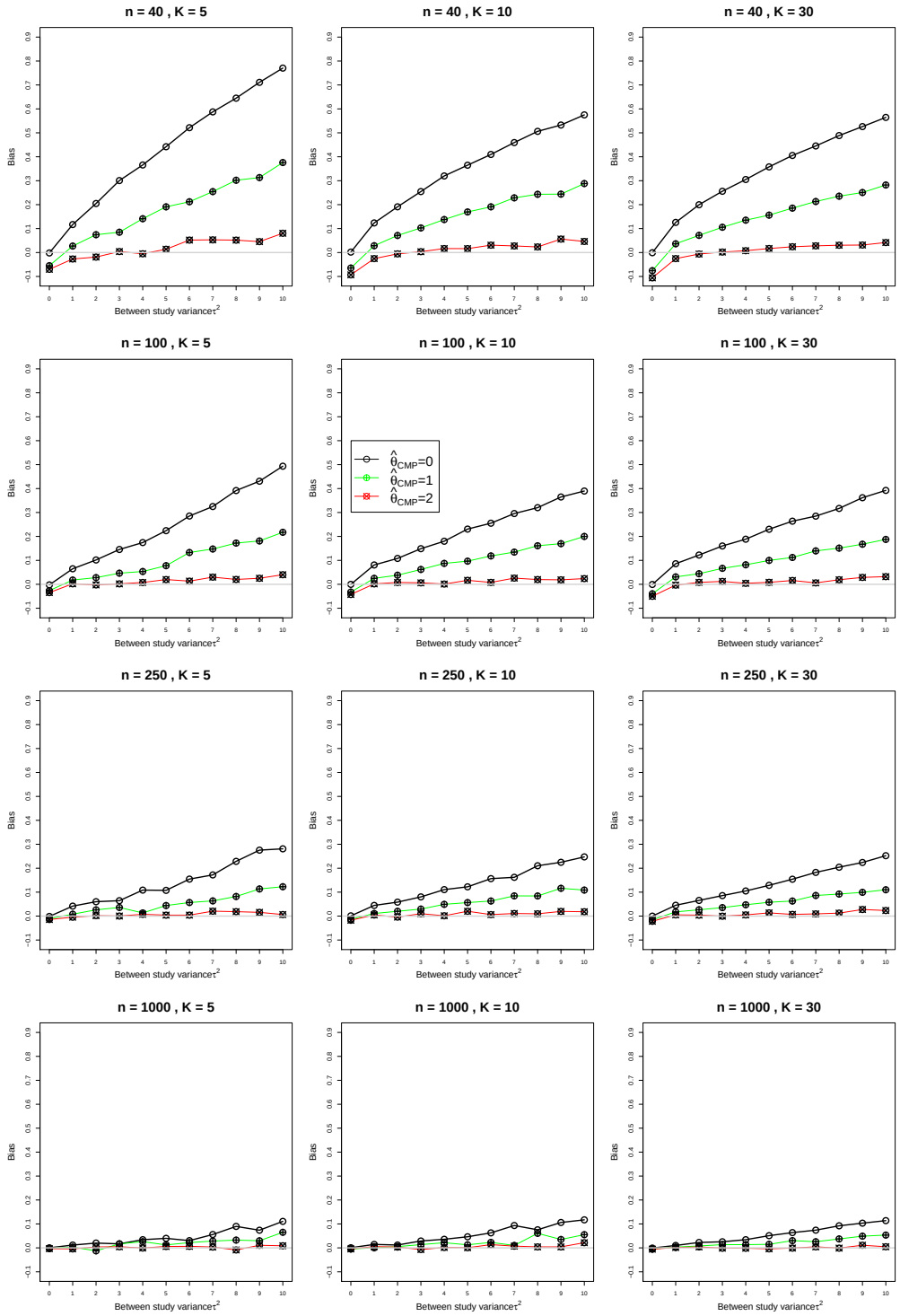


Figure 6.14: Bias of overall odds ratio θ_{IV} obtained from K studies by the inverse-variance method with the moment estimator $\hat{\tau^2}_{CMP}$ in the weights, for $p_{2i} = 0.1$, and $0 \leq \tau^2 \leq 10$. The biases are given for $\theta = 0$ (circles), $\theta = 1$ (circle plus), and $\theta = 2$ (circle cross). Light grey line at 0.

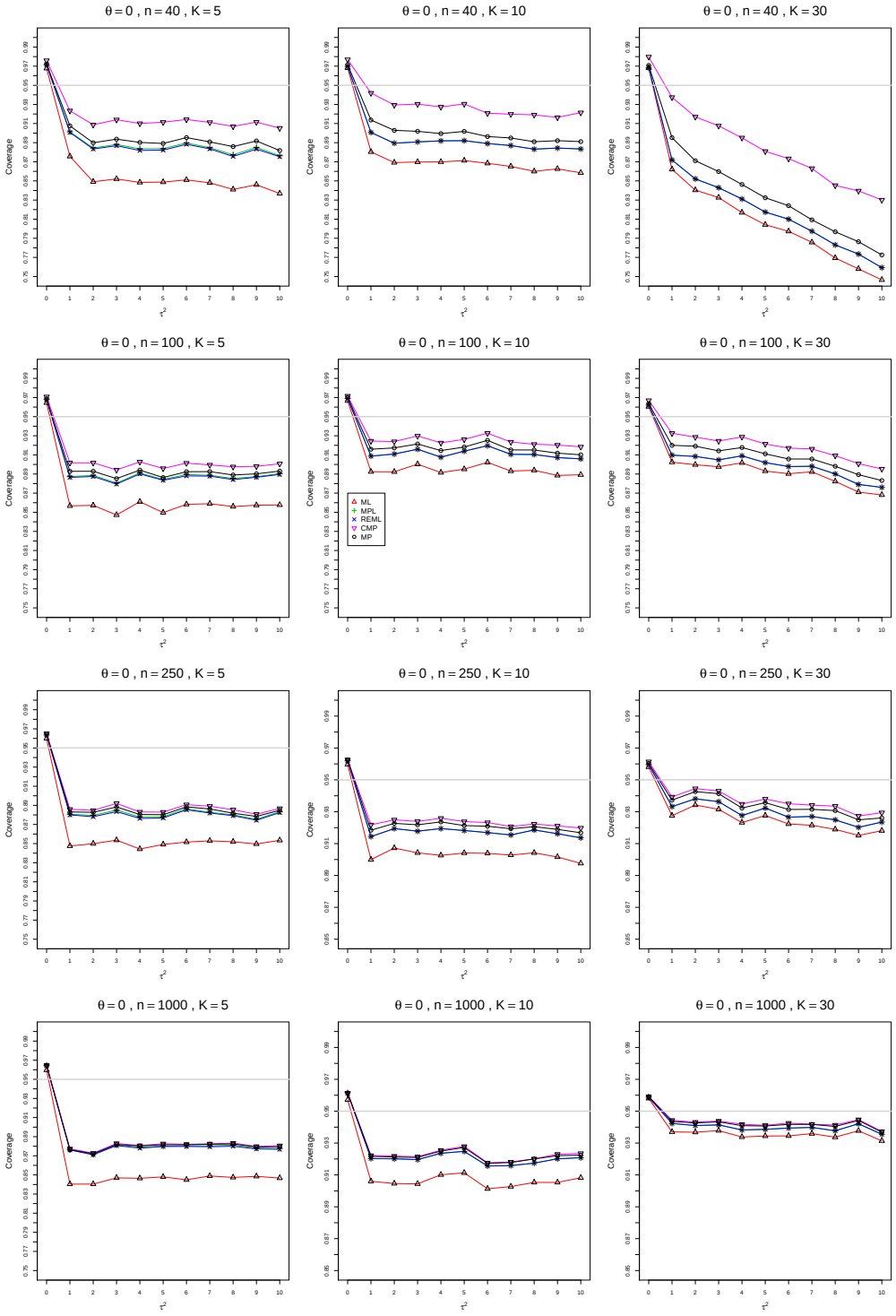


Figure 6.15: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_{RE}$ obtained from K studies by the inverse-variance method, for $p_{2i} = 0.1$, $\theta = 0$ and $0 \leq \tau^2 \leq 10$. The inverse-variance weights use the following estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

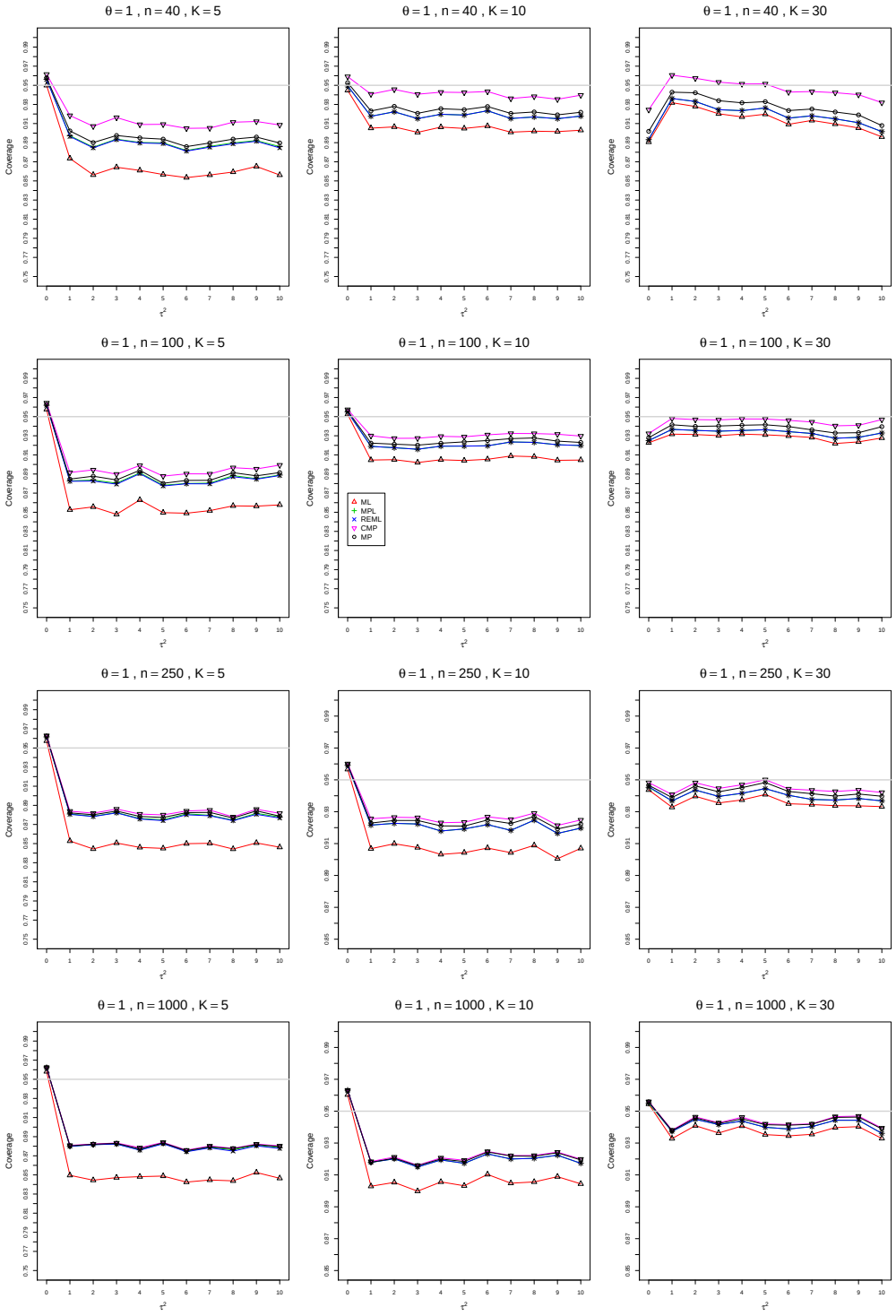


Figure 6.16: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_{RE}$ obtained from K studies by the inverse-variance method, for $p_{2i} = 0.1$, $\theta = 1$ and $0 \leq \tau^2 \leq 10$. The inverse-variance weights use the following estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

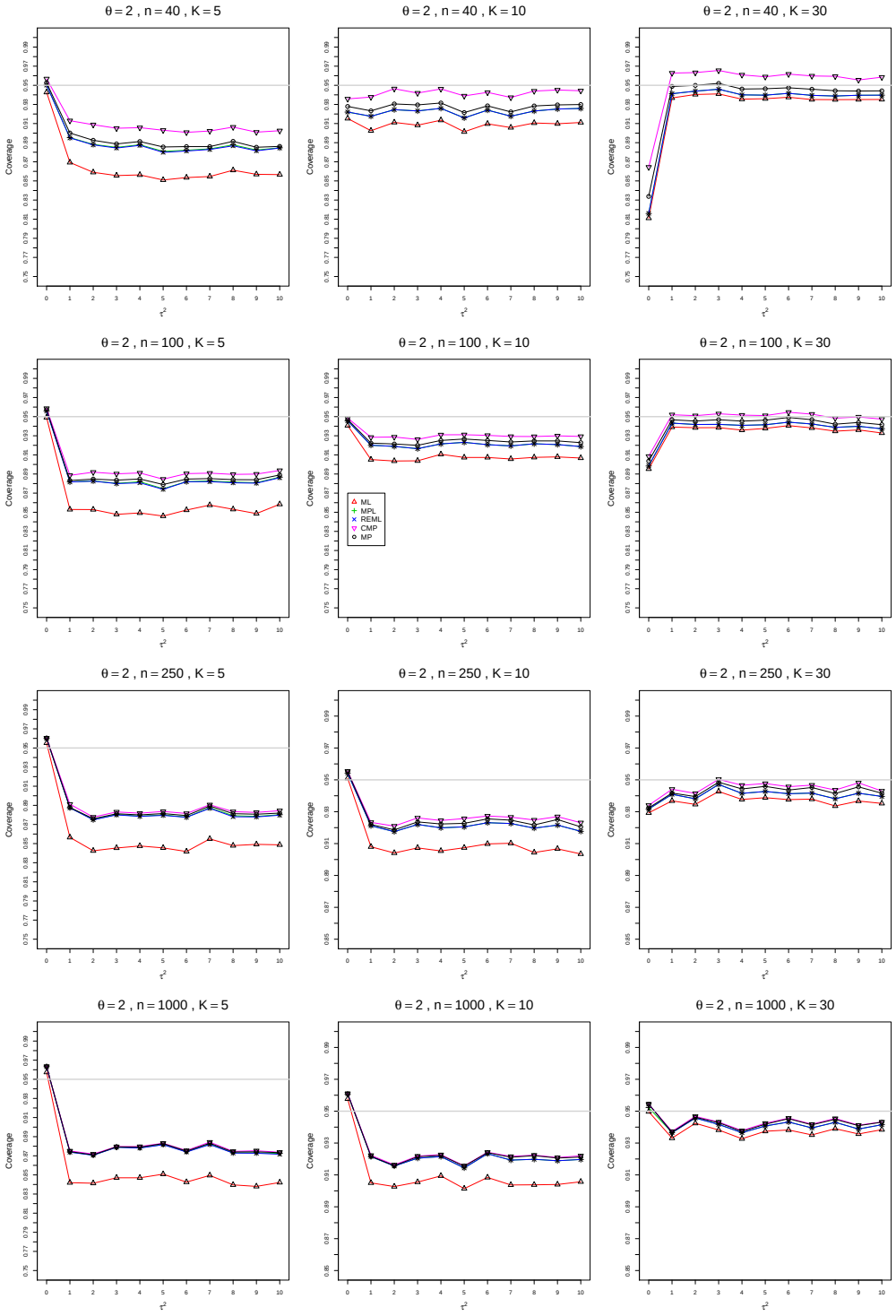


Figure 6.17: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_{RE}$ obtained from K studies by the inverse-variance method, for $p_{2i} = 0.1$, $\theta = 2$ and $0 \leq \tau^2 \leq 10$. The inverse-variance weights use the following estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

6.5 Simulation study for estimating τ^2 from a model with a pair of beta-binomial distributions

6.5.1 Simulation study

Similar to simulation study in Chapter 4, in this section we provide a simulation study to access the performance of point and interval estimators of random effect parameter τ^2 and the combined LOR θ in beta-binomial model of meta-analysis. Contrasting to simulation study in Chapter 4, we estimate between-study variance as in REM rather than the intra-cluster correlation as in ODM. Our main goal is to check how robust are the methods under misspecification between the standard additive REM and multiplicative ODM. In reality, we may never know which model is applicable to a particular dataset. We assess seven point estimators of τ^2 in respect to their bias: the DerSimonian and Laird method (6.3.1), the Mandel-Paule inspired method τ^2_{MP} - solution of equation 6.3.2, the corrected Mandel-Paule estimator based on the gamma approximation to Q distribution τ^2_{CMP} - solution of equation 6.3.6, the ML method (6.3.9), the REML method (6.3.11), the penalized ML method by Kosmidis et al. (2017) which maximize likelihood (6.3.13) and the BD-based method (6.3.14). We also assess four related confidence intervals for τ^2 (6.3.3), (6.3.7), (6.3.12) and (6.3.15) in respect to their coverage at the 95% confidence level. The combined odds ratio or its log is obtained by inverse-variance method $\hat{\theta}_w = \sum w_i(\tau^2)\hat{\theta}_i / \sum w_i(\tau^2)$. We assess inverse variance method of obtaining combined effect $\hat{\theta}$ for bias and for coverage.

6.5.2 Simulation design

Sizes of the control and treatment groups were taken equal $n_{1i} = n_{2i} = n_i$ across K studies. For a given probability p_{2i}, the number of cases in the control group X_{2i} was simulated from a beta-binomial (n_{2i}, p_{2i}, ρ) distribution using the R package *emdbook* (Bolker, 2011). The number of cases in the treatment group X_{1i} was generated from a beta-binomial (n_{1i}, p_{1i}, ρ) distribution with $p_{1i} = p_{2i}\exp(\theta)/(1 - p_{2i} + p_{2i}\exp(\theta))$ for a given LOR value of θ. When

$\rho = 0$, the numbers of events for treatment and control arm X_{ij} were generated from binomial distributions with sample size n_{ij} and probabilities p_{ij}, preserving the above relationship between the probabilities in the treatment and control arms.

The following configurations of parameters were included in the simulations. The number of studies $K = (5, 10, 30)$; average sample sizes in each arm are $n = (30, 100, 250, 1000)$; overdispersion parameter ρ varies between 0 and 0.1 (small to moderate heterogeneity) with steps 0.01, and between 0.1 and 0.3 in steps 0.05 (moderate to large heterogeneity). The corresponding true value of between-study variance is obtained through (6.2.7). The values of LOR θ vary from 0 to 2 in steps of 1. The probability in the control group p_{2i} takes values 0.1, 0.2, 0.4. A total of 10000 simulations were produced for each combination.

6.5.3 Simulation results

Figures 6.18 and 6.19 show the bias and coverage of τ^2 estimated by the five methods mentioned above for different combinations of K and n for the case of $p_{2i} \equiv 0.1$ and $\theta = 0$ and varying values of $0 \leq \rho \leq 0.3$. The corresponding true values of τ^2 can be obtained from relationship (6.2.7). The bias and coverage of true log odds ratio θ estimated by the inverse-variance (θ_{IV}) for values of $\theta = 0$, 1, 2, are shown in Figures 6.20 - 6.23, respectively.

Bias and coverage in estimation of between-study variance

All estimates of between-study variance have a non-linear bias that increases in τ^2, Figures 6.18 for $p_{2i} = 0.1$, C.63 and C.64 in Appendix for $p_{2i} = 0.2$ and $p_{2i} = 0.4$. The bias of between-study variance estimate obtained from Profiled Breslow-Day method and Corrected Mantel-Paule method is always positive. Whereas, the bias of estimates from standard methods methods based on maximising the likelihood vary with combination of N and K. The sign of the bias changes from switching from small-moderate sample sizes ($N = 40, 100$) to moderate large sample sizes ($N = 250, 1000$). The Der-Simonian and Laird method did not perform well when the assumptions of standard additive random effects model have been satisfied (see simulation results in section 6.4.1 and 6.4.1). Thus, we do not expect it perform well in beta-binomial

model when assumptions of standard additive random effects model are not satisfied (see Figures 6.18 for $p_{2i} = 0.1$, C.63 and C.64 in Appendix for $p_{2i} = 0.2$ and $p_{2i} = 0.4$). Coming to coverage of τ^2 (Figures 6.19 and C.65, C.66), the Q-profile and and Profile-likelihood based estimation seem to perform well for $N \leq 250$ and $K \leq 10$. However, the coverage of these methods deteriorates with increasing either N or K. The best scenario for the Q-profile and and Profile-likelihood based method is $N = 100$. However, we can not only rely on a particular set of simulations. In reality, we would never get the studies with the same samples studies across all studies. Considering asymptotics with increasing N or K, none of the four methods perform well. The Breslow method that performed well in Chapter 4 suffers from the transformation biases of probabilities that are used through relationship 6.2.7. Thus, it provides a biased point and interval estimate of between-study variance.

In summary, we can clearly see that the misspecification problems impact the estimation of between-study variance. $\hat{\tau}^2_{BD}$ is aimed at the beta-binomial model, so no surprise that it is good in ODM. However, when standard methods are used, all methods are not robust to misspecification of random effects model.

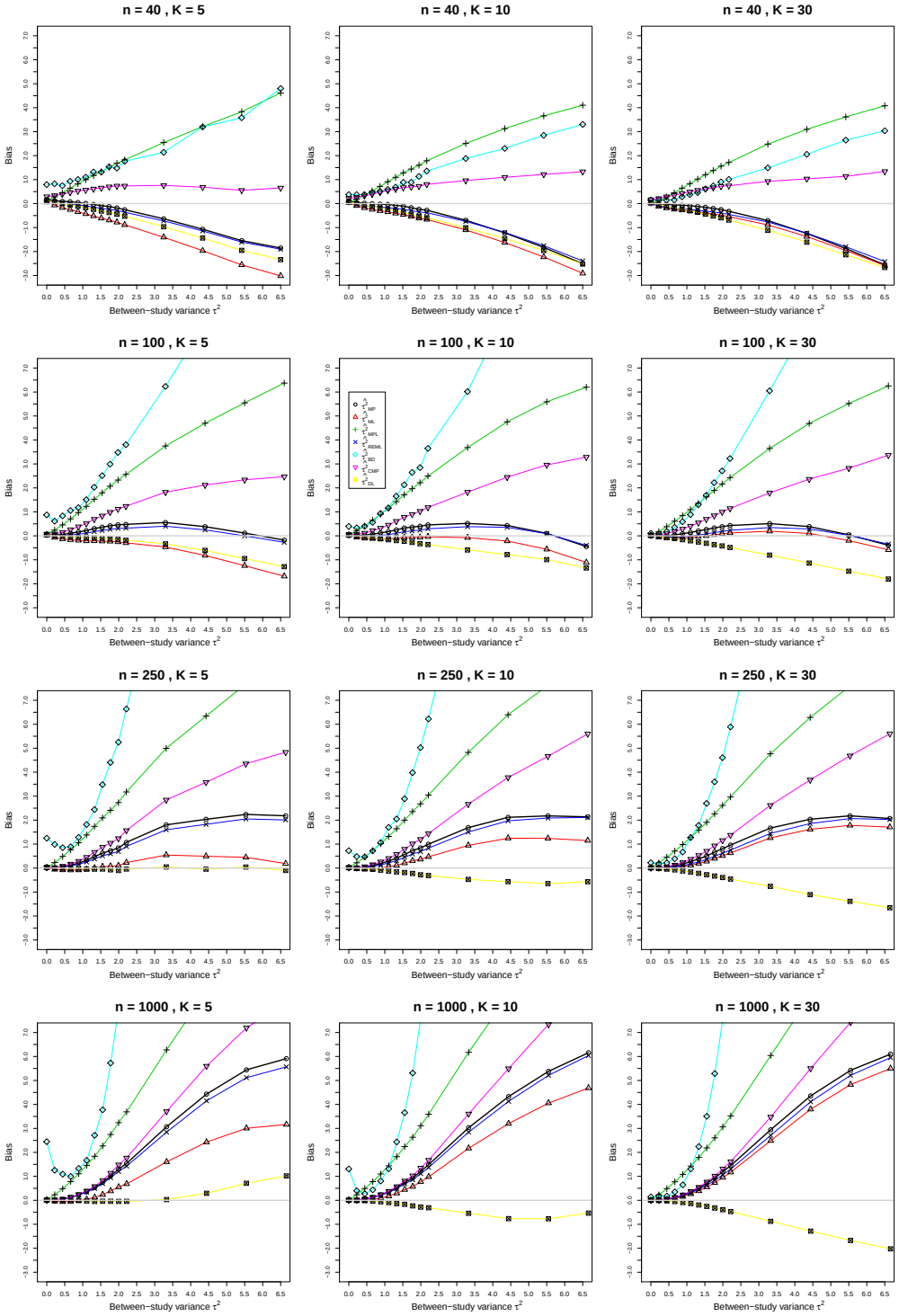


Figure 6.18: Bias of the between-study variance τ^2 obtained from K studies in beta-binomial model for $p_{2i} = 0.1$,$\theta = 0$ and $0 \leq \rho \leq 0.3$. The estimators of τ^2: yellow square with crosses – Der-Simonian and Laird estimator $\hat{\tau}^2_{DL}$, red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, dark blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, light blue diamonds – Profiled-Breslow-Day estimator $\hat{\tau}^2_{BD}$,pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

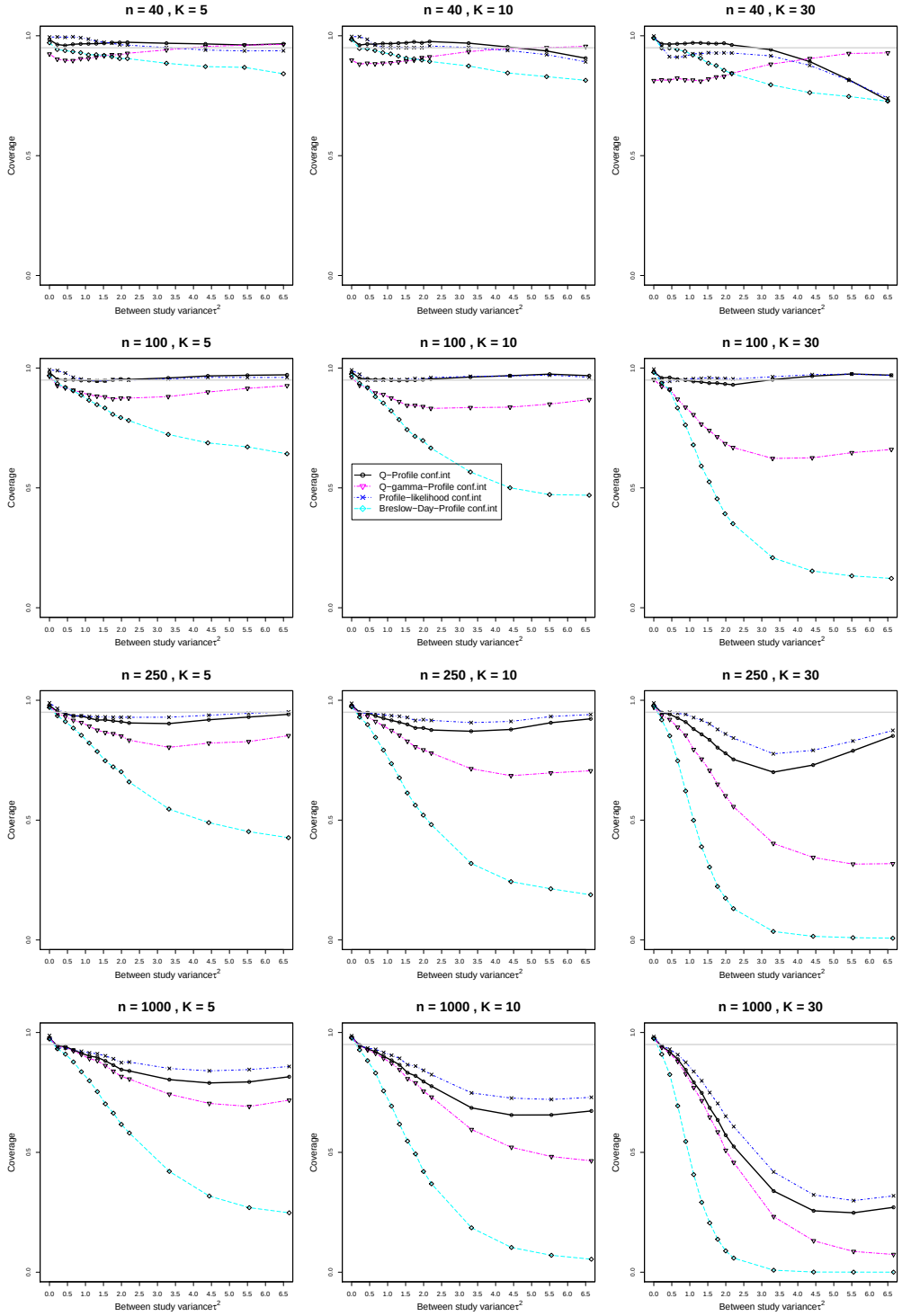


Figure 6.19: Coverage at the nominal confidence level of 0.95 of the between-study variance τ^2 estimated from K studies in beta-binomial model for $p_{2i} = 0.1$, $\theta = 0$ and $0 \leq \rho \leq 0.3$. Interval estimation methods: black circles – Q-profile confidence interval for τ^2 based on χ^2 distribution, pink inverse triangles – Q-profile confidence interval for τ^2 based on $\Gamma_{r(\tau^2),\lambda(\tau^2)}$ distribution), dark blue crosses – Profile likelihood confidence intervals, light blue diamonds – Breslow-Day-Profile confidence intervals. Light grey line at 0.95.

Bias and coverage in estimation of overall log-odds ratio θ

Similarly to results from Chapter 4, the bias of estimated odds ratio $\hat{\theta}$ was practically the same regardless of a method used for estimation of intra-class correlation τ^2. Without loss of generality, we plotted the results for bias of $\hat{\theta}$ obtained when using the moment estimator $\hat{\rho}_{CMP}$ in Figure 6.20 for values of log-odds $\theta = 0, 1$ and 2. We used the moment estimator $\hat{\rho}_{CMP}$, since it is the least unbiased estimator of between-study variance in standard additive random effects model (see section 6.4.1 and 6.4.1 for results of simulations from standard additive random effects model). Similarly to results from Chapter 4, there is no bias when $\theta = 0$, i.e. when the probabilities of an event in two arms are the same, but the bias clearly increases positively with increasing values of θ, and/or τ^2. This is conflicting with results of ODM, where the bias of θ increases negatively with increasing values of θ, and/or τ^2. Whenever applying the standard methods for estimation of between-study variance to beta-binomial model, estimates of between-study variance obtains positive bias which results in positive bias in θ. In ODM, the bias of random effect parameter is negative, hence the bias of θ is also negative. For more details see the Sections 4.6.2 and 4.6.2 in Chapter 4. Similar results for larger probabilities are provided in Appendix (see C.67 and C.71 for probabilities $p_{2i} = 0.2$ and $p_{2i} = 0.4$ with values of log-odds $\theta = 0, 1$ and 2).

For coverages of overall log-odds ratio θ, see the Figures 6.21, 6.22 and 6.23 for $\theta = 0$, 1 and 2 respectively. The method used for estimation of intra-class correlation τ^2 plays an important role in estimation of variance, and therefore the coverage of the odds-ratio. The Figures 6.21, 6.22 and 6.23 clearly show the reduced coverage of OR due to the transformation bias discussed in section 4.6.2 and 4.6.2 in Chapter 4. Among all the methods, Profiled Breslow-Day method shows the least reductions in the coverage of OR. This might be due to the very biased estimator of between-study variance $\hat{\tau}^2_{BD}$ which increases the variance hence provides the widest the confidence interval. The results of coverages for larger probabilities are provided in Appendix (see C.68, C.69 and C.70 and C.72, C.73 and C.74 for $p_{2i} = 0.2$ and $p_{2i} = 0.4$).

Overall, the standard methods for estimation of between-study variance are not robust to misspecification of REM models in meta-analysis. In standard additive random effects models, the performance of standard methods were studied by Viechtbauer (2005). However, estimating the between-study variance in beta-binomial model might leads to wrong inference about the size of heterogeneity, sizes of overall effect measure and its confidence interval. The transformation bias is an additional problem, which plays an important role when we have random effects model. $\hat{\tau}^2_{CMP}$ results in the best coverage of $\hat{\theta}_{RE}$ for $K \leq 10$, so the most robust.

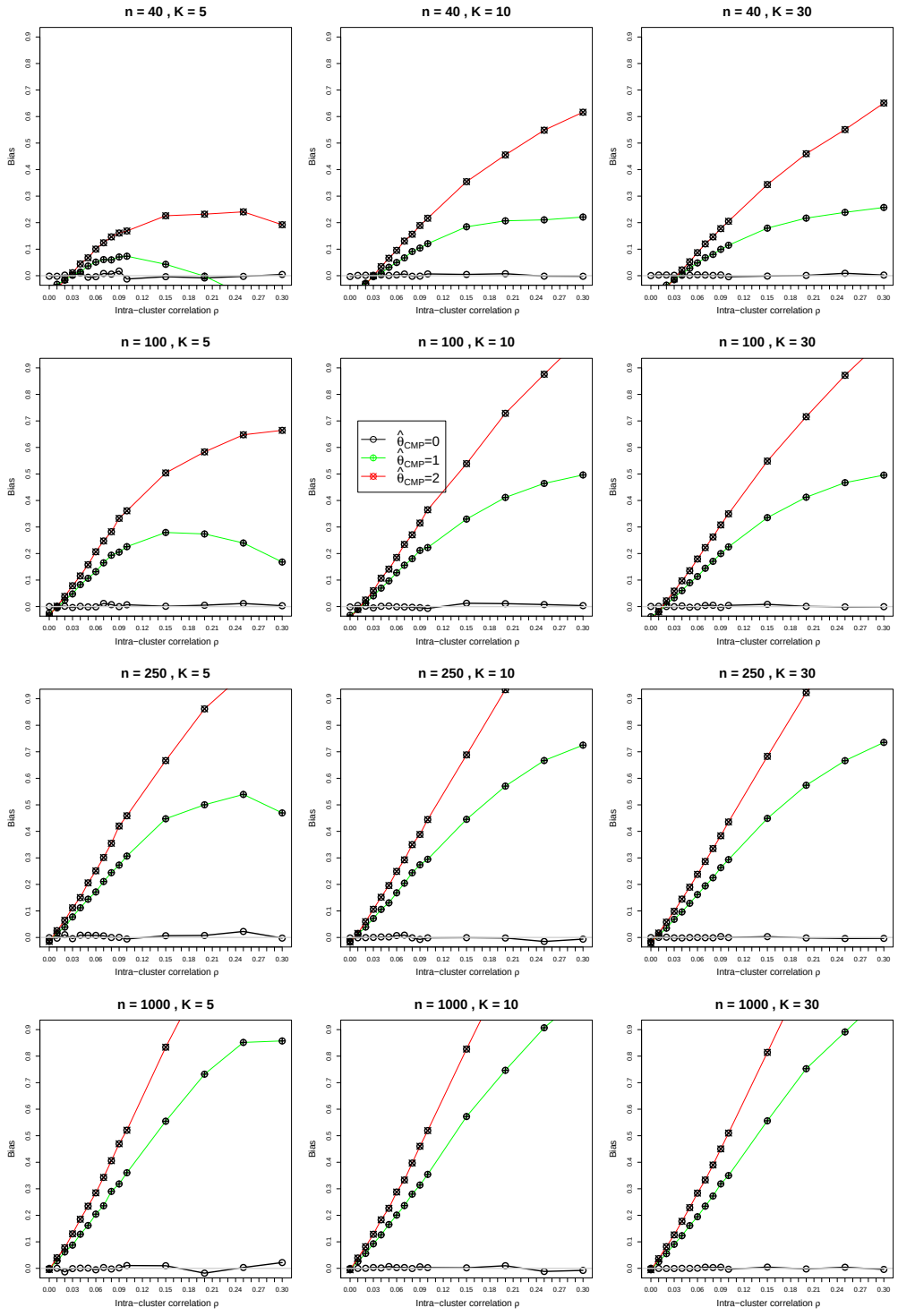


Figure 6.20: Bias of overall log odds ratio θ_{IV} obtained from K studies by the inverse-variance method with the moment estimator $\hat{\rho}_{CMP}$ in the weights, for $p_{2i} = 0.1$, and $0 \leq \rho \leq 0.3$. The biases are given for $\theta = 0$ (circles), $\theta = 1$ (circle plus), and $\theta = 2$ (circle cross). Light grey line at 0.

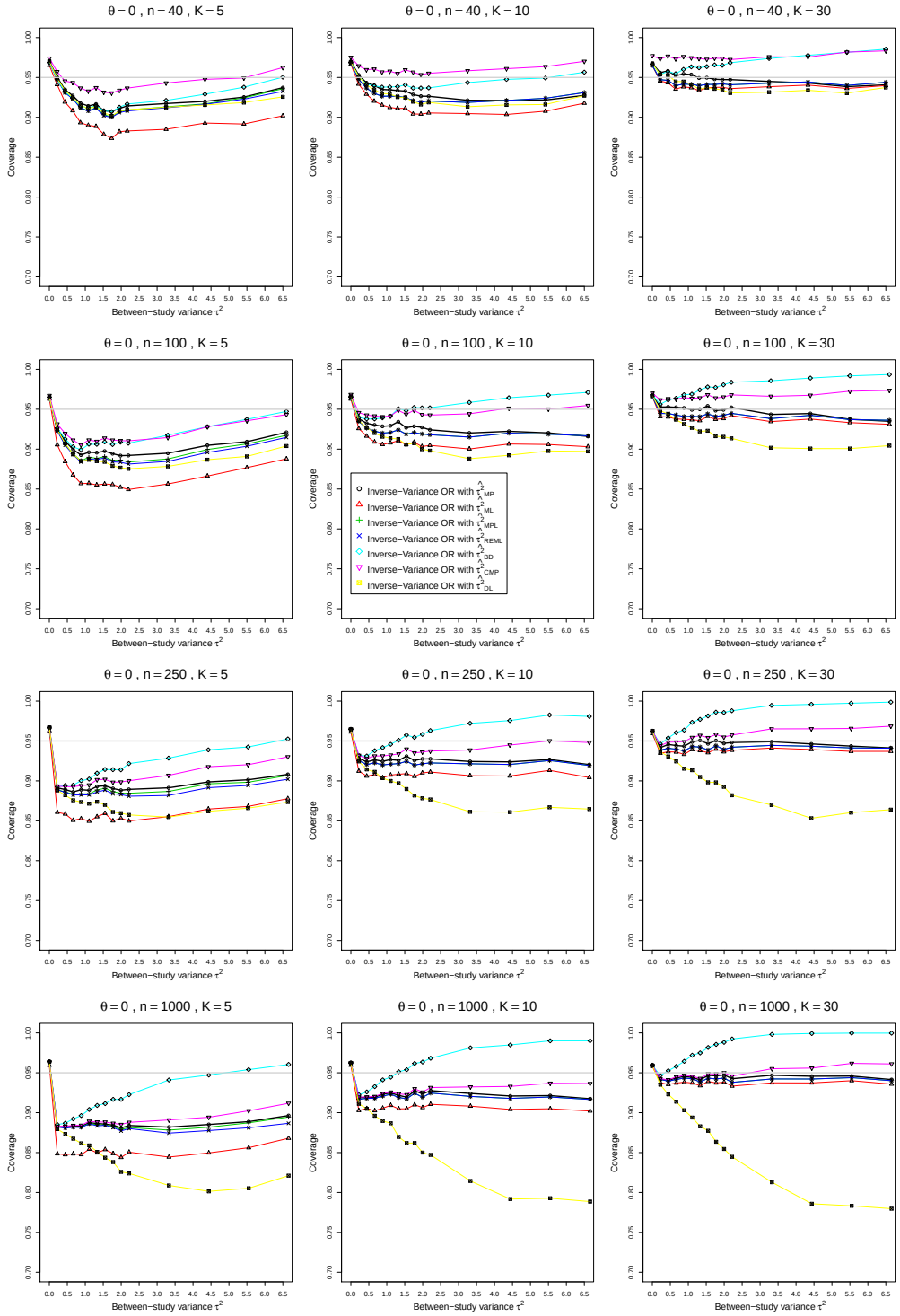


Figure 6.21: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio θ obtained from K studies by the inverse-variance method, for $p_{2i} = 0.1$, $\theta = 0$ and $0 \leq \rho \leq 0.3$ (equivalent to $0 \leq \tau^2 \leq 6.5$). The inverse-variance weights use the following estimators of τ^2: yellow square with crosses – Der-Simonian and Laird estimator $\hat{\tau}^2_{DL}$, red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, dark blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, light blue diamonds – Profiled-Breslow-Day estimator $\hat{\tau}^2_{BD}$,pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

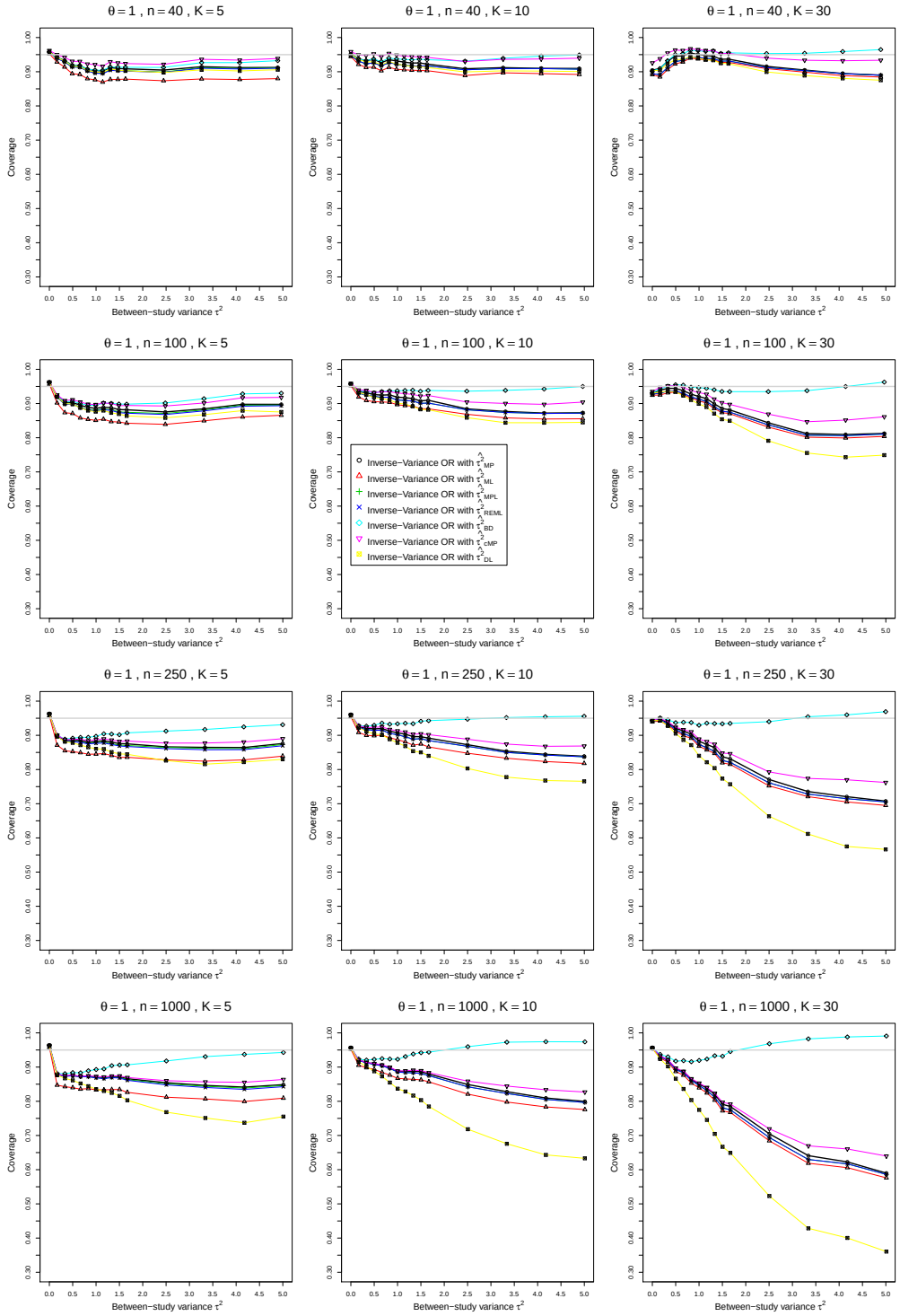


Figure 6.22: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio θ obtained from K studies by the inverse-variance method, for $p_{2i} = 0.1$, $\theta = 1$ and $0 \leq \rho \leq 0.3$ (equivalent to $0 \leq \tau^2 \leq 5$). The inverse-variance weights use the following estimators of τ^2: yellow square with crosses – Der-Simonian and Laird estimator $\hat{\tau}^2_{DL}$, red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, dark blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, light blue diamonds – Profiled-Breslow-Day estimator $\hat{\tau}^2_{BD}$,pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

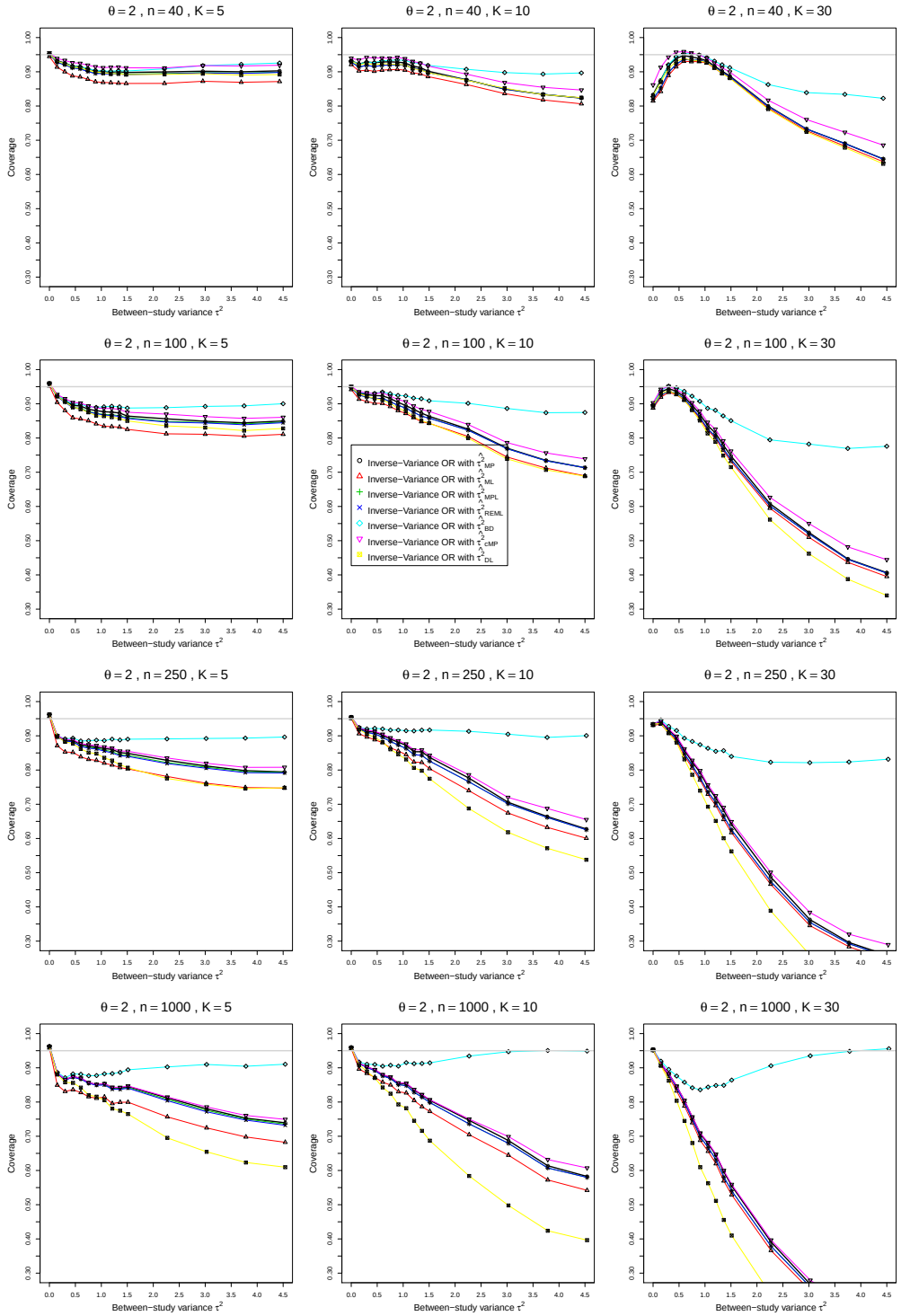


Figure 6.23: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio θ obtained from K studies by the inverse-variance method, for $p_{2i} = 0.1$, $\theta = 2$ and $0 \leq \rho \leq 0.3$ (equivalent to $0 \leq \tau^2 \leq 4.5$). The inverse-variance weights use the following estimators of τ^2: yellow square with crosses – Der-Simonian and Laird estimator $\hat{\tau}^2_{DL}$, red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, dark blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, light blue diamonds – Profiled-Breslow-Day estimator $\hat{\tau}^2_{BD}$,pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

6.6 Example: effects of diuretics on pre-eclampsia

A meta-analysis of nine trials on the effect of diuretics on pre-eclampsia (Collins et al., 1985) was studied in Chapters 4 and 5. The same example is re-analysed in this Chapter in order to compare the results using standard and new methods developed for standard additive random effects model, for overdispersed random effects model and generalized linear mixed effect models.

In addition to the results from Chapter 4 and Chapter 5, the results include maximum likelihood estimates of τ^2 and θ, penalized maximum likelihood estimates of τ^2 and θ proposed by Kosmidis et al. (2017), Mandel-Paule estimates of τ^2 and θ, Corrected Mandel-Paule estimates of τ^2 and θ and Profiled-Breslow-Day estimates of τ^2 and θ. For likelihood based estimates, the confidence interval for τ^2 is obtained by profile-likelihood method. All the results are shown in table 6.1.

From the results in table 6.1, we would not believe the estimate of between-study variance from Profiled-Breslow-Day method. Our simulations have shown that $\hat{\tau}^2_{BD}$ is very biased due to the bias in probabilities that are used through the correspondence 6.2.7. According to our simulation results the values of $\hat{\tau}^2_{BD}$ is too large on average. Similarly, the confidence interval of $\hat{\tau}^2$ based on Profiling the Breslow-Day test statistics is also biased. The penalized maximum likelihood estimate $\tau^2_{MPL} = 0.301$ of between-study variance is close to REML estimate $\tau^2_{MPL} = 0.300$. Hence the estimate of odds ratio and its confidence interval is pretty similar for MPL and REML. Kosmidis et al. (2017) states that $\hat{\tau}^2_{MPL}$ and $\hat{\tau}^2_{REML}$ are closely related. We have done simulations comparing $\hat{\tau}^2_{REML}$ and $\hat{\tau}^2_{MPL}$. The results which are not reported here have shown that $\hat{\tau}^2_{MPL}$ is just the same as $\hat{\tau}^2_{REML}$ for all combinations of N, K and τ^2. Also, our simulations have shown that $\hat{\tau}^2_{MPL}$ and $\hat{\tau}^2_{REML}$ are very similar. The bias of $\hat{\tau}^2_{REML}$ is similar to the bias of $\hat{\tau}^2_{MPL}$. Thus, the estimate of between-study variance $\hat{\tau}_{REML}$ is biased as well.

We have shown in the current Chapter that MPL method does not entirely eliminate the bias of $\hat{\tau}^2_{ML}$. The estimators $\hat{\tau}^2_{MPL}$ and $\hat{\tau}^2_{REML}$ still suffer from the negative biases which were visible

Table 6.1: Estimates and confidence intervals for the ICC ρ, for log odds ratios and for odds ratios diuretics in pre-eclampsia example; GLMM is the generalized linear mixed model, REM is the random effects and BB is the beta-binomial model. Heterogeneity parameters estimated are τ^2 in GLMM, and ρ in BB model. L and U are the lower and upper limits of the respective confidence intervals (CIs).

Model	Method	Hetero geneity	L	U	LOR	L	U	length of CI	OR	L	U
GLMM	UM.FS	0.254			-0.513	-0.923	-0.104	0.819	0.599	0.398	0.901
GLMM	UM.RS	0.264			-0.516	-0.930	-0.102	0.828	0.597	0.395	0.903
GLMM	CM.AL	0.165			-0.434	-0.777	-0.091	0.686	0.648	0.460	0.913
GLMM	CM.EL	0.260	-0.147(0)	0.667	-0.513	-0.927	-0.100	0.827	0.599	0.396	0.905
FEM		0.000			-0.398	-0.573	-0.223	0.530	0.672	0.564	0.800
REM	DL	0.230	0.072	2.202	-0.517	-0.916	-0.117	0.799	0.596	0.400	0.889
REM	REML	0.300	0.043	1.475	-0.518	-0.956	-0.080	0.876	0.596	0.384	0.923
REM	ML	0.239	0.041	1.499	-0.517	-0.921	-0.113	0.808	0.596	0.398	0.893
REM	MPL	0.301	0.043	1.475	-0.518	-0.956	-0.080	0.876	0.595	0.384	0.923
REM	MP	0.386	0.072	2.202	-0.518	-0.998	-0.037	0.961	0.596	0.368	0.963
REM	CMP	0.428	0.094	2.183	-0.517	-1.016	-0.018	0.998	0.596	0.362	0.983
REM	BD	0.478	0.098	37.78	-0.516	-1.038	0.005	1.033	0.597	0.354	1.005
BB	M&IV	0.008	0.002	0.095	-0.436	-0.792	-0.080	0.712	0.647	0.453	0.923
	M&MH				-0.427	-0.775	-0.080	0.695	0.652	0.461	0.923
BB	REML&IV	0.010	0.001	0.060	-0.447	-0.835	-0.059	0.776	0.640	0.434	0.942
	REML&MH				-0.431	-0.809	-0.053	0.756	0.650	0.445	0.949
BB	MP&IV	0.017	0.002	0.095	-0.469	-0.920	-0.018	0.902	0.626	0.399	0.982
	MP&MH				-0.459	-0.898	-0.020	0.879	0.632	0.407	0.981
BB	CMP&IV	0.018	0.003	0.094	-0.474	-0.942	-0.007	0.936	0.623	0.390	0.993
	CMP&MH				-0.472	-0.927	-0.016	0.911	0.624	0.396	0.984
BB	BD&IV	0.019	0.003	0.107	-0.475	-0.944	-0.006	0.938	0.622	0.389	0.994
	BD&MH				-0.463	-0.920	-0.021	0.899	0.630	0.399	0.980

from the results of simulations. Our simulations have shown that $\hat{\tau}^2_{MP}$ and $\hat{\tau}^2_{CMP}$ provide least biased estimates of between-study variance. Thus, in this example the values of $\hat{\tau}^2_{MP} = 0.386$ and $\hat{\tau}^2_{CMP} = 0.478$ are somewhat higher than the estimates from likelihood based methods. We believe that $\hat{\tau}^2_{CMP}$ corrects for the bias of $\hat{\tau}^2_{MP}$ and hence $\hat{\tau}^2_{CMP} > \hat{\tau}^2_{MP}$. Therefore, in addition to recommendation of the estimated ICC $\hat{\rho}_{BD} = 0.019$ and corresponding value of the pooled OR $\hat{\psi}_{IV} = 0.622$ with confidence interval $(0.389,\ 0.994)$ under beta-binomial model, we would recommend $\hat{\tau}^2_{CMP} = 0.428$ and corresponding value of the pooled OR $\hat{\psi}_{IV} = 0.596$ with confidence interval $(0.362,\ 0.983)$ as the least biased estimates of between-study variance in standard additive random effects model. These estimates are very similar. Generalized linear mixed effects models provide similar estimates of odds ratio with shorter confidence interval. Our recommendations are very theoretical. We still do not know which model to believe and which random effect component to estimate. It is important to develop the diagnostics against misspecification of dependency structure of Bernoulli variables.

6.7 Summary

In this Chapter, we compared the methods for estimation of random effect component from standard additive REM and multiplicative ODM. In standard additive REM, the binary data were generated similarly to the method of simulation study by Viechtbauer (2007) and to the method of simulation by Kosmidis et al. (2017) using logistic regression. In multiplicative ODM, the simulations were only generated for data simulated similarly to Viechtbauer (2007) with a pair of beta-binomial distributions instead of a pair of standard binomial distributions. Firstly, we compared the performance of methods for estimation of random effect component for both data simulated similarly to Viechtbauer (2007) and for data simulated similarly to Kosmidis et al. (2017) using logistic regression with binary data generated under assumptions of standard additive REM. Secondly, we compared the performance of methods for estimation of random effect component for data simulated similarly to Viechtbauer (2007) under assumptions of standard additive REM and multiplicative ODM. The two new methods for

estimation of between-study variance were studied through two different scenarios for data simulated similarly to Viechtbauer (2007). In the first scenario, the data were generated from a pair of binomial distributions with assumptions of standard REM. In the second scenario, instead of a pair of binomial distributions, we generated data from a pair of beta-binomial distributions. In addition to standard and new methods for estimation of between-study variance, we have also studied the bias correction to score function for maximum likelihood estimate of between-study variance proposed by Kosmidis et al. (2017). This bias correction is similar to penalization of the likelihood. Kosmidis et al. (2017) has shown by simulations that proposed penalized likelihood provides better coverage than maximum likelihood method for overall effect measure in case-control studies. Kosmidis et al. (2017) claims that the proposed method is universal in reducing the bias of maximum likelihood estimates in meta-analysis of continuous and binary outcomes. However, the behaviour of the bias reductions depends on the effect measure. The program written by Kosmidis et al. (2017) was used for analysis of log-odds ratio with the same structure of generating the data as in Viechtbauer (2007) and as in Abo-Zaid et al. (2013). The simulations for data generated by method similarly to Viechtbauer (2007) and by method simulated similarly to Kosmidis et al. (2017) using logistic regression were run for different scenarios in meta-analysis under assumptions of standard random effects model. The difference between the simulation method by Viechtbauer (2007) and simulation method by Kosmidis et al. (2017) is that the former is simulating data from the fixed intercept model and the latter from the random intercept model similar to Turner et al. (2000) and discussed in Chapter 5 of this thesis. Another difference between methods of simulation is that Kosmidis et al. (2017) uses logistic regression for estimation of study specific log-odds ratios. The results of both simulations have shown that bias correction improves the estimate of maximum likelihood by reducing its bias. However, the bias of between-study variance has not been completely eliminated. The bias is larger in the random intercept model than in fixed intercept model. This is due extra random effect and inconsistent within-study sample sizes in the random intercept model. Also, in both scenarios, the bias is larger for

sparse-data than for non-sparse data. The penalized likelihood is better than maximum likelihood estimate for between-study variance, but it is still biased. In our simulations, the bias of penalized maximum likelihood method and the bias of restricted maximum likelihood method are absolutely the same. The Profiled Breslow-Day method proposed in Chapter 4 did not result in a good estimator of between-study variance. This method is only recommended in ODM, but not in REM. The corrected Manted-Paule method has shown promising results providing the least estimate of between-study variance.

One of the important findings of this and previous Chapter is the importance of the misspecification between random effects models. All studied methods are not robust to misspecification of random effects model. The misspecification occurs due to wrong assumption of normality of random effects in beta-binomial model and wrong estimation of the random effect component. ODM model introduced in Chapter 4 is a better option. In ODM, we directly estimate intra-cluster correlation instead of between-study variance. However, for maximum-likelihood based methods, in ODM we still make assumption of normality across effects between studies. Comparing the maximum-likelihood based methods and methods based on method of moments, the latter are the better option since we avoid the distributional assumptions as in likelihood based methods. The similar misspecification appears in general case using GLMM. Litière et al. (2007) and Litière et al. (2008) studied the impact of misspecifying the random effects distribution on the maximum likelihood estimates in generalized linear mixed models. Litière et al. (2008) studied the replacement of normal random effects model by a non-parametric distribution. However, these models result in different consistency of MLE estimates. Other authors who studied misspecification problem in GLMM are Neuhaus et al. (1992), Verbeke and Lesaffre (1997), Agresti et al. (2004) and McCulloch and Neuhaus (2011). Which model to use in which scenario in meta-analysis of binary data is an open question ?! Robust diagnostics against misspecification of dependency structure of Bernoulli variables is required, and will be a subject of our future research.

Chapter 7

Conclusion

7.1 Summary of the thesis

The current thesis is motivated by the problems of overdispersion arising in modelling proportions for single studies and in meta-analysis. In meta-analysis, the proportions usually give rise to two sample effect measures (odds ratio, relative risk and risk difference). Due to ease of analysis following normalization and variance-stabilization, these measures are usually transformed from one scale to another. These transformations have been studied by many authors such as Cox (1983) and Bhaumik et al. (2012) for log-odds, or Kim and Taylor (1994) for arcsine transformation in the absence or the presence of overdispersion. Overdispersion is usually accounted for by correcting the variance of an effect measure such as log odds ratio or arcsine difference. However, correction of the variance does not solve all the problems of inference in single studies and in meta-analyses. The additional issues arise from the bias of order $O(1/n)$ for each effect measure. The bias appears from the non-linearity of the transformations such as log odds or arcsine. Transformed measures are meta-analysed by fixed or random effects models.

We studied the biases of order $O(1/n)$ for log-odds and arcsine transformation by theoretical derivations and by simulations. For arcsine transformation we managed to reduce the size of bias by adding the first and second order corrections. For log-odds, the corrections changes the magnitude of the bias but do not make it disappear. This is due to dependence

of corrections on probabilities which are biased themselves. The same bias is also present for log-odds ratio in standard random effects model. We have also studied the alternative representation of random effects model in multiplicative form for log odds ratio. We presented a model that used an intra-cluster correlation as a source of overdispersion. The intra-cluster correlation parameter is used as an inflation factor to the variance of FEM. We assumed the constant intra-cluster correlation parameter across the studies. This is the strong assumption, however common in meta-analysis with limited number of studies. We have also proposed new methods for estimation of intra-cluster correlation in ODM and between-study variance in REM by profiling the Breslow-Day test and correcting the Mandel-Paule method based on the improved approximation for Q statistics by gamma distribution.

The multiplicative overdispersed random effects model is an attractive counterpart to standard random effects model. However, the methods of the former model are still not fully unbiased. When the data is sparse in standard fixed effect model, the standard Mantel-Haenzsel method is superior to weighted inverse-variance approach (Breslow, 1981). In case of heterogeneous sparse data due to correlation within each arm, the inverse-variance method outperforms the corrected Mantel-Haenzsel method in terms of the bias and coverage. The low performance of Mantel-Haenzsel method in terms of the bias and coverage can be explained by the bias of estimators of the intra-cluster correlation. The same bias has lower effect on the bias of effect measure obtained by the inverse-variance method. When effect measure (log odds ratio) is far from zero, the inverse-variance and Mantel-Haenzsel methods also suffer from biases of order $O(1/n)$.

Generalized linear mixed effect models are alternatives to standard additive and multiplicative random effects models. Particularly, a conditional generalized linear mixed-effects model with exact likelihood that uses a non-central hypergeometric distribution within each study and normal distribution between studies is of interest. However, as we have shown in Chapter 5 it also suffers from the biases of order $O(1/n)$. In beta-binomial model (Chapter 5), this is due to violation of the non-central hypergeometric distribution assumption in the presence of

intra-cluster correlation and the bias of true unobserved conditioned odds ratio that is obtained by inverse of logit trasformation of the number of successes. The bias is smaller when the data was simulated from standard additive random effects model with a pair of binomial distributions within studies and normal distribution between studies. However for the latter model, the bias of order $O(1/n)$ does still exist. Thus, the use of NCHGN model is questionable. When the binary data is actually correlated, NCHGN might provide wrong inference. We need diagnostic tools to distinguish between correlated or independent binary data.

The approximation of non-central-hypergeometric distribution by binomial distribution does not perform well at all for both beta-binomial and additive random effects models. Hence, the use of this model is not recommended. The maximum likelihood estimates in standard random effects model do also suffer from the transformation biases of order $O(1/n)$. Kosmidis et al. (2017) believes that his bias correction eliminates the bias of order $O(1/n)$ and reduces the bias of maximum likelihood estimates. We assessed the bias correction proposed by Kosmidis et al. (2017) in standard REM and multiplicative ODM by simulation study. Our simulations show that in data generated through logistic regression, the methods for estimation of between-study variance produce larger biases than in data generated similar to simulations by Viechtbauer (2007). Probably this is because, in generation of data through logistic regression, two random effects are added to the model. One for intercept and one for covariate. In data generated similar to simulations by Viechtbauer (2007), we have only added one random effect to log-odds ratio. For both scenarios, the results from our simulations showed that the bias correction helps standard random effects model. However, the bias in estimates of between-study variance still exists in both sparse and non-sparse data. We have also assesed the standard and new methods for estimation of between-study variance in correlated binary data. The results of the latter simulations have shown that estimating the between-study variance when heterogeneity is introduced by dependence of Bernoulli variables leads to wrong inference about the overall effect measure. Thus, how to safeguard against misspecification problem in meta-analysis is not an easy question to answer. As we have shown,

use of generalized linear mixed effects model might not help when the log odds ratio is far from zero.

There are still so many problems that needs to be discovered within statistical inference in meta-analysis of binary data. In this thesis, we have concentrated our attention on transformations, overdispersion, methods of estimation and models for meta-analysis of binary data.

7.2 Practical issues and recommendations

The focus of this thesis is limited to the meta-analysis of binary data using log-odds and arcsine transformation. In practice, alternative effect measures such as relative risk or risk differences might be used instead. When choosing the effect measure, two most common aspects should be considered: firstly, the effect measure should be statistically appropriate and convenient to use, secondly, the effect measure should contain the useful clinical information (Sutton et al., 2000). The motivation of current thesis is that log odds ratio is commonly used in meta-analysis of binary data due to its attractive properties and the ease of interpretation. Alternatively, arcsine or difference of arcsine transformations is a variance stabilized counterpart to log-odds ratio and log-relative risk.

We have developed a multiplicative ODM model, which is a counterpart to standard REM. In this model, we have assumed a common intra-cluster correlation across all studies. This approach is common in randomized controlled trials. However, it becomes less efficient when there is a variation of intra-cluster correlations across the studies. If this is the case, the correlation should be estimated and then used within each study in adjusted inverse variance method. However, this leads to the question of estimation. We have used methods that assume a common intra-cluster correlation. Alternative methods are required. Our simulations also used equal sample sizes in treatment and control arm. The majority of studies are fairly balanced. However this is not true in general. We should not ignore the scenario when studies have unbalanced samples size. The inferential methods are expected to perform worse in

unbalanced case.

When using the standard methods for sparse data, the continuity corrections are added to cells with zero events. The most popular continuity corrections are 1/2 for log-odds and 3/8 for arcsine transformation are still preferred in most of the cases. In case of using the arcsine differences as an effect measure, we would recommend the use of our bias correction proposed in Chapter 3 for small values of intra-cluster correlation ($\rho \leq 0.06$). The log-odds transformation is biased of order $O(1/n)$ and we could not eliminate this bias. The further research is required for log-odds transformation in order to eliminate the higher order terms.

In meta-analysis of binary data, the standard REM and new ODM methods are the easiest and safest methods to implement for practitioner. Both models suffer from transformation biases and it is not clear which one to use for the data in hands. The random effect component in both models has to be estimated by either moment based method or likelihood based methods. Among all methods, we would recommend the proposed Corrected Mandel-Paule method for estimation of random effect component in both REM and ODM . The adjusted Mantel-Haenzsel method did not show any promising results. The inverse-variance method is recommended over the Mantel-Haenzsel method in ODM and REM. But both, the inverse variance and Mantel-Haenzsel method suffer from transformation bias when the log-odds ratio is far from zero.

An alternative is to use the exact non-central-hypergeometric likelihood. From the results of our simulations, we would not recommend the use of binomial approximation to non-central-hypergeometric distribution in GLMM. This method provided the worse case scenario results even when the assumption of binomially distributed variables within each study is satisfied. In practice, convergence problems might occur when trying to fit the saturated model in generalized linear mixed effect model with exact non-central-hypergeometric distribution. These convergence problems may result in singularity of variance-covariance matrices. Generalized linear mixed effect model with exact non-central-hypergeometric likelihood is also computationally difficult and time expensive. This is due to maximizing the marginal likelihood

obtained integrating out the random effects model.

7.3 Limitations and future research

When using log-odds ratio as an effect measure in meta-analysis, the standard methods for estimation of between-study variance are biased. The bias of between-study variance combined with the transformation bias of log-odds impacts the estimation of overall log-odds ratio and its confidence intervals. We have tried to eliminate the bias of order $O(1/N)$ in log-odds and arcsine difference. However, we have only succeed in bias reduction of arcsine differences. Kosmidis et al. (2017) proposed a method to eliminate the bias of order $O(1/N)$ in maximum likelihood based methods for various effect measures in meta-analysis. We have performed simulations with the method proposed by Kosmidis et al. (2017) for log-odds ratio. The bias correction improves the likelihood based estimates of between-study variance and overall effect measure. However, the bias of estimates have not been eliminated completely. The penalized maximum likelihood estimator is still biased. Further expansions of the score functions should be studied for bias reduction in maximum likelihood based estimates.

In ODM, multivariate extensions of the models and corresponding methods of estimation should be developed and studied for different effect measures. Moreover, in ODM, it would be worthwhile to develop an improved methods for estimation of intra-cluster correlation within each arm across the studies. This is possible if the distribution of Q or Breslow-Day statistics is obtained under random effects model. It is also important to create diagnostic tools to analyze the random effects distribution. Extension of ODM to regression models would be advantageous to incorporate covariates in the analysis. The intra-cluster dependence between Bernoulli variables may also be modelled through the covariates.

Recently, Chen et al. (2016) proposed a marginal beta-binomial model based on the composite likelihood approach to a pair of beta-binomial distributions. Similar to our ODM model, this model assumes a pair of beta-binomial distributions in meta-analysis of binary data. However,

the main difference between our ODM model and the model proposed by Chen et al. (2016) is the use of inverse variance methods versus composite maximum likelihood method. Similarly the model by Chen et al. (2016), Kuss (2015) suggests to use beta-binomial regression model with maximum likelihood estimation of composite likelihood from a pair of beta-binomial distributions as generalized linear model family with logit link function for log-odds ratio. The future research should include the comparison of model proposed and studied by Chen et al. (2016), model proposed by Kuss (2015) and our ODM model proposed in Chapter 4 in meta-analysis of binary outcomes from case-control studies.

In conclusion, the misspecification of a model is a very important issue. We have proposed ODM model, an alternative to standard REM model. We have also studied different methods for estimation of random effect component in ODM and REM. A method robust against misspecification of dependency between Bernoulli variables would be preferable. Generalised linear mixed models are a logical choice in meta-analysis of discrete data, when the misspecification of models is present.

Bibliography

Abo-Zaid, G., Guo, B., Deeks, J. J., Debray, T. P. A., Steyerberg, E. W., Moons, K. G., and Riley, R. D. (2013). Individual participant data meta-analyses should not ignore clustering. *Journal of clinical epidemiology*, 66(8):865–873.

Agresti, A., Caffo, B., and Ohman-Strickland, P. (2004). Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Computational Statistics & Data Analysis*, 47(3):639–653.

Agresti, A. and Hartzel, J. (2005). Analysis: Strategies for comparing treatments on a binary response with multi-centre data. *Tutorials in Biostatistics: Statistical Methods in Clinical Studies*, 1:397–421.

Akritas, M. G. and Papadatos, N. (2004). Heteroscedastic one-way anova and lack-of-fit tests. *Journal of the American Statistical Association*, 99(466):368–382.

Alanko, T. and Duffy, J. (1996). Compound binomial distributions for modelling consumption data. *The Statistician*, 45:269–286.

Anscombe, F. J. (1948). The transformation of poisson, binomial and negative-binomial data. *Biometrika*, 35:246–254.

Ashby, D., Hutton, J. L., and McGee, M. A. (1993). Simple bayesian analyses for case-control studies in cancer epidemiology. *The Statistician*, 42:385–397.

Bagheri, Z., Ayatollahi, S. M. T., and Jafari, P. (2011). Comparison of three tests of homogeneity of odds ratios in multicenter trials with unequal sample sizes within and among centers. *BMC medical research methodology*, 11(1):58.

Beijer, U., Wolf, A., and Fazel, S. (2012). Prevalence of tuberculosis, hepatitis c virus, and hiv in homeless people: a systematic review and meta-analysis. *The Lancet infectious diseases*, 12(11):859–870.

Berkey, C. S., Hoaglin, D. C., Mosteller, F., and Colditz, G. A. (1995). A random-effects regression model for meta-analysis. *Statistics in medicine*, 14(4):395–411.

Bhaumik, D. K., Amatya, A., Normand, S. L. T., Greenhouse, J., Kaizar, E., Neelon, B., and Gibbons, R. D. (2012). Meta-analysis of rare binary adverse event data. *Journal of the American Statistical Association*, 107(498):555–567.

Bibby, B. and Væth, M. (2011). The two-dimensional beta binomial distribution. *Statistics and Probability Letters*, 81:884–891.

Biggerstaff, B. and Tweedie, R. (1997). Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. *Statistics in Medicine*, 16:753–768.

Biggerstaff, B. J. and Jackson, D. (2008). The exact distribution of cochran's heterogeneity statistic in one-way random effects meta-analysis. *Statistics in medicine*, 27(29):6093–6110.

Biggerstaff, B. J., Tweedie, R. L., et al. (1997). Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. *Statistics in medicine*, 16(7):753–768.

Biswas, A. and Hwang, J. S. (2010). Distribution of odds ratio in $2\times$ 2 contingency table: Adjustment for correlation. *Journal of biopharmaceutical statistics*, 21(1):136–150.

Böhning, D., Malzahn, U., Dietz, E., Schlattmann, P., Viwatwongkasem, C., and Biggeri, A. (2002). Some general points in estimating heterogeneity variance with the dersimonian–laird estimator. *Biostatistics*, 3(4):445–457.

Böhning, D. and Mylona, K.and Kimber, A. (2015). Meta-analysis of clinical trials with rare events. *Biometrical Journal*, 57(4):633–648.

Böhning, D. and Sarol, J. (2000). Estimating risk difference in multicenter studies under baseline-risk heterogeneity. *Biometrics*, 56(1):304–308.

Böhning, D. and Viwatwongkasem, C. (2005). Revisiting proportion estimators. *Statistical methods in medical research*, 14(2):147–169.

Bolker, B. (2011). emdbook: Support functions and data for book "ecological models and data". *R package, version 2011*, 1(3).

Boos, D. D. and Stefanski, L. (2013). *Large Sample Results for Likelihood-Based Methods*. Springer, New York, NY, USA.

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

Breslow, N. (1981). Odds ratio estimators when the data are sparse. *Biometrika*, 68(1):73–84.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*, 88(421):9–25.

Breslow, N. E. and Day, N. E. (1980). *Statistical methods in cancer research. Vol. 1. The analysis of case-control studies.*, volume 1, No. 32. Distributed for IARC by WHO, Geneva, Switzerland.

Breslow, N. E. and Liang, K. Y. (1982). The variance of the mantel-haenszel estimator. *Biometrics*, 38:943–952.

Brockwell, S. E. and Gordon, I. R. (2001). A comparison of statistical methods for meta-analysis. *Statistics in medicine*, 20(6):825–840.

Cai, T., Parast, L., and Ryan, L. (2010). Meta-analysis for rare events. *Statistics in medicine*, 29(20):2078–2089.

Chaganty, N. R. and Joe, H. (2004). Efficiency of generalized estimating equations for binary responses. *J. R. Statist. Soc. B*, 66(4):851–860.

Chen, S. F. (2012). *Heterogeneity issues in the meta-analysis of cluster randomization trials.* PhD thesis, Electronic Thesis and Dissertation Repository. 572., http://ir.lib.uwo.ca/etd/572, The University of Western Ontario.

Chen, X. and Xie, M.-g. (2014). A split-and-conquer approach for analysis of extraordinarily large data. *Statistica Sinica*, 24:1655–1684.

Chen, Y., Hong, C., Ning, Y., and Su, X. (2016). Meta-analysis of studies with bivariate binary outcomes: a marginal beta-binomial model approach. *Statistics in medicine*, 35(1):21–40.

Chu, H., Nie, L., Chen, Y., Huang, Y., and Sun, W. (2012). Bivariate random effects models for meta-analysis of comparative studies with binary outcomes: Methods for the absolute risk difference and relative risk. *Statistical Methods in Medical Research*, 21(6):621–633.

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

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

Cohen, J. M. (1988). *Statistical power analysis for the behavioral sciences, 2nd ed.* Lawrence Erlbaum Associates, Hillsdale, NJ.

Collett, D. (1991). *Modelling binary data, 2nd ed.* London: Chapman & Hall/CRC.

Collins, R., Yusuf, S., and Peto, R. (1985). Overview of randomised trials of diuretics in pregnancy. *Br Med J (Clin Res Ed)*, 290(6461):17–23.

Connolly, M. A. and Liang, K. Y. (1988). Conditional logistic regression models for correlated binary data. *Biometrika*, 75(3):501–506.

Cox, D. (1983). Some remarks on overdispersion. *Biometrika*, 70(2):269–274.

Crowder, M. J. (1978). Beta-binomial anova for proportions. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 27:34–37.

Crowder, M. J. (1979). Inference about the intraclass correlation coefficient in the beta-binomial anova for proportions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41:230–234.

Darlington, G. A. and Donner, A. (2007). Meta-analysis of community-based cluster randomization trials with binary outcomes. *Clinical Trials*, 4(5):491–498.

Demidenko, E. (2004). *Mixed models: theory and applications*. John Wiley & Sons; Hoboken, NJ.

Demirtas, H., Hedeker, D., and Kapur, K. (2009). A comparative study on most commonly used correlated binary data generation methods. *Advances and Applications in Statistical Sciences*, 1(1):44–55.

DerSimonian, R. and Kacker, R. (2007). Random-effects model for meta-analysis of clinical trials: an update. *Contemporary clinical trials*, 28(2):105–114.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled clinical trials*, 7(3):177–188.

Donald, A. and Donner, A. (1987). Adjustments to the mantel–haenszel chi-square statistic and odds ratio variance estimator when the data are clustered. *Statistics in Medicine*, 6(4):491–499.

Donner, A. and Klar, N. (2002). Issues in the meta-analysis of cluster randomized trials. *Statistics in medicine*, 21(19):2971–2980.

Donner, A., Piaggio, G., and Villar, J. (2001). Statistical methods for the meta-analysis of cluster randomization trials. *Statistical methods in medical research*, 10(5):325–338.

Eldridge, S. M., Ashby, D., Feder, G. S., Rudnicka, A. R., and Ukoumunne, O. C. (2004). Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clinical trials*, 1(1):80–90.

Eldridge, S. M., Ukoumunne, O. C., and Carlin, J. B. (2009). The intra-cluster correlation coefficient in cluster randomized trials: A review of definitions. *International Statistical Review*, 77(3):378–394.

Emerson, J. D. (1994). Combining estimates of the odds ratio: the state of the art. *Statistical Methods in Medical Research*, 3(2):157–178.

Emrich, L. J. and Piedmonte, M. R. (1991). A method for generating high-dimensional multivariate binary variates. *The American Statistician*, 45:302–304.

Erez, A., Bloom, M. C., and Wells, M. T. (1996). Using random rather than fixed effects models in meta-analysis: Implications for situational specificity and validity generalization. *Personnel Psychology*, 49(2):275–306.

Fisher, R. A. (1932). Statistical methods for research workers: Biological monographs and manuals, no. v. *Edinburgh Oliver and Boyd*.

Fleiss, J. L., Levin, B., and Paik, M. C. (2003). *Statistical methods for rates and proportions*. John Wiley & Sons.

Fog, A. and Fog, M. A. (2013). The biasedurn package in r. *The Comprehensive R Archive Network. Package 'BiasedUrn'. http://cran.r-project.org/web/packages/BiasedUrn/BiasedUrn.pdf*.

Follmann, D. A. and Proschan, M. A. (1999). Valid inference in random effects meta-analysis. *Biometrics*, 55(3):732–737.

Friedman, L. (2000). Estimators of random effects variance components in meta-analysis. *Journal of Educational and Behavioral Statistics*, 25(1):1–12.

Friedrich, J. O., Adhikari, N. K. J., and Beyene, J. (2007). Inclusion of zero total event trials in meta-analyses maintains analytic consistency and incorporates all available data. *BMC medical research methodology*, 7(1):1.

Ganio-Gibbons, L. M. (1989). *Diagnostic tools for overdispersion in generalized linear models*. PhD thesis, Electronic Theses and Dissertations, Oregon State University.

Gart, J. J., Pettigrew, H. M., and Thomas, D. G. (1985). The effect of bias, variance estimation, skewness and kurtosis of the empirical logit on weighted least squares analyses. *Biometrika*, 72(1):179–190.

Gavaghan, D. J., Moore, R. A., and McQuay, H. J. (2000). An evaluation of homogeneity tests in meta-analyses in pain using simulations of individual patient data. *Pain*, 85(3):415–424.

Gelman, A. and Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press: Cambridge, England. URL http://www.stat.columbia.edu/ gelman/arm/.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher*, 5:3–8.

Guilbaud, O. and Hauck, W. (1983). On the large-sample distribution of the mantel-haenszel odds-ratio estimator. *Biometrics*, 39(2):523–525.

Gulliford, M., Adams, G., Ukoumunne, O., Latinovic, R., Chinn, S., and Campbell, M. (2005). Intraclass correlation coefficient and outcome prevalence are associated in clustered binary data. *Journal of clinical epidemiology*, 58(3):246–251.

Guolo, A. and Varin, C. (2012). The r package metalik for likelihood inference in meta-analysis. *Journal of Statistical Software*, 50(7):1–14.

Hamza, T. H., Van Houwelingen, H. C., and Stijnen, T. (2008). The binomial distribution of meta-analysis was preferred to model within-study variability. *Journal of clinical epidemiology*, 61(1):41–51.

Hanfelt, J. J. and Liang, K. Y. (1998). Inference for odds ratio regression models with sparse dependent data. *Biometrics*, 54:136–147.

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

Hardy, R. J. and Thompson, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. *Statistics in medicine*, 17(8):841–856.

Hartung, J. and Knapp, G. (2001). A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine*, 20(24):3875–3889.

Hartung, J. and Makambi, K. H. (2003). Reducing the number of unjustified significant results in meta-analysis. *Communications in Statistics-Simulation and Computation*, 32(4):1179–1190.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72(358):320–338.

Hasselblad, V. I. C. and McCrory, D. C. (1995). Meta-analytic tools for medical decision making: a practical guide. *Medical Decision Making*, 15(1):81–96.

Hauck, W. W. (1979). The large sample variance of the mantel-haenszel estimator of a common odds ratio. *Biometrics*, 35(4):817–819.

Hauck, W. W. (1984). A comparative study of conditional maximum likelihood estimation of a common odds ratio. *Biometrics*, 40(4):1117–1123.

Hedges, L. V. (1983). A random effects model for effect sizes. *Psychological Bulletin*, 93(2):388–395.

Hedges, L. V. and Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press, Orlando.

Hedges, L. V. and Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological methods*, 6(3):203.

Henmi, M. and Copas, J. B. (2010). Confidence intervals for random effects meta-analysis and robustness to publication bias. *Statistics in medicine*, 29(29):2969–2983.

Higgins, J. and Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, 21(11):1539–1558.

Higgins, J., Thompson, S. G., and Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(1):137–159.

Hinde, J. and Demétrio, C. G. B. (1998). Overdispersion: models and estimation. *Computational Statistics & Data Analysis*, 27(2):151–170.

Hoaglin, D. C. (2016). Misunderstandings about q and cochran's q test' in meta-analysis. *Statistics in Medicine*, 35(4):485–495. sim.6632.

Hunter, J. E. and Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage Publications, Inc, California.

Hwang, J.-S. and Biswas, A. (2008). Odds ratio for a single 2× 2 table with correlated binomials for two margins. *Statistical Methods and Applications*, 17(4):483–497.

Jackson, D. (2006). The power of the standard test for the presence of heterogeneity in meta-analysis. *Statistics in medicine*, 25(15):2688–2699.

Jackson, D., Bowden, J., and Baker, R. (2010). How does the dersimonian and laird procedure for random effects meta-analysis compare with its more efficient but harder to compute counterparts? *Journal of Statistical Planning and Inference*, 140(4):961–970.

James, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika*, 34(3/4):324–329.

Jones, M. P., O'Gorman, T. W., Lemke, J. H., and Woolson, R. F. (1989). A monte-carlo investigation of homogeneity tests of the odds ratio under various sample size configurations. *Biometrics*, 45(1):171–181.

Kim, D. K. and Taylor, J. M. G. (1994). Transform-both-sides approach for overdispersed binomial data when n is unobserved. *Journal of the American Statistical Association*, 89(427):833–845.

Kosmidis, I., Guolo, A., and Varin, C. (2017). Improving the accuracy of likelihood-based inference in meta-analysis and meta-regression. *Biometrika*, 104(2):489–496.

Kulinskaya, E. and Dollinger, M. B. (2015). An accurate test for homogeneity of odds ratios based on cochran¿ s q-statistic. *BMC medical research methodology*, 15(1):49.

Kulinskaya, E. and Dollinger, M. B. (2016). Commentary on misunderstandings about q and cochran's q test in meta analysis. *Statistics in Medicine*, 35(4):501–502. SIM-15-0668.

Kulinskaya, E., Dollinger, M. B., and Bjørkestøl, K. (2011a). On the moments of cochran's q statistic under the null hypothesis, with application to the meta-analysis of risk difference. *Research Synthesis Methods*, 2(4):254–270.

Kulinskaya, E., Dollinger, M. B., and Bjørkestøl, K. (2011b). Testing for homogeneity in meta-analysis i. the one-parameter case: Standardized mean difference. *Biometrics*, 67(1):203–212.

Kulinskaya, E., Morgenthaler, S., and Staudte, R. G. (2008). *Meta analysis: a guide to calibrating and combining statistical evidence*. Wiley Series in Probability and Statistics, John Wiley & Sons Ltd.

Kulinskaya, E., Morgenthaler, S., and Staudte, R. G. (2014). Combining statistical evidence. *International Statistical Review*, 82(2):214–242.

Kulinskaya, E. and Olkin, I. (2014). An overdispersion model in meta-analysis. *Statistical Modelling*, 14(1):49–76.

Kulinskaya, E., Staudte, R. G., and Gao, H. (2003). Power approximations in testing for unequal means in a one-way anova weighted for unequal variances. *Communications in Statistics-Theory and Methods*, 32(12):2353–2371.

Kuss, O. (2015). Statistical methods for meta-analyses including information from studies without any eventsadd nothing to nothing and succeed nevertheless. *Statistics in medicine*, 34(7):1097–1116.

Leonard, T. and Duffy, J. C. (2002). A bayesian fixed effects analysis of the mantel–haenszel model applied to meta-analysis. *Statistics in medicine*, 21(16):2295–2312.

Levin, B. (1984). Simple improvements on cornfield's approximation to the mean of a non-central hypergeometric random variable. *Biometrika*, 71(3):630–632.

Li, Y., Shi, L., and Daniel, R. H. (1994). The bias of the commonly-used estimate of variance in meta-analysis. *Communications in Statistics-Theory and Methods*, 23(4):1063–1085.

Liang, K. Y. (1985). Odds ratio inference with dependent data. *Biometrika*, 72(3):678–682.

Liang, K. Y. and Self, S. G. (1985). Tests for homogeneity of odds ratio when the data are sparse. *Biometrika*, 72(2):353–358.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Lipsey, W. M. and Wilson, D. B. (2001). *Practical meta-analysis*, volume 49. Sage publications Thousand Oaks, CA.

Lipsitz, S. R., Dear, K. B. G., Laird, N. M., and Molenberghs, G. (1998). Tests for homogeneity of the risk difference when data are sparse. *Biometrics*, 54(1):148–160.

Litière, S., Alonso, A., and Molenberghs, G. (2007). Type i and type ii error under random-effects misspecification in generalized linear mixed models. *Biometrics*, 63(4):1038–1044.

Litière, S., Alonso, A., and Molenberghs, G. (2008). The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in medicine*, 27(16):3125–3144.

Littenberg, B. and MacLean, C. (2006). Intra-cluster correlation coefficients in adults with diabetes in primary care practices: the vermont diabetes information system field survey. *BMC medical research methodology*, 6(1):20.

Liu, Q. and Pierce, D. A. (1993). Heterogeneity in mantel-haenszel-type models. *Biometrika*, 80(3):543–556.

Lunn, A. D. and Davies, S. J. (1998). A note on generating correlated binary variables. *Biometrika*, 85(2):487–490.

Madsen, L. and Birkes, D. (2013). Simulating dependent discrete data. *Journal of Statistical Computation and Simulation*, 83(4):677–691.

Malzahn, U., Böhning, D., and Holling, H. (2000). Nonparametric estimation of heterogeneity variance for the standardised difference used in meta-analysis. *Biometrika*, 87(3):619–632.

Mandel, J. and Paule, R. C. (1970). Interlaboratory evaluation of a material with unequal numbers of replicates. *Analytical chemistry*, 42(11):1194–1197.

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22(4):719.

Mantel, N. and Haenszel, W. (2004). Statistical aspects of the analysis of data from retrospective studies of disease. *The Challenge of Epidemiology: Issues and Selected Readings*, 1(1):533–553.

Martin, A. D., Quinn, K. M., Park, J. H., and Park, M. J. H. (2016). Package mcmcpack. *R package*.

McCulloch, C. E. and Neuhaus, J. M. (2011). Misspecifying the shape of a random effects distribution: why getting it wrong may not matter. *Statistical science*, 26(3):388–402.

Morris, C. N. (1983). Parametric empirical bayes inference: theory and applications. *Journal of the American Statistical Association*, 78(381):47–55.

Mosteller, F. and Colditz, G. A. (1996). Understanding research synthesis (meta-analysis). *Annual review of public health*, 17(1):1–23.

Nemes, S., Jonasson, J., Genell, A., and Steineck, G. (2009). Bias in odds ratios by logistic regression modelling and sample size. *BMC medical research methodology*, 9(1):1.

Neuhaus, J. M., Hauck, W. W., and Kalbfleisch, J. D. (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika*, 79(4):755–762.

Normand, S. T. (1999). Tutorial in biostatistics meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in medicine*, 18(3):321–359.

O'Gorman, T. W., Woolson, R. F., Jones, M. P., and Lemke, J. H. (1990). Statistical analysis of k 2 x 2 tables: a comparative study of estimators/test statistics for association and homogeneity. *Environmental health perspectives*, 87:103.

Olkin, I. and Gleser, L. (2009). Stochastically dependent effect sizes. *The handbook of research synthesis and meta-analysis*, pages 357–376.

Olkin, I. and Trikalinos, T. (2015). Constructions for a bivariate beta distribution. *Statistics and Probability Letters*, 96:54–60.

Palmer, S., Vecchio, M., Craig, J. C., Tonelli, M., Johnson, D. W., Nicolucci, A., Pellegrini, F., Saglimbene, V., Logroscino, G., Fishbane, S., et al. (2013). Prevalence of depression in chronic kidney disease: systematic review and meta-analysis of observational studies. *Kidney international*, 84(1):179–191.

Paul, S. R. and Donner, A. (1989). A comparison of tests of homogeneity of odds ratios in k 2× 2 tables. *Statistics in medicine*, 8(12):1455–1468.

Paul, S. R. and Donner, A. (1992). Small sample performance of tests of homogeneity of odds ratios in k 2× 2 tables. *Statistics in medicine*, 11(2):159–165.

Phillips, A. and Holland, P. W. (1987). Estimators of the variance of the mantel-haenszel log-odds-ratio estimate. *Biometrics*, 43(2):425–431.

Platt, R. W., Leroux, B. G., and Breslow, N. (1999). Generalized linear mixed models for meta-analysis. *Statistics in medicine*, 18(6):643–654.

Prentice, R. L. (1986). Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. *Journal of the American Statistical Association*, 81(394):321–327.

Qaqish, B. F., Zink, R. C., and Preisser, J. S. (2012). Orthogonalized residuals for estimation of marginally specified association parameters in multivariate binary data. *Scandinavian Journal of Statistics*, 39(3):515–527.

Raghunathan, T. and Yoichi, I. (1993). Analysis of binary data from a multicentre clinical trial. *Biometrika*, 80(1):127–139.

Rao, C. R. (2009). *Linear statistical inference and its applications*, volume 22. John Wiley & Sons.

Raudenbush, S. W. and Bryk, A. S. (1985). Empirical bayes meta-analysis. *Journal of Educational and Behavioral Statistics*, 10(2):75–98.

Reis, I. M., Hirji, K. F., and Afifi, A. A. (1999). Exact and asymptotic tests for homogeneity in several 2× 2 tables. *Statistics in medicine*, 18(8):893–906.

Ridout, M. S., Demetrio, C., and Firth, D. (1999). Estimating intraclass correlation for binary data. *Biometrics*, 55(1):137–148.

Robins, J., Breslow, N., and Greenland, S. (1986). Estimators of the mantel-haenszel variance consistent in both sparse data and large-strata limiting models. *Biometrics*, pages 311–323.

Rosenthal, R. (1994). Parametric measures of effect size. *The Handbook of Research Synthesis*, pages 231–244.

Rücker, G., Schwarzer, G., Carpenter, J., and Olkin, I. (2009). Why add anything to nothing? the arcsine difference as a measure of treatment effect in meta-analysis with zero cells. *Statistics in medicine*, 28(5):721–738.

Rücker, G., Schwarzer, G., Carpenter, J. R., and Schumacher, M. (2008). Undue reliance on i2 in assessing heterogeneity may mislead. *BMC medical research methodology*, 8(1):79.

Rukhin, A. L. (2003). Two procedures of meta-analysis in clinical trials and interlaboratory studies. *Tatra Mt. Math. Publ*, 26(155):155–168.

Rukhin, A. L. (2009). Weighted means statistics in interlaboratory studies. *Metrologia*, 46(3):323.

Rukhin, A. L. (2013). Estimating heterogeneity variance in meta-analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(3):451–469.

Rukhin, A. L., Biggerstaff, B. J., and Vangel, M. G. (2000). Restricted maximum likelihood estimation of a common mean and the mandel–paule algorithm. *Journal of Statistical Planning and Inference*, 83(2):319–330.

Rukhin, A. L. and Vangel, M. G. (1998). Estimation of a common mean and weighted means statistics. *Journal of the American Statistical Association*, 93(441):303–308.

Sánchez-Meca, J., Marín-Martínez, F., and Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological methods*, 8(4):448.

Sankey, S. S., Weissfeld, L. A., Fine, M. J., and Kapoor, W. (1996). An assessment of the use of the continuity correction for sparse data in meta-analysis. *Communications in Statistics-Simulation and Computation*, 25(4):1031–1056.

Sato, T. (1990). Confidence limits for the common odds ratio based on the asymptotic distribution of the mantel-haenszel estimator. *Biometrics*, 46:71–80.

Schmidt, F. L. and Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage publications, Inc, California.

Scott, A. J. and Holt, D. (1982). The effect of two-stage sampling on ordinary least squares methods. *Journal of the American Statistical Association*, 77(380):848–854.

Sharma, G. and Mathew, T. (2011). Higher order inference for the consensus mean in interlaboratory studies. *Biometrical Journal*, 53(1):128–136.

Shuster, J. J. (2010). Empirical vs natural weighting in random effects meta-analysis. *Statistics in medicine*, 29(12):1259–1265.

Shuster, J. J., Jones, L. S., and Salmon, D. A. (2007). Fixed vs random effects meta-analysis in rare event studies: The rosiglitazone link with myocardial infarction and cardiac death. *Statistics in medicine*, 26(24):4375–4385.

Shuster, J. J. and Walker, M. A. (2016). Low-event-rate meta-analyses of clinical trials: implementing good practices. *Statistics in Medicine*, 35(14):2467–78. SIM-14-0953.R4.

Sidik, K. and Jonkman, J. N. (2002). A simple confidence interval for meta-analysis. *Statistics in medicine*, 21(21):3153–3159.

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

Sidik, K. and Jonkman, J. N. (2008). Estimation using non-central hypergeometric distributions in combining 2× 2 tables. *Journal of Statistical Planning and Inference*, 138(12):3993–4005.

Silcocks, P. (2005). An easy approach to the robins-breslow-greenland variance estimator. *Epidemiologic Perspectives & Innovations*, 2(1):9.

Simpson, R. J. S. and Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *The British Medical Journal*, pages 1243–1246.

Song, J. X. (2004). Adjusted homogeneity tests of odds ratios when data are clustered. *Pharmaceutical Statistics*, 3(2):81–87.

Sørensen, H. (2008). Small sample distribution of the likelihood ratio test in the random effects model. *Journal of Statistical Planning and Inference*, 138(6):1605–1614.

Stijnen, T., Hamza, T. H., and Özdemir, P. (2010). 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.

Sutton, A. J., Abrams, K. R., Jones, D. R., Jones, D. R., Sheldon, T. A., and Song, F. (2000). *Methods for meta-analysis in medical research*, volume 1. Wiley Chichester.

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

Takkouche, B., Cadarso-Suárez, C., and Spiegelman, D. (1999). Evaluation of old and new tests of heterogeneity in epidemiologic meta-analysis. *American Journal of Epidemiology*, 150(2):206–215.

Tang, J. L. (2000). Weighting bias in meta-analysis of binary outcomes. *Journal of clinical epidemiology*, 53(11):1130–1136.

Tarone, R. E. (1985). On heterogeneity tests based on efficient scores. *Biometrika*, 72(1):91–95.

Thompson, S. (1994). Why source of heterogeneity in meta-analysis should be investigated:a review. *British Medical Journal*, 399:1351–1355.

Thompson, S. G. and Higgins, J. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in medicine*, 21(11):1559–1573.

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

Tippett, L. H. C. (1931). *The Methods of Statistics.* London: Williams & Norgate Ltd.

Trikalinos, T. A., Trow, P., and Schmid, C. H. (2013). Simulation-based comparison of methods for meta-analysis of proportions and rates. *Agency for Healthcare Research and Quality (US)*.

Turner, R. M., Omar, R. Z., Yang, M., Goldstein, H., and Thompson, S. G. (2000). A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Statistics in medicine*, 19(24):3417–3432.

Valentine, J. C., Pigott, T. D., and Rothstein, H. R. (2010). How many studies do you need? a primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, 35(2):215–247.

Van Houwelingen, H. C., Zwinderman, K. H., and Stijnen, T. (1993). A bivariate approach to meta-analysis. *Statistics in medicine*, 12(24):2273–2284.

Vangel, M. G. and Rukhin, A. L. (1999). Maximum likelihood analysis for heteroscedastic one-way random effects anova in interlaboratory studies. *Biometrics*, 55(1):129–136.

Verbeke, G. and Lesaffre, E. (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics & Data Analysis*, 23(4):541–556.

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J., Langan, D., and Salanti, G. (2015). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research synthesis methods*, 7(1):55–79.

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

Viechtbauer, W. (2007). Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in medicine*, 26(1):37–52.

Viechtbauer, W. (2015). Package metafor. *The Comprehensive R Archive Network. Package 'metafor'. http://cran. r-project. org/web/packages/metafor/metafor.pdf*.

Viechtbauer, W. et al. (2010). Conducting meta-analyses in r with the metafor package. *Journal of Statistical Software*, 36(3):1–48.

Welch, B. L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika*, 38(3/4):330–336.

Whitehead, A. and Whitehead, J. (1991). A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in medicine*, 10(11):1665–1677.

Williams, D. A. (1982). Extra-binomial variation in logistic linear models. *Applied statistics*, 31(2):144–148.

Williams, D. A. (1996). Overdispersion in logistic linear models. *Statistics in Toxicology (ed. B. J. T. Morgan). Oxford University Press, Oxford, UK*, pages 75–84.

Woolf, B. et al. (1955). On estimating the relation between blood group and disease. *Ann Hum Genet*, 19(4):251–253.

Xekalaki, E. (2014). On the distribution theory of over-dispersion. *Journal of Statistical Distributions and Applications*, 1(1):1–22.

Yates, F. and Cochran, W. G. (1938). The analysis of groups of experiments. *The Journal of Agricultural Science*, 28(04):556–580.

Young-Xu, Y. and Chan, K. A. (2008). Pooling overdispersed binomial data to estimate event rate. *BMC medical research methodology*, 8(1):58.

Yusuf, S., Peto, R.and Lewis, J., Collins, R., and Sleight, P. (1985). Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Progress in cardiovascular diseases*, 27(5):335–371.

Zelen, M. and Parker, R. A. (1986). Case-control studies and bayesian inference. *Statistics in Medicine*, 5(3):261–269.

Zhang, J. and Boos, D. D. (1997). Mantel-haenszel test statistics for correlated binary data. *Biometrics*, 53(4):1185–1198.

Zhou, X. H., Brizendine, E. J., and Pritz, M. B. (1999). Methods for combining rates from several studies. *Statistics in medicine*, 18(5):557–566.

Zhu, M. and Lu, A. Y. (2004). The counter-intuitive non-informative prior for the bernoulli family. *Journal of Statistics Education*, 12(2):1–10.

Appendix A

A.1 Bayesian setting

This problem may not exist in the Bayesian setting.

The binomial likelihood for $X|p \sim B(n,p)$ is proportionate to $p^X(1-p)^{n-X}$. Given that $p \sim B(r,s)$, the full likelihood is proportionate to $p^{X+r-1}(1-p)^{n-X+s-1}$. Thus the posterior distribution $\theta|X$ is $B(X+r, n-X+s)$, and (as is well known) Beta is the conjugate distribution for Binomial. We need $r>0, s>0$ for the beta-binomial density to be defined, i.e. for a proper prior, but this is not that necessary. The posterior mean and variance of this distribution are, respectively,

$$\hat{p}_{Bayes} = \frac{r+X}{r+s+n} \text{ and } \text{Var}(\hat{p}_{Bayes}) = \frac{(r+x)(s+n-X)}{(r+s+n)^2(r+s+n+1)}.$$

Given a density $g(p)$, find the density for log-odds θ, given that $p = h(\theta) = [1+exp(-\theta)]^{-1}$. Using the standard change of variables formula, the density is $g(p) = f(\theta(p))h'(p)$, where $h'(\theta) = h(\theta)(1-h(\theta))$ is the derivative of the inverse log-odds transformation. :

$$g(p) = f(\theta(p))h'(p) = \frac{f(\theta(p))}{p(1-p)}.$$

Similarly,

$$g(\theta) = f(p(\theta))[h'(p)]^{-1} = f(p(\theta))p(1-p).$$

So, for a prior $\pi(p) = B(r,s)$ on p,

$$\pi(\theta) = B_{r,s}(p(\theta))p(1-p) = B_{r+1,s+1}(p(\theta)).$$

We need $r>-1, s>-1$ for this density to be defined, i.e. for a proper prior on the log-odds scale. Important particular cases of the beta priors for p are: The non-informative

Jeffreys prior for p is $\pi(p) \propto p^{-1/2}(1-p)^{-1/2} = B(1/2, 1/2)$ (Zhu and Lu, 2004), (Boos and Stefanski, 2013, ch. 4 p.175, formula 4.15) provides corresponding Jeffreys prior for log-odds as $\pi(\theta) = e^{\theta/2}/(1+e^{\theta}) = e^{-\theta/2}/(1+e^{-\theta})$. This corresponds to $s = r = 3/2$. The uniform prior $U(0,1) = B(1,1)$ would result in the prior $B(2,2)$ for log-odds. Haldane's improper prior $p^{-1}(1-p)^{-1} \propto B(0,0)$ results in the flat uniform prior for log-odds (Zhu and Lu, 2004). All these cases agree with the general result above. The gist of this is that non-informative prior for p often results in a very informative prior for log-odds and vice-versa. The likelihood on the log-odds scale is

$$\Big[\frac{1}{1+e^{-\theta}}\Big]^{X}\Big[\frac{e^{-\theta}}{1+e^{-\theta}}\Big]^{n-X}\pi(\theta) = B_{X+r+1,n-X+s+1}(p(\theta)).$$

A.2 Results of simulation

A.2.1 Simulation(without bias correction) for larger probabilities

Figures A.1 and A.3 show the bias of the arcsine transformation for $p = 0.2$ and $p = 0.4$. Figures A.2 and A.4 show the coverage of the arcsine transformation for $p = 0.2$ and $p = 0.4$

A.3 Bias correction of arcsine transformation with known ρ and p

Results for $p = 0.1$, $p = 0.2$ and $p = 0.4$ with theoretical bias correction are given in Figures A.5, A.6 and A.7, respectively.

Coverage after bias correction (for known p and ρ) is depicted in Figures A.5, A.6 and A.7, respectively. It can be seen that the coverage is greatly improved.

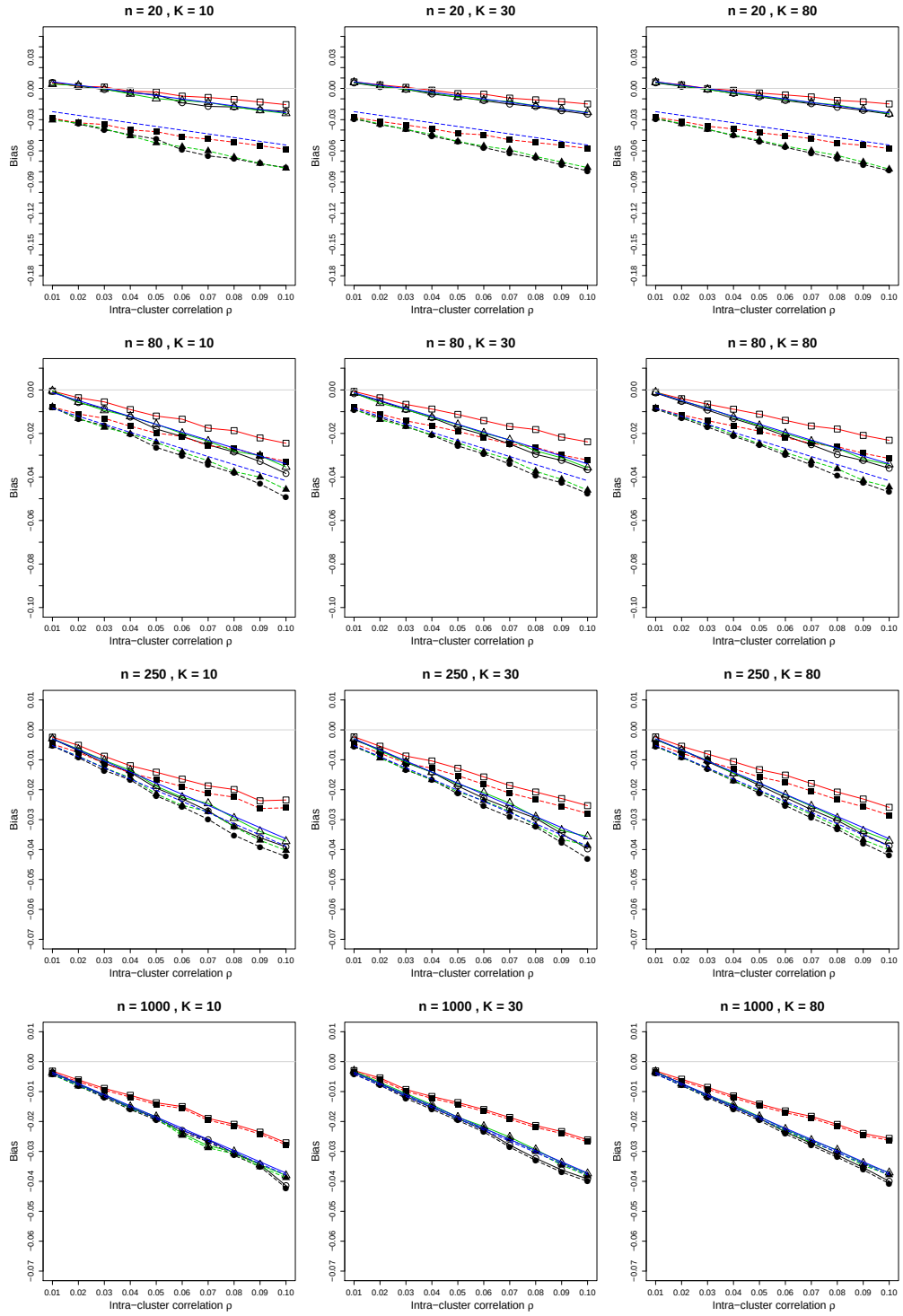


Figure A.1: Bias on the arcsine scale of the meta-analysis of arcsine transformations from K studies in overdispersed binomial model for $p = 0.2$ and $0 \leq \rho \leq 0.1$. Simulations (10000 for each values of ρ) from beta-binomial distribution (black), from Lunn and Davies (1998) model (red) and from Gaussian copula (Emrich and Piedmonte, 1991) (green) with and without the continuity correction (solid and dashed lines, respectively). Also the first order bias terms given by the first two terms of equation (3.2.1) and plotted for known p and ρ (solid or dashed blue lines). Light grey line at zero.

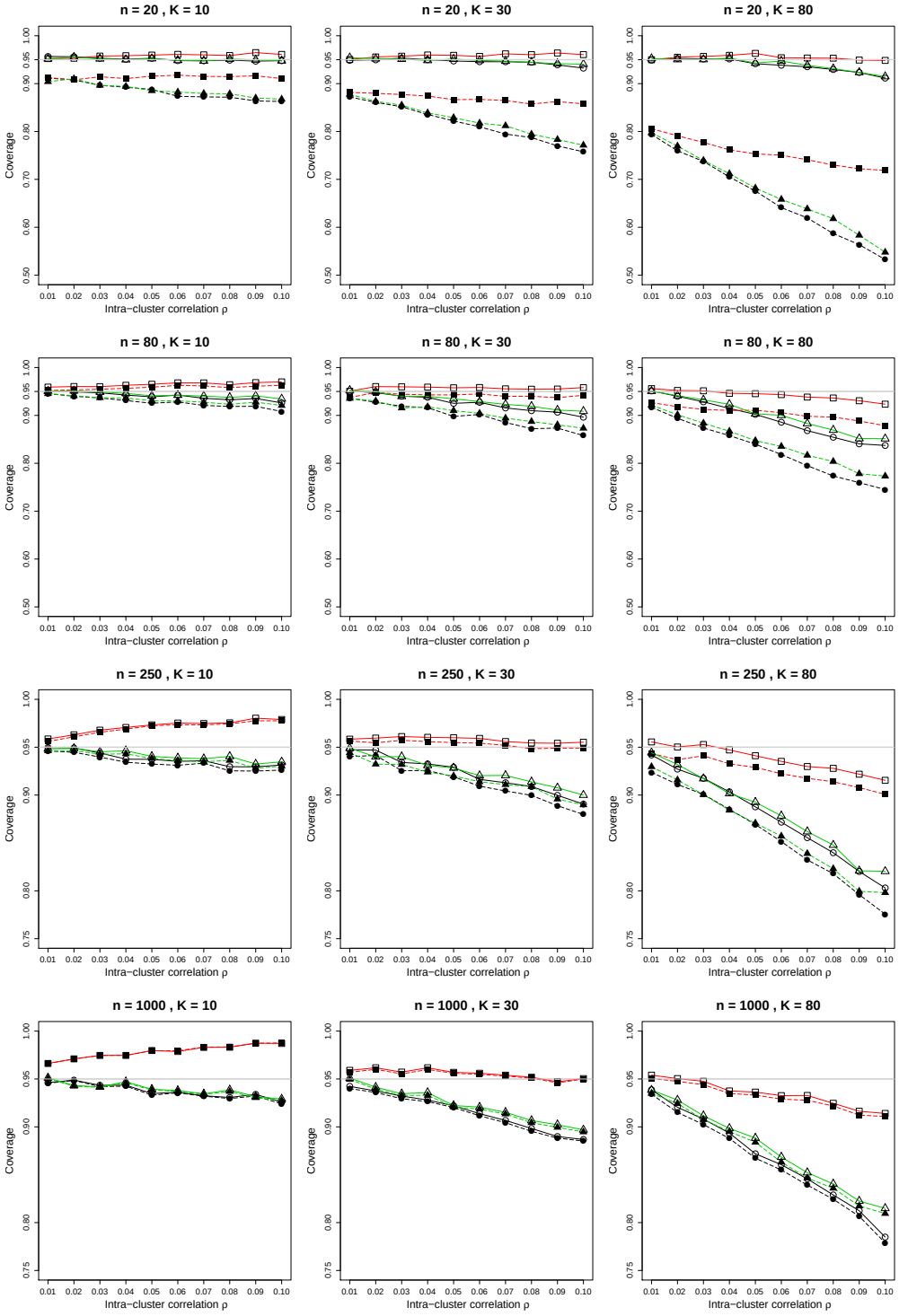


Figure A.2: Coverage (for a known value of ρ) at the nominal 95% level of the true value of p using the meta-analysis of acrsine transformation from K studies in overdispersed binomial model for $p = 0.2$ and $0 \leq \rho \leq 0.1$. Simulations (10000 for each values of ρ) from beta-binomial distribution (black lines), from Lunn and Davies (1998) model (red lines) and from Gaussian copula (Emrich and Piedmonte, 1991) (green) with and without the continuity correction (solid or dashed lines, respectively). Light grey line at 0.95.

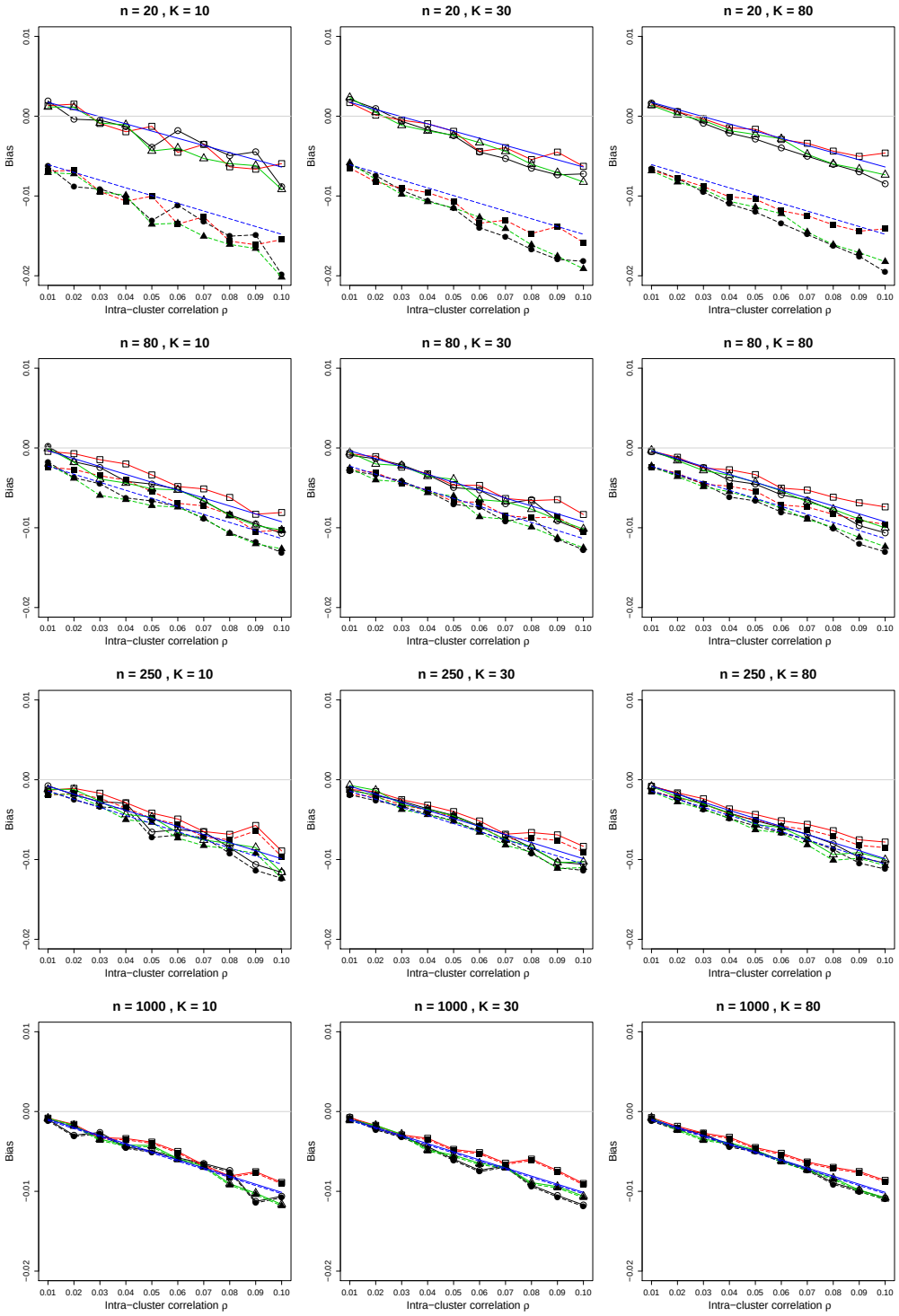


Figure A.3: Bias on the arcsine scale of the meta-analysis of arcsine transformations from K studies in overdispersed binomial model for $p = 0.4$ and $0 \leq \rho \leq 0.1$. Simulations (10000 for each values of ρ) from beta-binomial distribution (black), from Lunn and Davies (1998) model (red) and from Gaussian copula (Emrich and Piedmonte, 1991) (green) with and without the continuity correction (solid and dashed lines, respectively). Also the first order bias terms given by the first two terms of equation (3.2.1) and plotted for known p and ρ (solid or dashed blue lines). Light grey line at zero.

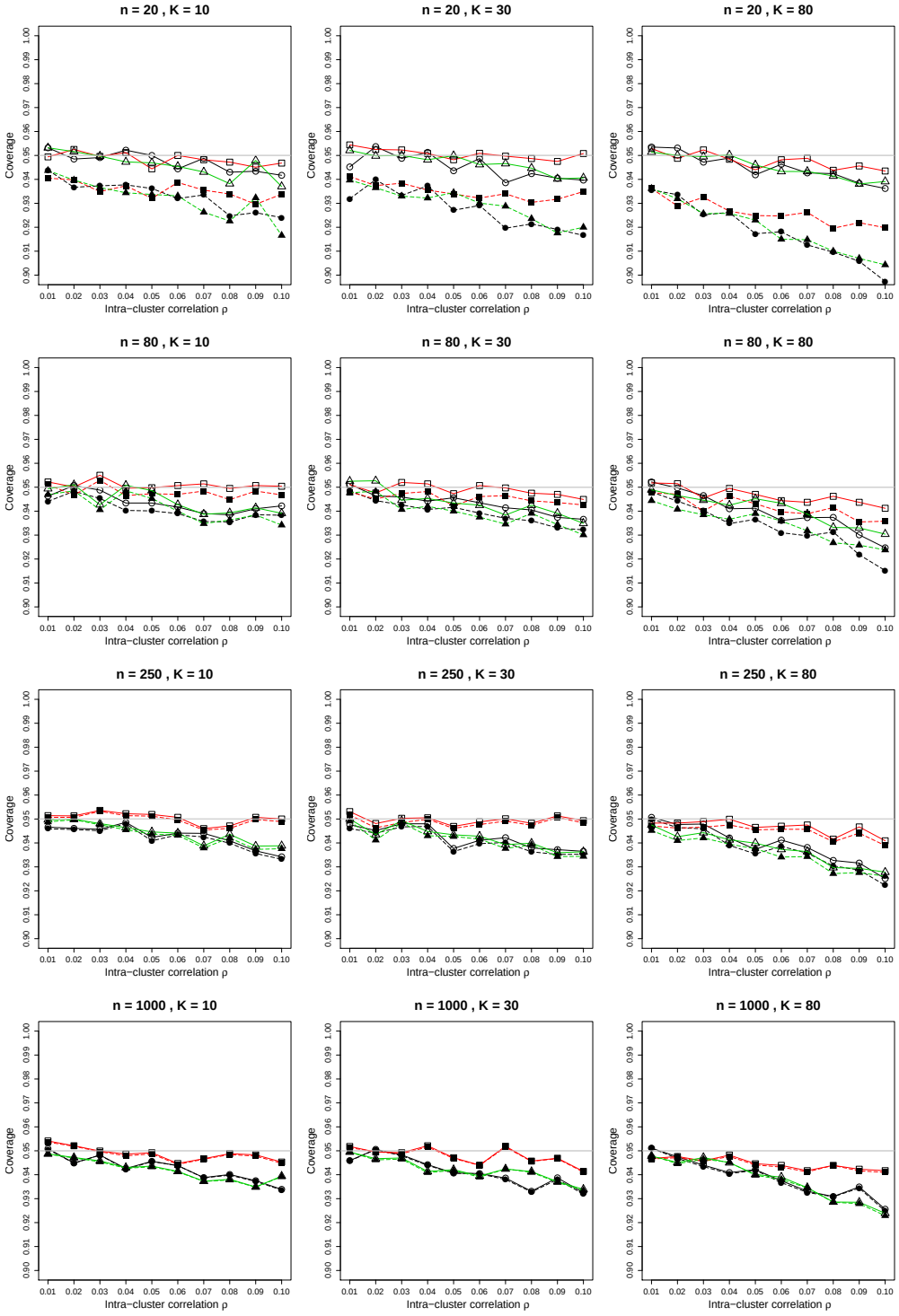


Figure A.4: Coverage (for a known value of ρ) at the nominal 95% level of the true value of p using the meta-analysis of acrsine transformation from K studies in overdispersed binomial model for $p = 0.4$ and $0 \leq \rho \leq 0.1$. Simulations (10000 for each values of ρ) from beta-binomial distribution (black lines), from Lunn and Davies (1998) model (red lines) and from Gaussian copula (Emrich and Piedmonte, 1991) (green) with and without the continuity correction (solid or dashed lines, respectively). Light grey line at 0.95.

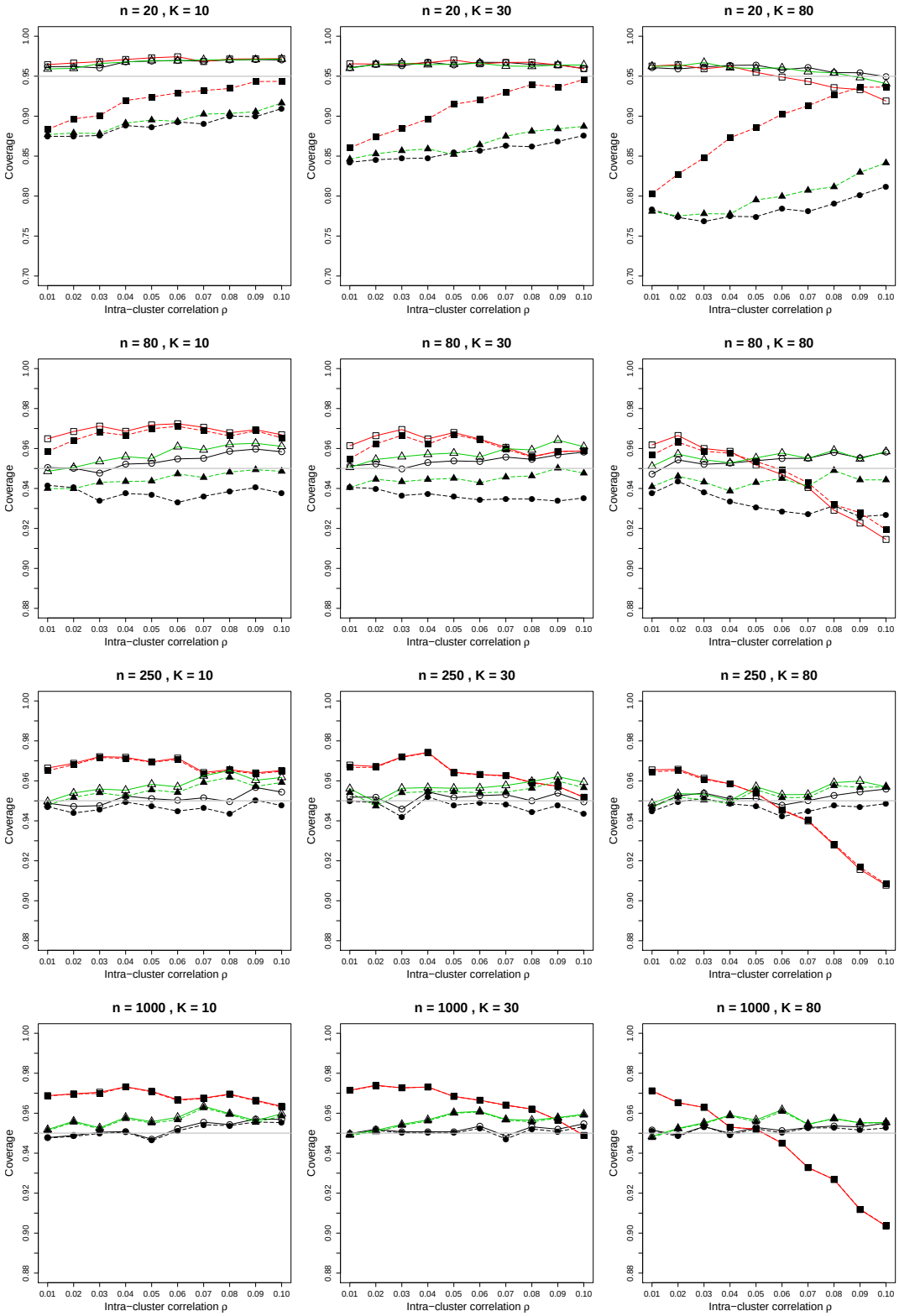


Figure A.5: Coverage in meta-analysis at the nominal 95% level of the true value of p using the acrsine transformation with bias-correction in overdispersed binomial model for $p = 0.1$ and $0 \leq \rho \leq 0.1$; n sample size; k number of studies. Simulations (10000 for each values of ρ) from beta-binomial distribution (solid lines), from Lunn and Davies (1998) model (dashed lines) and from model with Gaussian Copula (Emrich and Piedmonte, 1991) with and without the continuity correction (black, red and green colour, respectively). Light grey line at 0.95.

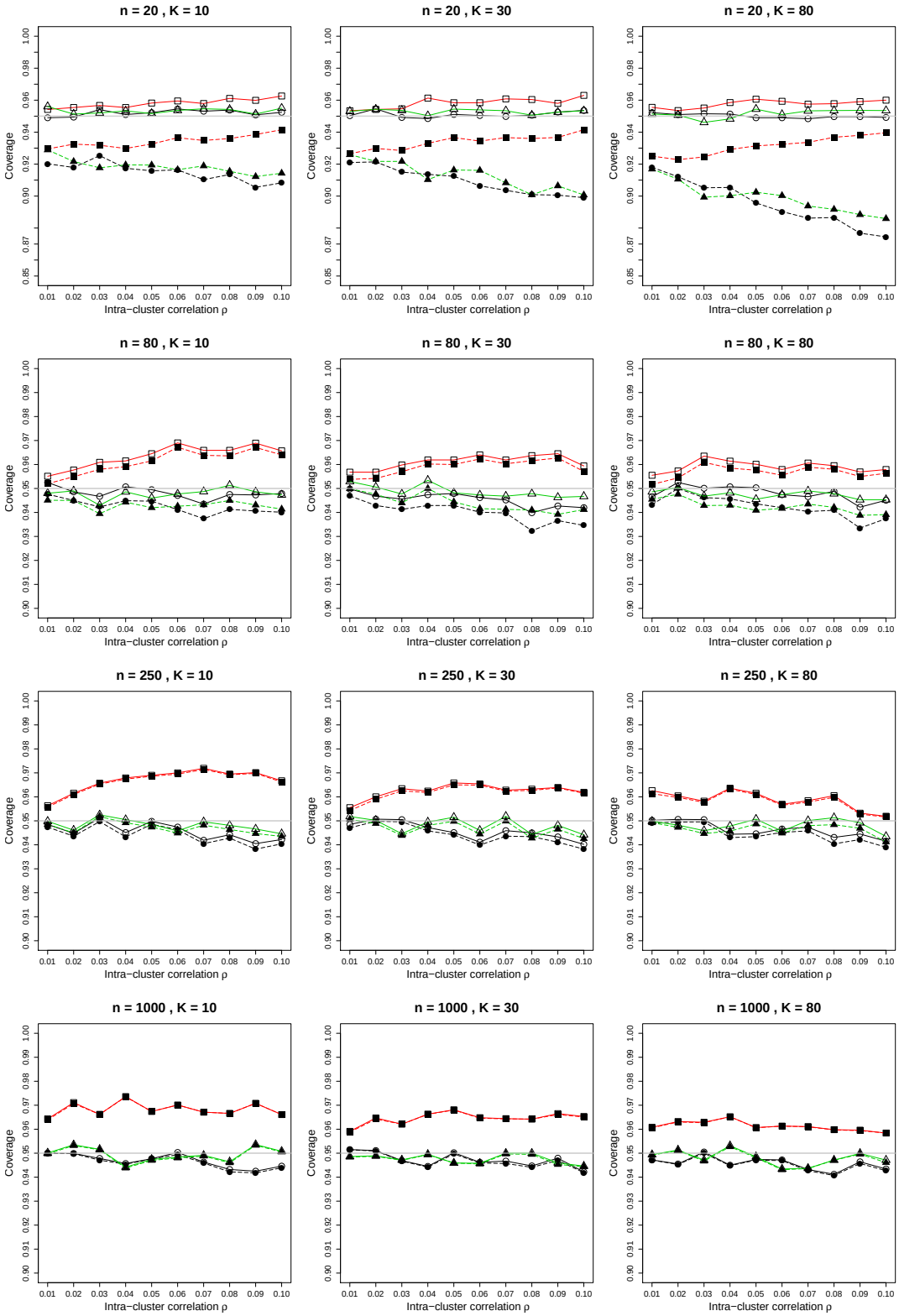


Figure A.6: Coverage at the nominal 95% level of the true value of p using the acrsine transformation with bias-correction in overdispersed binomial model for $p = 0.2$ and $0 \leq \rho \leq 0.1$; n sample size; k number of studies. Simulations (10000 for each values of ρ) from beta-binomial distribution (solid lines), from Lunn and Davies (1998) model (dashed lines) and from model with Gaussian Copula (Emrich and Piedmonte, 1991) with and without the continuity correction (black, red and green colour respectively). Light grey line at 0.95.

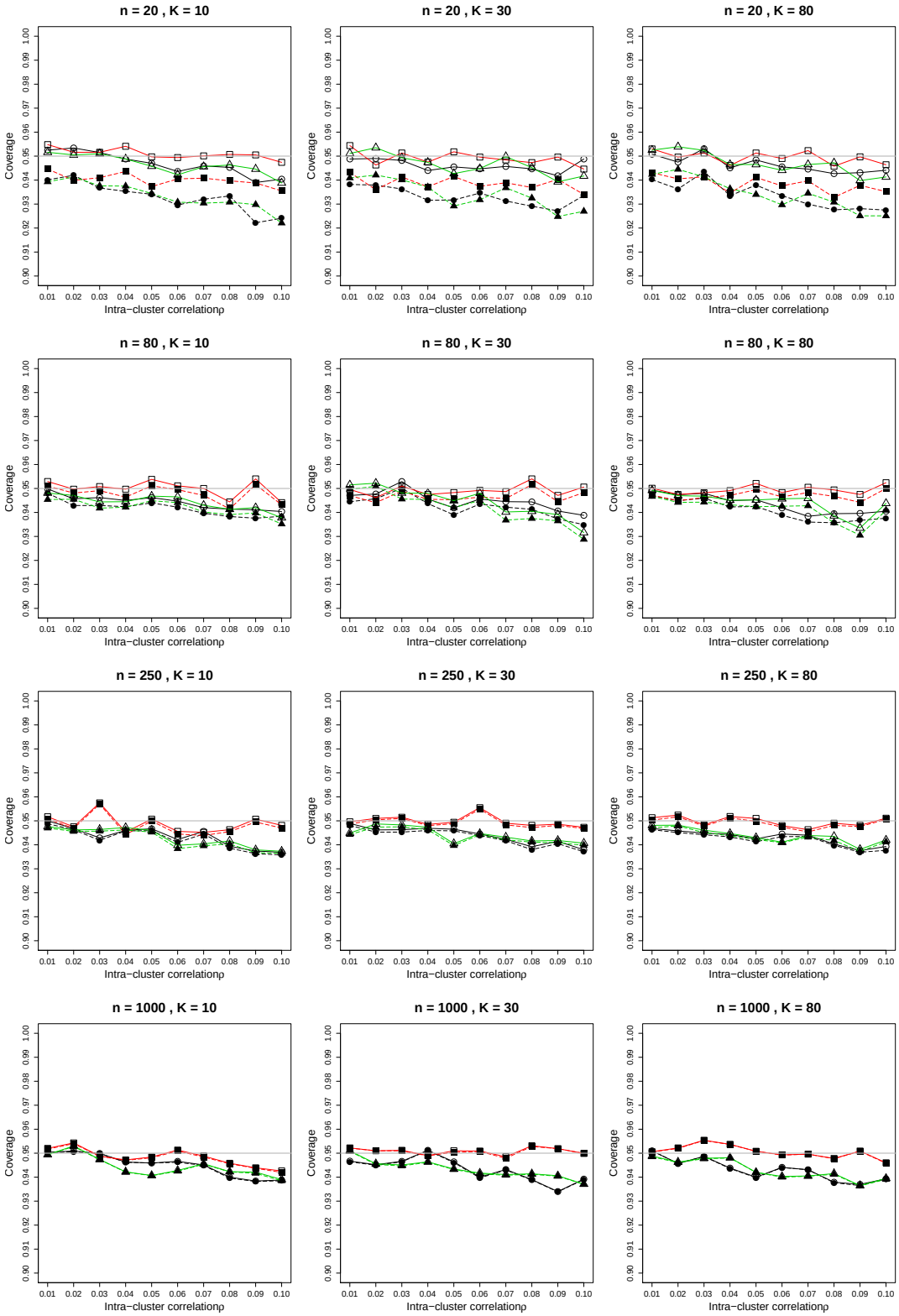


Figure A.7: Coverage at the nominal 95% level of the true value of p using the acrsine transformation with bias-correction in overdispersed binomial model for $p = 0.4$ and $0 \leq \rho \leq 0.1$; n sample size; k number of studies. Simulations (10000 for each values of ρ) from beta-binomial distribution (solid lines), from Lunn and Davies (1998) model (dashed lines) and from model with Gaussian Copula (Emrich and Piedmonte, 1991) with and without the continuity correction (black, red and green colour, respectively). Light grey line at 0.95.

Appendix B

B.1 Variance of corrected Mantel-Haenzsel odds ratio

Following the derivation from Silcocks (2005), the variance of LOR similar to 2.4.10 can be obtained as derived by Robins et al. (1986) and Phillips and Holland (1987) as

$$\mathrm{Var}(\hat{\psi}_{MH}) = \frac{\sum_{j=1}^{K} W_{jC}^2 \mathrm{Var}(\hat{\psi}_j)}{(\sum_{j=1}^{K} W_{jC})^2}$$

where W_{jC} is

$$W_{jC} = \frac{(n_{1j} - X_{1j})X_{2j}}{C_{1j}n_{2j} + C_{2j}n_{1j}}$$

and the variance of individual odds ratio is

$$\mathrm{Var}(\hat{\psi}_j) = \frac{C_{1j}}{X_{1j}} + \frac{C_{1j}}{n_{1j} - X_{1j}} + \frac{C_{2j}}{X_{2j}} + \frac{C_{2j}}{n_{2j} - X_{2j}}.$$

Here we assumed that W_{jC} and $\hat{\psi}_j$ are independent from each other. However, in general this assumption is wrong. W_{jC} and $\hat{\psi}_j$ are correlated with each other, since both of these variables depend on the observed number of cases in treatment and control arm.
Substituting the weights and variances of each individual odds ratio, we have

$$\mathrm{Var}(\log(\hat{\psi}_{MH})) = \frac{\sum_{j=1}^{K} (\frac{(n_{1j}-X_{1j})X_{2j}}{C_{1j}n_{2j}+C_{2j}n_{1j}})^2 (\frac{C_{1j}}{X_{1j}} + \frac{C_{1j}}{n_{1j}-X_{1j}} + \frac{C_{2j}}{X_{2j}} + \frac{C_{2j}}{n_{2j}-X_{2j}})}{(\sum_{j=1}^{K} \frac{(n_{1j}-X_{1j})X_{2j}}{C_{1j}n_{2j}+C_{2j}n_{1j}})^2}$$

Assuming a common odds ratio across K studies estimated by $\hat{\psi}_{MH}$, the equation for variance can be re-written as

$$\text{Var}(\log(\hat{\psi}_{MH})) = \frac{\hat{\psi}^2_{MH}\sum_{j=1}^{K}(\frac{(n_{1j}-X_{1j})X_{2j}}{C_{1j}n_{2j}+C_{2j}n_{1j}})^2(\frac{C_{1j}}{X_{1j}}+\frac{C_{1j}}{n_{1j}-X_{1j}}+\frac{C_{2j}}{X_{2j}}+\frac{C_{2j}}{n_{2j}-X_{2j}})}{(\sum_{j=1}^{K}\frac{(n_{1j}-X_{1j})X_{2j}}{C_{1j}n_{2j}+C_{2j}n_{1j}})^2}$$

which is the same as equation derived by Donald and Donner (1987) with corrections factors (4.3.1) taken into account. The variance can be further re-written as

$$\text{Var}(\log(\hat{\psi}_{MH})) = \frac{\sum_{j=1}^{K}(\frac{(n_{1j}-X_{1j})X_{2j}}{C_{1j}n_{2j}+C_{2j}n_{1j}})^2(\frac{C_{1j}(n_{2j}-X_{2j})+C_{2j}X_{1j}}{X_{1j}(n_{2j}-X_{2j})}+\frac{C_{2j}(n_{1j}-X_{1j})+C_{1j}X_{2j}}{X_{2j}(n_{1j}-X_{1j})})}{(\sum_{j=1}^{K}\frac{(n_{1j}-X_{1j})X_{2j}}{C_{1j}n_{2j}+C_{2j}n_{1j}})^2}$$

or

$$\text{Var}(\log(\hat{\psi}_{MH})) = \frac{\sum_{j=1}^{K}\frac{(n_{1j}-X_{1j})X_{2j}}{(C_{1j}n_{2j}+C_{2j}n_{1j})^2}(n_{1j}-X_{1j})X_{2j}(\frac{C_{1j}(n_{2j}-X_{2j})+C_{2j}X_{1j}}{X_{1j}(n_{2j}-X_{2j})}+\frac{C_{2j}(n_{1j}-X_{1j})+C_{1j}X_{2j}}{X_{2j}(n_{1j}-X_{1j})})}{(\sum_{j=1}^{K}\frac{(n_{1j}-X_{1j})X_{2j}}{C_{1j}n_{2j}+C_{2j}n_{1j}})^2}$$

By re-defining the term $\psi_{MH} = \frac{X_{2j}(n_{2j}-X_{2j})}{(n_{1j}-X_{1j})X_{2j}}$, we have

$$\text{Var}(\log(\hat{\psi}_{MH})) = \frac{\sum_{j=1}^{K}\frac{(n_{1j}-X_{1j})X_{2j}}{(C_{1j}n_{2j}+C_{2j}n_{1j})^2}(\frac{C_{1j}(n_{2j}-X_{2j})+C_{2j}X_{1j}}{\psi_{MH}}+C_{2j}(n_{1j}-X_{1j})+C_{1j}X_{2j})}{(\sum_{j=1}^{K}\frac{(n_{1j}-X_{1j})X_{2j}}{C_{1j}n_{2j}+C_{2j}n_{1j}})^2}$$

Now as it was done by Silcocks (2005), if the rows of the 2×2 contingency table are interchanged, the variance should not change. Hence, this is the same as the statement about the variance of interchanged table such as

$$\text{Var}(\log(\hat{\psi}^*_{MH})) = \frac{\sum_{j=1}^{K}\frac{(n_{2j}-X_{2j})X_{1j}}{(C_{1j}n_{2j}+C_{2j}n_{1j})^2}(\frac{C_{2j}(n_{1j}-X_{1j})+C_{1j}X_{2j}}{\psi^*_{MH}}+C_{1j}(n_{2j}-X_{2j})+C_{2j}X_{1j})}{(\sum_{j=1}^{K}\frac{(n_{2j}-X_{2j})X_{1j}}{C_{1j}n_{2j}+C_{2j}n_{1j}})^2}$$

such that $\psi_{MH} = 1/(\psi^*_{MH})$.

Let $R = \sum_{j=1}^{K}\frac{(n_{2j}-X_{2j})X_{1j}}{C_{1j}n_{2j}+C_{2j}n_{1j}}$ and $S = \sum_{j=1}^{K}\frac{(n_{1j}-X_{1j})X_{2j}}{C_{1j}n_{2j}+C_{2j}n_{1j}}$, thus taking the mean of two estimates above results in

$$\text{Var}(\log(\hat{\psi}_{MH})) =$$

$$\frac{\sum_{j=1}^{K}(\frac{R^2(n_{1j}-X_{1j})X_{2j}}{(C_{1j}n_{2j}+C_{2j}n_{1j})^2\hat{\psi}_{MH}}+\frac{S^2(n_{2j}-X_{2j})X_{1j}}{(C_{1j}n_{2j}+C_{2j}n_{1j})^2})(C_{1j}(n_{2j}-X_{2j})+C_{2j}X_{1j}+\hat{\psi}_{MH}(C_{2j}(n_{1j}-X_{1j})+C_{1j}X_{2j}))}{2R^2S^2}.$$

Dividing the numerator and denominator by S^2, we have

$$\mathrm{Var}(\log(\hat{\psi}_{MH})) =$$

$$\frac{\sum_{j=1}^{K}\frac{(\hat{\psi}_{MH}(n_{1j}-X_{1j})X_{2j}+(n_{2j}-X_{2j})X_{1j})}{(C_{1j}n_{2j}+C_{2j}n_{1j})^2}(C_{1j}(n_{2j}-X_{2j})+C_{2j}X_{1j}+\hat{\psi}_{MH}(C_{2j}(n_{1j}-X_{1j})+C_{1j}X_{2j}))}{2R^2}$$

Now let $P_j = (C_{1j}(n_{2j}-X_{2j})+C_{2j}X_{1j})/(C_{1j}n_{2j}+C_{2j}n_{1j})$ and $Q_j = (C_{2j}(n_{1j}-X_{1j})+C_{1j}X_{2j})/(C_{1j}n_{2j}+C_{2j}n_{1j})$ with $R_j = (n_{2j}-X_{2j})X_{1j}/(C_{1j}n_{2j}+C_{2j}n_{1j})$ and $S_j = (n_{1j}-X_{1j})X_{2j}/(C_{1j}n_{2j}+C_{2j}n_{1j})$ then

$$\mathrm{Var}(\log(\hat{\psi}_{MH})) = \frac{\sum_{j=1}^{K}(\psi_{MH}S_j+R_j)(P_j+\psi_{MH}Q_j)}{2R^2}$$

multiplying the brackets out, the variance can be expressed in the form of standard variance derived by Robins et al. (1986) and Phillips and Holland (1987) such that

$$\mathrm{Var}(\log(\hat{\psi}_{MH})) = \frac{\sum_{j=1}^{K}R_jP_j}{2R^2}+\frac{\sum_{j=1}^{K}(P_jS_j+Q_jR_j)}{2RS}+\frac{\sum_{j=1}^{K}S_jQ_j}{2S^2}.$$

B.2 Limit of $\hat{\psi}_{MH}$ on ρ

The corrected Mantel-Haenzsel odds ratio (4.3.2) can be rewritten in the form

$$\hat{\psi}_{CMH} = \frac{\sum_{j=1}^{K}\frac{n_{1j}n_{2j}p_{1j}(1-p_{2j})}{C_{1j}n_{2j}+C_{2j}n_{1j}}}{\sum_{j=1}^{K}\frac{n_{1j}n_{2j}p_{2j}(1-p_{1j})}{C_{1j}n_{2j}+C_{2j}n_{1j}}}$$

When $\rho \to -1/\max(a_j)$, $C_{ij} = 1-(n_{ij}-1)/\max(a_j)$. Then the the correction factors are

$$C_{1j} = 1-(n_{1j}-1)/\max(a_j) \quad \text{and} \quad C_{2j} = 1-(n_{2j}-1)/\max(a_j)$$

For balanced studies $n_{1j} = n_{2j} = n_j$, $a_j = n_j - 1$ and $C_{1j} = C_{2j} = C(\rho) = 1+(n_j-1)\rho$. For $\rho \to -1/\max(a_j)$

$$C(\rho) = 1-(n_j-1)/\max(n_j-1).$$

The corrected Mantel-Haenzsel odds ratio (4.3.2) is

$$\hat{\psi}_{CMH} = \frac{\sum_{n_j=\max(n_j)} \frac{n_{1j}n_{2j}p_{1j}(1-p_{2j})}{2C(\rho)n_j} + \sum_{n_j\neq\max(n_j)} \frac{n_{1j}n_{2j}p_{1j}(1-p_{2j})}{C_{1j}n_{2j}+C_{2j}n_{1j}}}{\sum_{n_j=\max(n_j)} \frac{n_{1j}n_{2j}p_{2j}(1-p_{1j})}{2C(\rho)n_j} + \sum_{n_j\neq\max(n_j)} \frac{n_{1j}n_{2j}p_{1j}(1-p_{2j})}{C_{1j}n_{2j}+C_{2j}n_{1j}}}$$

which is the same as

$$\hat{\psi}_{CMH} = \frac{\sum_{n_j=\max(n_j)} n_{1j}n_{2j}p_{1j}(1-p_{2j}) + 2C(\rho)n_j \sum_{n_j\neq\max(n_j)} \frac{n_{1j}n_{2j}p_{1j}(1-p_{2j})}{C_{1j}n_{2j}+C_{2j}n_{1j}}}{\sum_{n_j=\max(n_j)} n_{1j}n_{2j}p_{2j}(1-p_{1j}) + 2C(\rho)n_j \sum_{n_j\neq\max(n_j)} \frac{n_{1j}n_{2j}p_{1j}(1-p_{2j})}{C_{1j}n_{2j}+C_{2j}n_{1j}}}$$

When $\rho \to -1/\max(a_j)$, $C(\rho) \to 0$, so

$$\hat{\psi}_{CMH} = \frac{\sum_{n_j=\max(n_j)} n_{1j}n_{2j}p_{1j}(1-p_{2j})}{\sum_{n_j=\max(n_j)} n_{1j}n_{2j}p_{2j}(1-p_{1j})} = \frac{\sum_{n_j=\max(n_j)} n_j^2 p_{1j}(1-p_{2j})}{\sum_{n_j=\max(n_j)} n_j^2 p_{2j}(1-p_{1j})}$$

$$\hat{\psi}_{CMH} = \frac{\sum_{n_j=\max(n_j)} p_{1j}(1-p_{2j})}{\sum_{n_j=\max(n_j)} p_{2j}(1-p_{1j})}$$

B.3 Limit of $\hat{\theta}_{IV}$ on ρ

The inverse variance odds ratio with weight $w_j(\rho)$ is

$$\hat{\theta}_w = \frac{\sum_{j=1}^{K} w_j(\rho)\hat{\theta}_j}{\sum_{j=1}^{K} w_j(\rho)} = \frac{\sum_{j=1}^{K} \frac{\hat{\theta}_j}{\sigma_j^2(1+a_j\rho)}}{\sum_{j=1}^{K} \frac{1}{\sigma_j^2(1+a_j\rho)}}$$

Similarly to $\hat{\psi}_{MH}$, when $\rho \to -1/\max(a_j)$,$\hat{\theta}_w$ should be subdivided into 2 components

$$\hat{\theta}_w = \frac{\sum_{n_j=\max(n_j)} \frac{\hat{\theta}_j}{\sigma_j^2(1+a_j\rho)} + \sum_{n_j\neq\max(n_j)} \frac{\hat{\theta}_j}{\sigma_j^2(1+a_j\rho)}}{\sum_{n_j=\max(n_j)} \frac{1}{\sigma_j^2(1+a_j\rho)} + \sum_{n_j\neq\max(n_j)} \frac{1}{\sigma_j^2(1+a_j\rho)}}$$

For balanced studies $n_{1j} = n_{2j} = n_j$, $a_j = n_j - 1$. So, when $\rho \to -1/\max(a_j)$, $1 + a_j\rho \to 0$. Thus,

$$\hat{\theta}_w = \frac{\sum_{n_j=\max(n_j)} \frac{\hat{\theta}_j}{\sigma_j^2} + (1+a_j\rho)|_{n_j=\max(n_j)} \sum_{n_j\neq\max(n_j)} \frac{\hat{\theta}_j}{\sigma_j^2(1+a_j\rho)}}{\sum_{n_j=\max(n_j)} \frac{1}{\sigma_j^2} + (1+a_j\rho)|_{n_j=\max(n_j)} \sum_{n_j\neq\max(n_j)} \frac{1}{\sigma_j^2(1+a_j\rho)}}$$

and

$$\hat{\theta}_w = \frac{\sum_{n_j=\max(n_j)} \frac{\hat{\theta}_j}{\sigma_j^2}}{\sum_{n_j=\max(n_j)} \frac{1}{\sigma_j^2}}$$

Table B.1: Correspondence between θ_j and ψ_j

θ_j	0	1	2	3
ψ_j	1	2.718282	7.389056	20.08554

B.4 Analysis of probabilities

In the simulations we are only interested in three values of effect measure. The correspondence between the effect of interest log odds ratio and odds ratio is

Since the probability of response for treatment arm p_{jT} depends on effect measure θ_j and probability of control p_{jC}, the correspondence between these three variables is

Table B.2: Correspondence between p_{jC} and p_{jT}

	The values for log odds ratio θ_j		
p_{jC}	1	2	3
0.1	0.2319693	0.45085306	0.690567858
0.2	0.4046097	0.648785644	0.83392523
0.3	0.5381015	0.760004128	0.895921012
0.4	0.644405	0.831253174	0.930509025
0.5	0.7310586	0.880797078	0.952574127
0.6	0.8030497	0.917243097	0.967874898
0.7	0.8638095	0.945178838	0.979108454
0.8	0.9157762	0.967273444	0.98770625
0.9	0.9607297	0.985185515	0.994498537

B.5 Analysis of correspondence between τ^2 and ρ

The correspondence between τ^2 and ρ is important for multiplicative random effects model in simulations. Below we show the different value for a_j in the correspondence (6.2.7) with different configurations. From tables below for different values of sample sizes $n = (10, 20, 40, 80, 160, 250, 640, 1000)$, we show that the values of a_j is increasing monotonically with probabilities for control arm p_{iC}. Due to correspondence (6.2.7) between τ^2 and ρ, any increase in a_j influence the increase in values for τ^2.

	$N = 10$			$N = 20$		
	$\theta = 0$	$\theta = 1$	$\theta = 2$	$\theta = 0$	$\theta = 1$	$\theta = 2$
p_{iC}						
0.1	15.05	13.64	14.21	15.89	14.39	15.00
0.2	9.36	9.57	12.12	9.88	10.11	12.80
0.3	7.91	9.22	13.94	8.35	9.73	14.71
0.4	7.68	10.17	17.67	8.10	10.73	18.65
0.5	8.18	12.17	23.52	8.63	12.85	24.83
0.6	9.44	15.61	32.70	9.96	16.47	34.51
0.7	11.94	21.65	48.28	12.60	22.86	50.97
0.8	17.29	34.06	79.74	18.25	35.95	84.17
0.9	33.85	71.66	174.50	35.74	75.65	184.19

Table B.3: The values of a_j for $N = (10, 20)$

	$N = 40$			$N = 80$		
	$\theta = 0$	$\theta = 1$	$\theta = 2$	$\theta = 0$	$\theta = 1$	$\theta = 2$
p_{iC}						
0.1	16.31	14.77	15.40	16.52	14.96	15.59
0.2	10.14	10.37	13.13	10.27	10.51	13.30
0.3	8.57	9.99	15.10	8.68	10.12	15.29
0.4	8.32	11.01	19.14	8.42	11.15	19.39
0.5	8.86	13.19	25.48	8.97	13.36	25.81
0.6	10.23	16.91	35.42	10.36	17.12	35.87
0.7	12.93	23.46	52.31	13.10	23.76	52.98
0.8	18.73	36.89	86.39	18.97	37.37	87.50
0.9	36.68	77.64	189.04	37.15	78.63	191.46

Table B.4: The values of a_j for $N = (40, 80)$

	$N = 160$			$N = 250$		
	$\theta = 0$	$\theta = 1$	$\theta = 2$	$\theta = 0$	$\theta = 1$	$\theta = 2$
p_{iC}						
0.1	16.62	15.06	15.69	16.66	15.09	15.73
0.2	10.34	10.57	13.39	10.36	10.60	13.42
0.3	8.73	10.18	15.39	8.75	10.20	15.42
0.4	8.48	11.23	19.51	8.50	11.25	19.55
0.5	9.03	13.44	25.97	9.05	13.47	26.03
0.6	10.42	17.23	36.10	10.45	17.27	36.18
0.7	13.18	23.91	53.31	13.21	23.96	53.43
0.8	19.10	37.60	88.05	19.14	37.69	88.25
0.9	37.38	79.13	192.67	37.47	79.31	193.11

Table B.5: The values of a_j for $N = (160, 250)$

	$N = 640$			$N = 1000$		
	$\theta = 0$	$\theta = 1$	$\theta = 2$	$\theta = 0$	$\theta = 1$	$\theta = 2$
p_{iC}						
0.1	16.70	15.13	15.77	16.71	15.13	15.78
0.2	10.38	10.62	13.45	10.39	10.63	13.46
0.3	8.77	10.23	15.46	8.78	10.23	15.47
0.4	8.52	11.28	19.60	8.52	11.28	19.61
0.5	9.07	13.50	26.09	9.08	13.51	26.11
0.6	10.47	17.31	36.27	10.48	17.32	36.29
0.7	13.24	24.02	53.57	13.25	24.04	53.60
0.8	19.19	37.78	88.47	19.20	37.80	88.52
0.9	37.56	79.50	193.58	37.58	79.55	193.69

Table B.6: The values of a_j for $N = (640, 1000)$

B.6 Simulation results

B.6.1 Bias and coverage in estimation of intra-cluster correlation ρ

Fixed ρ

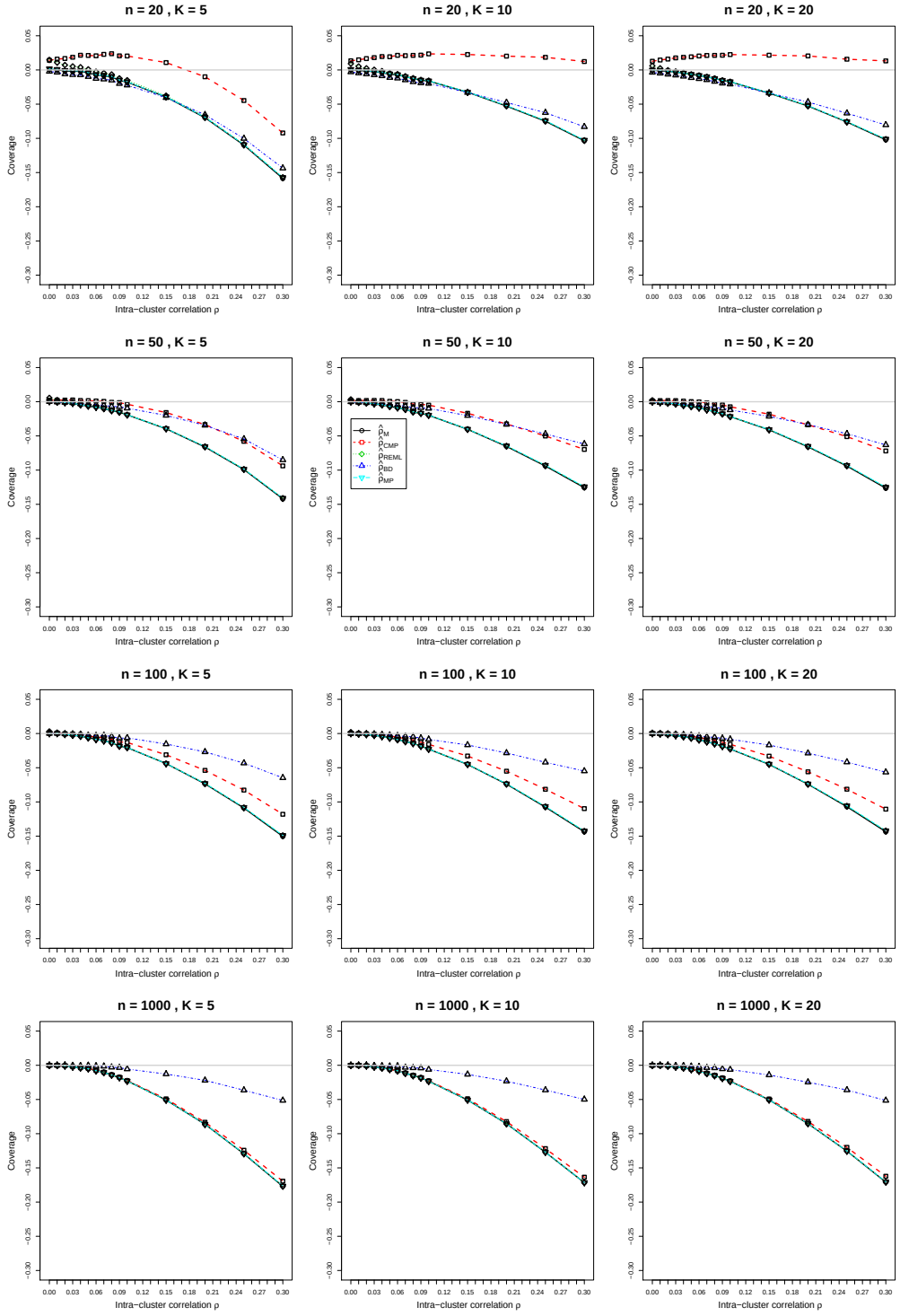


Figure B.1: Bias of the intra-cluster correlation ρ from K studies in beta-binomial model for $p_{2j} = 0.2$, $\theta = 0$ and $0 \leq \rho \leq 0.3$. The method of estimators for ρ:circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on $\Gamma_{r(\rho),\lambda(\rho)}$ distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0 for bias.

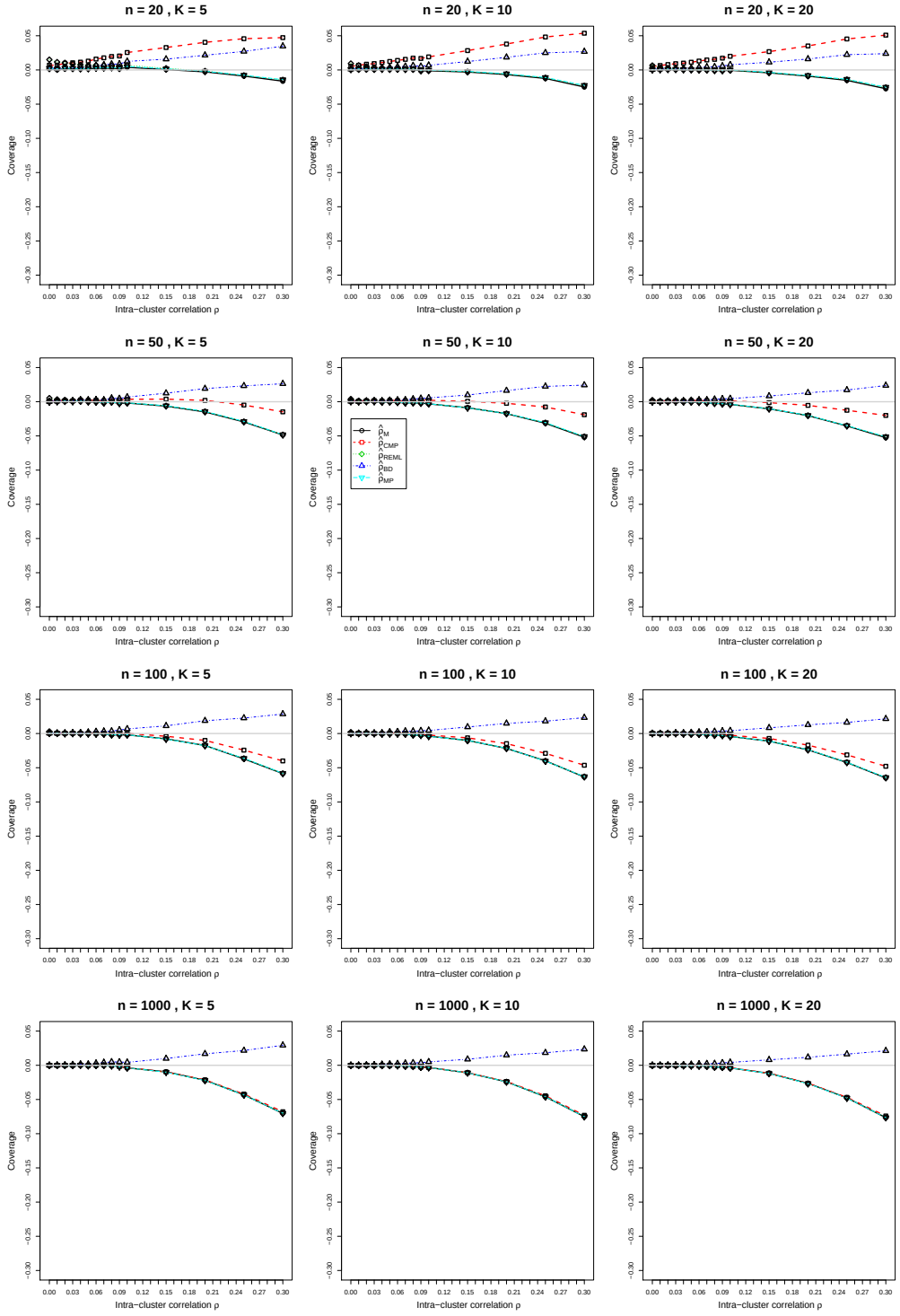


Figure B.2: Bias of the intra-cluster correlation ρ from K studies in beta-binomial model for $p_{2j} = 0.4$, $\theta = 0$ and $0 \leq \rho \leq 0.3$. The method of estimators for ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0 for bias.

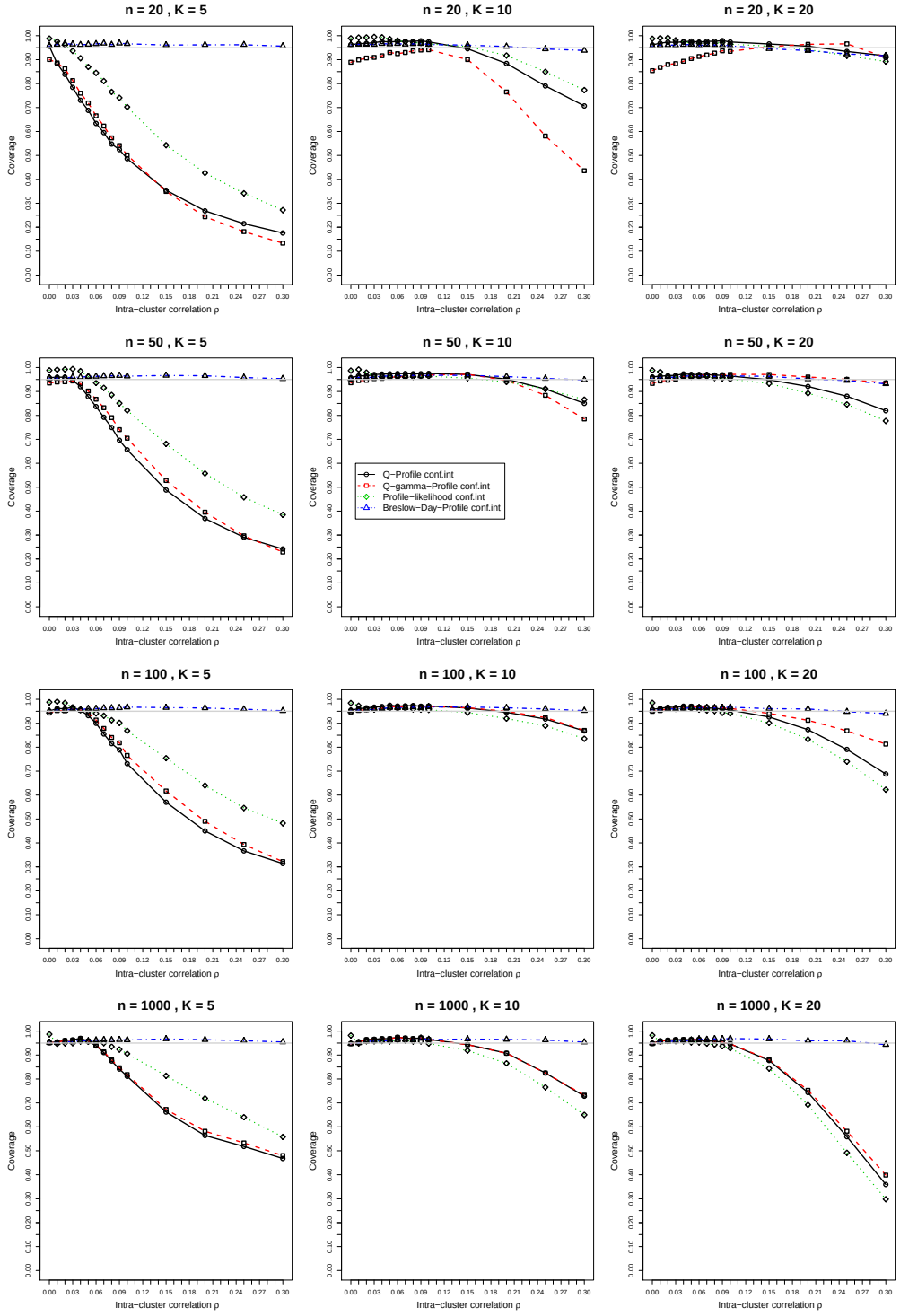


Figure B.3: Coverage at the nominal confidence level of 0.95 of the intra-cluster correlation ρ from K studies in beta-binomial model for $p_{2j} = 0.2$, $\theta = 0$ and $0 \leq \rho \leq 0.3$. The method for obtaining the confidence interval are shown as follows: circles (Q-profile confidence interval for ρ based on χ^2 distribution), squares (Q-profile confidence interval for ρ based on $\Gamma_{r(\rho),\lambda(\rho)}$ distribution), diamonds (Profile likelihood confidence intervals) and triangles (Breslow-Day-Profile confidence interval for ρ based on χ^2 distribution). Light grey line at 0.95 for coverage.

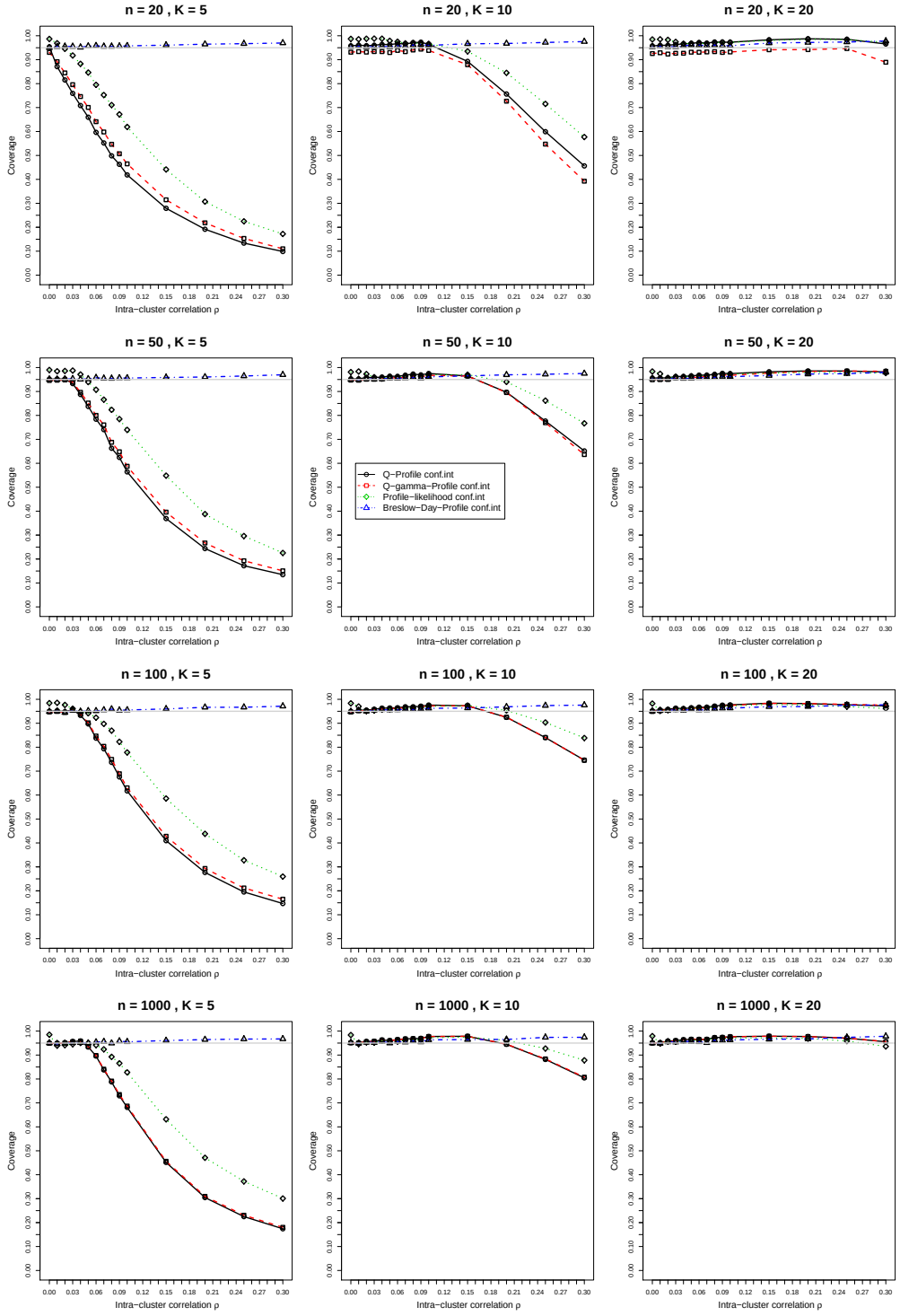


Figure B.4: Coverage at the nominal confidence level of 0.95 of the intra-cluster correlation ρ from K studies in beta-binomial model for $p_{2j} = 0.4$, $\theta = 0$ and $0 \leq \rho \leq 0.3$. The method for obtaining the confidence interval are shown as follows: circles (Q-profile confidence interval for ρ based on χ^2 distribution), squares (Q-profile confidence interval for ρ based on $\Gamma_{r(\rho),\lambda(\rho)}$ distribution), diamonds (Profile likelihood confidence intervals) and triangles (Breslow-Day-Profile confidence interval for ρ based on χ^2 distribution) Light grey line at 0.95 for coverage.

Fixed p_{2j}

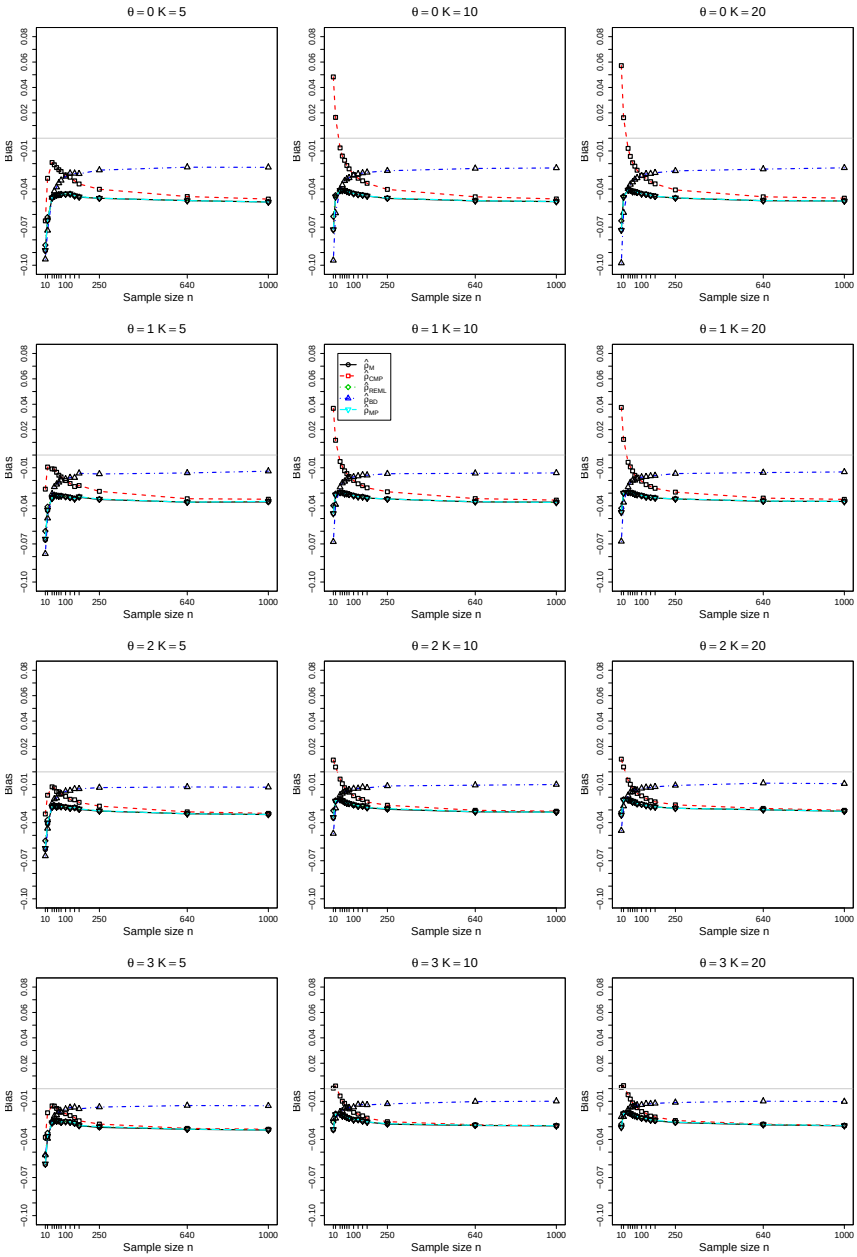


Figure B.5: Bias of the intra-cluster correlation ρ from K studies in beta-binomial model for $p_{2j} = 0.1$, $0 \leq \theta \leq 3$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The method of estimators for ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on $\Gamma_{r(\rho),\lambda(\rho)}$ distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0 for bias.

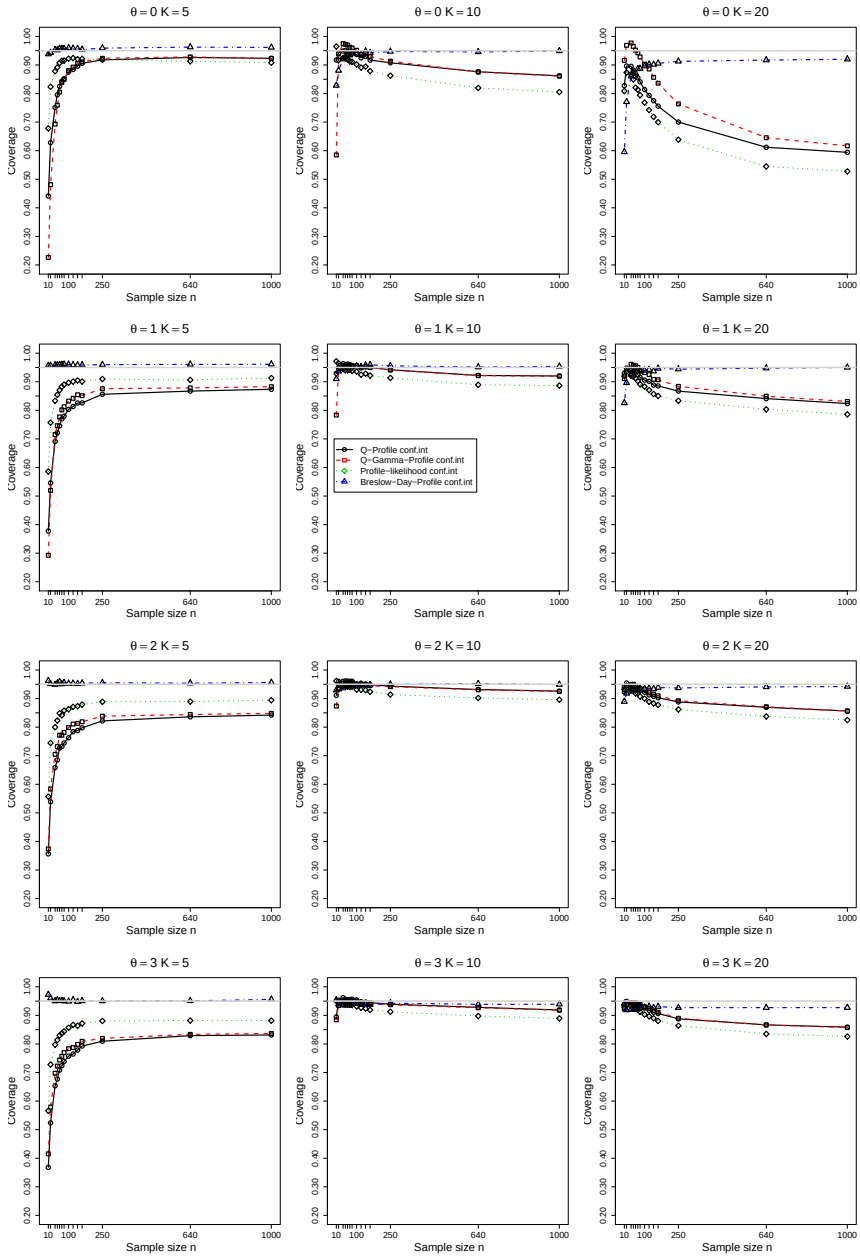


Figure B.6: Coverage at the nominal confidence level of 0.95 of the intra-cluster correlation ρ from K studies in beta-binomial model for $p_{2j} = 0.1$, $0 \leq \theta \leq 3$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The method for obtaining the confidence interval are shown as follows: circles (Q-profile confidence interval for ρ based on χ^2 distribution), squares (Q-profile confidence interval for ρ based on $\Gamma_{r(\rho),\lambda(\rho)}$ distribution), diamonds (Profile likelihood confidence intervals) and triangles (Breslow-Day-Profile confidence interval for ρ based on χ^2 distribution). Light grey line at 0.95 for coverage.

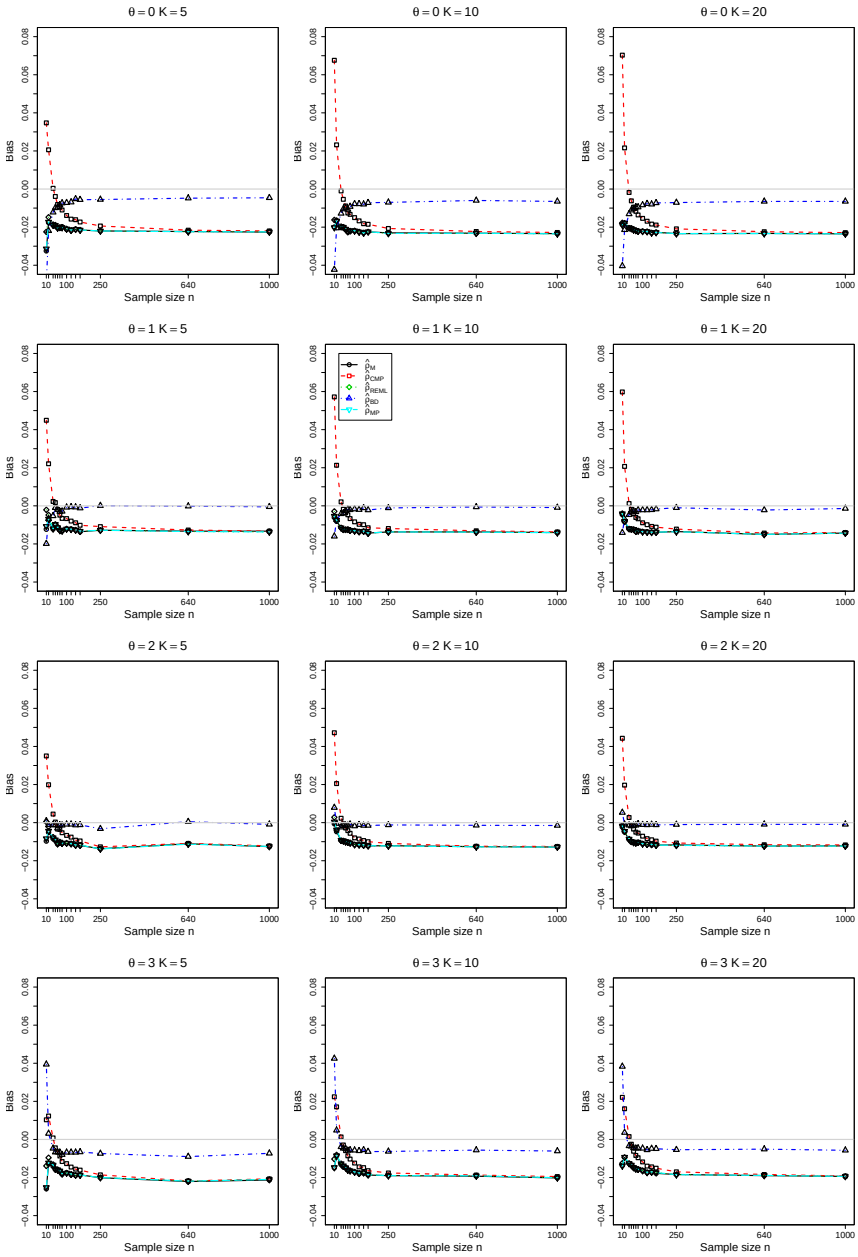


Figure B.7: Bias of the intra-cluster correlation ρ from K studies in beta-binomial model for $p_{2j} = 0.2$, $0 \leq \theta \leq 3$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The method of estimators for ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on $\Gamma_{r(\rho),\lambda(\rho)}$ distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0 for bias.

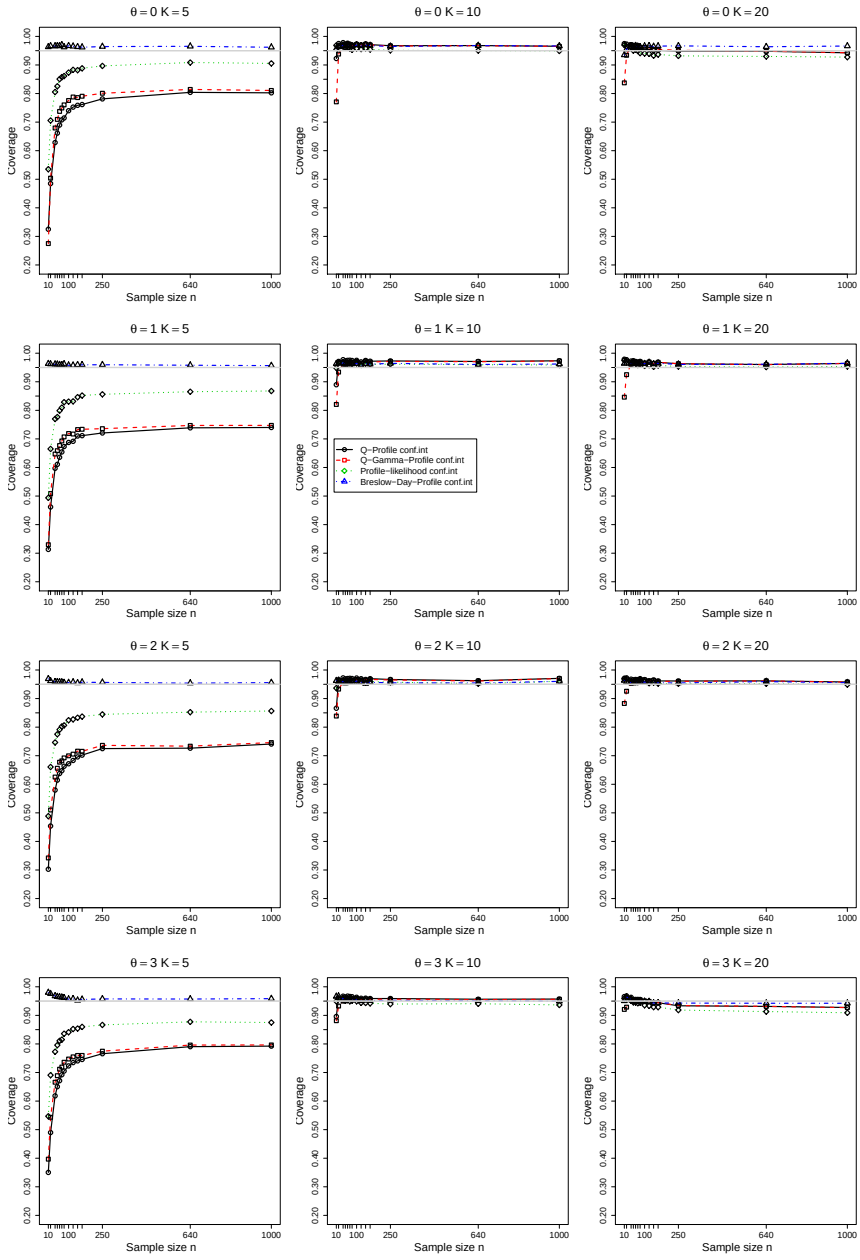


Figure B.8: Coverage at the nominal confidence level of 0.95 of the intra-cluster correlation ρ from K studies in beta-binomial model for $p_{2j} = 0.2$, $0 \leq \theta \leq 3$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The method for obtaining the confidence interval are shown as follows: circles (Q-profile confidence interval for ρ based on χ^2 distribution), squares (Q-profile confidence interval for ρ based on $\Gamma_{r(\rho),\lambda(\rho)}$ distribution), diamonds (Profile likelihood confidence intervals) and triangles (Breslow-Day-Profile confidence interval for ρ based on χ^2 distribution). Light grey line at 0.95 for coverage.

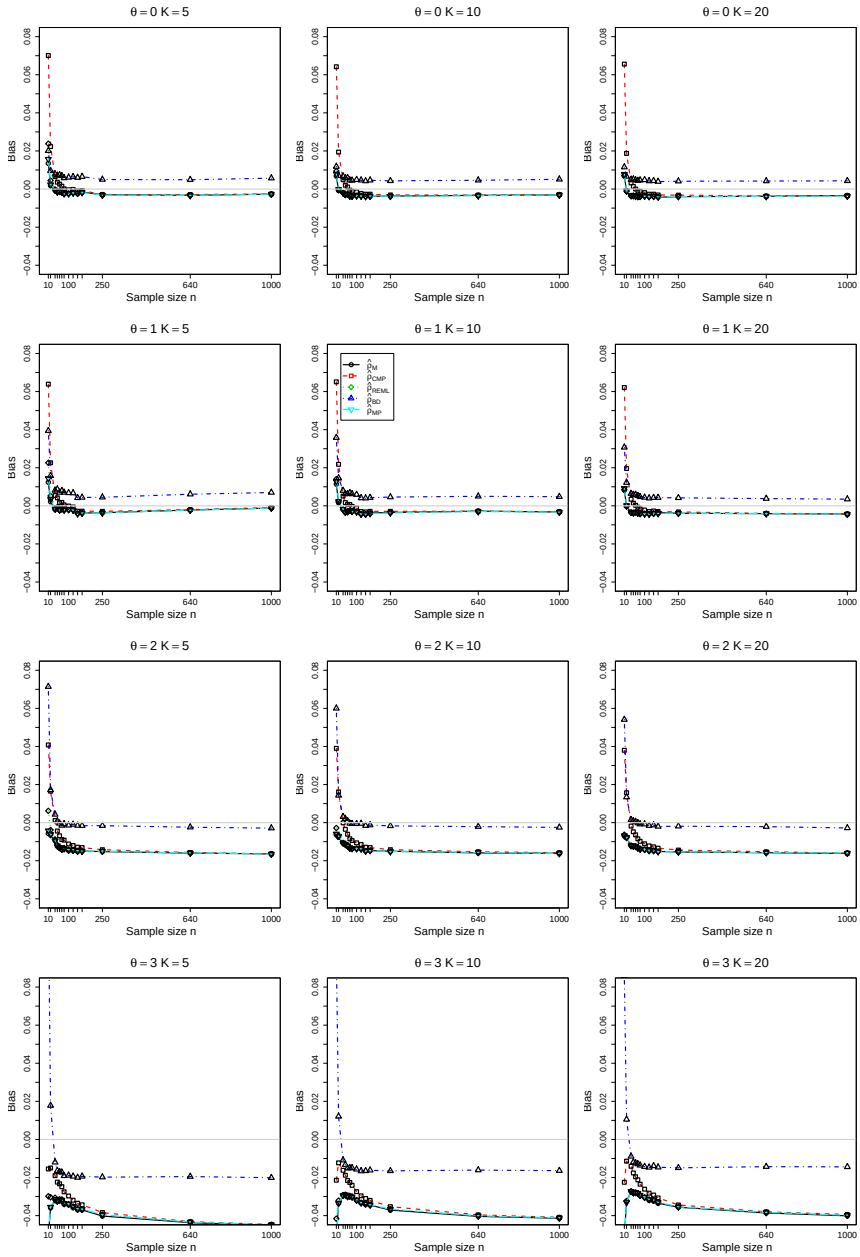


Figure B.9: Bias of the intra-cluster correlation ρ from K studies in beta-binomial model for $p_{2j} = 0.4$, $0 \leq \theta \leq 3$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The method of estimators for ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on $\Gamma_{r(\rho),\lambda(\rho)}$ distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0 for bias.

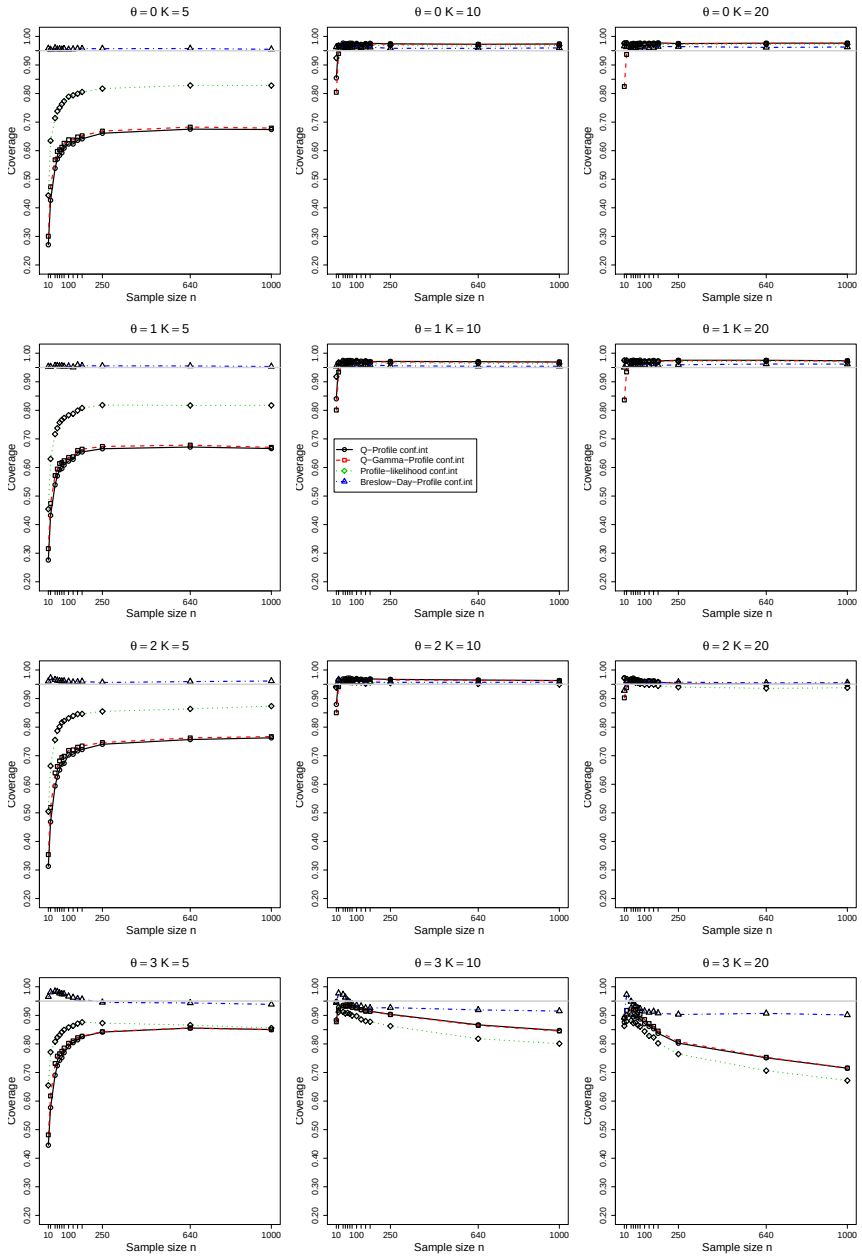


Figure B.10: Coverage at the nominal confidence level of 0.95 of the intra-cluster correlation ρ from K studies in beta-binomial model for $p_{2j} = 0.4$, $0 \leq \theta \leq 3$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The method for obtaining the confidence interval are shown as follows: circles (Q-profile confidence interval for ρ based on χ^2 distribution), squares (Q-profile confidence interval for ρ based on $\Gamma_{r(\rho),\lambda(\rho)}$ distribution), diamonds (Profile likelihood confidence intervals) and triangles (Breslow-Day-Profile confidence interval for ρ based on χ^2 distribution). Light grey line at 0.95 for coverage.

Fixed θ

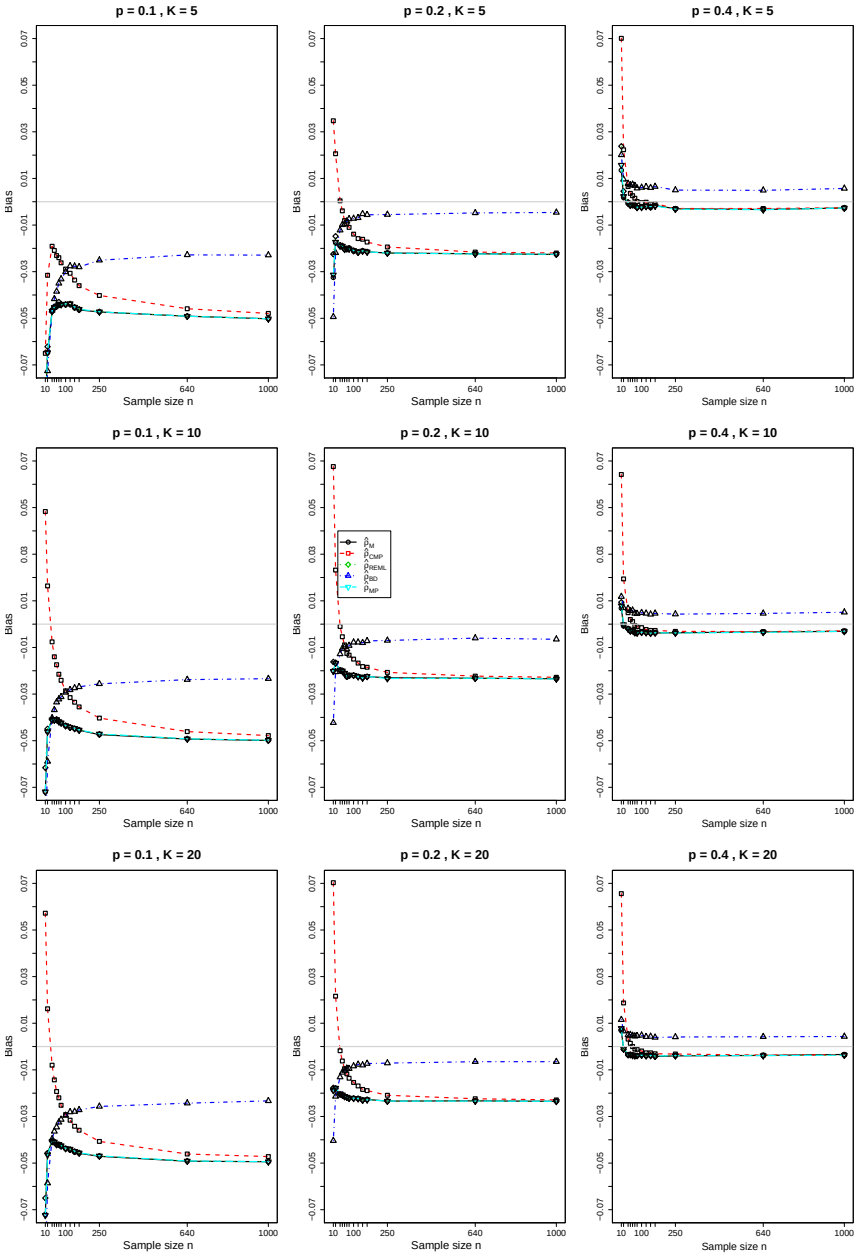


Figure B.11: Bias of the intra-cluster correlation ρ from K studies in beta-binomial model for $0.1 \leq p_{2j} \leq 0.4$, $\theta = 0$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The method of estimators for ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on $\Gamma_{r(\rho),\lambda(\rho)}$ distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0 for bias.

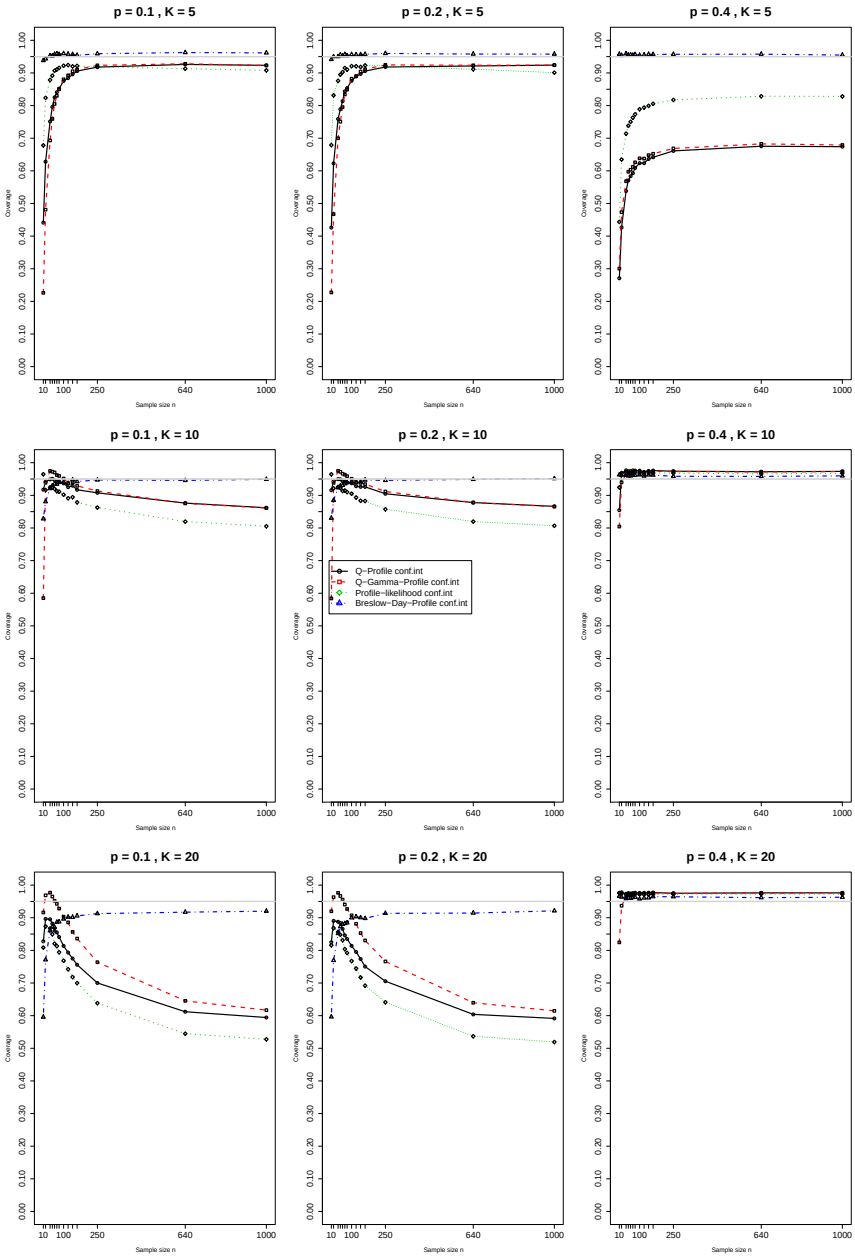


Figure B.12: Coverage at the nominal confidence level of 0.95 of the intra-cluster correlation ρ from K studies in beta-binomial model for $0.1 \leq p_{2j} \leq 0.4$, $\theta = 0$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The method for obtaining the confidence interval are shown as follows: circles (Q-profile confidence interval for ρ based on χ^2 distribution), squares (Q-profile confidence interval for ρ based on $\Gamma_{r(\rho),\lambda(\rho)}$ distribution), diamonds (Profile likelihood confidence intervals) and triangles (Breslow-Day-Profile confidence interval for ρ based on χ^2 distribution). Light grey line at 0.95 for coverage.

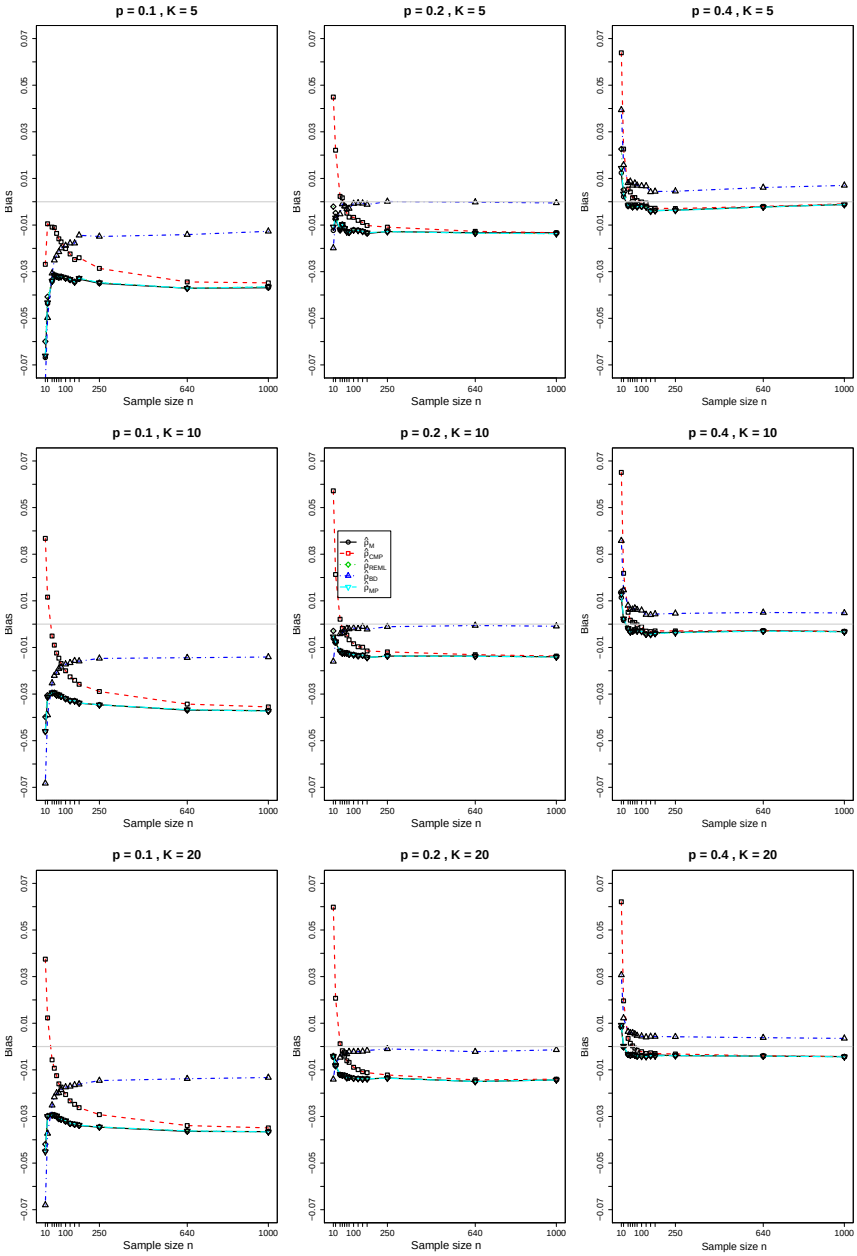


Figure B.13: Bias of the intra-cluster correlation ρ from K studies in beta-binomial model for $0.1 \leq p_{2j} \leq 0.4$, $\theta = 1$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The method of estimators for ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on $\Gamma_{r(\rho),\lambda(\rho)}$ distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0 for bias.

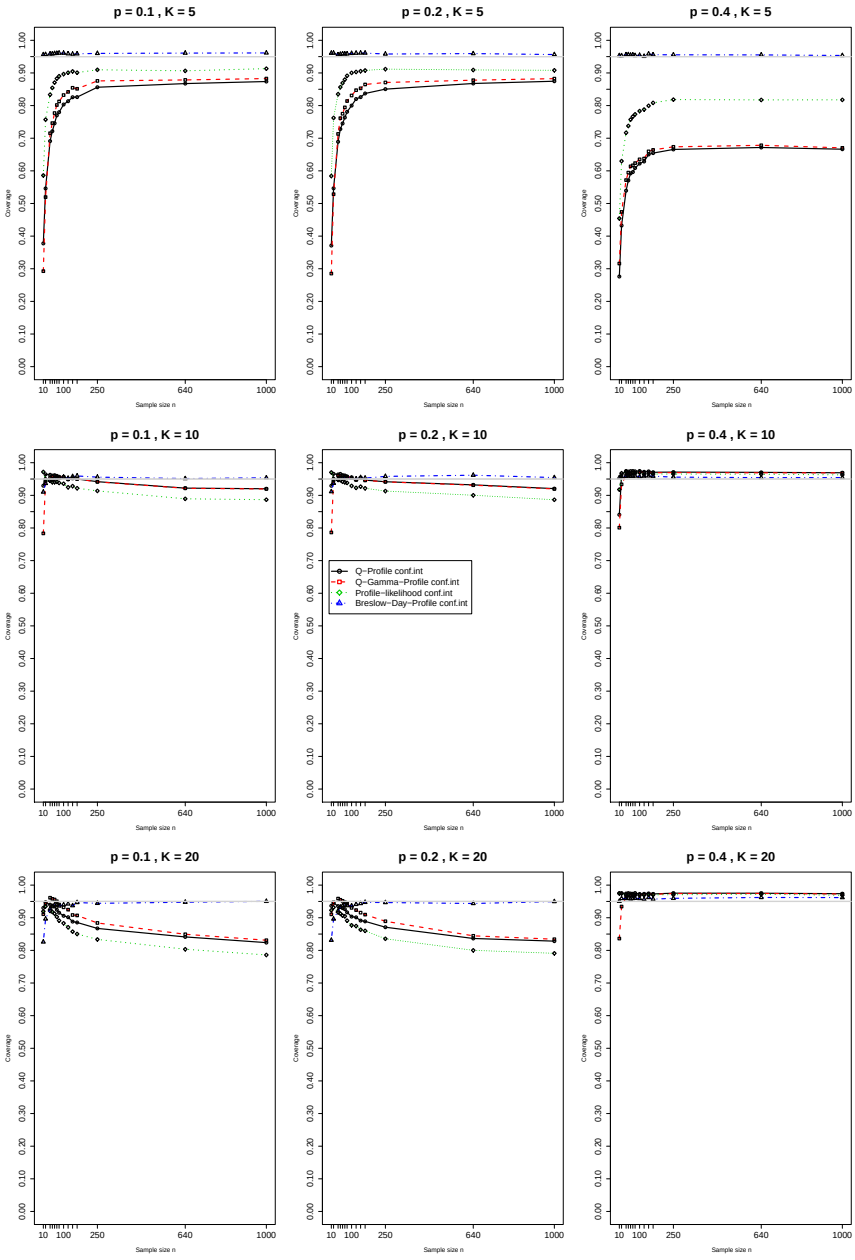


Figure B.14: Coverage at the nominal confidence level of 0.95 of the intra-cluster correlation ρ from K studies in beta-binomial model for $0.1 \leq p_{2j} \leq 0.4$, $\theta = 1$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The method for obtaining the confidence interval are shown as follows: circles (Q-profile confidence interval for ρ based on χ^2 distribution), squares (Q-profile confidence interval for ρ based on $\Gamma_{r(\rho),\lambda(\rho)}$ distribution), diamonds (Profile likelihood confidence intervals) and triangles (Breslow-Day-Profile confidence interval for ρ based on χ^2 distribution). Light grey line at 0.95 for coverage.

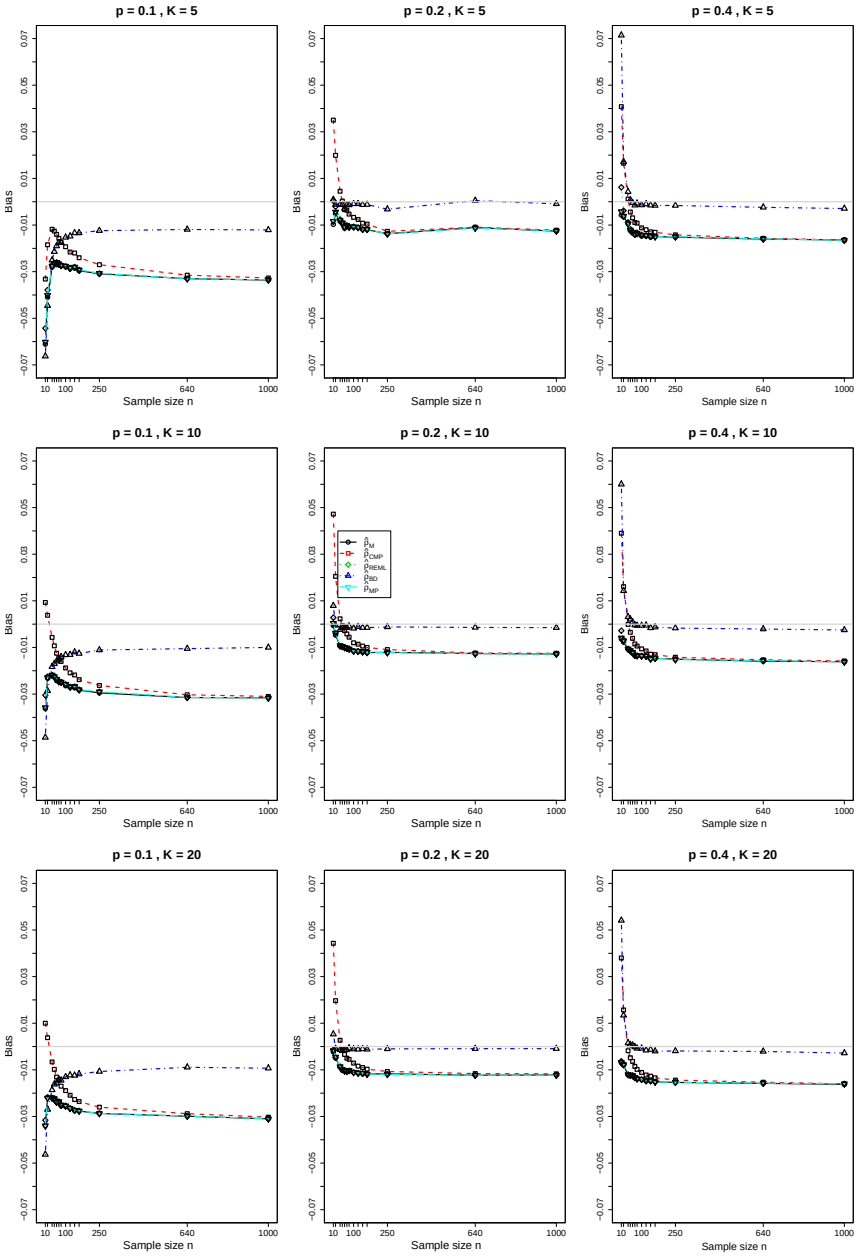


Figure B.15: Bias of the intra-cluster correlation ρ from K studies in beta-binomial model for $0.1 \leq p_{2j} \leq 0.4$, $\theta = 2$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The method of estimators for ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on $\Gamma_{r(\rho),\lambda(\rho)}$ distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0 for bias.

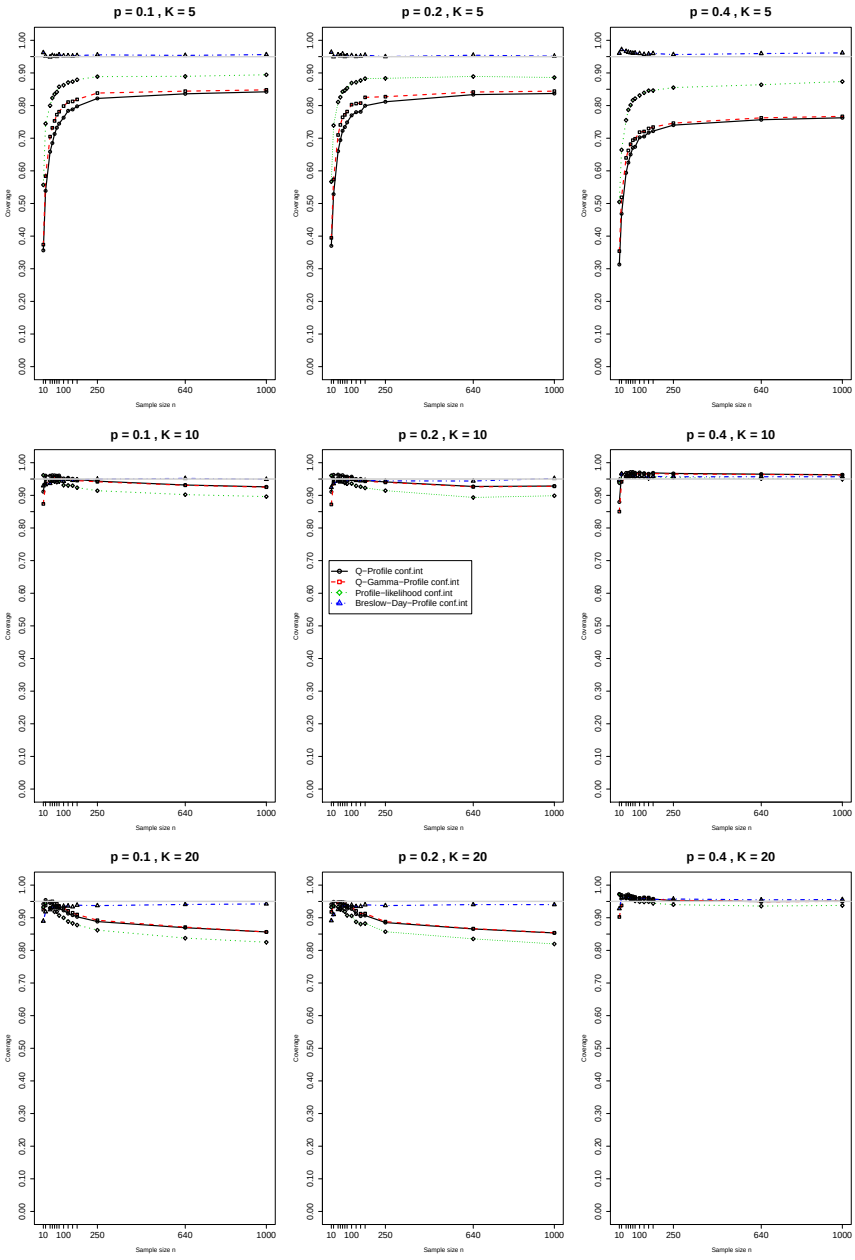


Figure B.16: Coverage at the nominal confidence level of 0.95 of the intra-cluster correlation ρ from K studies in beta-binomial model for $0.1 \leq p_{2j} \leq 0.4$, $\theta = 2$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The method for obtaining the confidence interval are shown as follows: circles (Q-profile confidence interval for ρ based on χ^2 distribution), squares (Q-profile confidence interval for ρ based on $\Gamma_{r(\rho),\lambda(\rho)}$ distribution), diamonds (Profile likelihood confidence intervals) and triangles (Breslow-Day-Profile confidence interval for ρ based on χ^2 distribution). Light grey line at 0.95 for coverage.

B.6.2 Bias and coverage in estimation of overall effect measure θ_{IV}

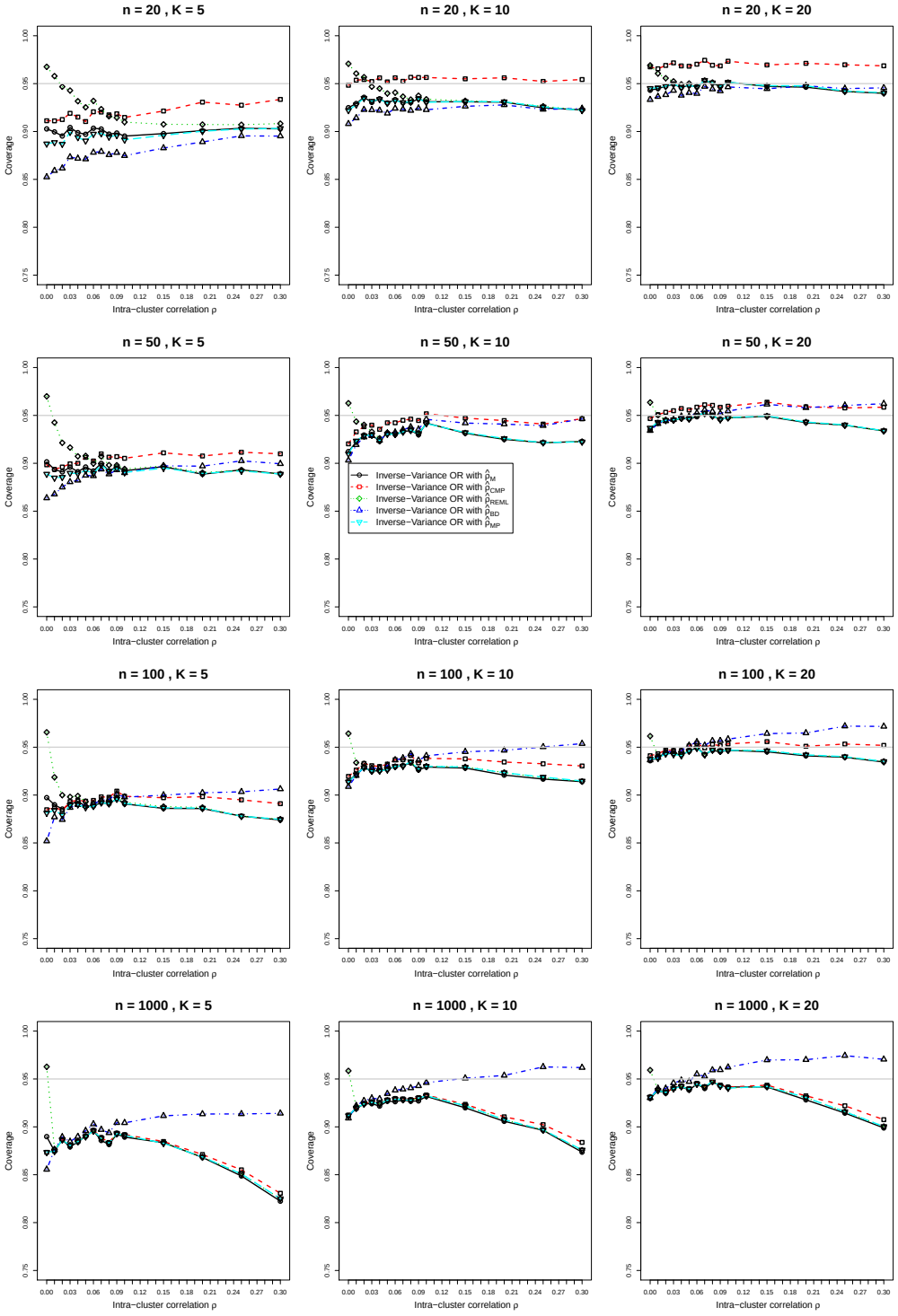


Figure B.17: Coverage at the nominal confidence level of 0.95 of the inverse-variance overall effect measure θ_{IV} from K studies in beta-binomial model for $p_{2j} = 0.2$, $\theta = 0$ and $0 \leq \rho \leq 0.3$. The inverse-variance weights use the estimators of ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0.95 for coverage.

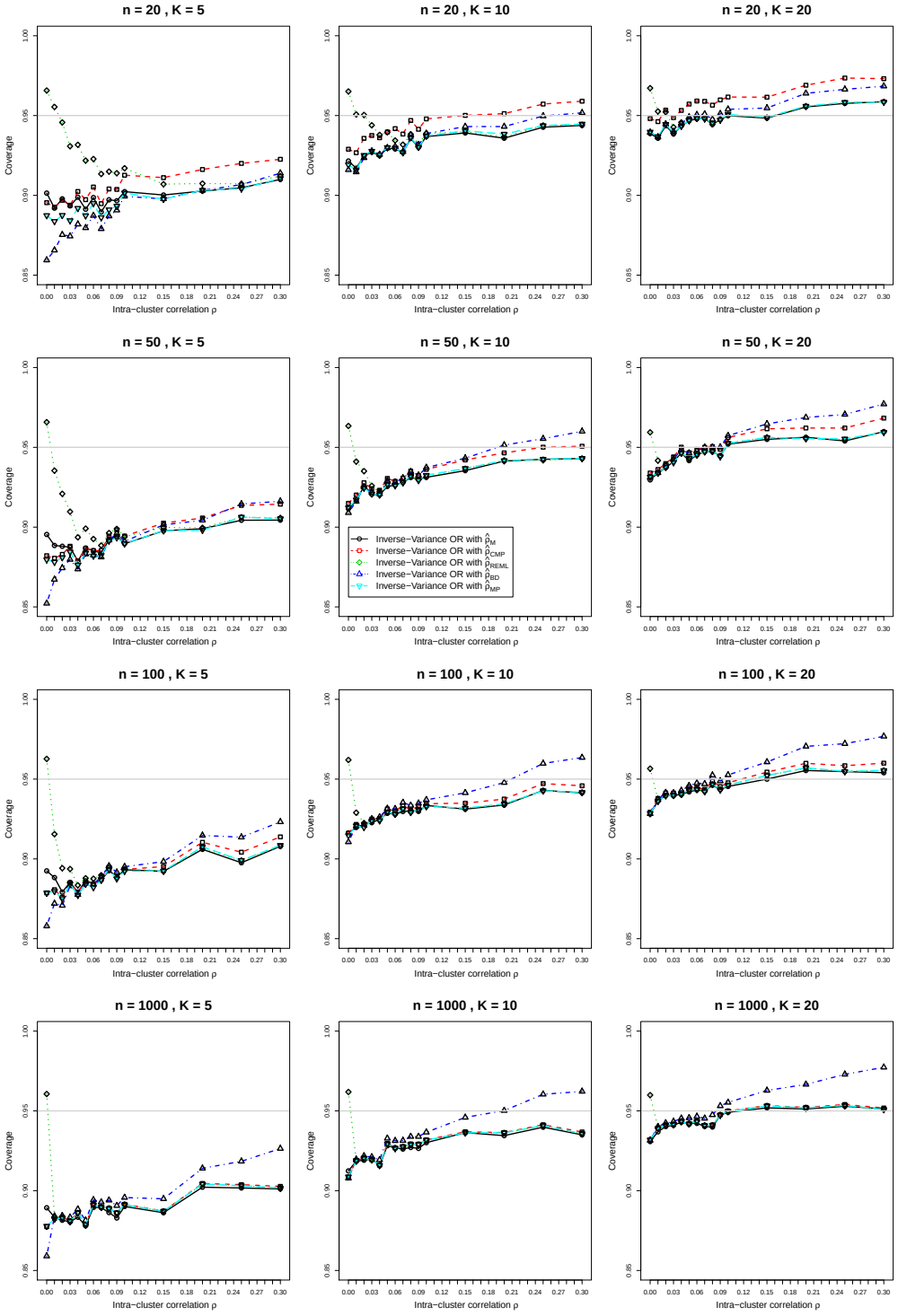


Figure B.18: Coverage at the nominal confidence level of 0.95 of the inverse-variance overall effect measure θ_{IV} from K studies in beta-binomial model for $p_{2j} = 0.4$, $\theta = 0$ and $0 \leq \rho \leq 0.3$. The inverse-variance weights use the estimators of ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0.95 for coverage.

B.6.3 Fixed θ

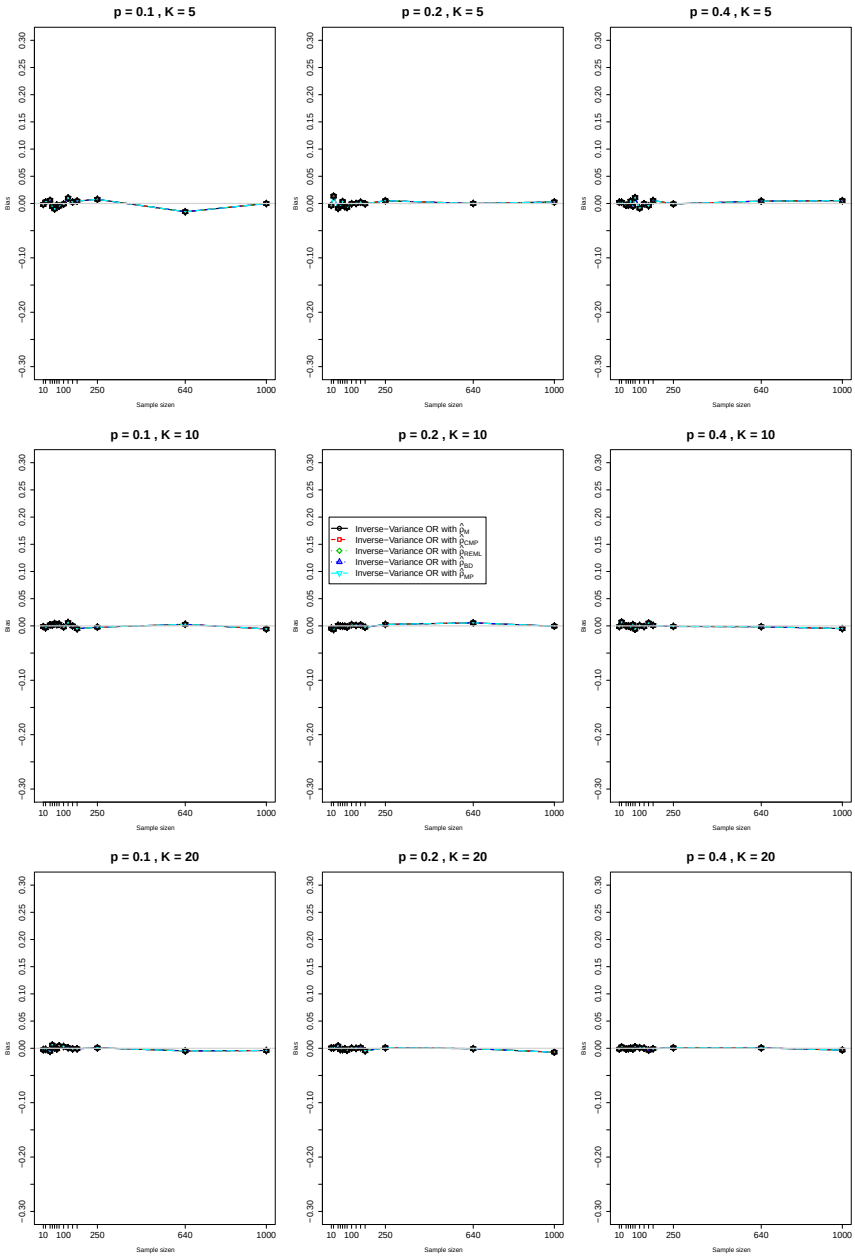


Figure B.19: Bias of the inverse-variance overall effect measure ψ_{IV} from K studies in beta-binomial model for $0.1 \leq p_{2j} \leq 0.4$, $\theta = 0$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The inverse-variance weights use the estimators of ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0 for bias.

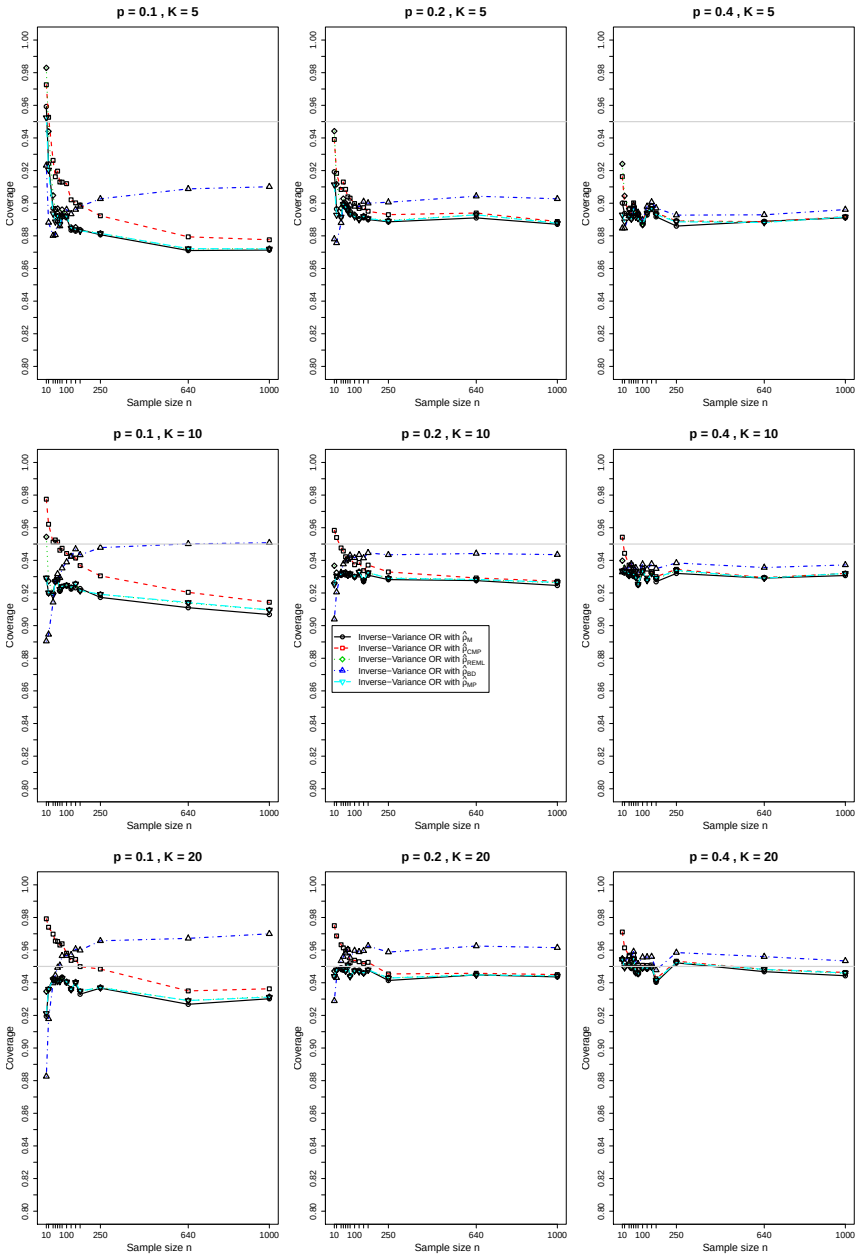


Figure B.20: Coverage at the nominal confidence level of 0.95 of the Inverse-Variance overall effect measure θ_{IV} from K studies in beta-binomial model for $0.1 \leq p_{2j} \leq 0.4$, $\theta = 0$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The weights of the Mandel-Haenzsel odds ratio use the estimators of ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0.95 for coverage.

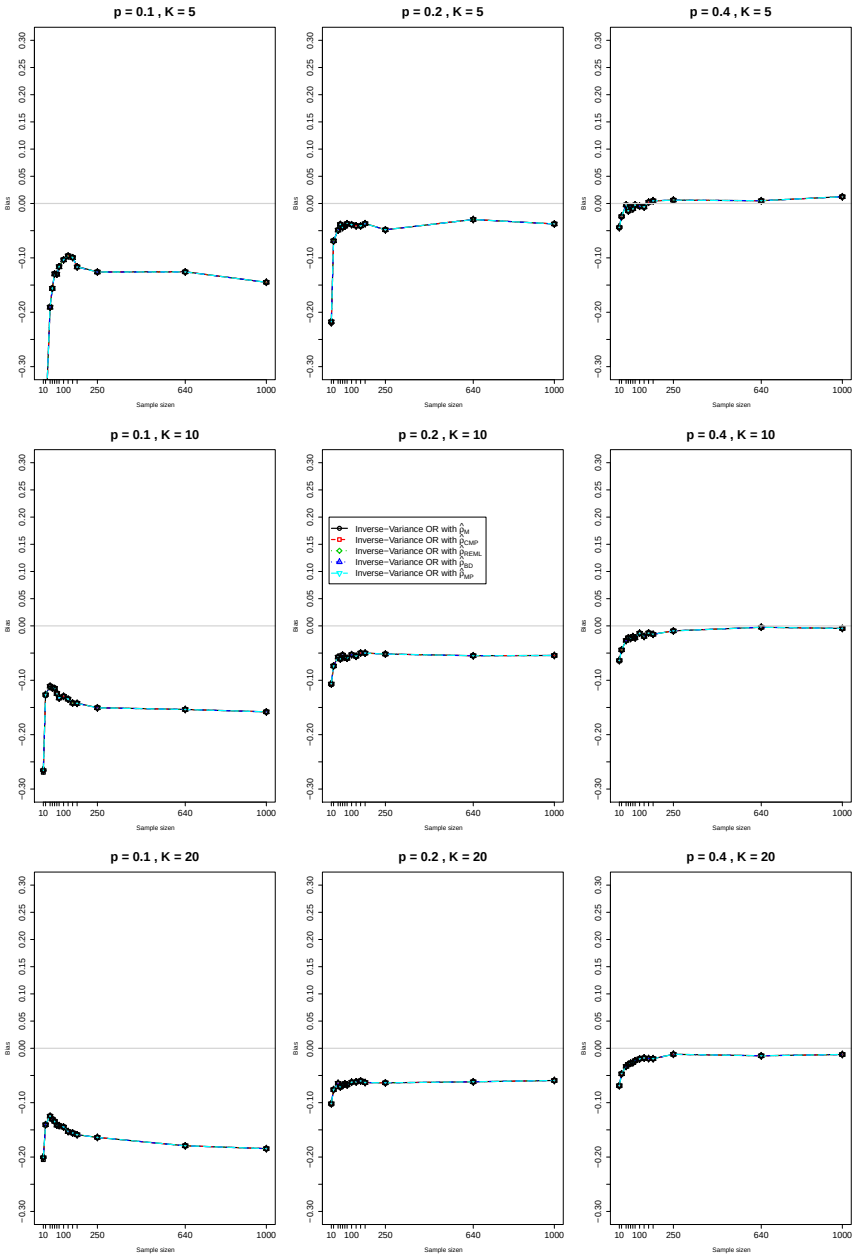


Figure B.21: Bias of the inverse-variance overall effect measure ψ_{IV} from K studies in beta-binomial model for $0.1 \leq p_{2j} \leq 0.4$, $\theta = 1$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The inverse-variance weights use the estimators of ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0 for bias.

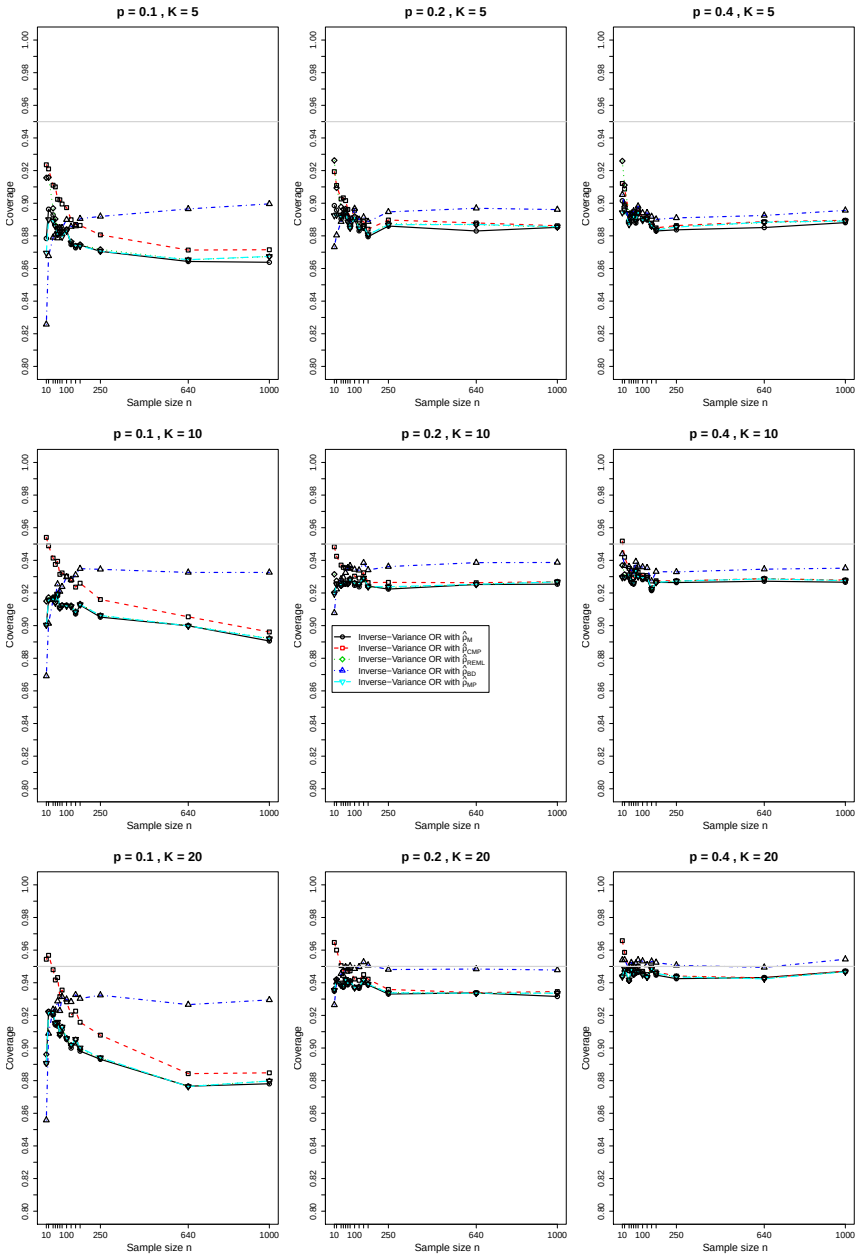


Figure B.22: Coverage at the nominal confidence level of 0.95 of the Inverse-Variance overall effect measure θ_{IV} from K studies in beta-binomial model for $0.1 \leq p_{2j} \leq 0.4$, $\theta = 1$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The weights of the Mandel-Haenzsel odds ratio use the estimators of ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0.95 for coverage.

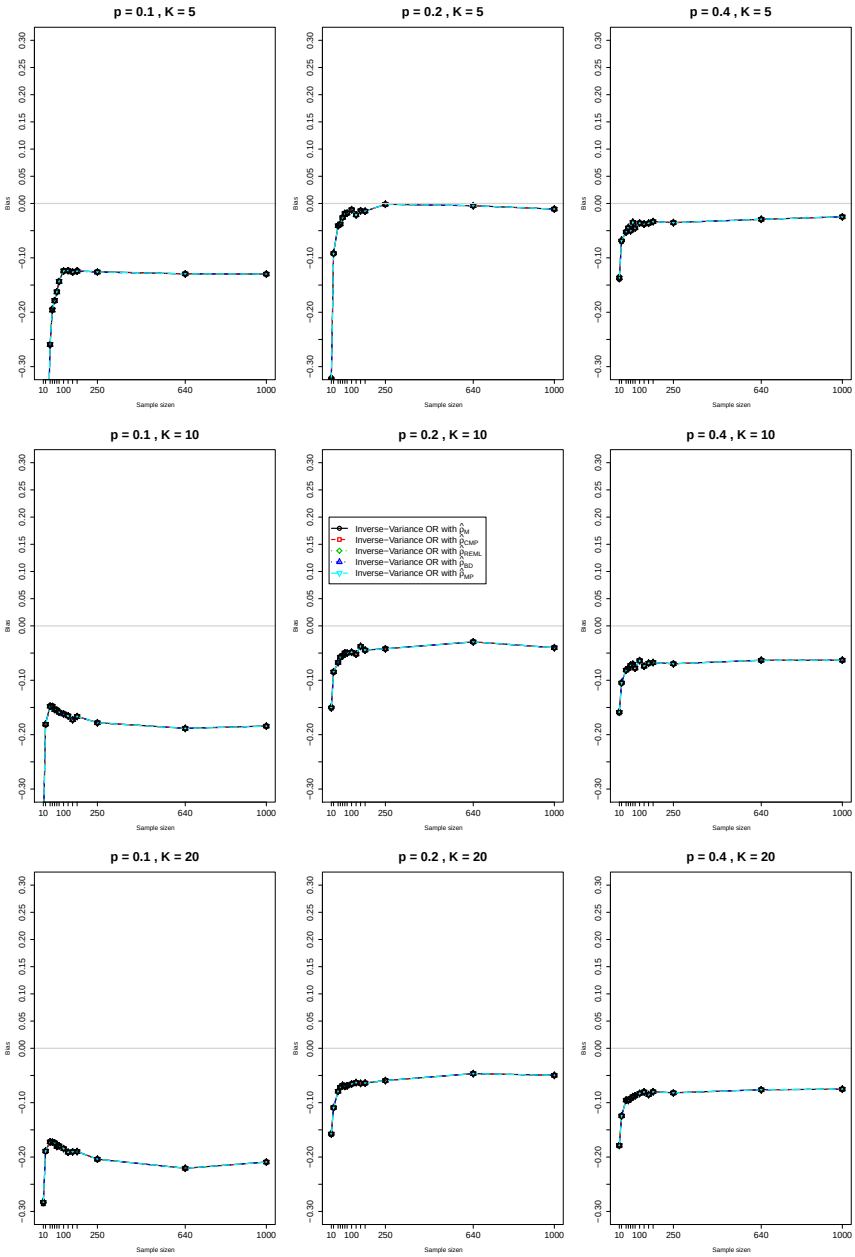


Figure B.23: Bias of the inverse-variance overall effect measure ψ_{IV} from K studies in beta-binomial model for $0.1 \leq p_{2j} \leq 0.4$, $\theta = 2$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The inverse-variance weights use the estimators of ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0 for bias.

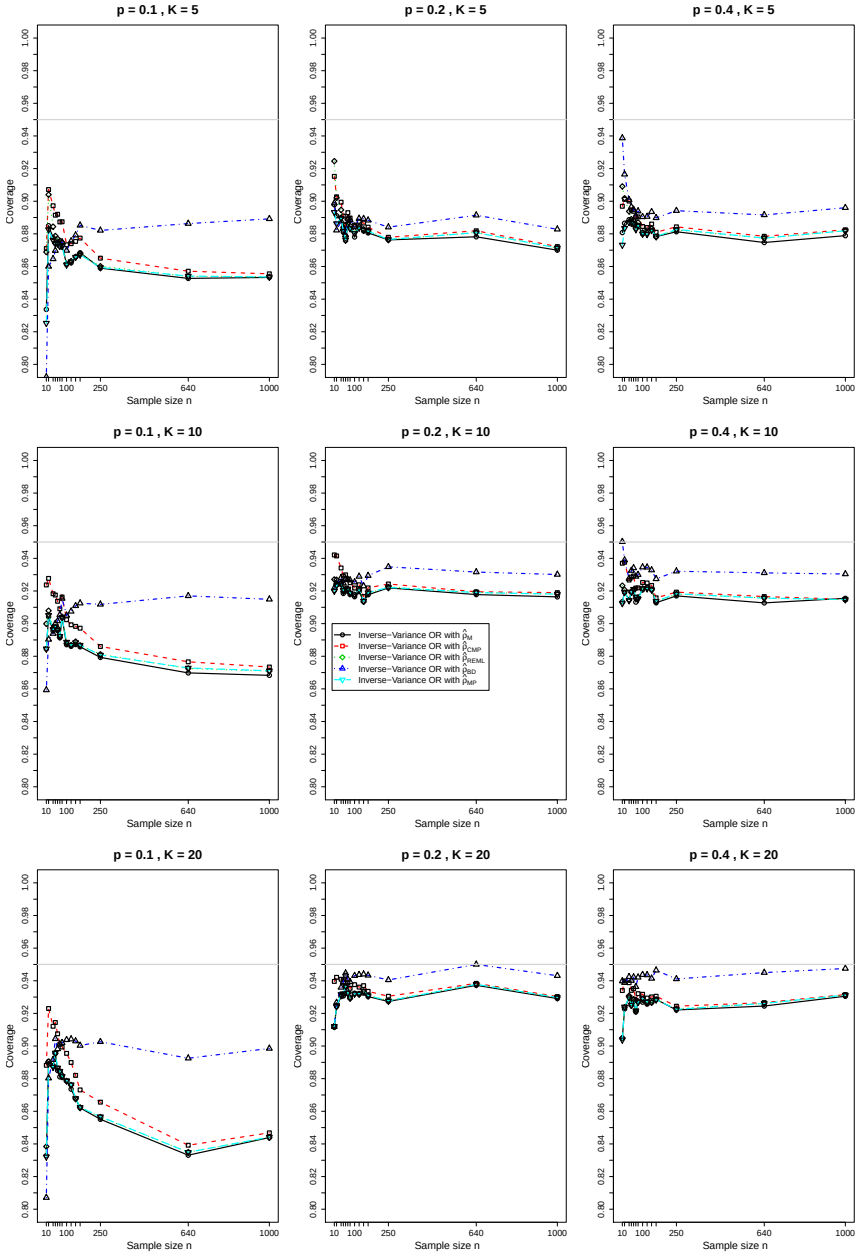


Figure B.24: Coverage at the nominal confidence level of 0.95 of the Inverse-Variance overall effect measure θ_{IV} from K studies in beta-binomial model for $0.1 \leq p_{2j} \leq 0.4$, $\theta = 2$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The weights of the Mandel-Haenzsel odds ratio use the estimators of ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0.95 for coverage.

B.6.4 Bias and coverage in estimation of overall effect measure θ_{MH}

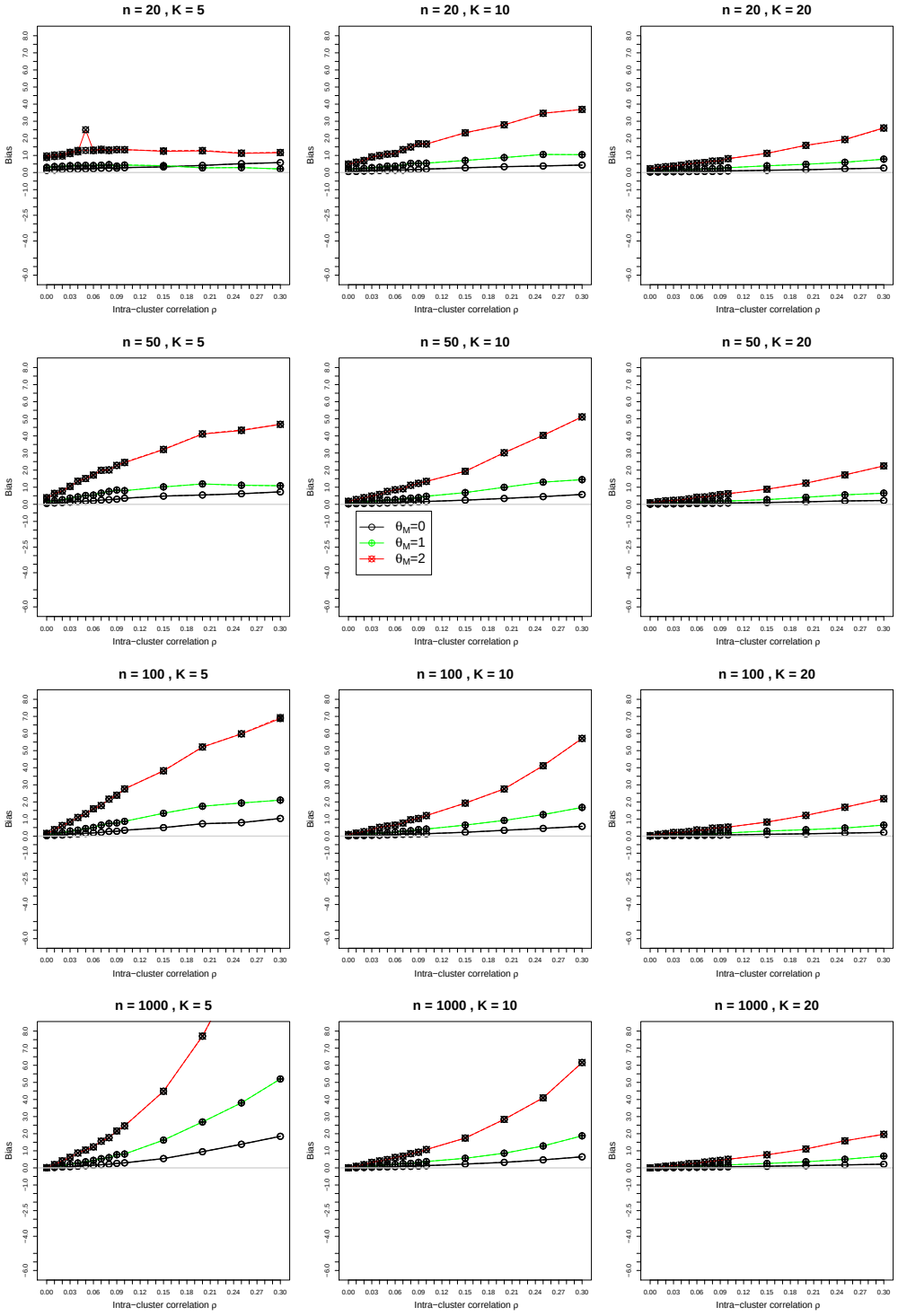


Figure B.25: Bias of the Mantel-Haenzsel overall effect measure psi_{MH} from K studies in beta-binomial model for $p_{2j} = 0.1$, $\theta = 0$ and $0 \leq \rho \leq 0.3$. The weights of the Mantel-Haenzsel odds ratio use the estimators of ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0 for bias.

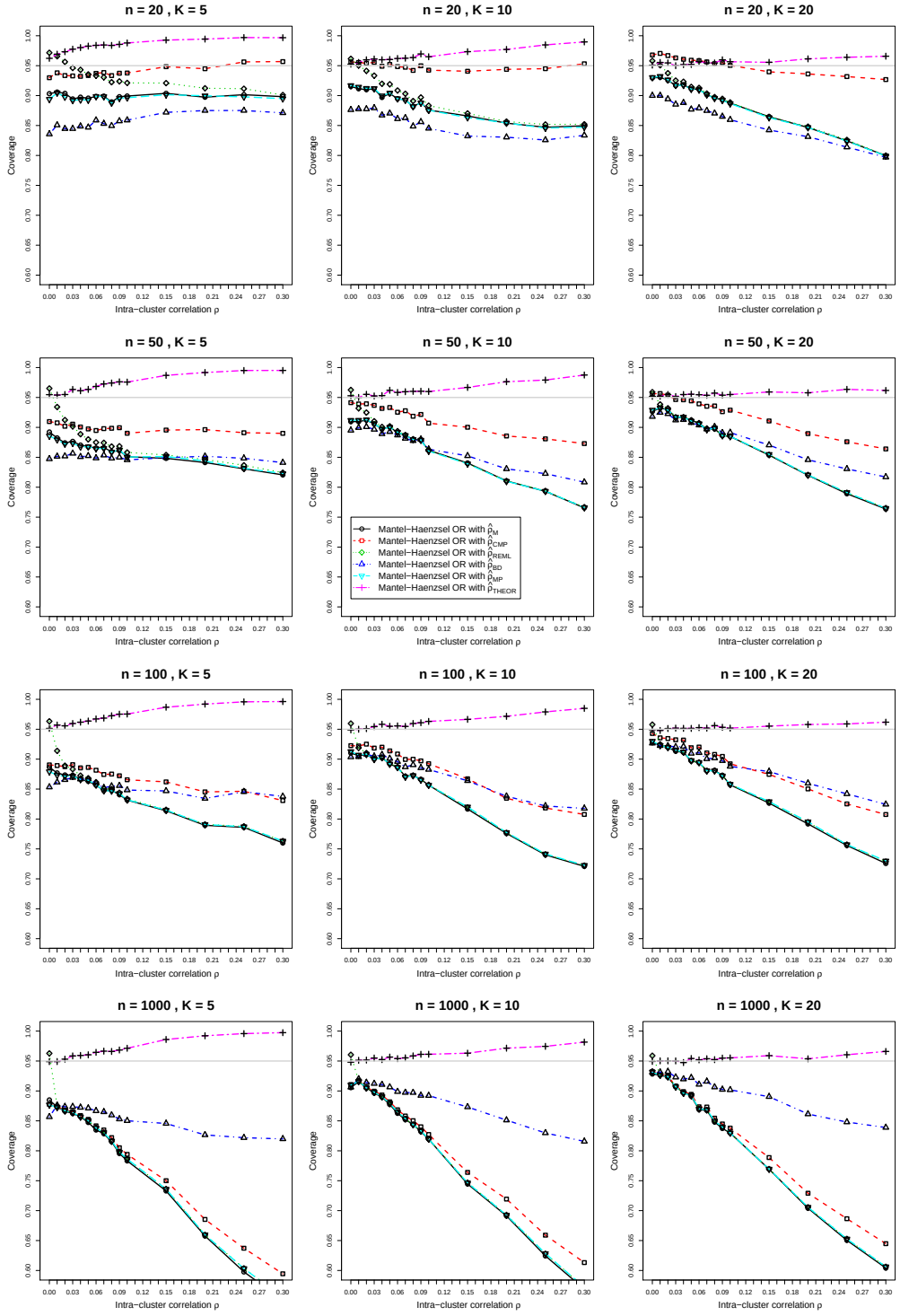


Figure B.26: Coverage at the nominal confidence level of 0.95 of the Mantel-Haenzsel overall effect measure θ_{MH} from K studies in beta-binomial model for $p_{2j} = 0.1$, $\theta = 0$ and $0 \leq \rho \leq 0.3$. The weights of the Mandel-Haenzsel odds ratio use the estimators of ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0.95 for coverage.

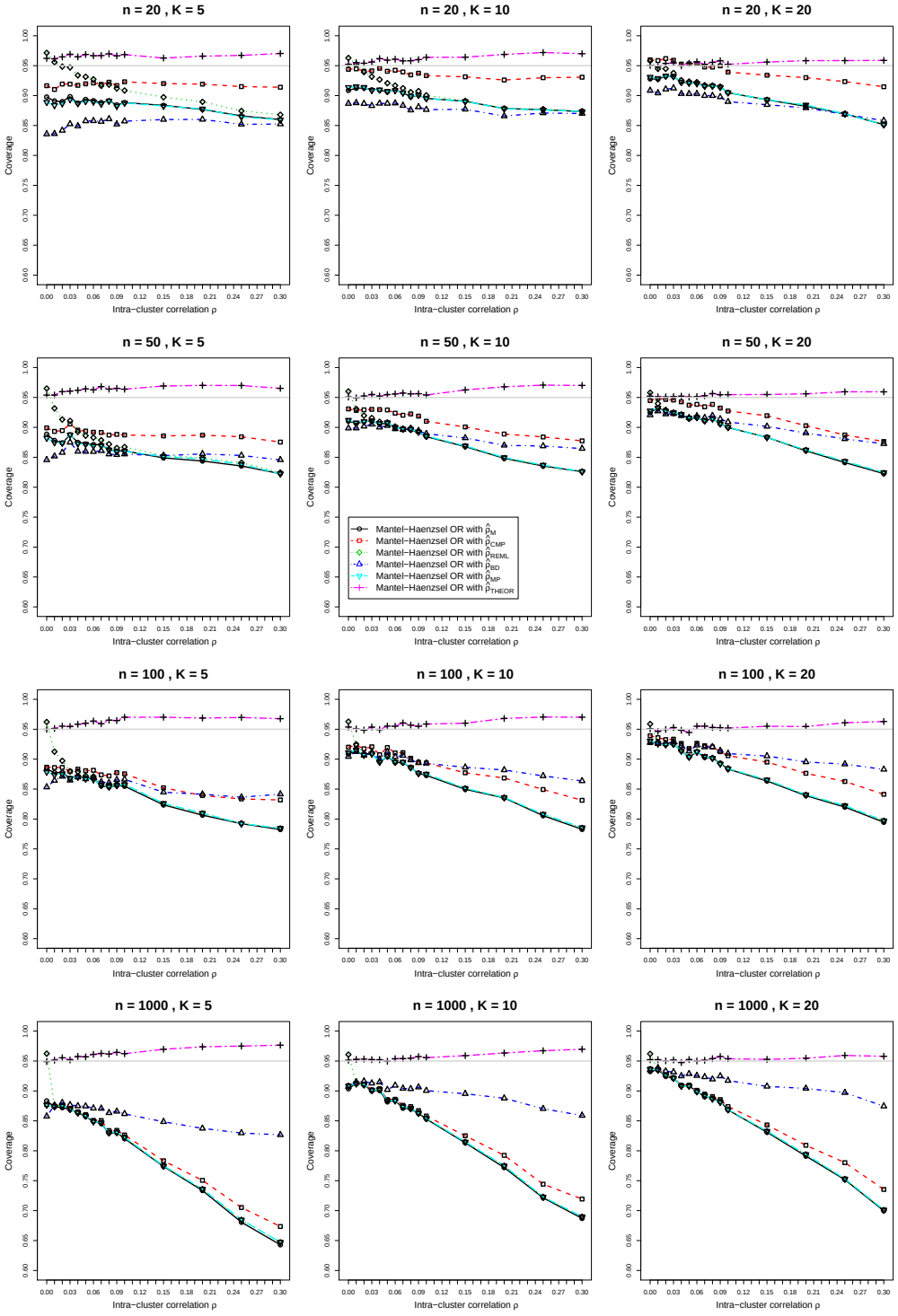


Figure B.27: Coverage of the Mantel-Haenzsel overall effect measure ψ_{MH} from K studies in beta-binomial model for $p_{2j} = 0.1$, $\theta = 1$ and $0 \leq \rho \leq 0.3$. The weights of the Mantel-Haenzsel odds ratio use the estimators of ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0 for bias.

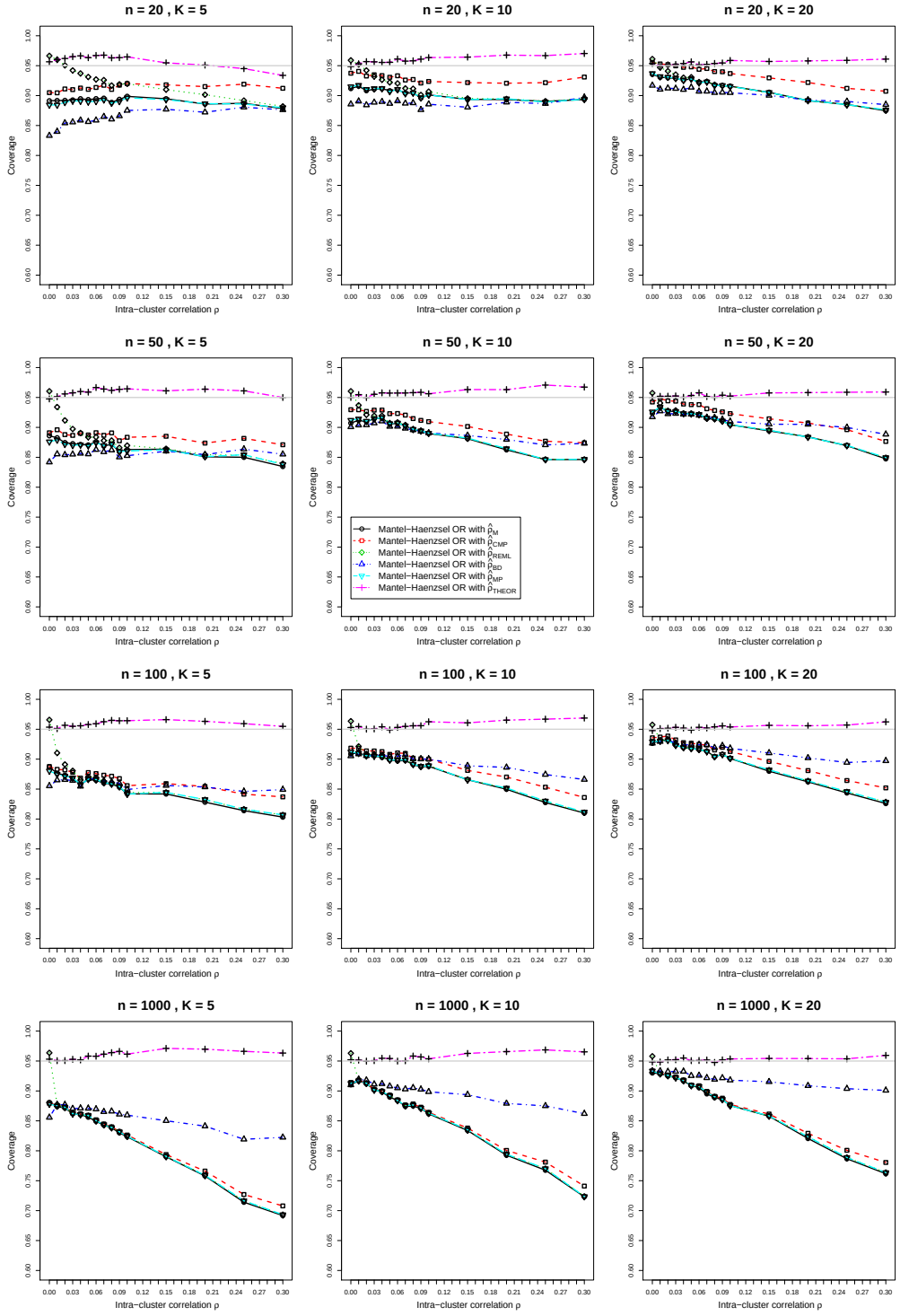


Figure B.28: Coverage of the Mantel-Haenzsel overall effect measure ψ_{MH} from K studies in beta-binomial model for $p_{2j} = 0.1$, $\theta = 2$ and $0 \leq \rho \leq 0.3$. The weights of the Mantel-Haenzsel odds ratio use the estimators of ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0 for bias.

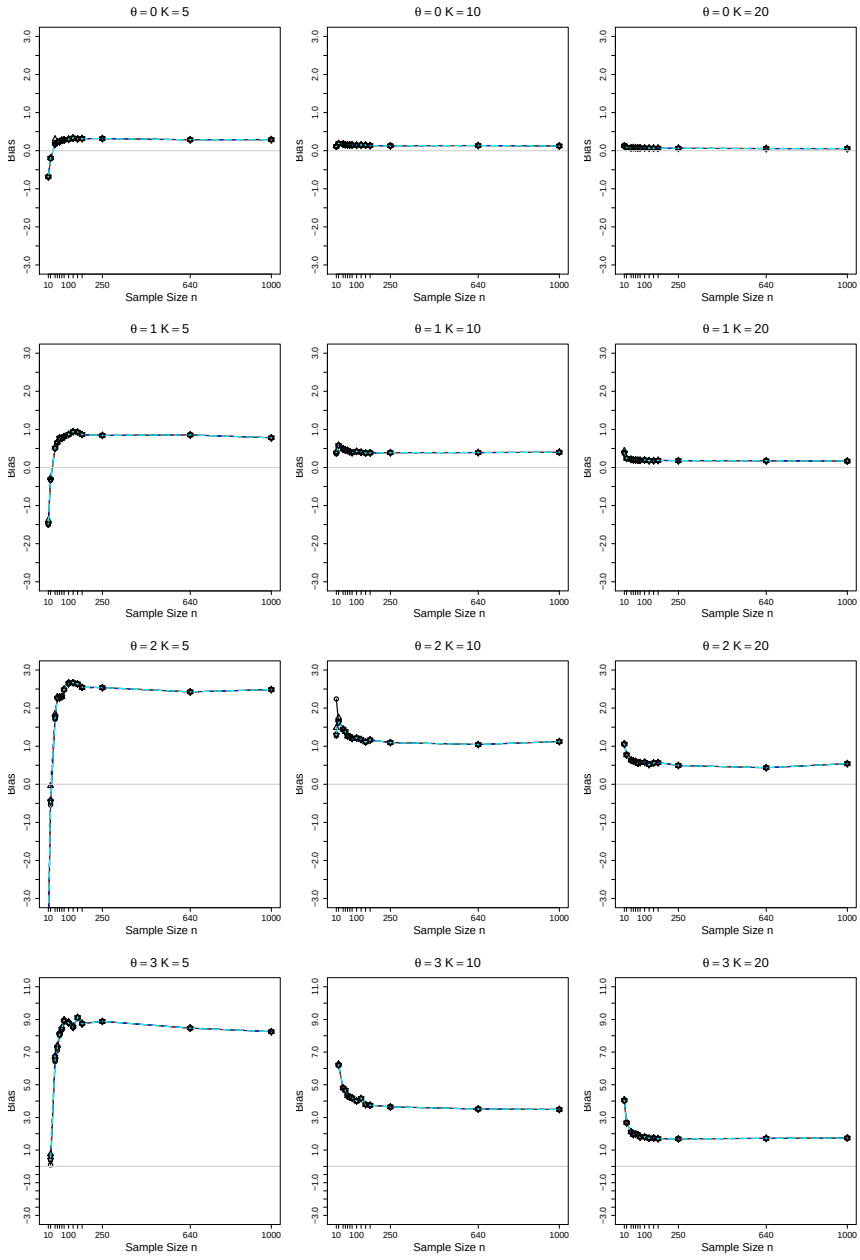


Figure B.29: Bias of the Mantel-Haenzsel overall effect measure θ_{MH} from K studies in beta-binomial model for $p_{2j} = 0.1$, $0 \leq \theta \leq 3$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The weights of the Mantel-Haenzsel odds ratio use the estimators of ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0 for bias.

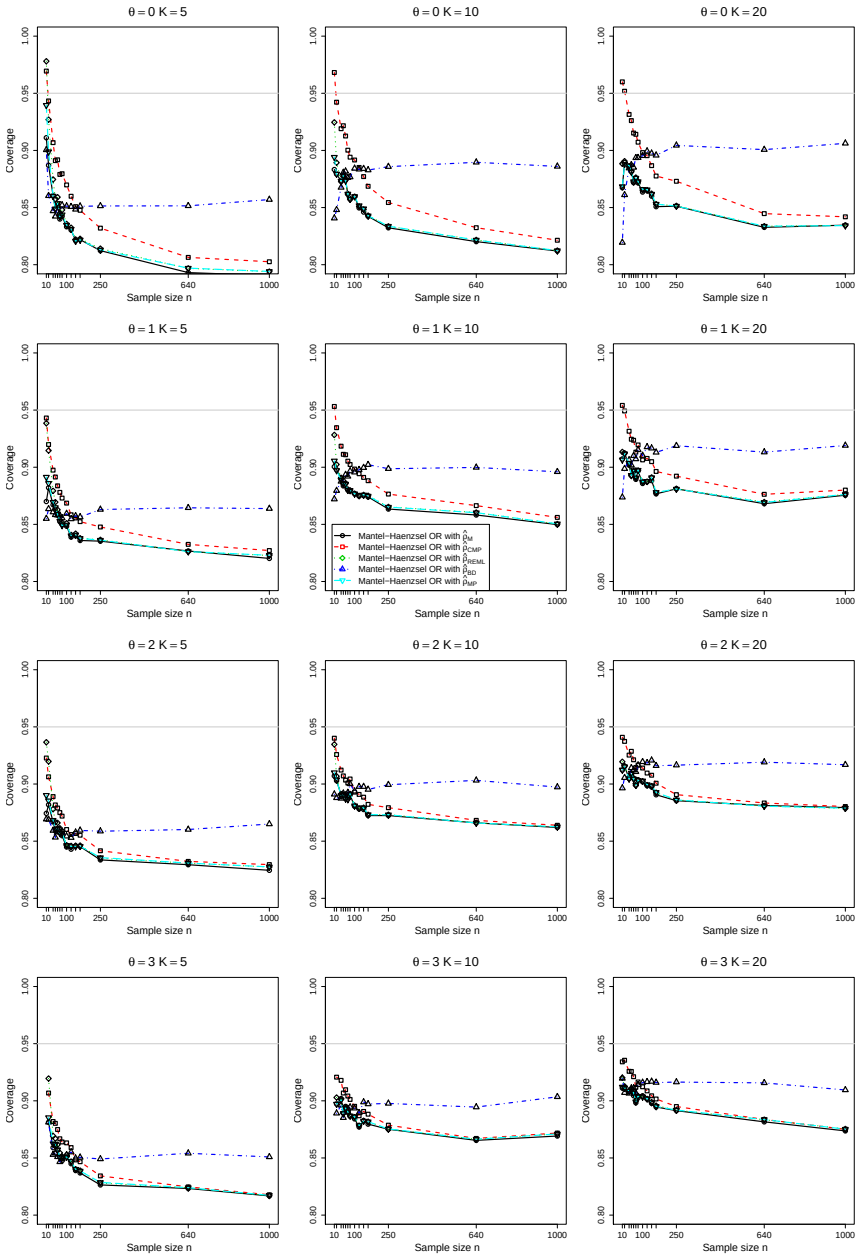


Figure B.30: Coverage at the nominal confidence level of 0.95 of the Mantel-Haenzsel overall effect measure θ_{MH} from K studies in beta-binomial model for $p_{2j} = 0.1$,$0 \leq \theta \leq 3$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The weights of the Mandel-Haenzsel odds ratio use the estimators of ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0.95 for coverage.

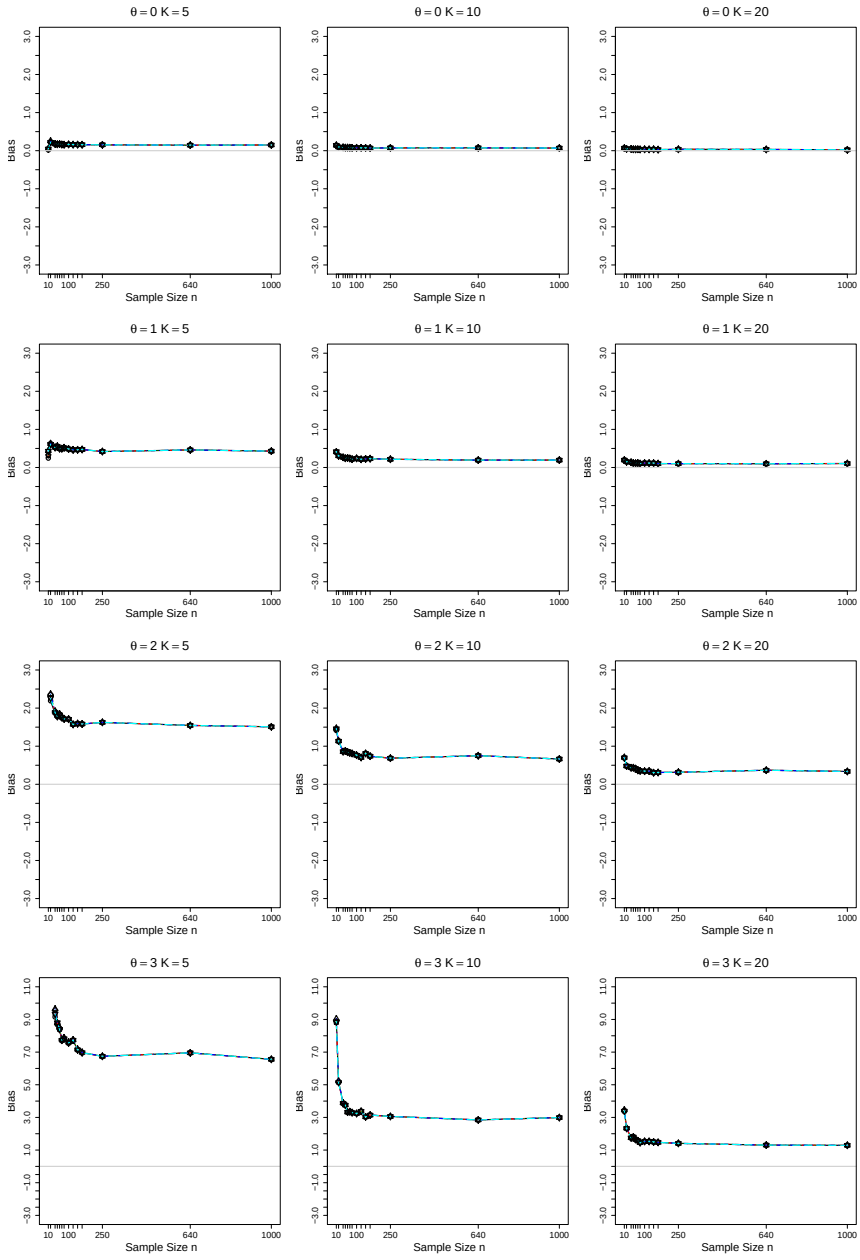


Figure B.31: Bias of the Mantel-Haenzsel overall effect measure ψ_{MH} from K studies in beta-binomial model for $p_{2j} = 0.2$, $0 \leq \theta \leq 3$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The weights of the Mantel-Haenzsel odds ratio use the estimators of ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0 for bias.

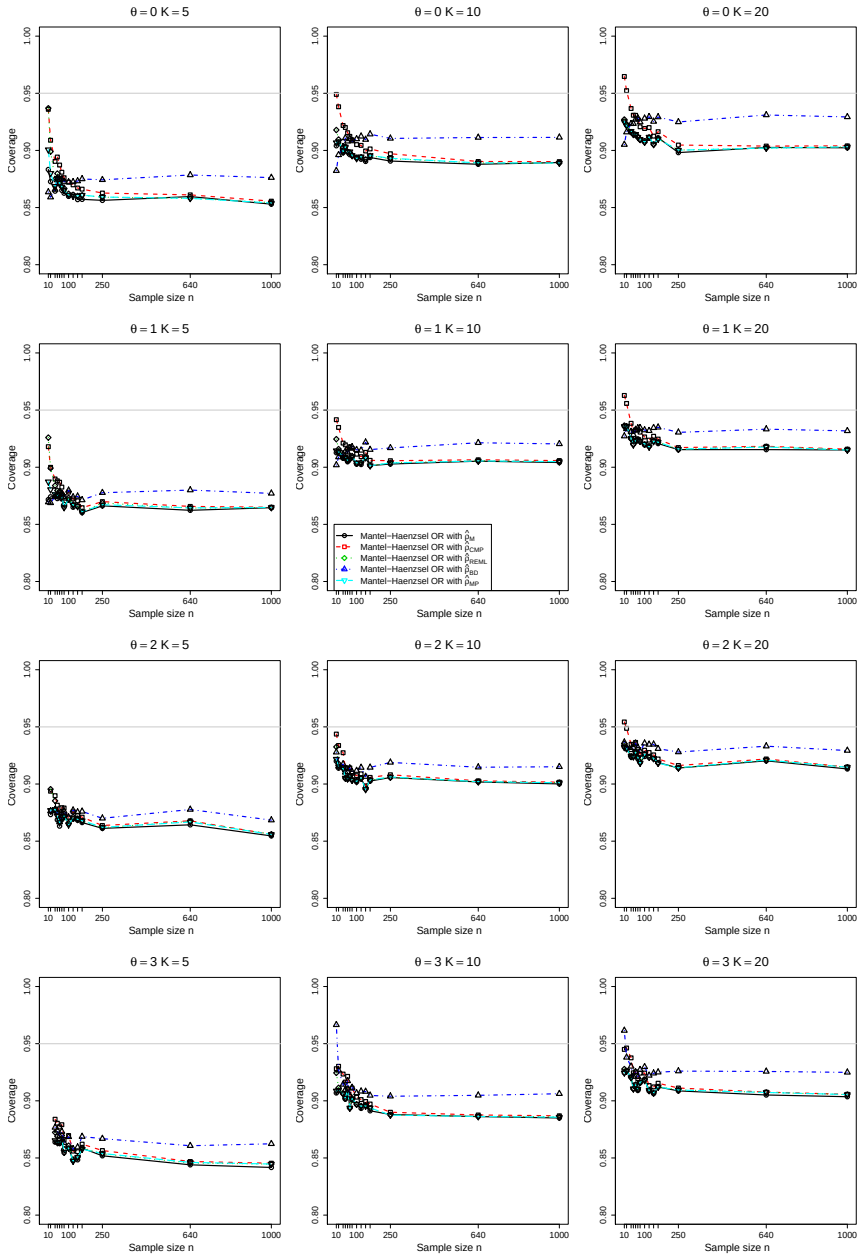


Figure B.32: Coverage at the nominal confidence level of 0.95 of the Mantel-Haenzsel overall effect measure θ_{MH} from K studies in beta-binomial model for $p_{2j} = 0.2$, $0 \leq \theta \leq 3$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The weights of the Mandel-Haenzsel odds ratio use the estimators of ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0.95 for coverage.

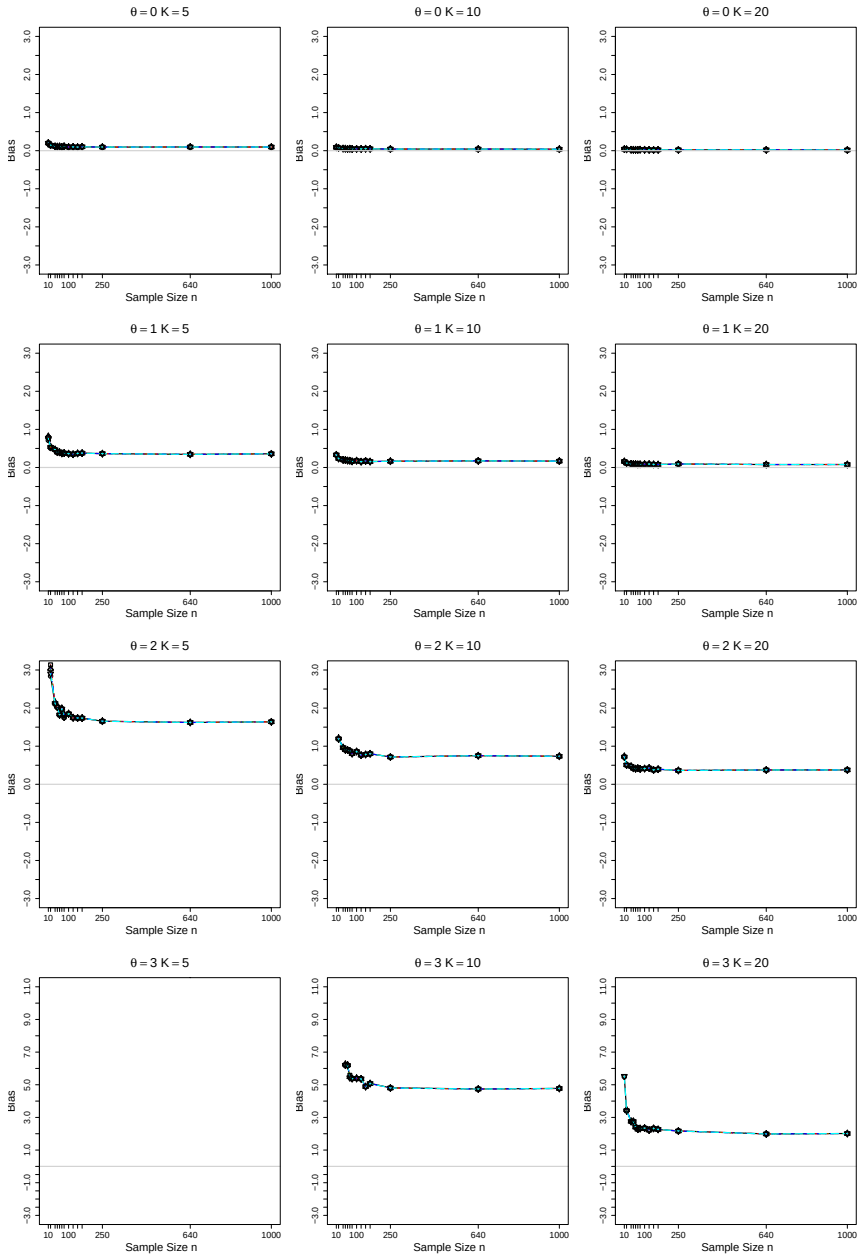


Figure B.33: Bias of the Mantel-Haenzsel overall effect measure ψ_{MH} from K studies in beta-binomial model for $p_{2j} = 0.4$, $0 \leq \theta \leq 3$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The weights of the Mantel-Haenzsel odds ratio use the estimators of ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0 for bias.

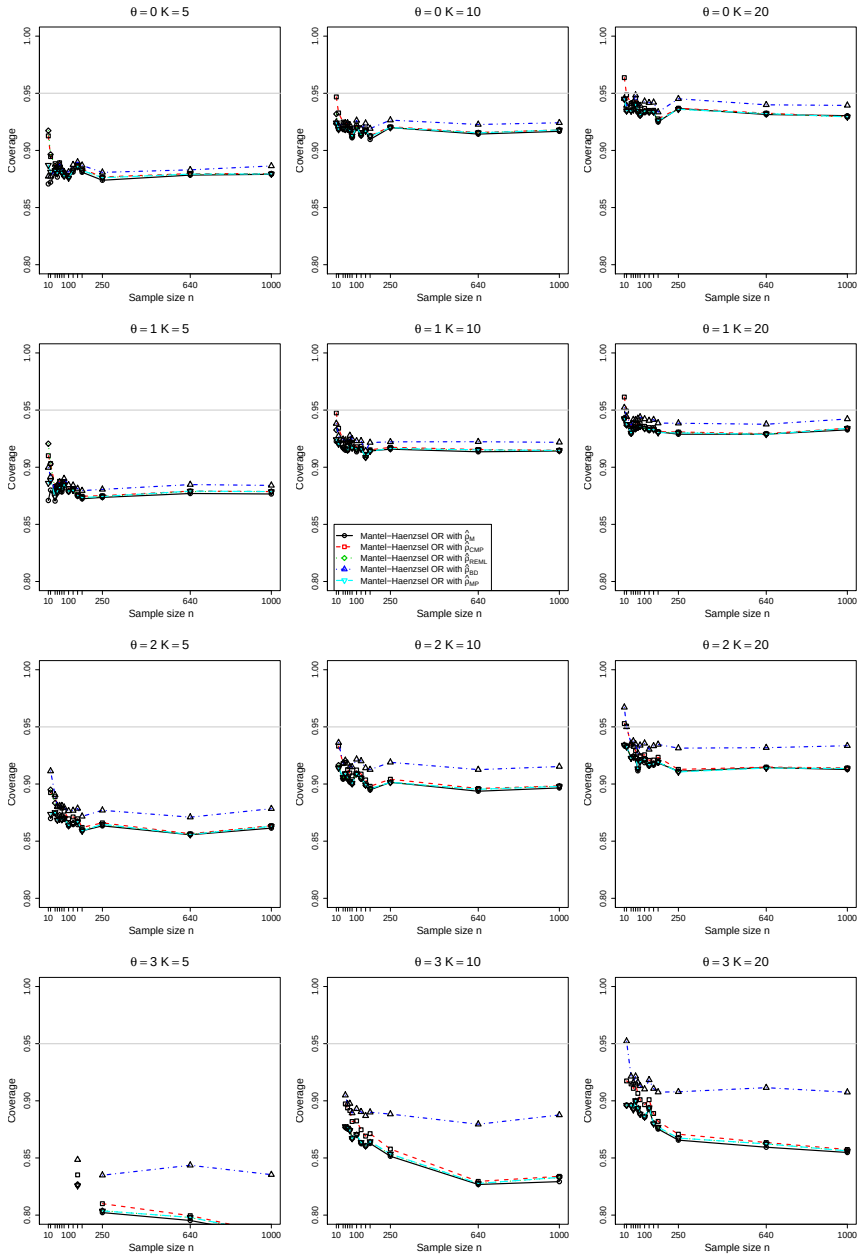


Figure B.34: Coverage at the nominal confidence level of 0.95 of the Mantel-Haenzsel overall effect measure θ_{MH} from K studies in beta-binomial model for $p_{2j} = 0.4$, $0 \leq \theta \leq 3$, $\rho = 0.1$ and $10 \leq n \leq 1000$. The weights of the Mandel-Haenzsel odds ratio use the estimators of ρ: circles (Moment estimator of ρ - $\hat{\rho}_M$), squares (Corrected Mandel-Paule moment estimator for ρ based on gamma approximation for Q distribution - $\hat{\rho}_{CMP}$), diamonds (Restricted maximum likelihood estimator for ρ - $\hat{\rho}_{REML}$), triangles (Breslow-Day estimator for ρ based on χ^2 distribution - $\hat{\rho}_{BD}$) and reverse-triangles (Mandel-Paule estimator of ρ - $\hat{\rho}_{MP}$). Light grey line at 0.95 for coverage.

B.7 Transformation Bias of $\hat{\theta}$

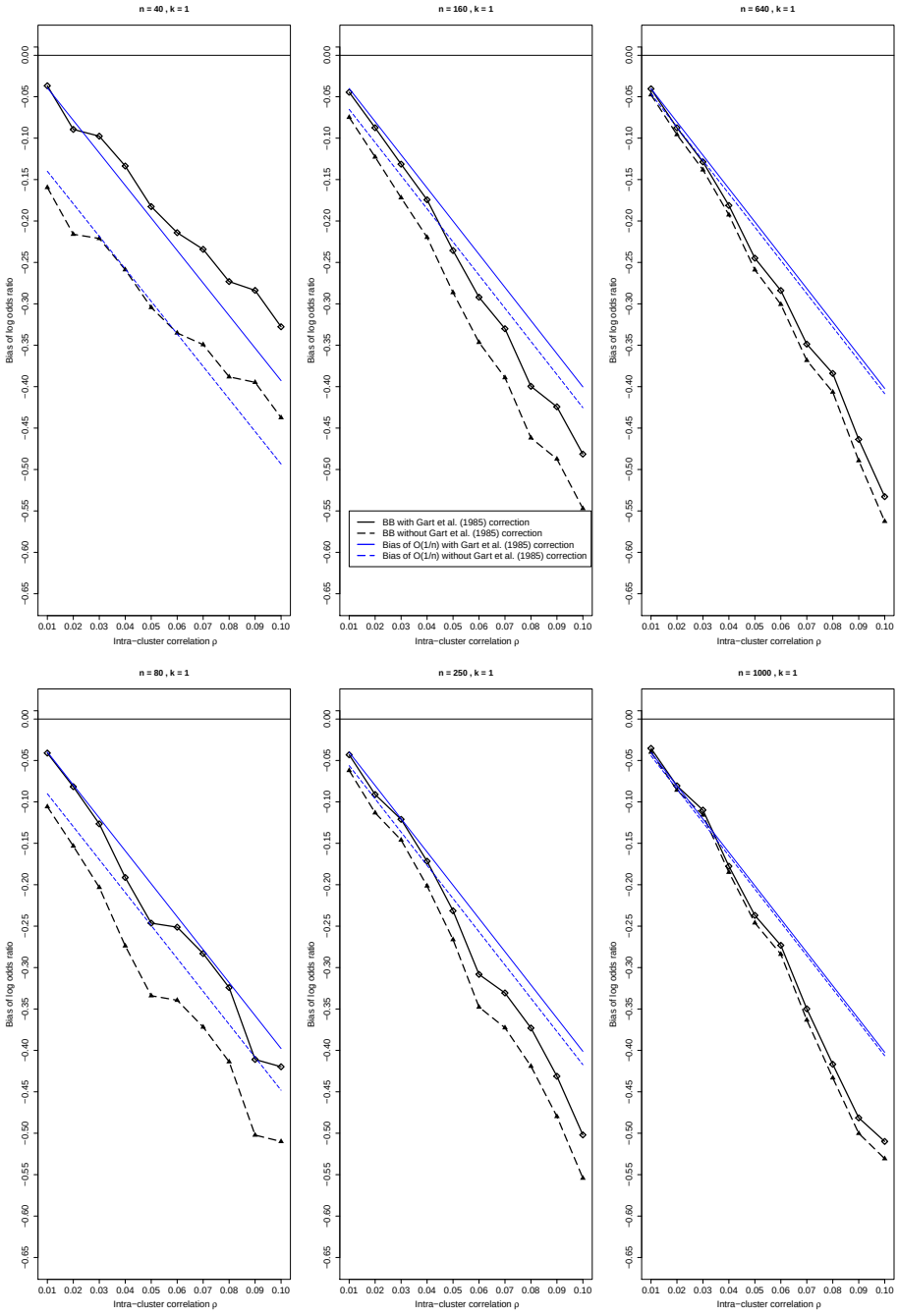


Figure B.35: Bias of log-odds ratio in overdispersed binomial model for $p_{1j} = 0.1$ ($\log(p_{1j}/(1-p_{1j})) = -2.20$) and $p_{2j} = 0.4$ ($\log(p_{2j}/(1-p_{2j})) = -0.40$) and $0 \leq \rho \leq 0.1$. 10000 simulations for each value of ρ from the beta-binomial distribution (black); the first order bias term given by the first two terms of equation (4.6.2) with known values of p and ρ (blue), with and without the continuity correction (solid and dashed lines, respectively)

Appendix C

C.1 Results for method of simulation similar to Viechtbauer (2007) and Kosmidis et al. (2017)

C.1.1 Full results for small-moderate heterogeneity with $p_{2j} = 0.1$, $p_{2j} = 0.2$ and $p_{2j} = 0.4$ with the method of simulation similar to Viechtbauer (2007)

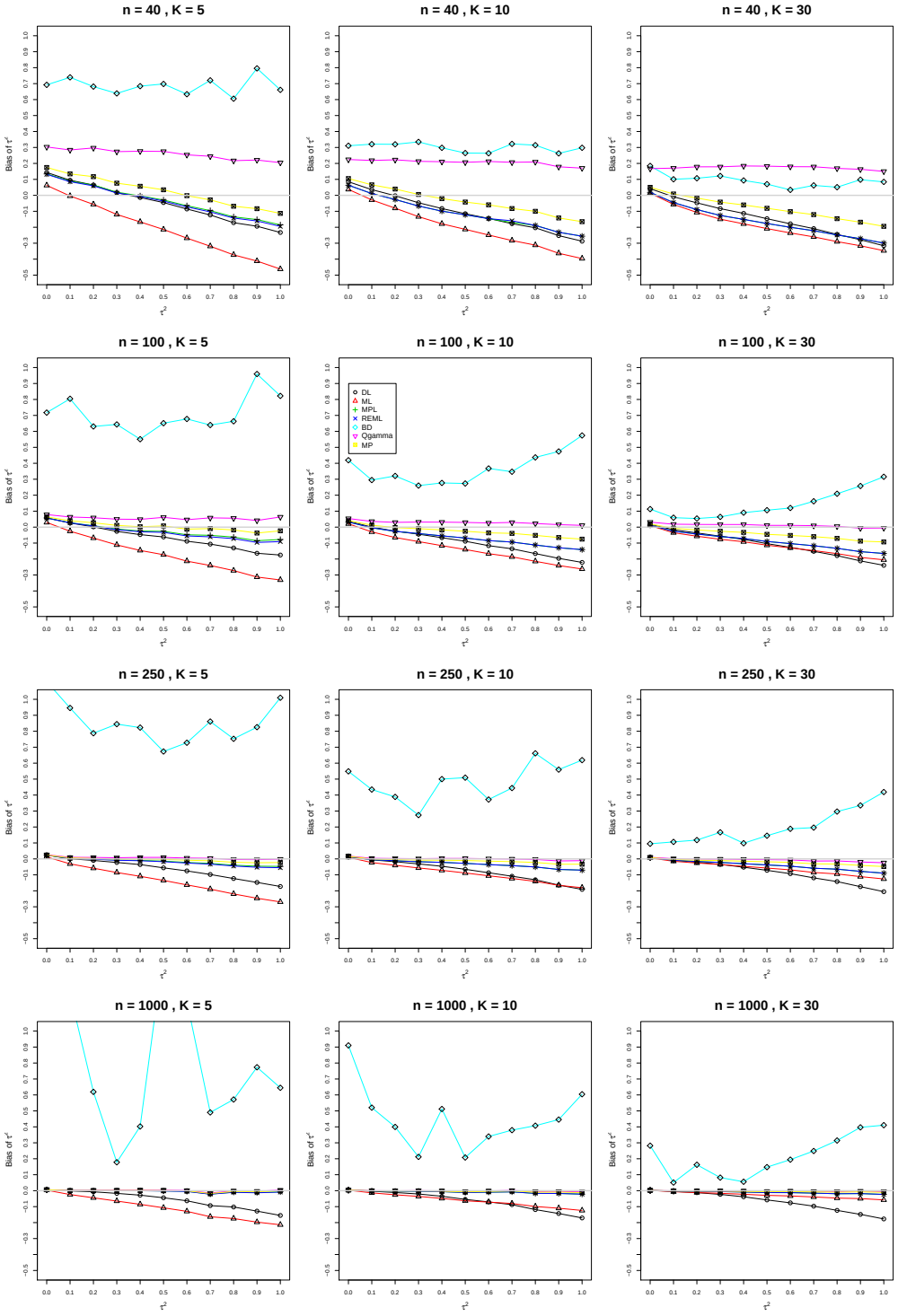


Figure C.1: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 0$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: black circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood $\hat{\tau}^2_{MPL}$, dark blue crosses – Restricted Maximum Likelihood $\hat{\tau}^2_{REML}$, light blue diamond – Profiled-Breslow-Day $\hat{\tau}^2_{BD}$, pink reverse triangles – Corrected Mandel-Paule $\hat{\tau}^2_{CMP}$ and yellow circles – standard Mandel-Paule $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

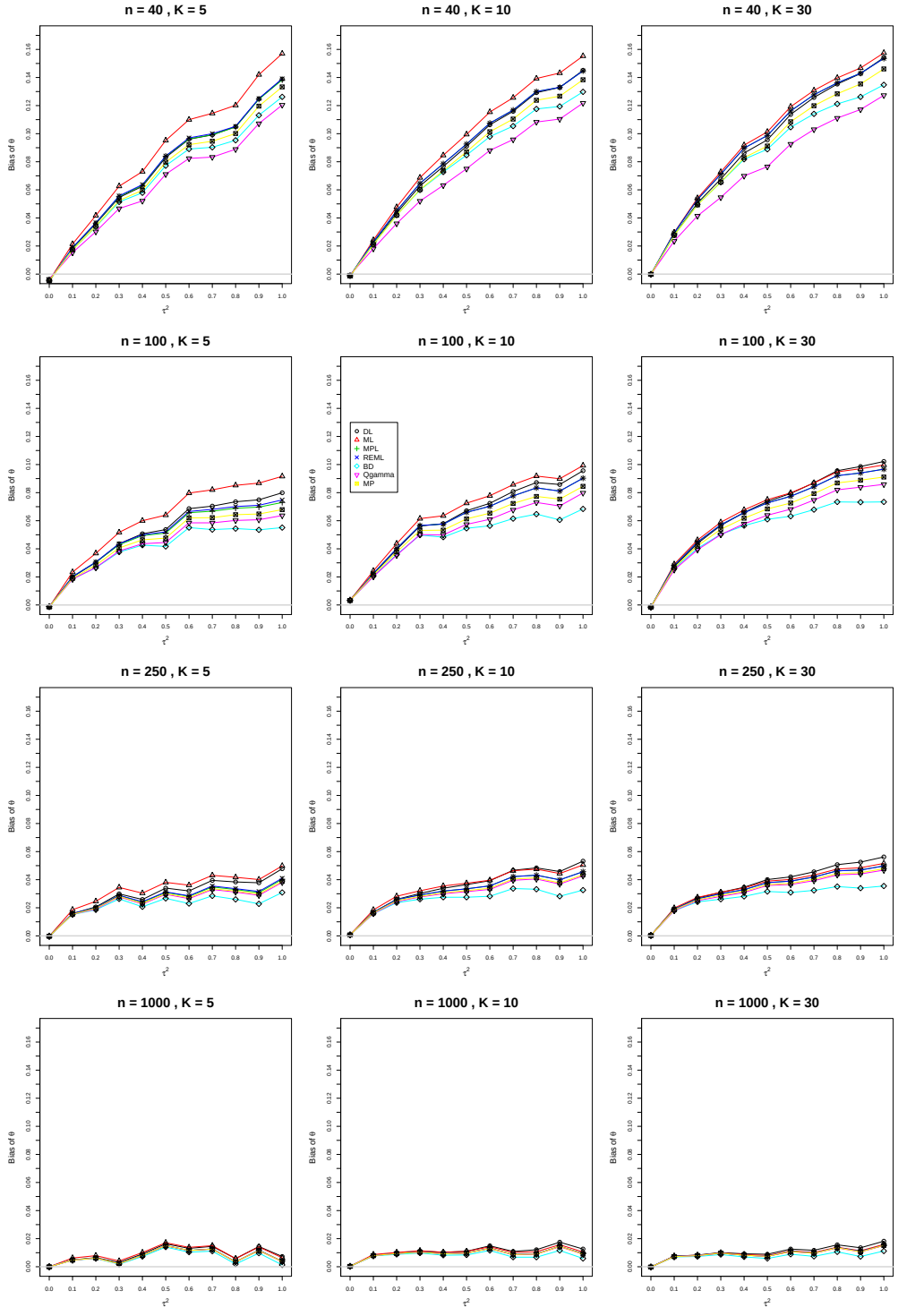


Figure C.2: Bias of the estimated overall effect measure $\hat{\theta}_w$ obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 0$ and $0 \leq \tau^2 \leq 1$. The estimators of $\hat{\theta}_w$ include the estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

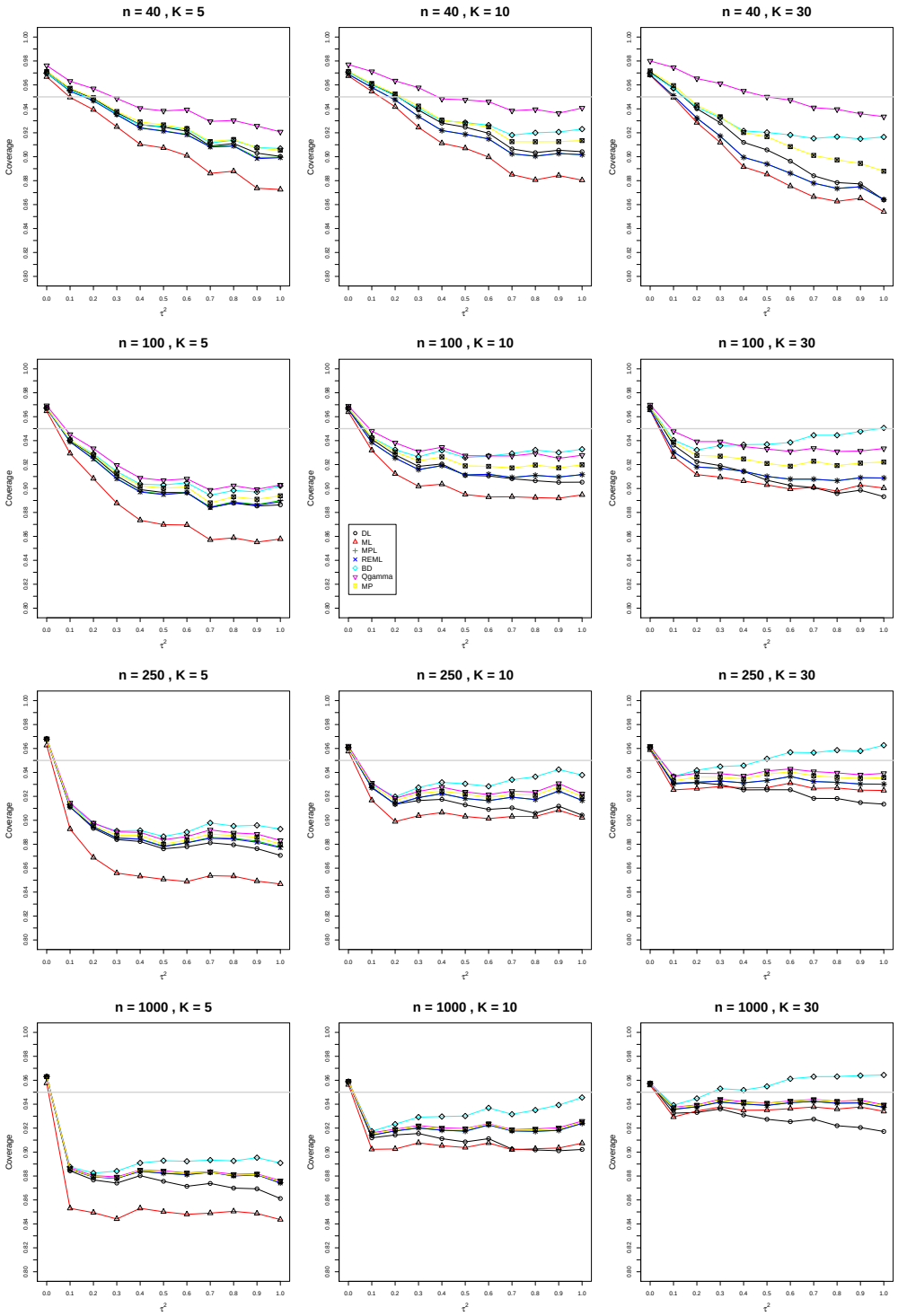


Figure C.3: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_w$ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.1$, $\theta_w = 0$ and $0 \leq \tau^2 \leq 1$. The inverse-variance weights use the following estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

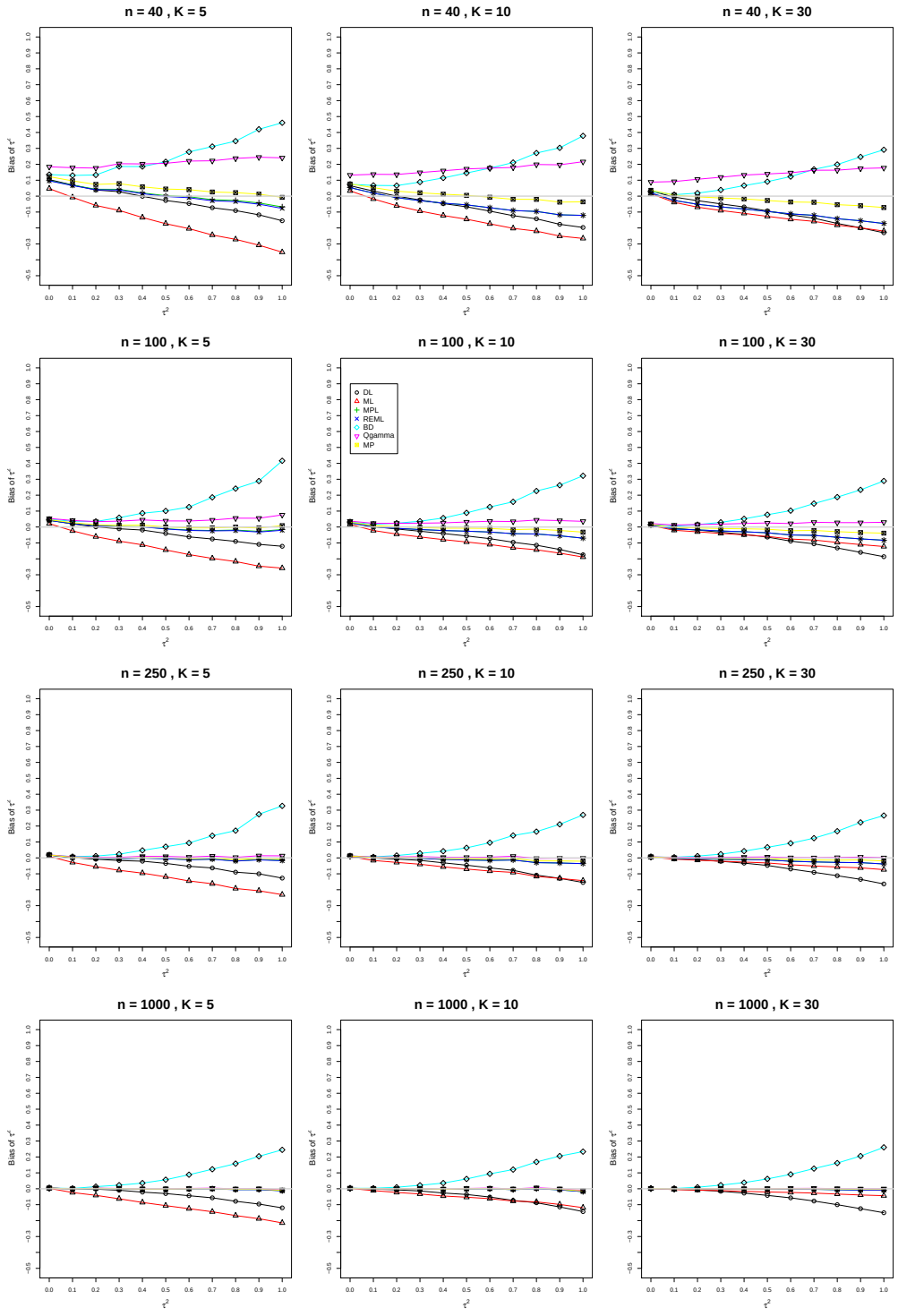


Figure C.4: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 1$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

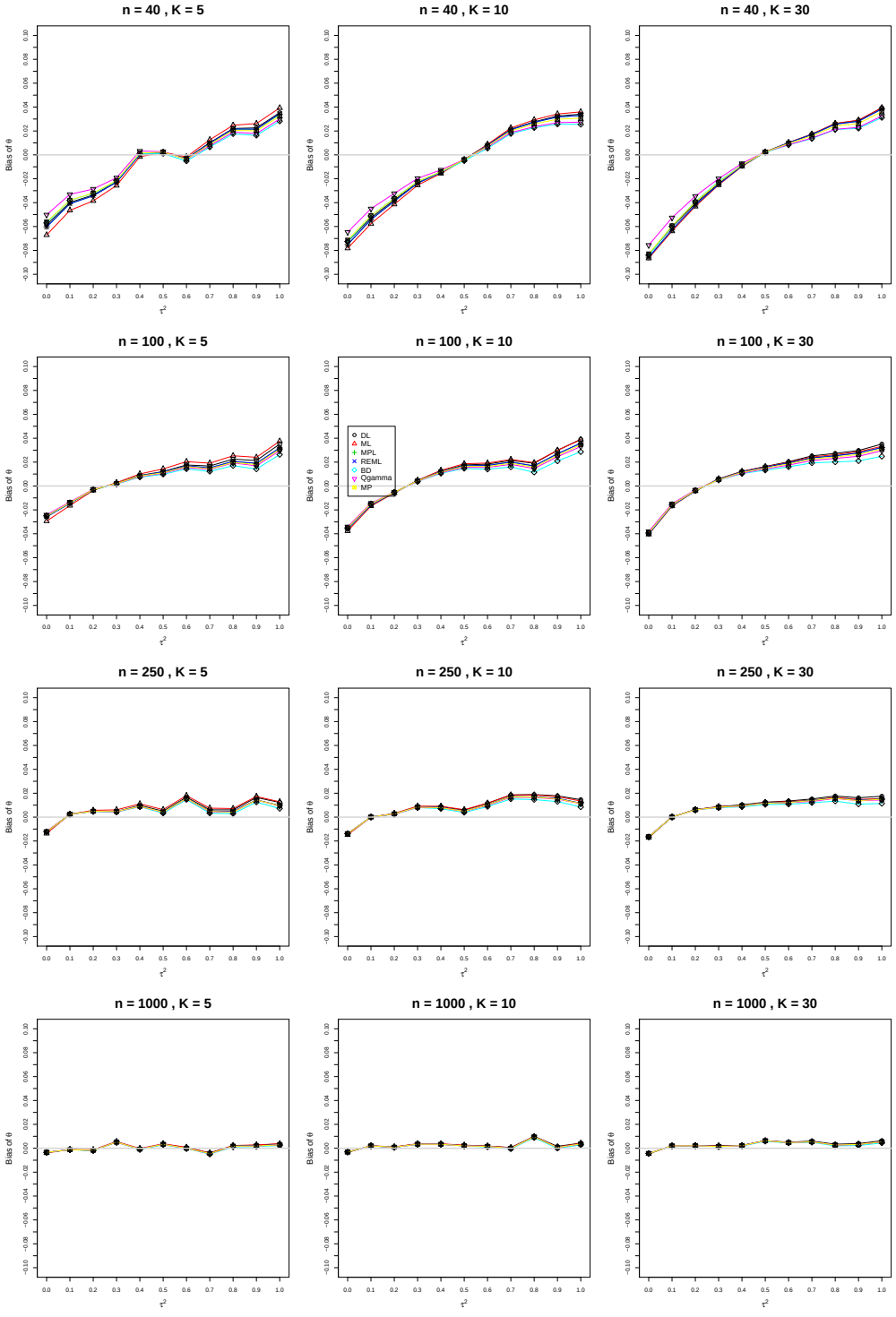


Figure C.5: Bias of the estimated overall effect measure $\hat{\theta}_w$ obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 1$ and $0 \leq \tau^2 \leq 1$. The estimators of $\hat{\theta}_w$ include the estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

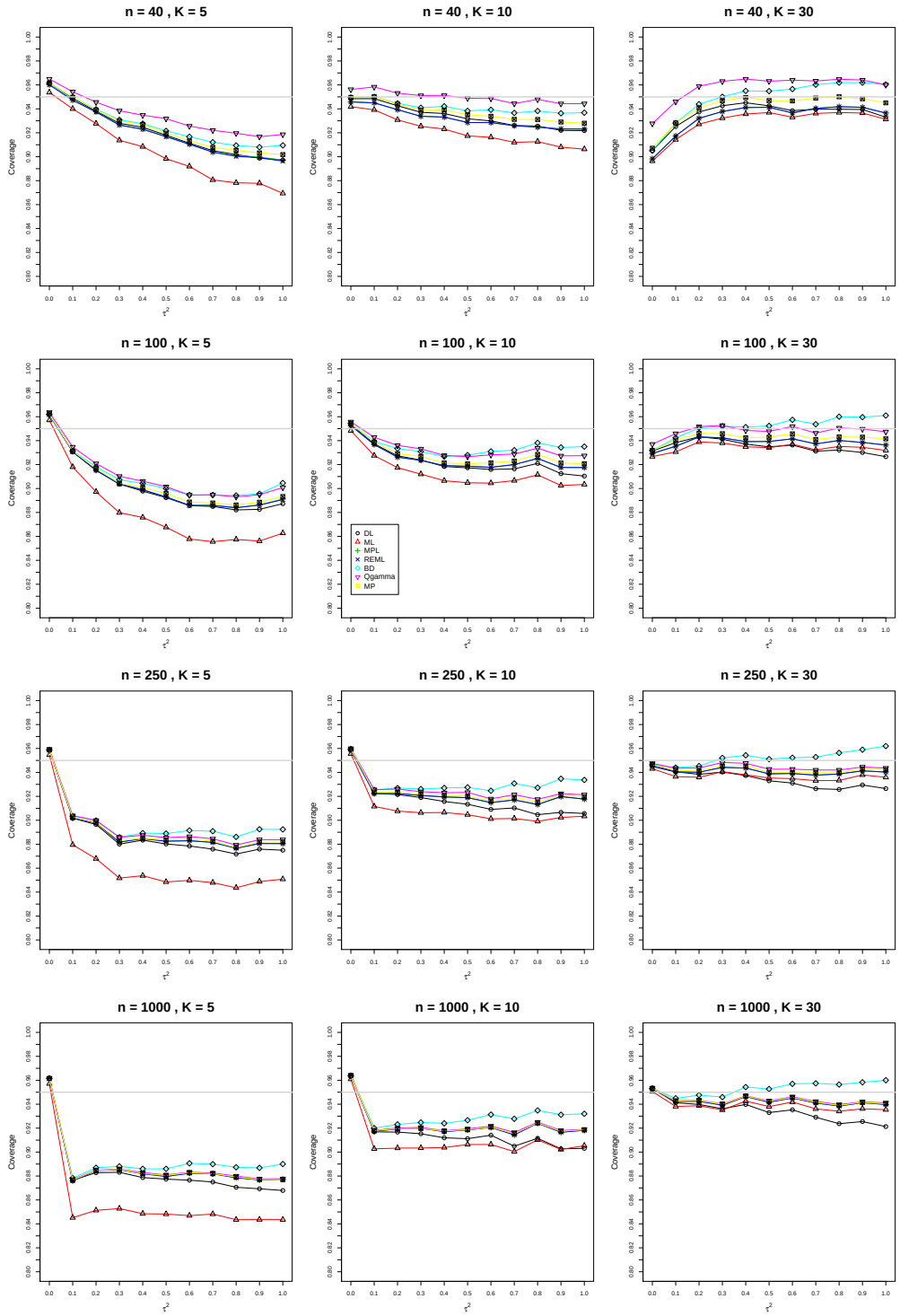


Figure C.6: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_w$ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.1$, $\theta = 1$ and $0 \leq \tau^2 \leq 1$. The inverse-variance weights use the following estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

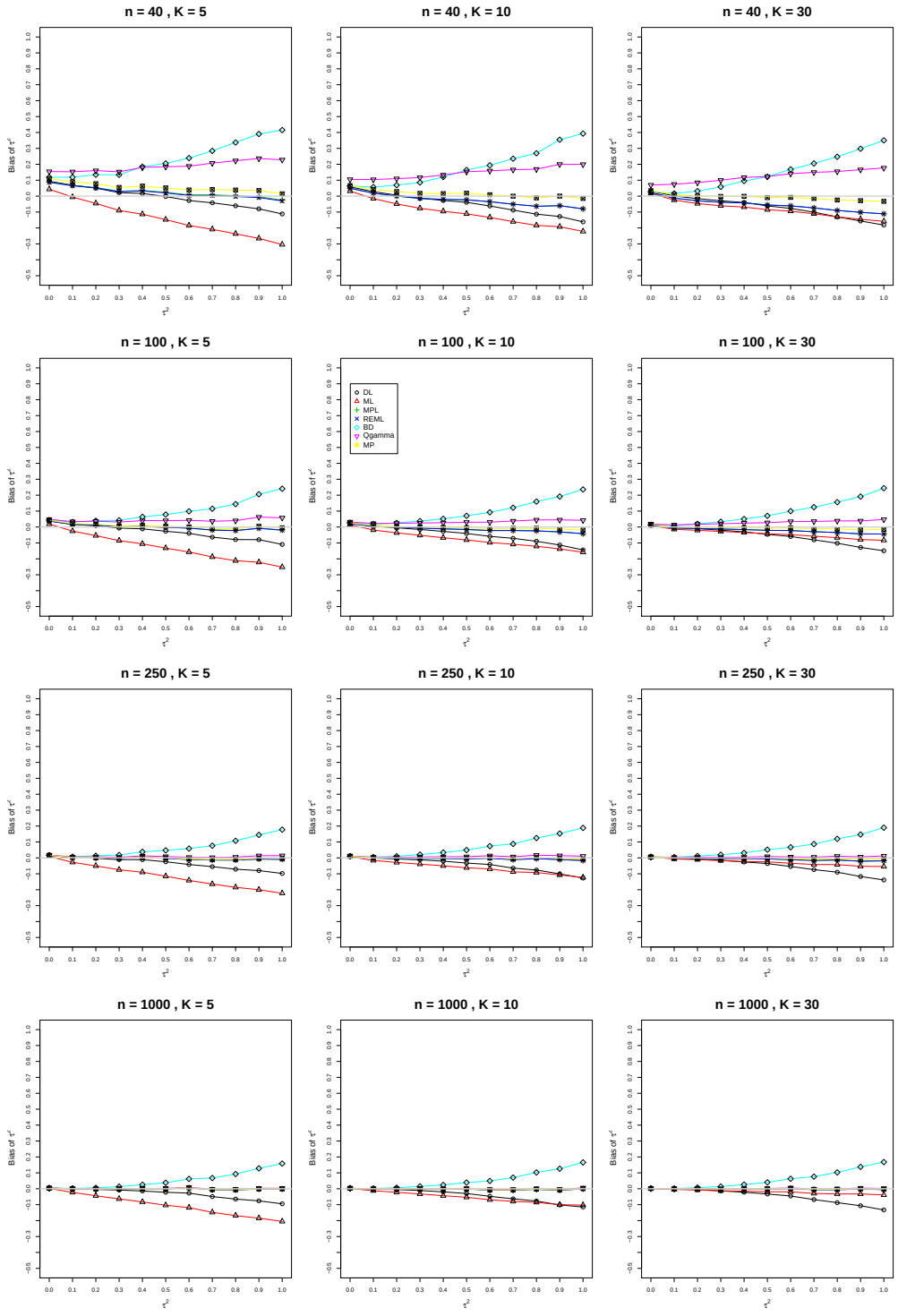


Figure C.7: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

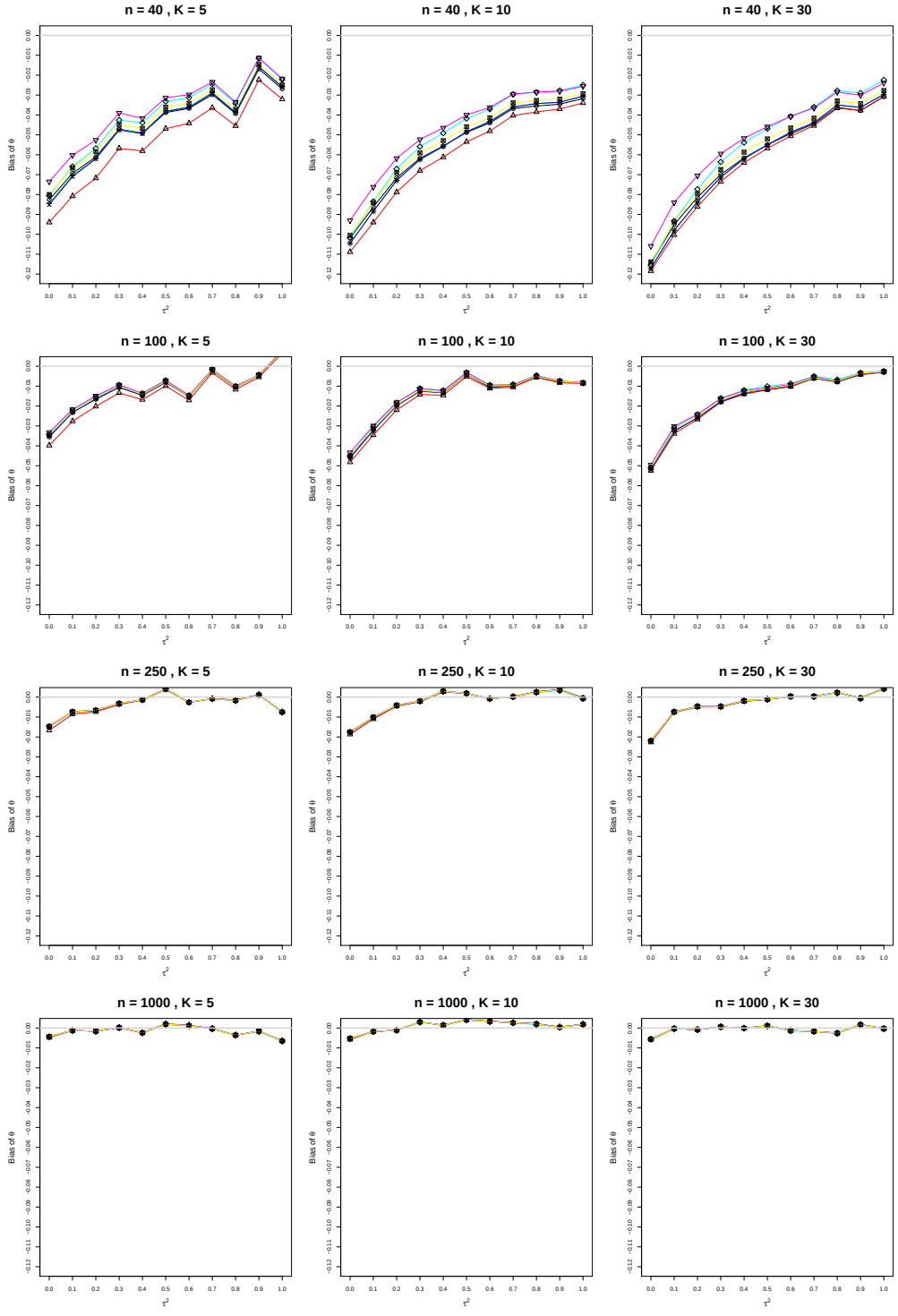


Figure C.8: Bias of the estimated overall effect measure $\hat{\theta}_w$ obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 1$. The estimators of $\hat{\theta}_w$ include the estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

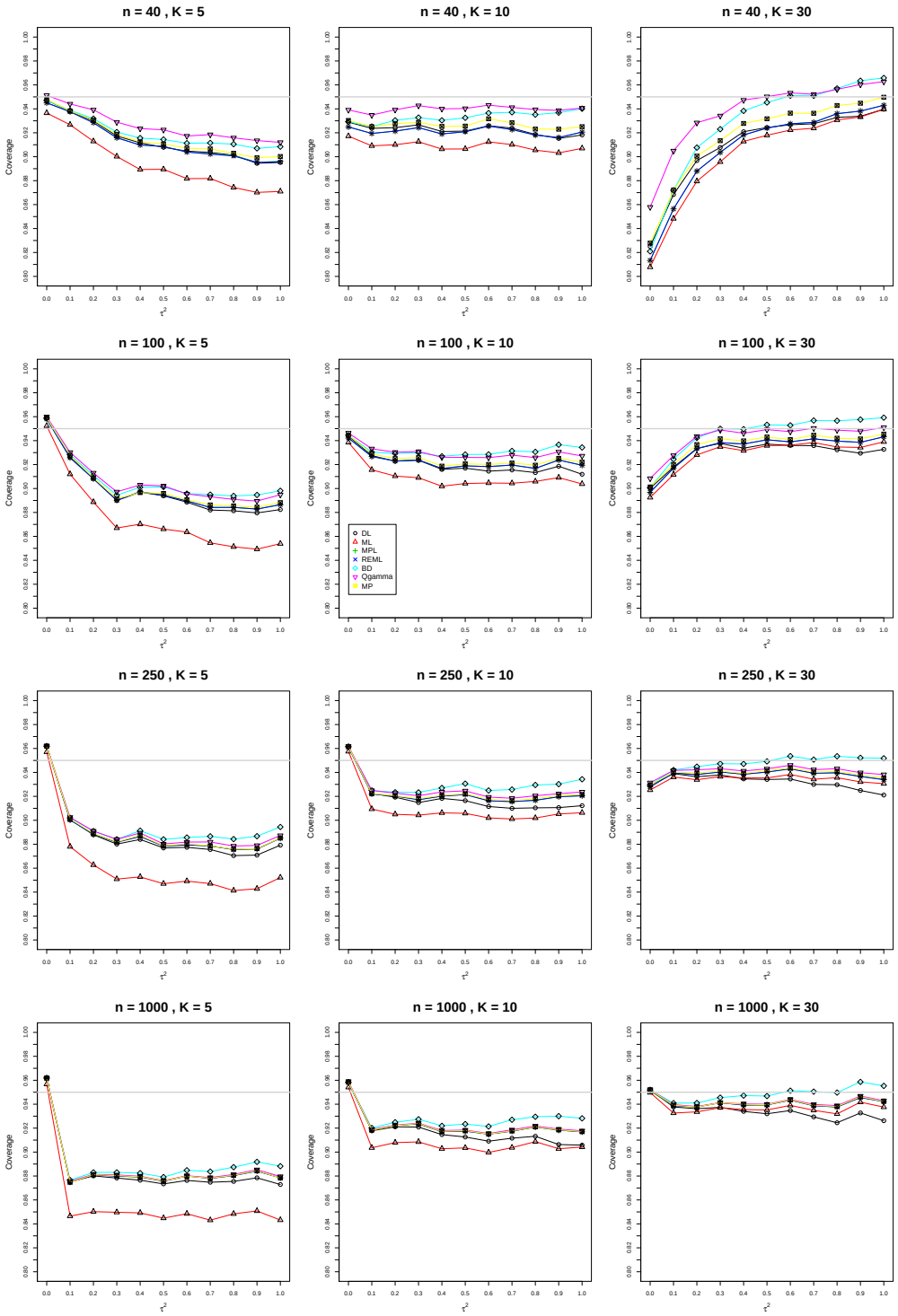


Figure C.9: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_w$ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.1$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 1$. The inverse-variance weights use the following estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

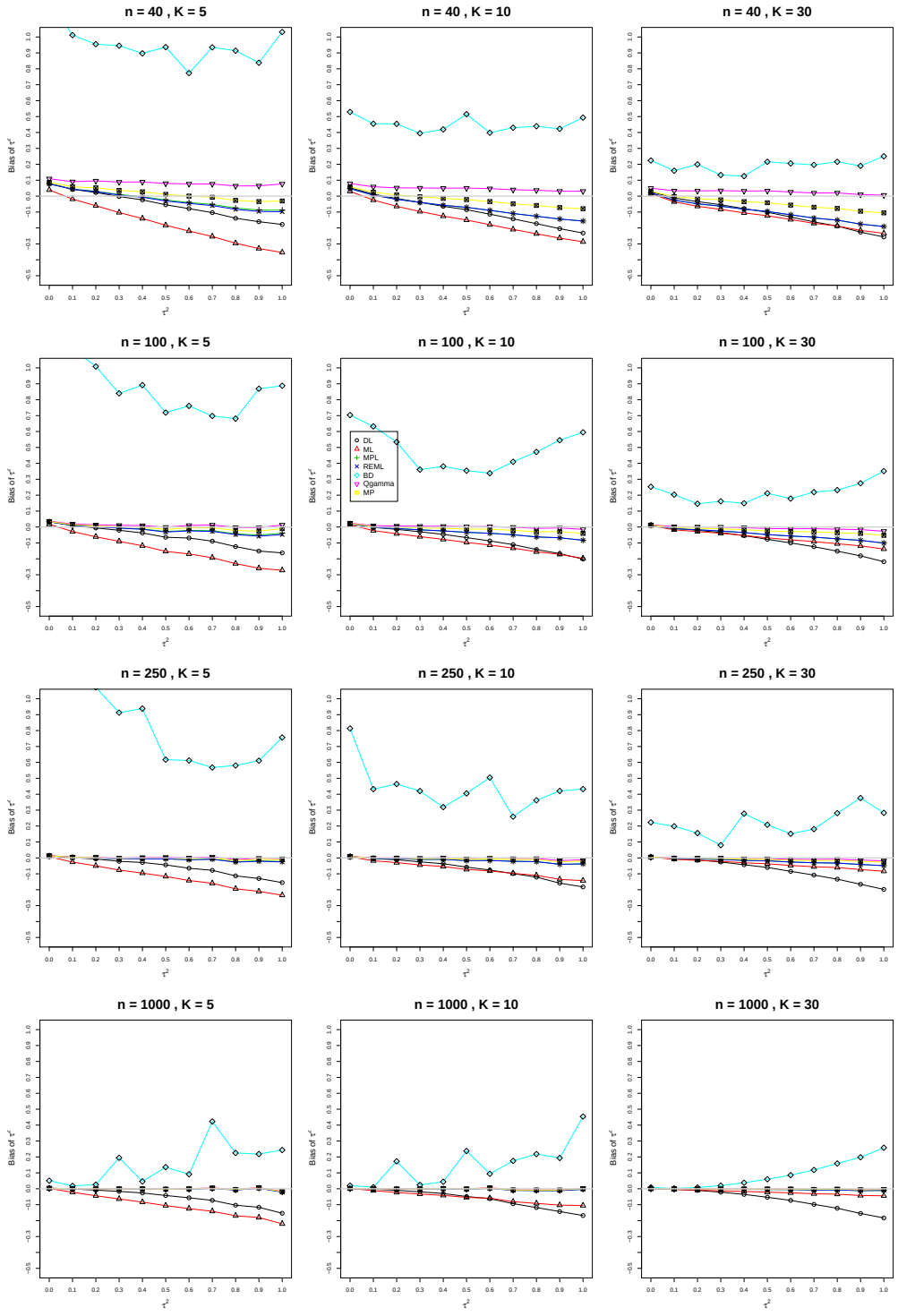


Figure C.10: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.2$, $\theta_w = 0$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

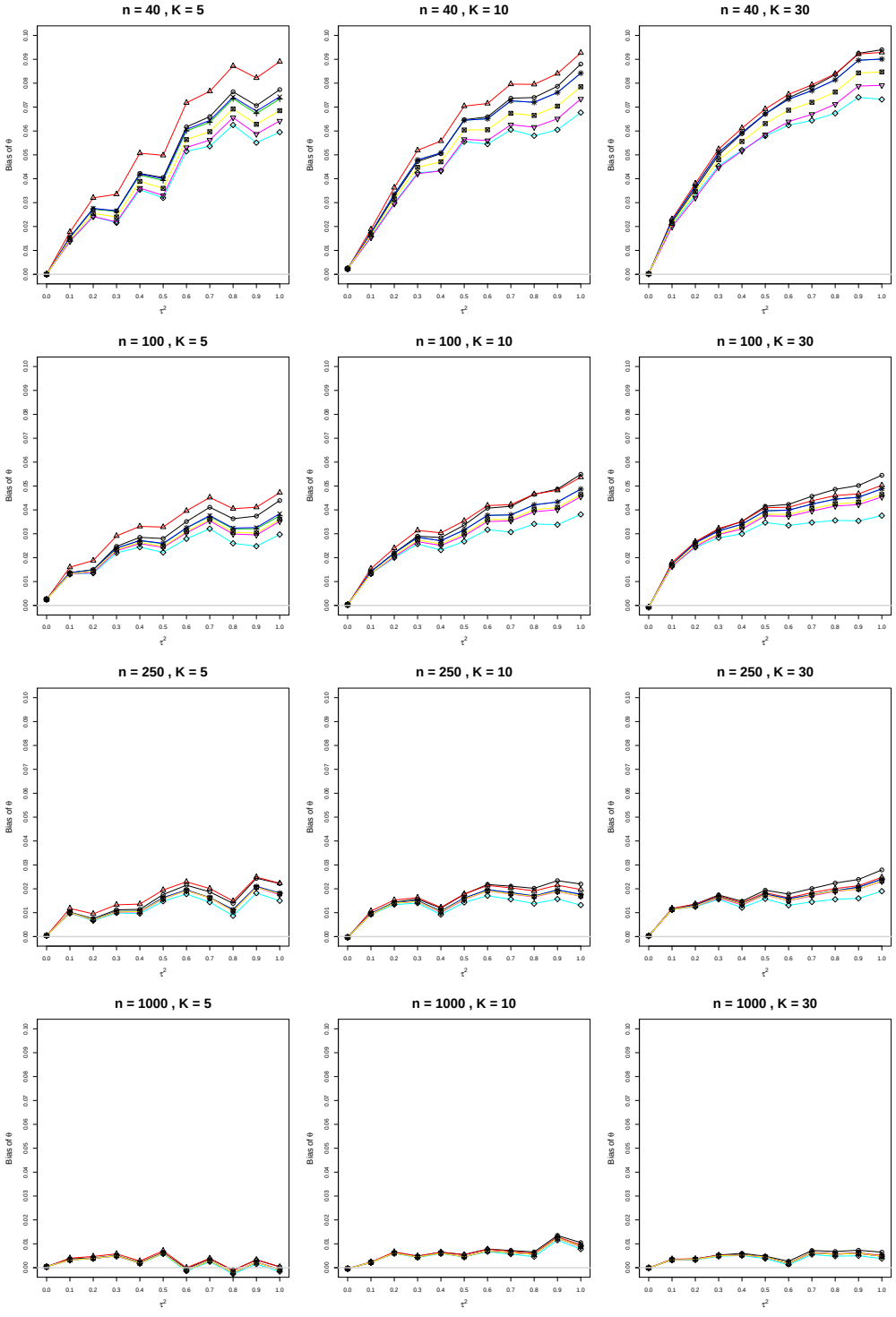


Figure C.11: Bias of the estimated overall effect measure $\hat{\theta}_w$ obtained from K studies for $p_{2j} = 0.2$, $\theta = 0$ and $0 \leq \tau^2 \leq 1$. The estimators of $\hat{\theta}_w$ include the estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

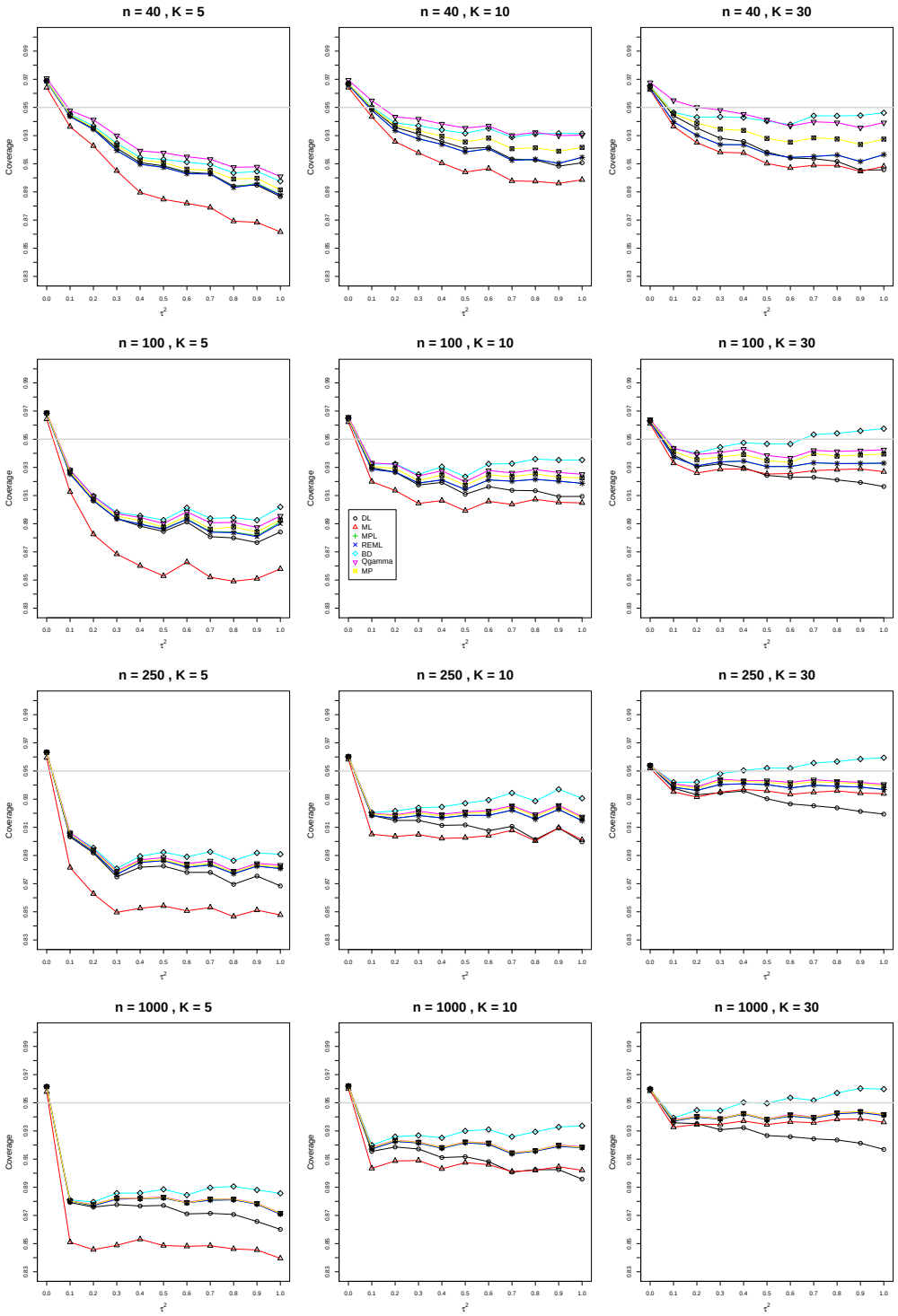


Figure C.12: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_w$ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.2$, $\theta_w = 0$ and $0 \leq \tau^2 \leq 1$. The inverse-variance weights use the following estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

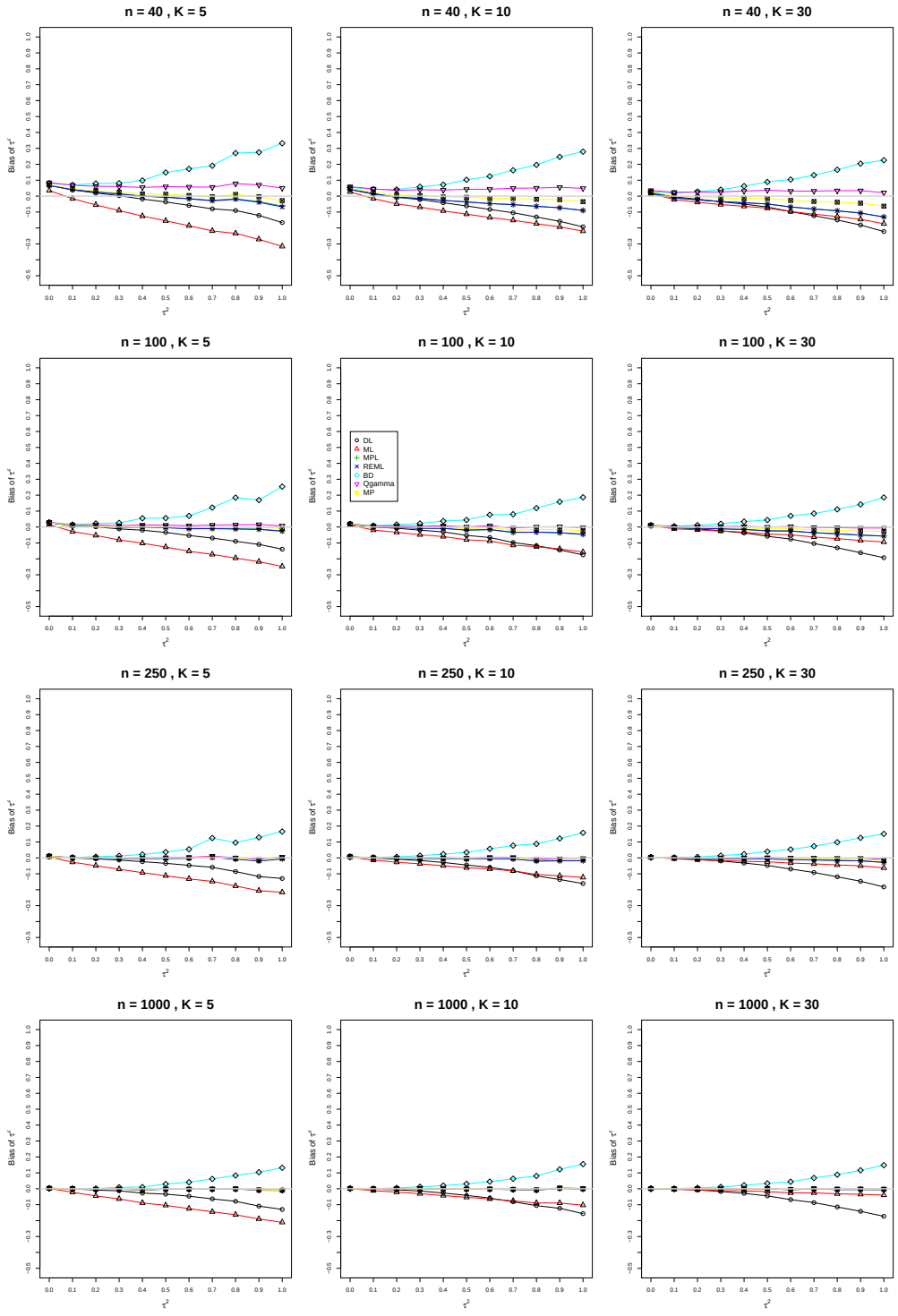


Figure C.13: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.2$, $\theta_w = 1$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

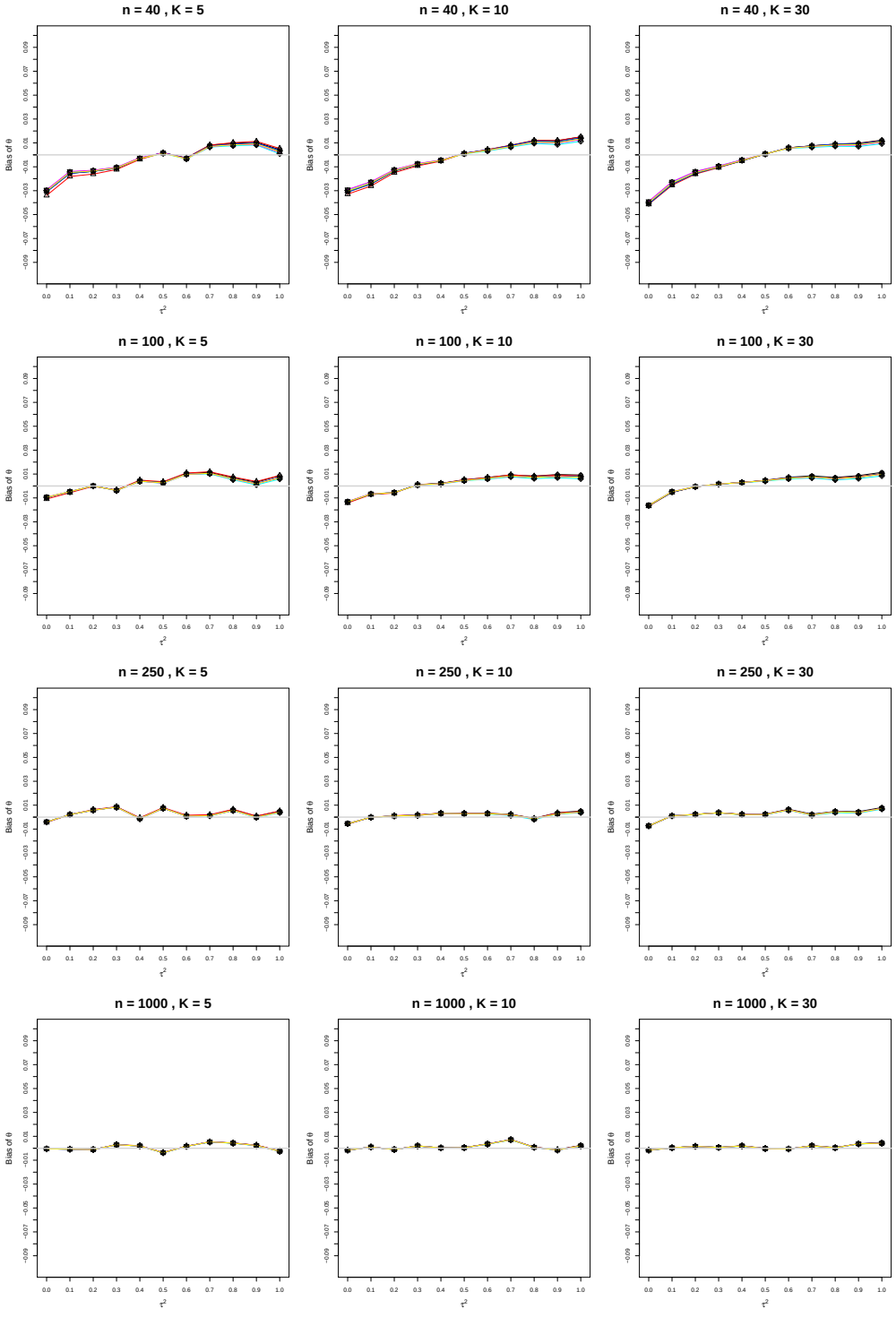


Figure C.14: Bias of the estimated overall effect measure $\hat{\theta}_w$ obtained from K studies for $p_{2j} = 0.2$, $\theta_w = 1$ and $0 \leq \tau^2 \leq 1$. The estimators of $\hat{\theta}_w$ include the estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

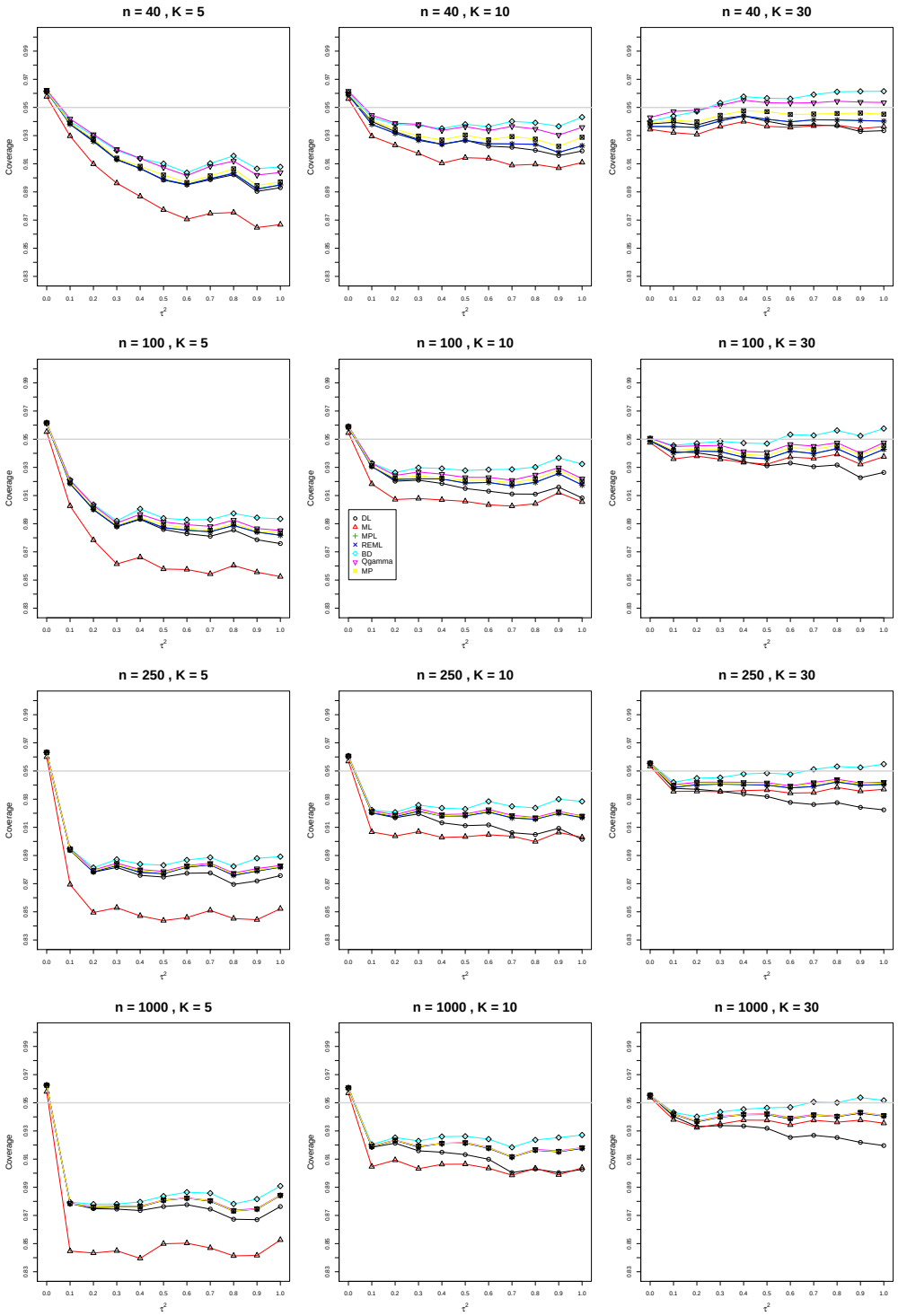


Figure C.15: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_w$ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.2$, $\theta = 1$ and $0 \leq \tau^2 \leq 1$. The inverse-variance weights use the following estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

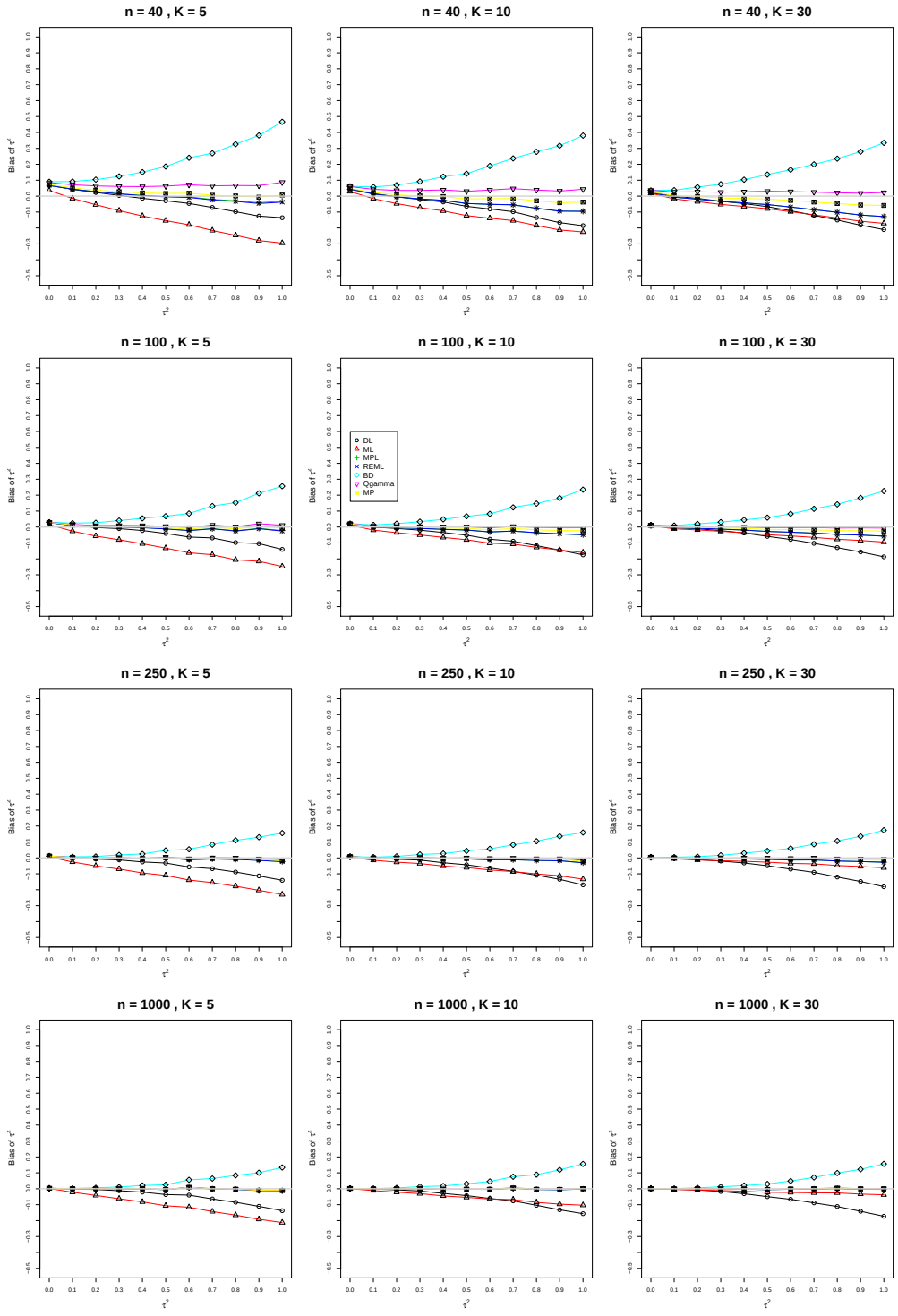


Figure C.16: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.2$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

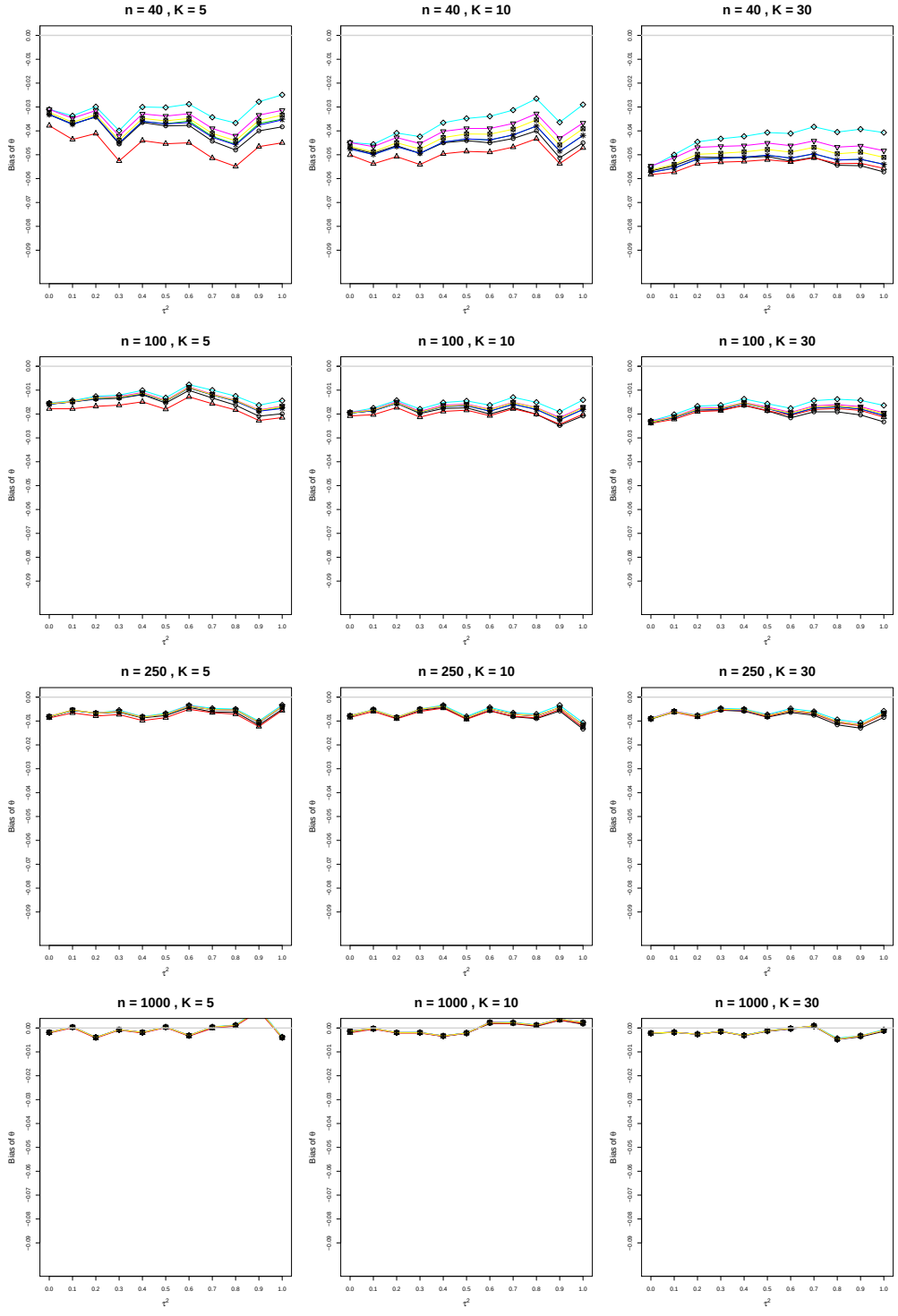


Figure C.17: Bias of the estimated overall effect measure $\hat{\theta}_w$ obtained from K studies for $p_{2j} = 0.2$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 1$. The estimators of $\hat{\theta}_w$ include the estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

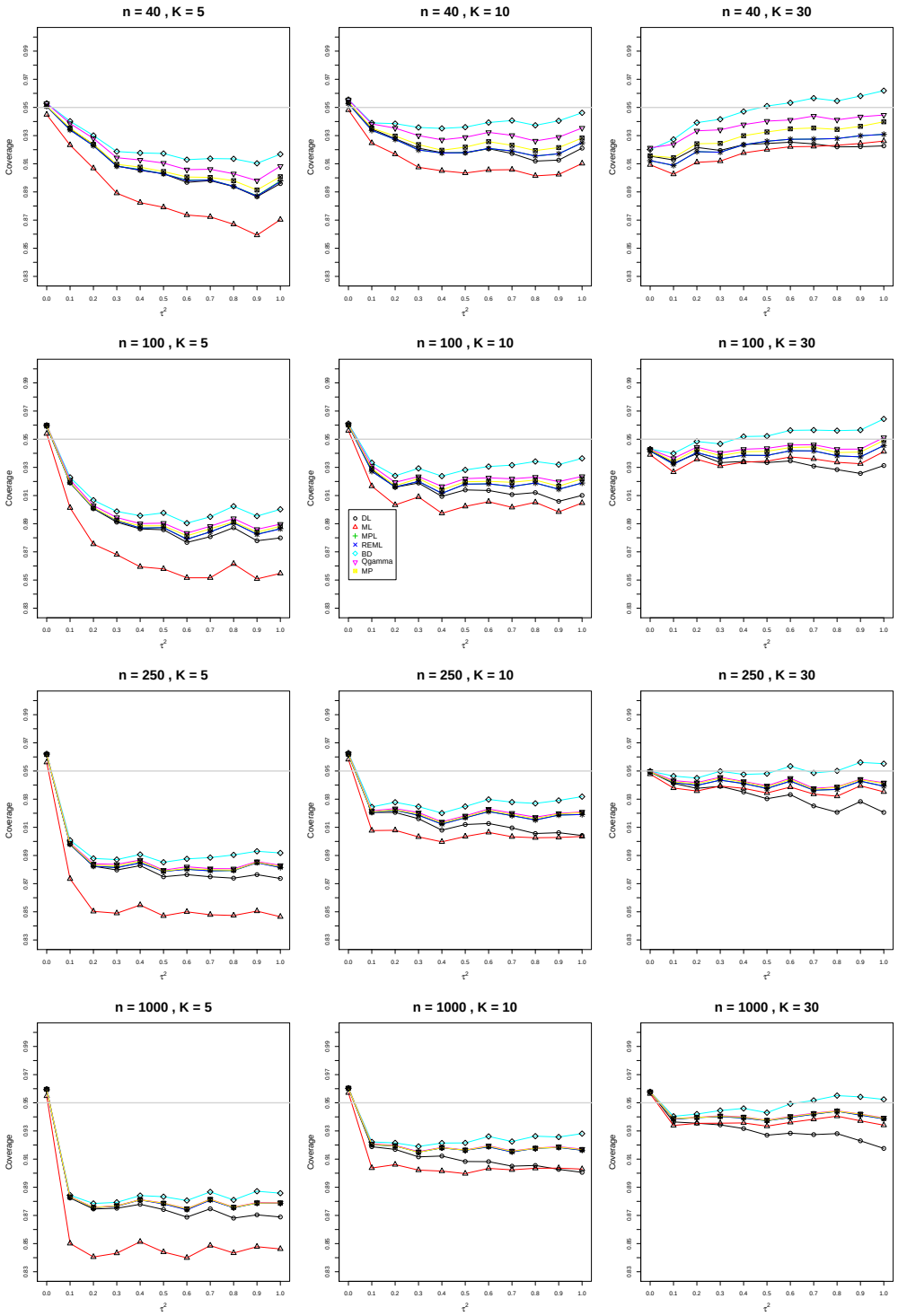


Figure C.18: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_w$ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.2$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 1$. The inverse-variance weights use the following estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

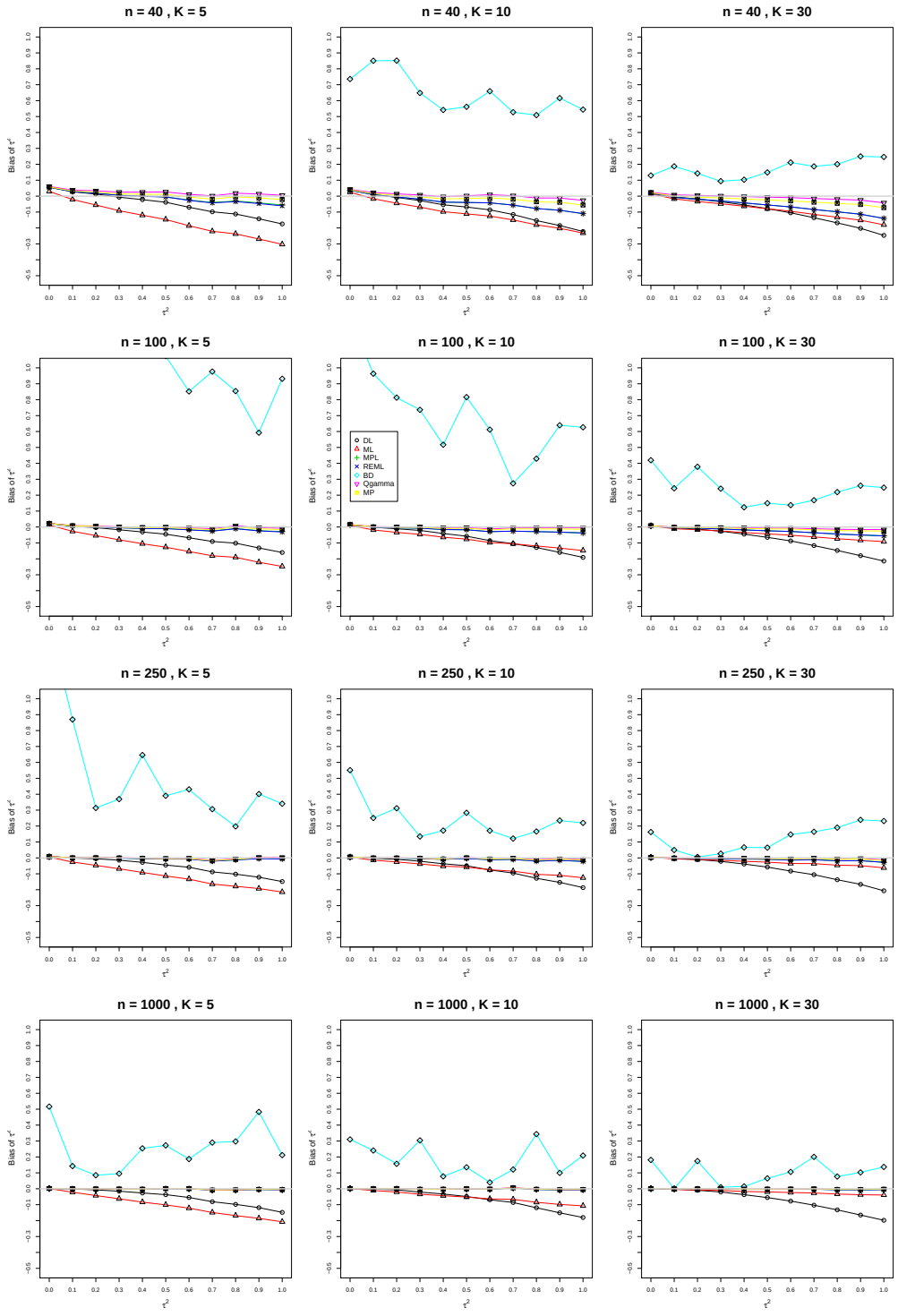


Figure C.19: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.4$, $\theta_w = 0$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

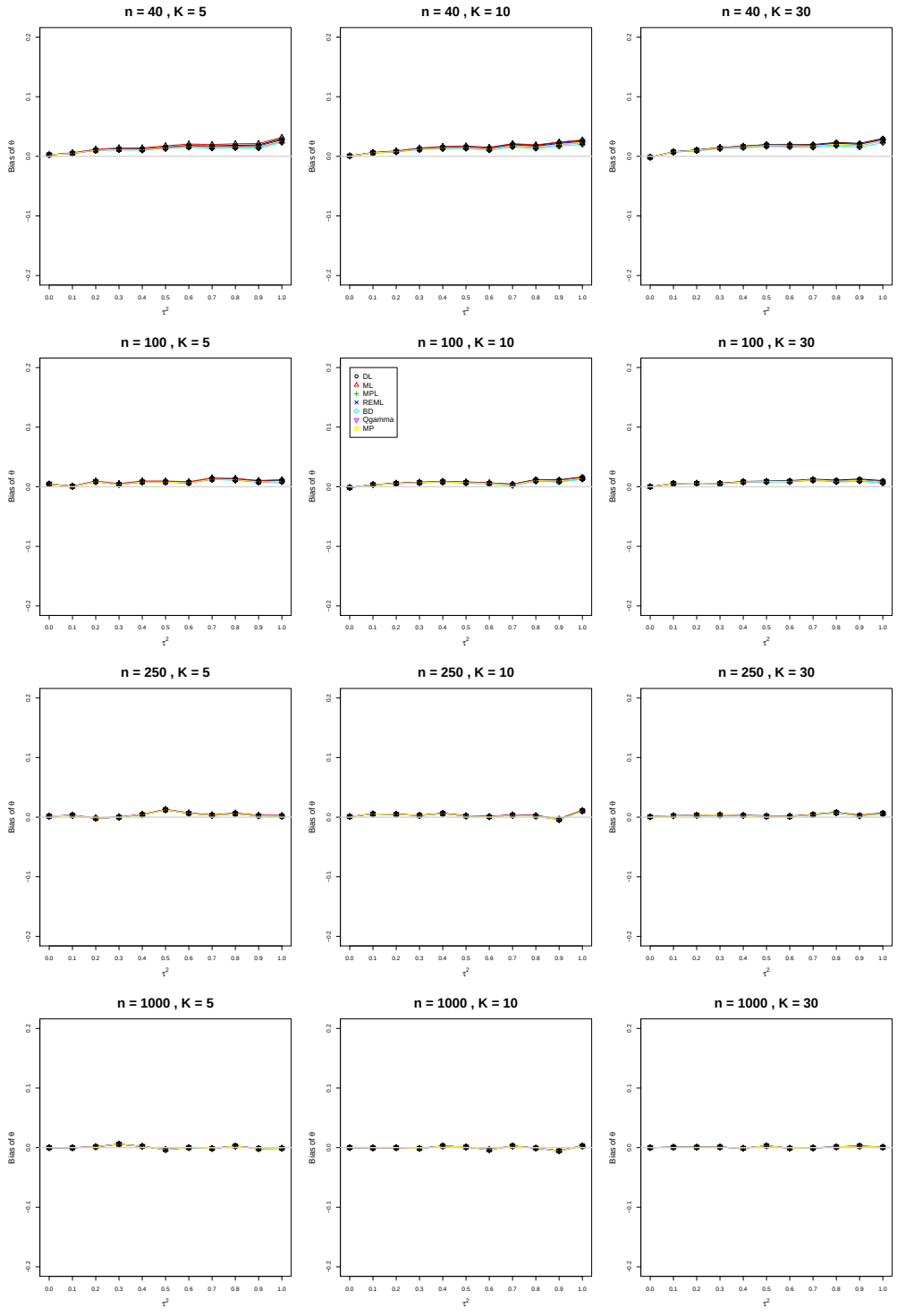


Figure C.20: Bias of the estimated overall effect measure $\hat{\theta}_w$ obtained from K studies for $p_{2j} = 0.4$, $\theta = 0$ and $0 \leq \tau^2 \leq 1$. The estimators of $\hat{\theta}_w$ include the estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

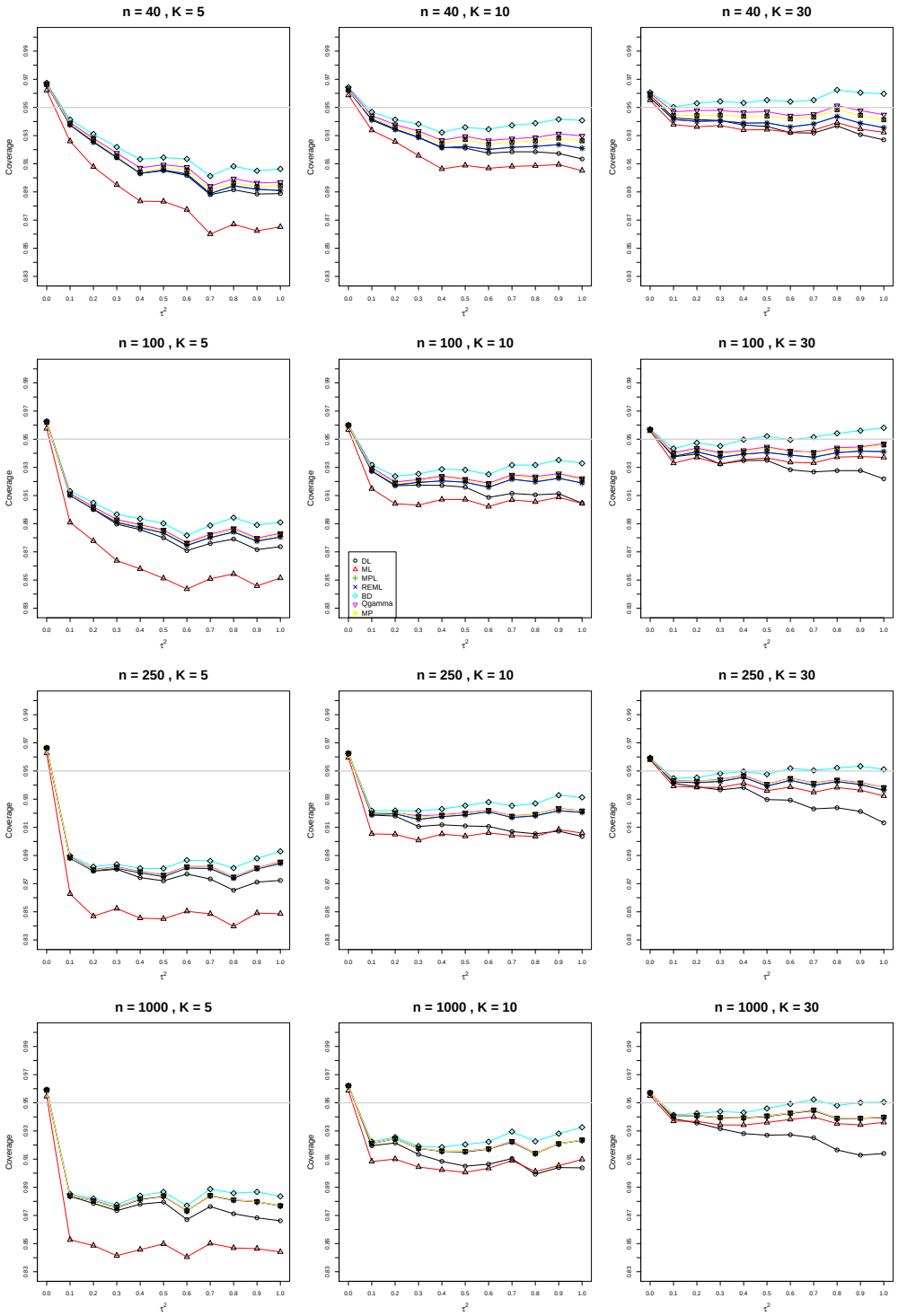


Figure C.21: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_w$ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.4$, $\theta_w = 0$ and $0 \leq \tau^2 \leq 1$. The inverse-variance weights use the following estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

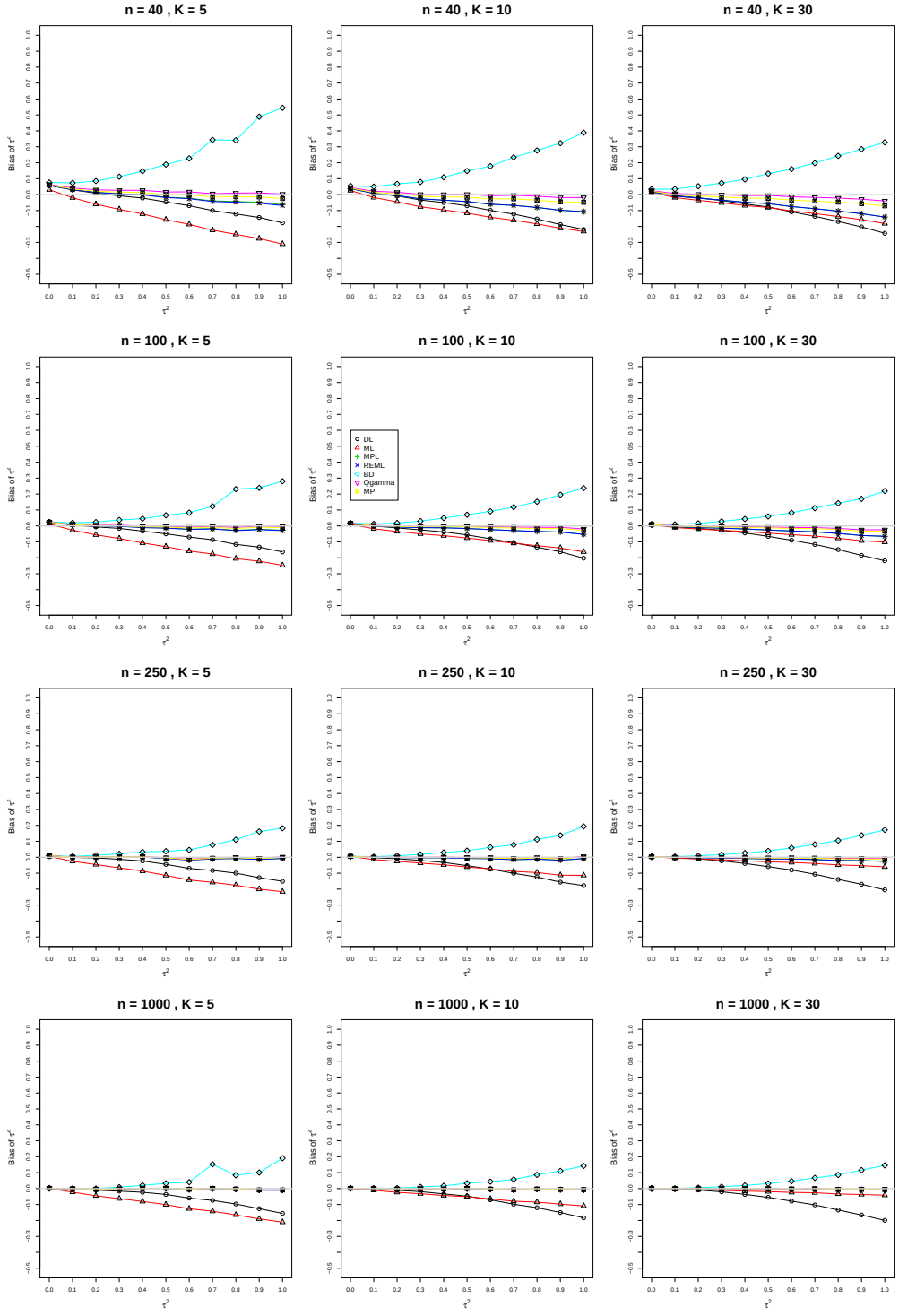


Figure C.22: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.4$, $\theta_w = 1$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

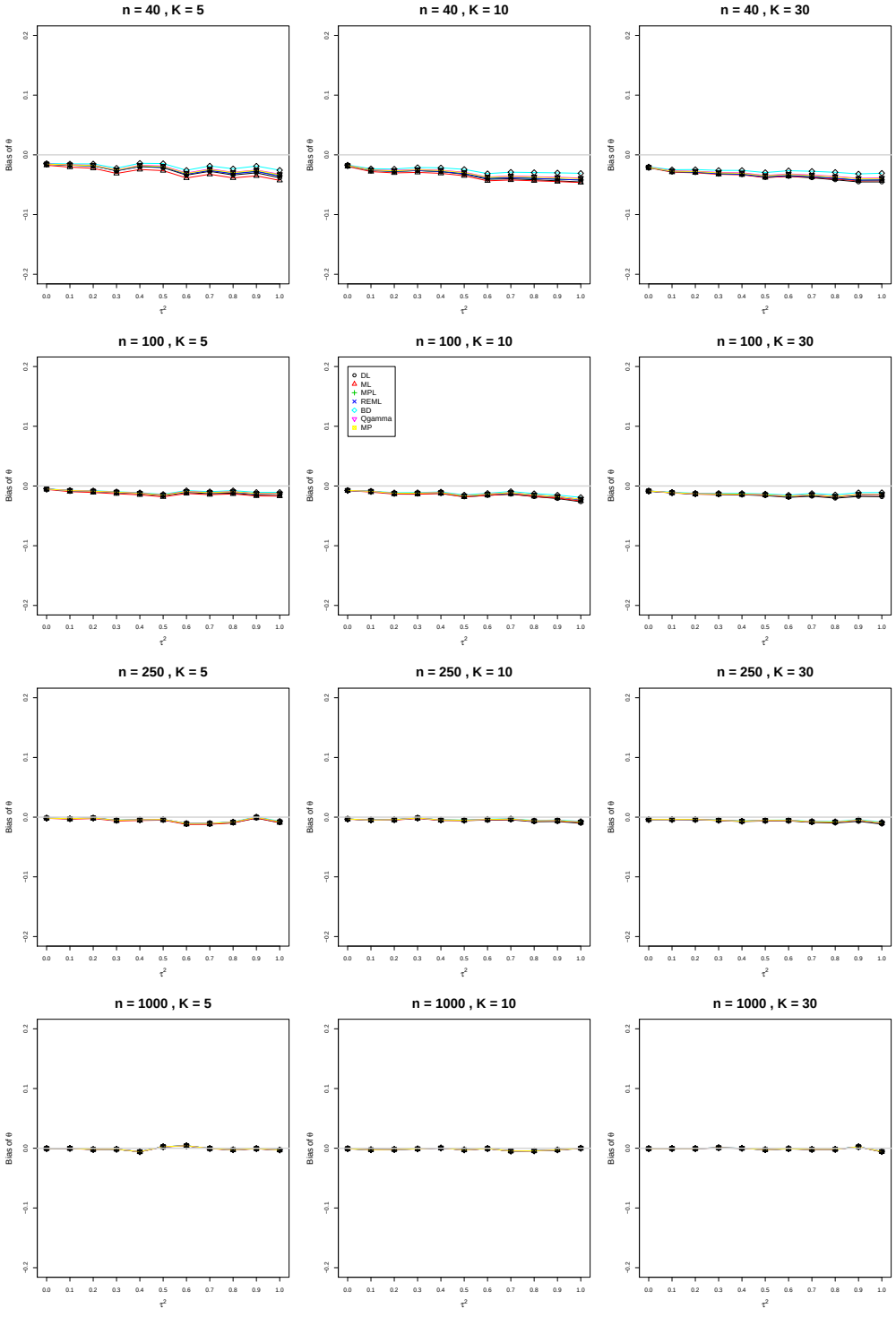


Figure C.23: Bias of the estimated overall effect measure $\hat{\theta}_w$ obtained from K studies for $p_{2j} = 0.4$, $\theta_w = 1$ and $0 \leq \tau^2 \leq 1$. The estimators of $\hat{\theta}_w$ include the estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

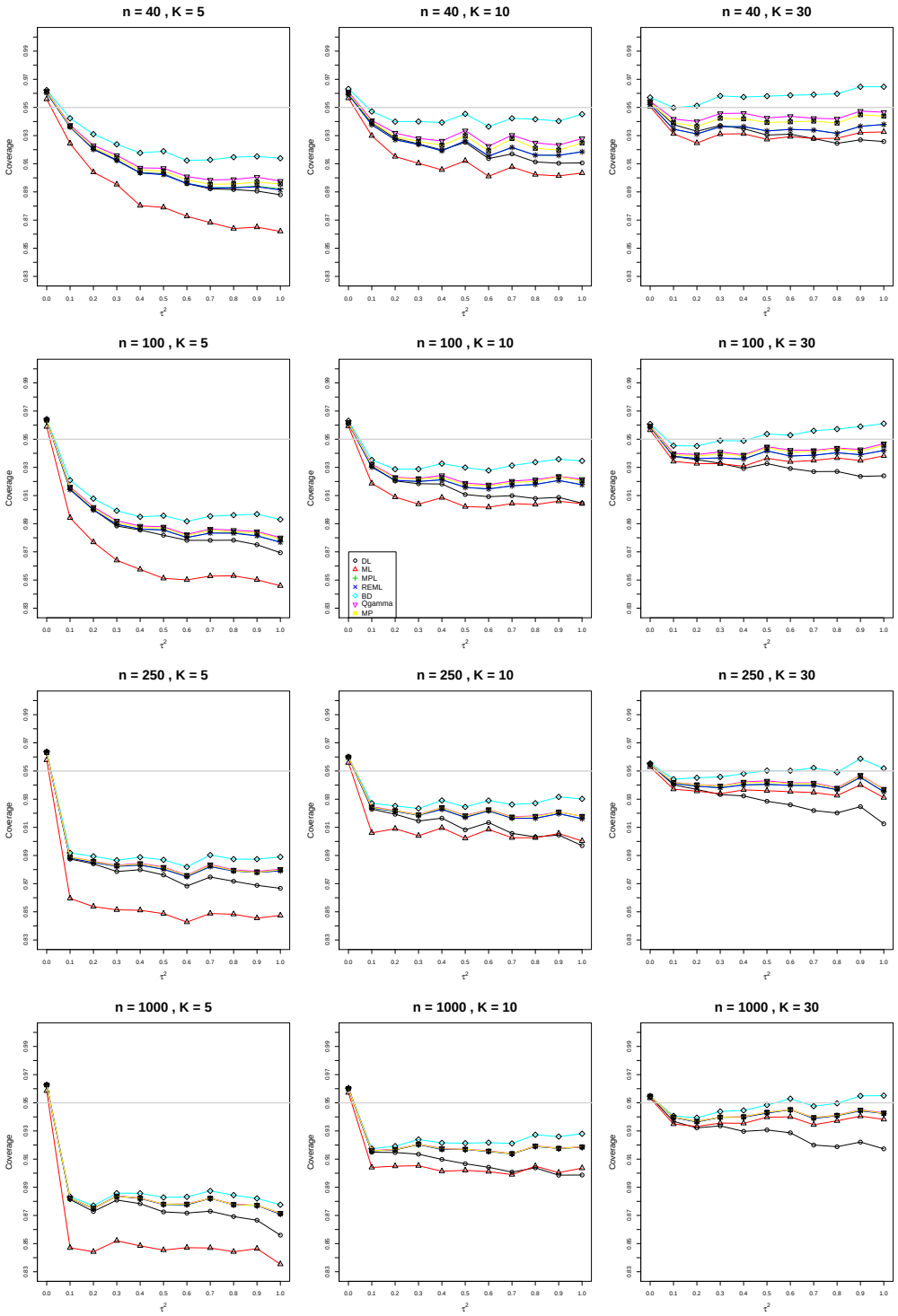


Figure C.24: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_w$ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.4$, $\theta = 1$ and $0 \leq \tau^2 \leq 1$. The inverse-variance weights use the following estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

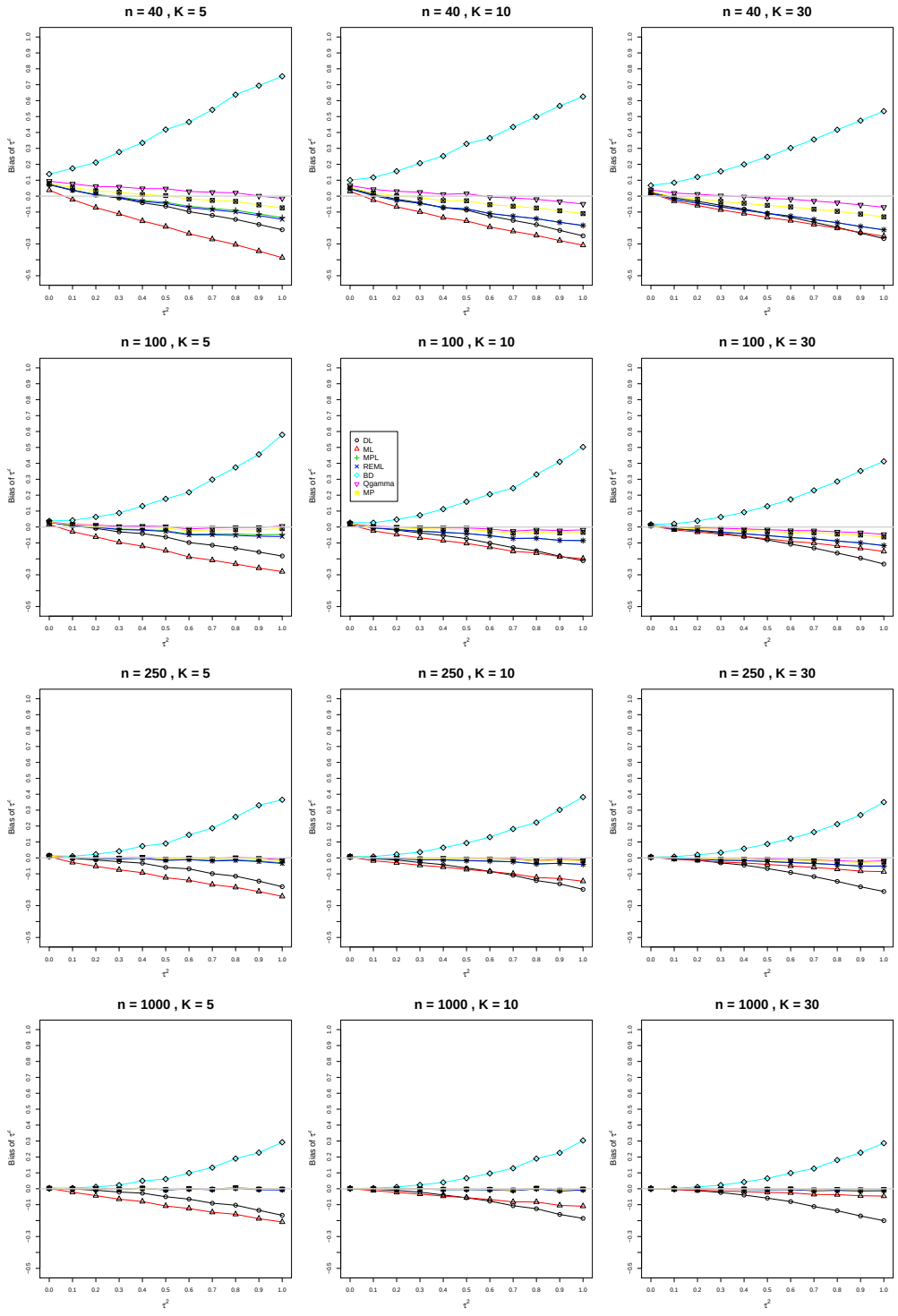


Figure C.25: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.4$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

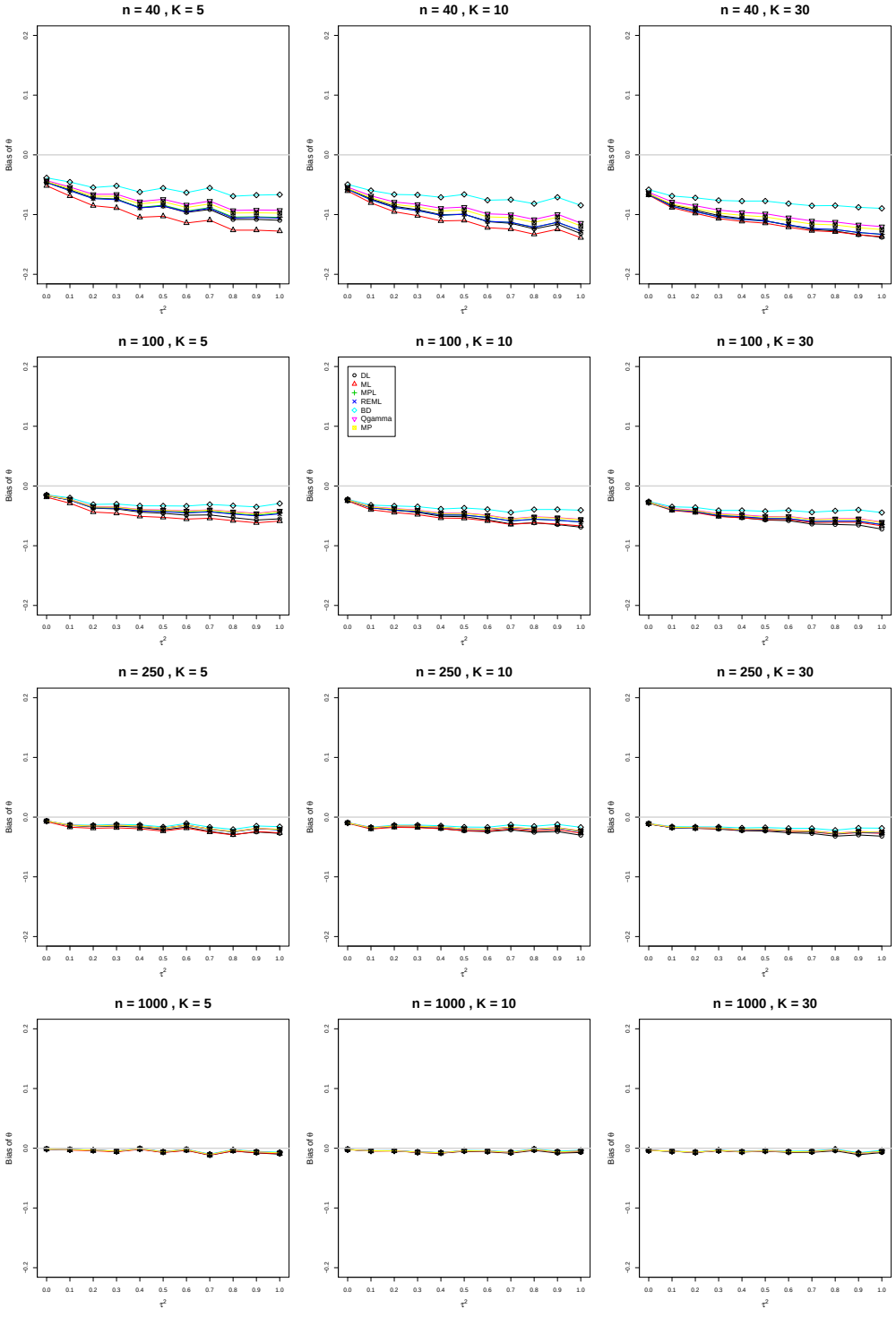


Figure C.26: Bias of the estimated overall effect measure $\hat{\theta}_w$ obtained from K studies for $p_{2j} = 0.4$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 1$. The estimators of $\hat{\theta}_w$ include the estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

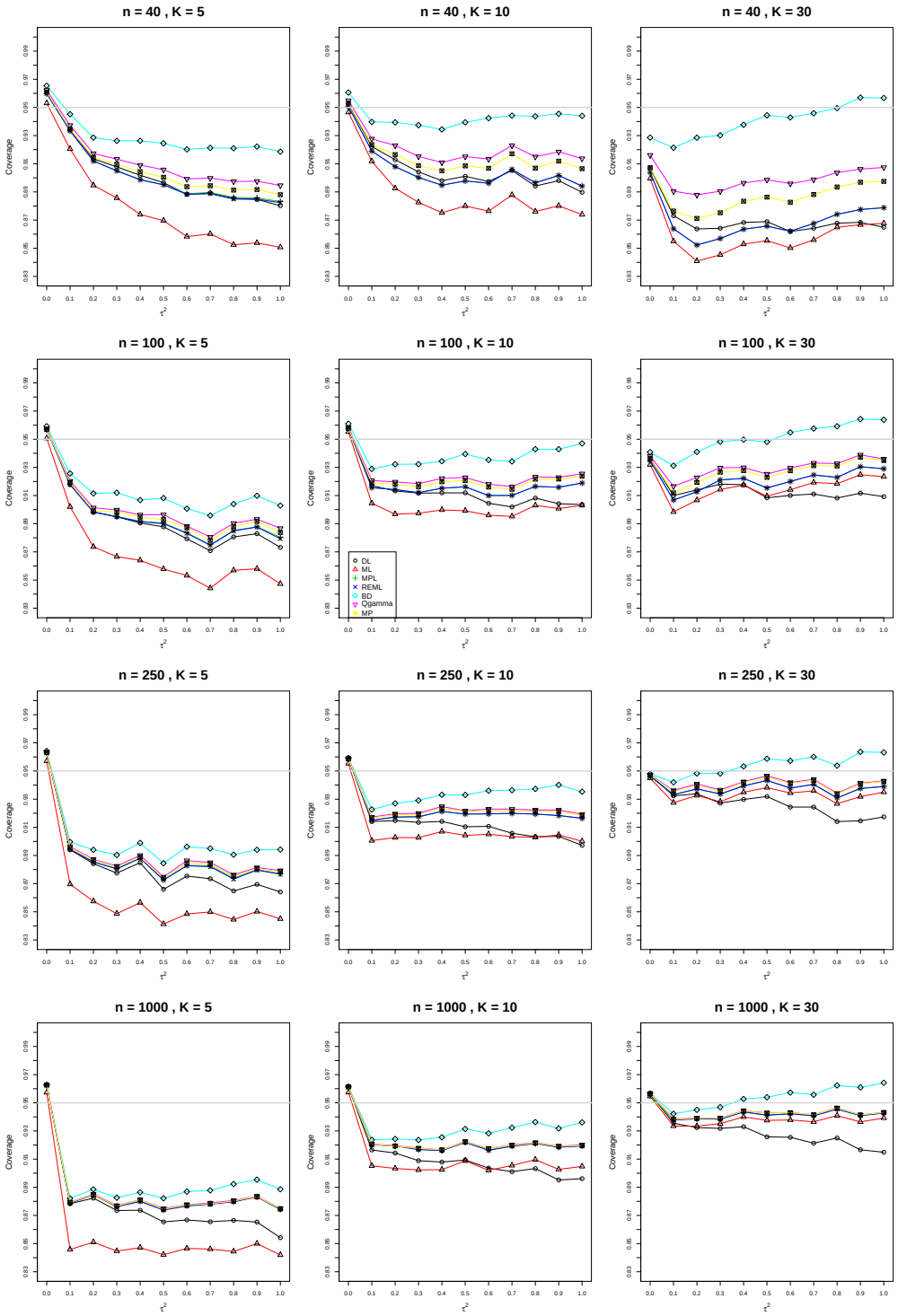


Figure C.27: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_w$ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.4$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 1$. The inverse-variance weights use the following estimators of τ^2: circles – DerSimonian and Laird estimator $\hat{\tau}^2_{DL}$, triangles – maximum likelihood estimator $\hat{\tau}^2_{ML}$ and pluses – maximum penalized likelihood $\hat{\tau}^2_{MPL}$. Light grey line at 0.95.

C.1.2 Full simulation study for method of simulation similar to Kosmidis et al. (2017) when $p_{2j} = 0.1$, $p_{2j} = 0.2$ and $p_{2j} = 0.4$

Results of simulation study for bias and coverage of τ^2 in case of small-moderate heterogeneity

The results of the simulations for data generated using the method of simulation as in Kosmidis et al. (2017) is as following. Figures C.28 - C.33 show the bias and coverage of between study variance for $p_{2j} = 0.1$ and different values of effect measure $\theta = 0, 1, 2$. From Figures C.28,C.30,C.32, we can clearly see that the estimate of τ^2 from all four methods are biased. The bias of τ^2 varies with the method of estimation. The DerSimonian and Laird method performs worse than maximum likelihood and penalized likelihood estimators. Hence, similarly to results for the method of simulation by Viechtbauer (2007), we did not include this method into the figures. Among all the methods, Mandel-Paule method performs the best being a least biased estimator of between study variance. Our primary interest lies in maximum likelihood and penalized likelihood estimators. Similarly to the method of simulation by Viechtbauer (2007), the penalized likelihood estimator of τ^2 performs identical to restricted maximum likelihood estimator. Overall, the bias of maximum likelihood estimator of τ^2 for sparse data ($p_{2j} = 0.1$) varies between 4.13-98.8% for different combination of samples sizes N, number of studies K and overall effect measure θ. Whereas, the overall bias of penalized maximum likelihood estimator of τ^2 ranges between 0-98.2%. The penalized likelihood estimator τ^2 is noticeably better than maximum likelihood estimator of τ^2. However, the bias does still exists. Also for some combinations of samples sizes N, number of studies K and overall effect measure θ, the reductions in bias from $\hat{\tau}^2_{MPL}$ to $\hat{\tau}^2_{ML}$ is very small. Similarly to the method of simulation by Viechtbauer (2007), the bias is larger when probabilities in both arms are equal and data is sparse ($p_{2j} = 0.1$ and $\theta = 0$). For the case when $p_{2j} = 0.1$ and $\theta = 0$, the bias of maximum likelihood estimator is about 8.35-98.8% and the bias of penalized maximum likelihood estimator is about 0.03-98.2%. When the probabilities in both arms are not equal, the bias of maximum likelihood estimator is about 4.9-92.8% for $\theta = 1$ and 4.13-82.8% for $\theta = 2$. For the same scenario, the bias of penalized maximum likelihood estimator is about 0.01-89.8% for $\theta = 1$ and 0-76.9% for $\theta = 2$. Thus, the bias of $\hat{\tau}^2_{ML}$ and $\hat{\tau}^2_{MPL}$ are higher in case of sparse data in both arms ($p_{2j} = 0.1$ and $\theta = 0$) than sparse data in only control arm ($p_{2j} = 0.1$ and $\theta = 1$, $\theta = 2$). The overall improvement in the bias

of maximum likelihood estimate of τ^2 varies between 0.6-52.4%. The improvement in bias is bigger for small number of studies, $K = 5$ (7.7-52.4% improvement) and $K = 10$ (8.175-29.1% improvement), in comparison to $K = 30$ (0.6-9% improvement). Comparing the simulation method by Viechtbauer (2007) and the simulation method by Kosmidis et al. (2017), the bias of estimators of τ^2 are somewhat larger in the simulation method by Kosmidis et al. (2017). Coming to the coverage, the coverage of τ^2 from Q-profiled confidence intervals and Profiled-likelihood confidence intervals are shown in Figures C.29, C.31 and C.33. Both methods perform very well for the data simulated using the method of simulation by Kosmidis et al. (2017) apart from the case when coverage deteriorate when $N \leq 250$ and $K = 30$ for sparse data ($p_{2j} = 0.1$ and $\theta = 0$).

For $p_{2j} > 0.1$, the bias of between-study variance is similar to the bias when $p_{2j} = 0.1$. Figure C.51 C.53 C.55 and C.57 C.59 C.61 in this Appendix show the bias of τ^2 when $p_{2j} = 0.2$ and $p_{2j} = 0.4$ respectively. Again, the Mandel-Paule method outperform maximum likelihood, penalized maximum likelihood and restricted maximum likelihood method in estimation of between study variance τ^2. The bias of maximum likelihood estimator $\hat{\tau}^2_{ML}$ in non-sparse data varies between 4-87.3% for $p_{2j} = 0.2$ and 4-74.8% for $p_{2j} = 0.4$. Whereas the bias of penalized maximum likelihood estimator $\hat{\tau}^2_{MPL}$ varies between 0-81.9% for $p_{2j} = 0.2$ and 0-67.95% $p_{2j} = 0.4$. We can clearly see the reduction of the bias between maximum likelihood estimator $\hat{\tau}^2_{ML}$ and penalized maximum likelihood estimator $\hat{\tau}^2_{MPL}$. However, the bias has not been eliminated completely. Also, the absolute value of the bias is less in non-sparse data ($p_{2j} = 0.2$ and $p_{2j} = 0.4$ with $\theta = 0, 1, 2$) than in sparse data ($p_{2j} = 0.1$ with $\theta = 0, 1, 2$). Coverages for τ^2 are shown in the Figures C.52, C.54 C.56 and C.58, C.60 C.62 in Appendix for $p_{2j} = 0.2$ and $p_{2j} = 0.4$ are respectively. When $p_{2j} = 0.2$ and $p_{2j} = 0.4$ the coverages of τ^2 are similar $p_{2j} = 0.1$ with the Q-profile confidence intervals being to conservative for $N \geq 100$ and Profile-likehood confidence interval being too liberate. In order to be safe in choose intervals for τ^2, we recommend Q-profile confidence intervals. Comparing the simulation study by Kosmidis et al. (2017) and our simulations, the results are as following. Kosmidis et al. (2017) considered only particular cases of meta-analysis. One of the cases is when $p_{1j} = 0.40$ and $p_{2j} = 0.219$ resulting in true overall odds ratio $\psi = 2.377$ and log odds ratio $\theta = 0.9$. Another case is when $p_{1j} = 0.3$ and $p_{2j} = 0.1$ resulting in true overall odds ratio $\psi = 3.85$ and log odds ratio $\theta = 1.35$. We considered the general case when $p_{2j} = 0.1, 0.2, 0.4$ with $\theta = 0, 1, 2$. This

results in twelve combinations of p_{2j} and θ. So one of the simulation results by Kosmidis et al. (2017) is a nearly a particular case of our simulations when $p_{2j} = 0.2$ and $\theta = 1$ (Figure C.53). In addition to simulations of Kosmidis et al. (2017), we added the Mandel-Paule method and restricted maximum likelihood method for point estimator of τ^2. We have also added the interval estimation of τ^2 with Q-profile and Profile-likehood method based of $\hat{\tau}^2_{REML}$. Since $\hat{\tau}^2_{REML}$ $\hat{\tau}^2_{MPL}$ perform similarly, their confidence intervals would similar too.

Kosmidis et al. (2017) limited his simulations with $0 \leq \tau^2 \leq 2.5$. In addition to values $0 \leq \tau^2 \leq 2.5$, we considered simulations with small-moderate heterogeneity ($0 \leq \tau^2 \leq 1$) and moderate-large heterogeneity $0 \leq \tau^2 \leq 10$.

Our simulations and simulations Kosmidis et al. (2017) provide similar results. The results of our simulations show that bias of maximum likelihood estimate $\hat{\tau}_{ML}$ is reduced by penalization of the likelihood. The proportion of the bias reduced for small-moderate values of heterogeneity and moderate-large values of heterogeneity are similar. For sparse and non-sparse data, our results show that the penalization of the likelihood does not entirely remove the bias of maximum likelihood. The estimates of between study variance still remain biased for both sparse and non-sparse data. The bias of between-study variance is larger for sparse data than for non-sparse data.

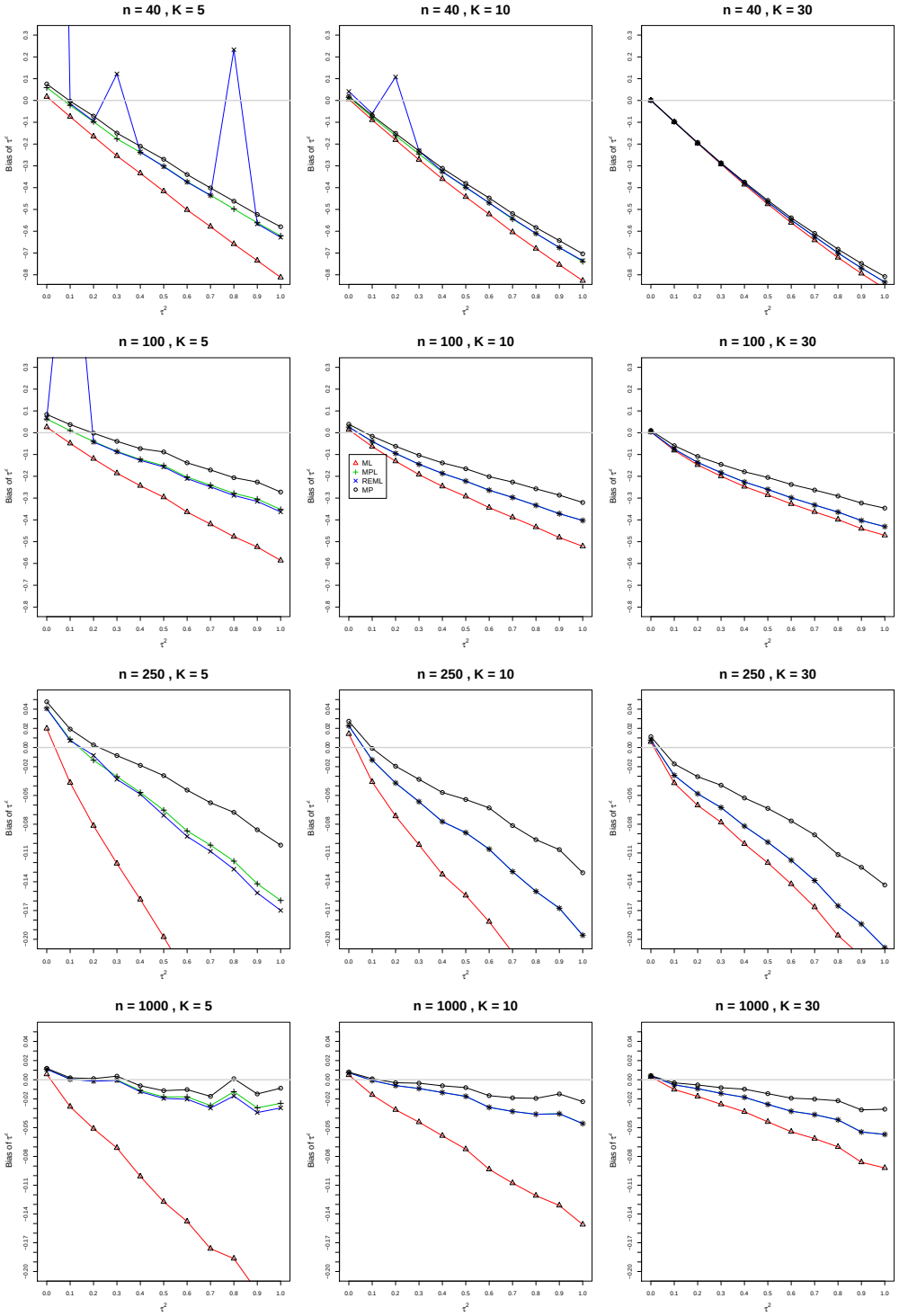


Figure C.28: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 0$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

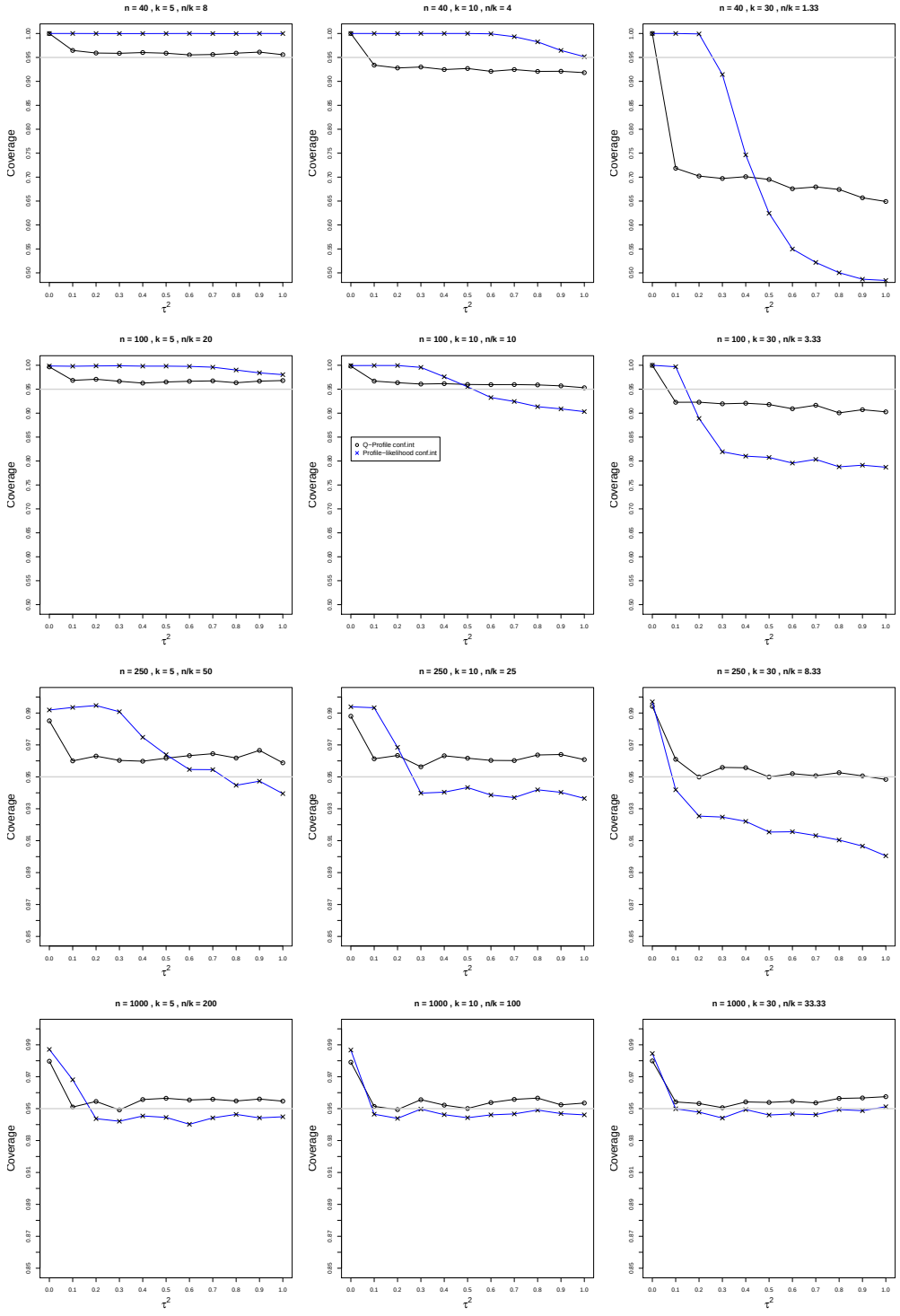


Figure C.29: Coverage at the nominal confidence level of 0.95 of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 0$ and $0 \leq \tau^2 \leq 1$. Interval estimation methods: black circles – Q-profile confidence interval for τ^2 based on χ^2 distribution and dark blue crosses – Profile likelihood confidence intervals. Light grey line at 0.95.

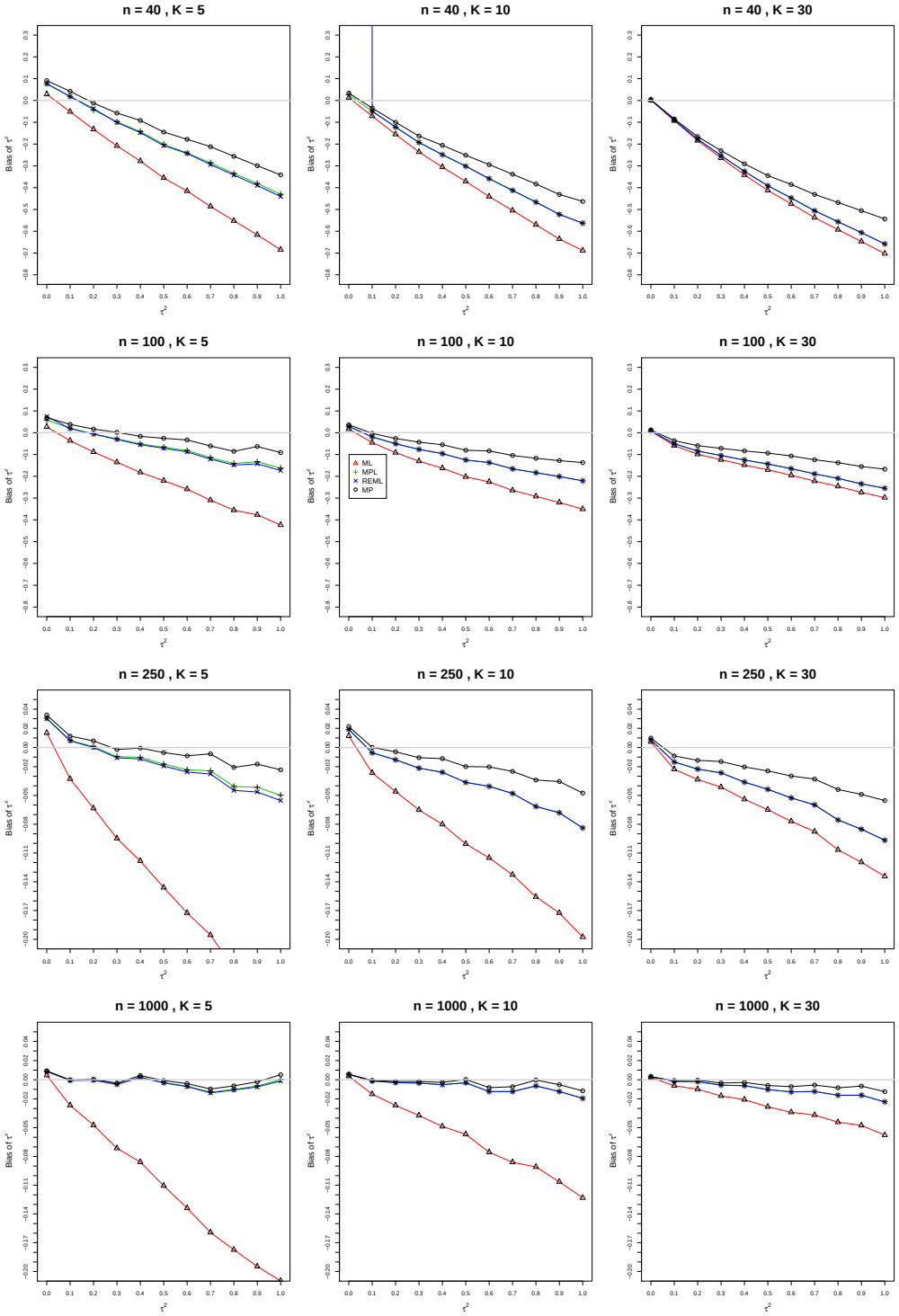


Figure C.30: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 1$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

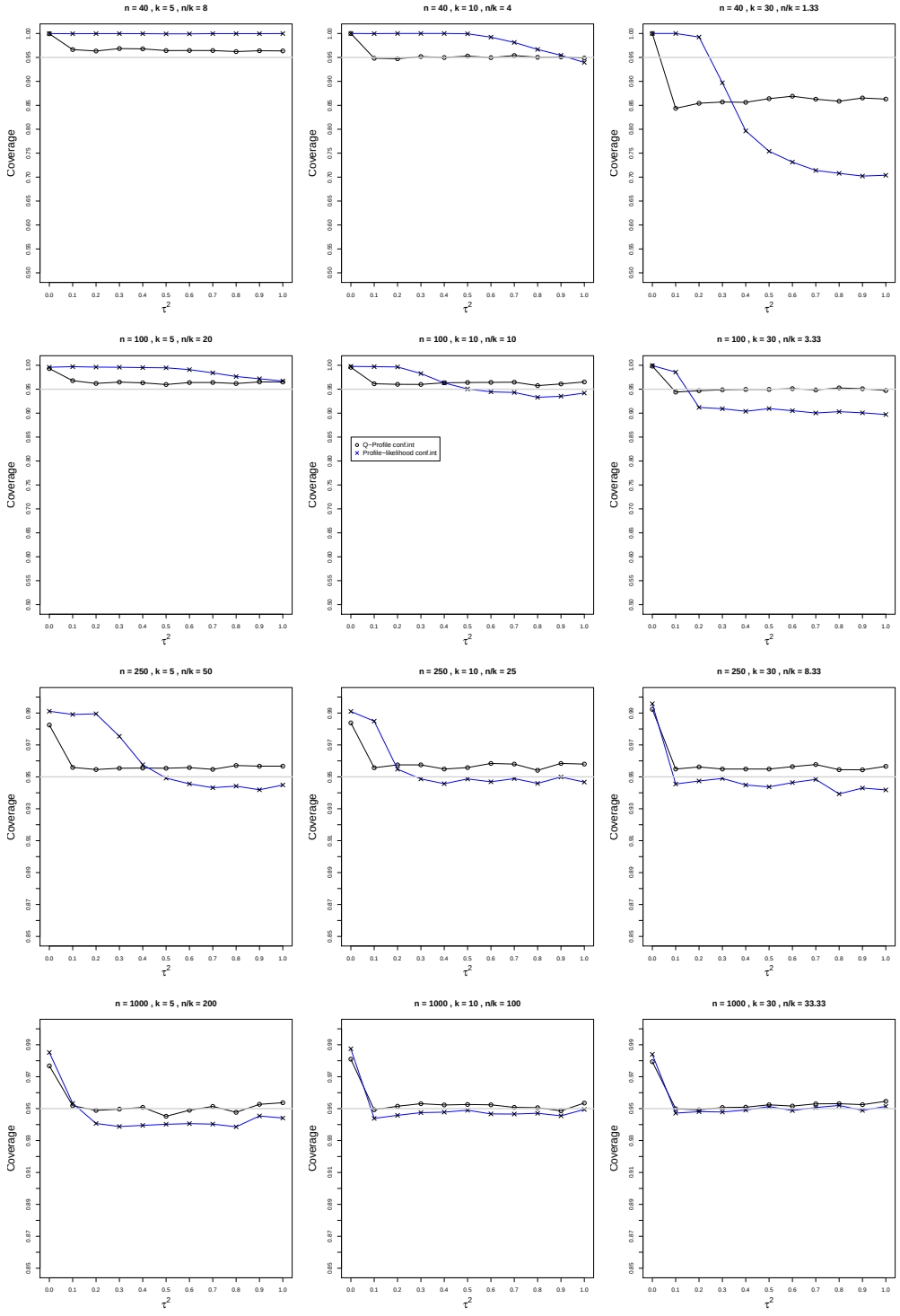


Figure C.31: Coverage at the nominal confidence level of 0.95 of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 1$ and $0 \leq \tau^2 \leq 1$. Interval estimation methods: black circles – Q-profile confidence interval for τ^2 based on χ^2 distribution dark blue crosses – Profile likelihood confidence intervals. Light grey line at 0.95.

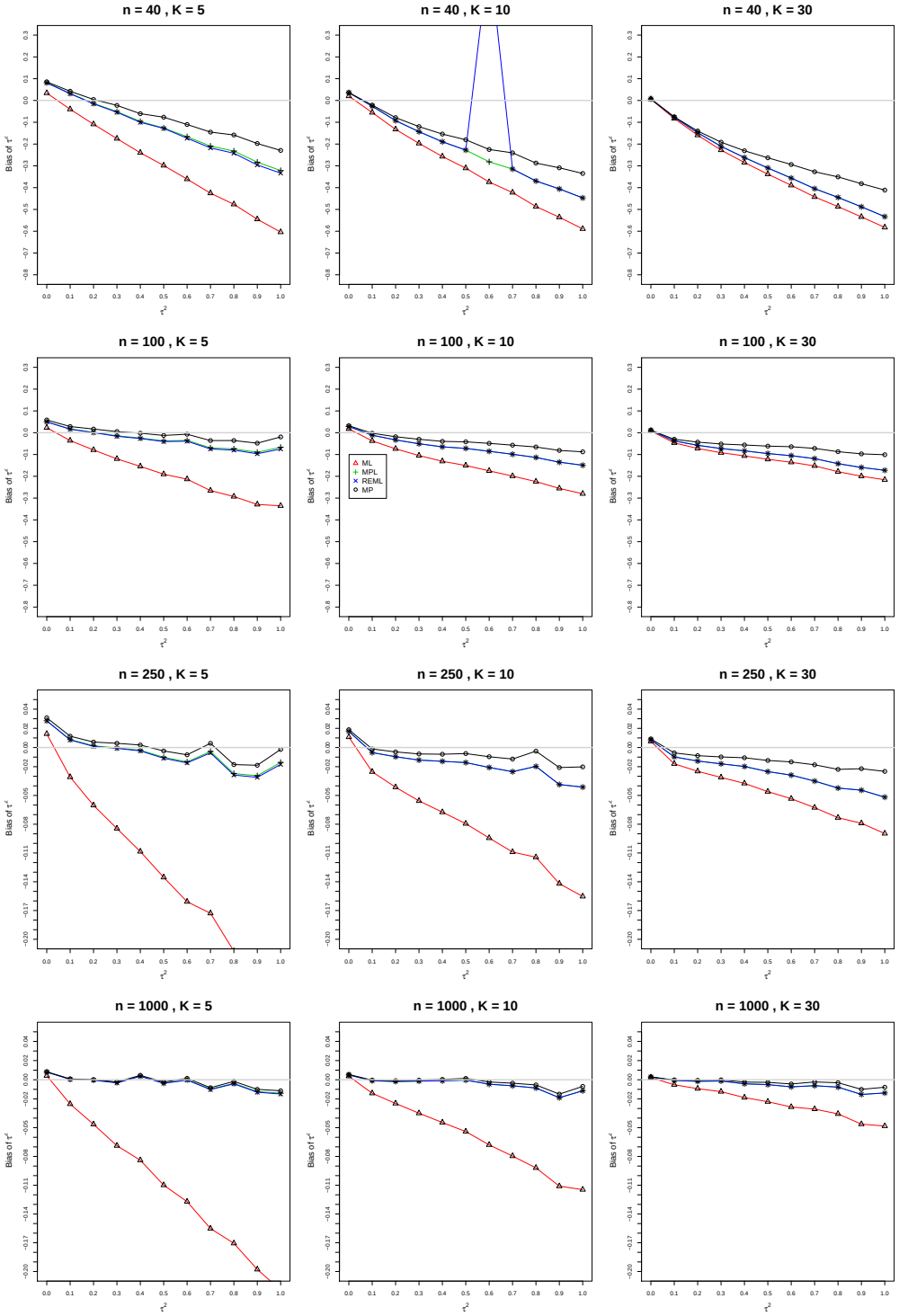


Figure C.32: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

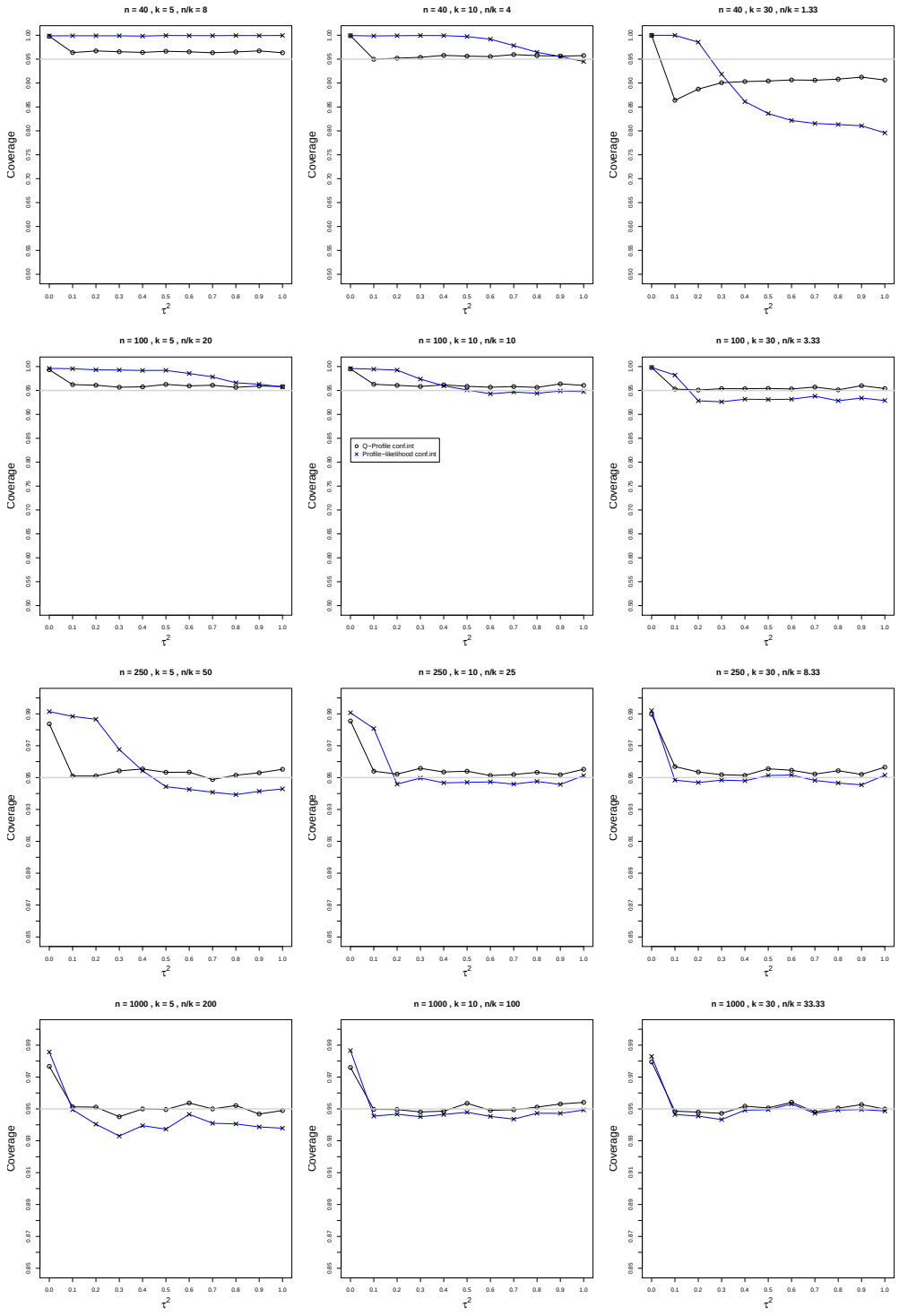


Figure C.33: Coverage at the nominal confidence level of 0.95 of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 1$. Interval estimation methods: black circles – Q-profile confidence interval for τ^2 based on χ^2 distribution and dark blue crosses – Profile likelihood confidence intervals. Light grey line at 0.95.

Results of simulation study for bias and coverage of θ in case of small-moderate heterogeneity

The improvement in the bias of τ^2 has a positive impact on the bias and coverage of overall effect measure. Regardless of the method, the bias of overall effect measure is indistinguishable. The bias of θ is dissimilar for $\theta = 0$, 1 and 2 and $p_{2j} = 0.1$ (see Figure C.34 for bias of $\hat{\theta}_{MPL}$ for $\theta = 0$, 1 and 2. The bias of $\hat{\theta}_{MPL}$ is large when $N \leq 100$ and decreases with increase in sample size from $N = 100$ to $N = 1000$. The biases from all methods are shown in the Figures C.35, C.37 and C.39 for $\theta = 0$, 1 and 2. Figures C.35 - C.40 shows that the penalized maximum likelihood method outperforms maximum likelihood in terms of bias and coverage. However when data is sparse, $\theta = 0$ and $p_{2j} = 0.1$, the coverage deteriorate with increasing K. The coverage improves when control arm only has a low probability ($p_{2j} = 0.1$) and treatment arm has probabilities $p_{1j} > 0.1$.

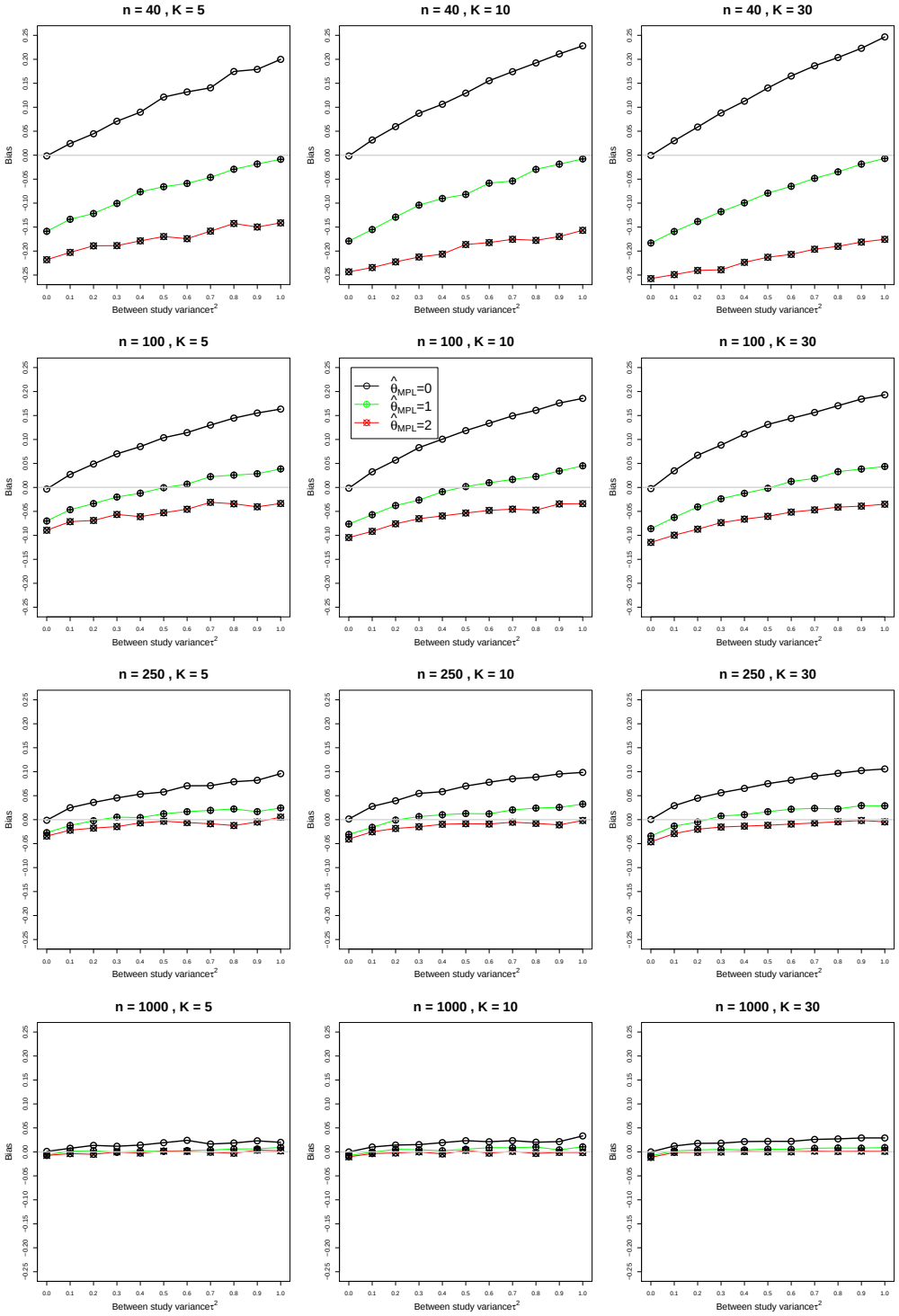


Figure C.34: Bias of overall odds ratio θ_{IV} obtained from K studies by the inverse-variance method with the moment estimator $\hat{\tau}^2_{MPL}$ in the weights, for $p_{2j} = 0.1$, and $0 \leq \tau^2 \leq 1$. The biases are given for $\theta = 0$ (circles), $\theta = 1$ (circle plus), and $\theta = 2$ (circle cross). Light grey line at 0.

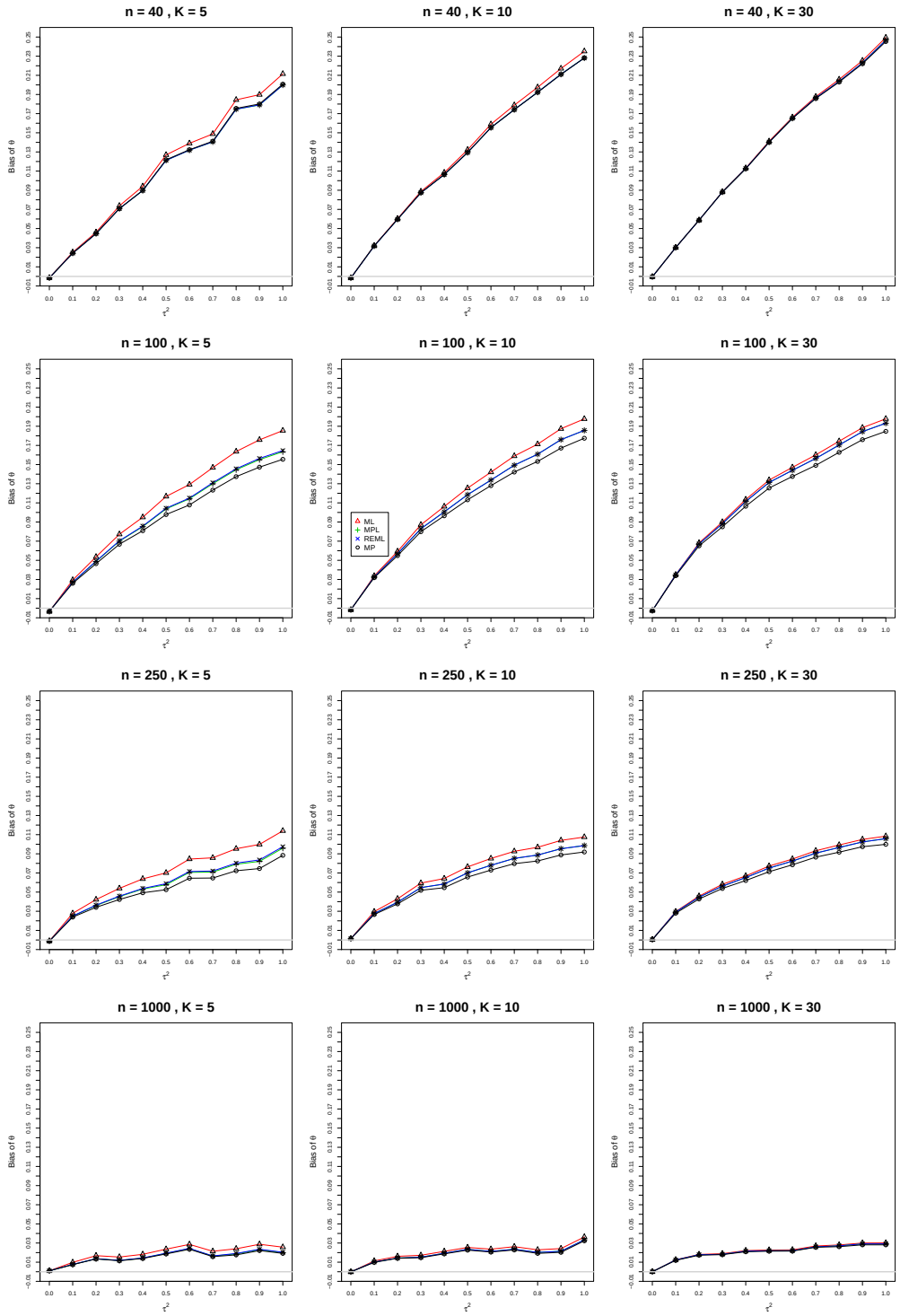


Figure C.35: Bias of the estimated overall effect measure $\hat{\theta}_w$ obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 0$ and $0 \leq \tau^2 \leq 1$. The estimators of $\hat{\theta}_w$ include the estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

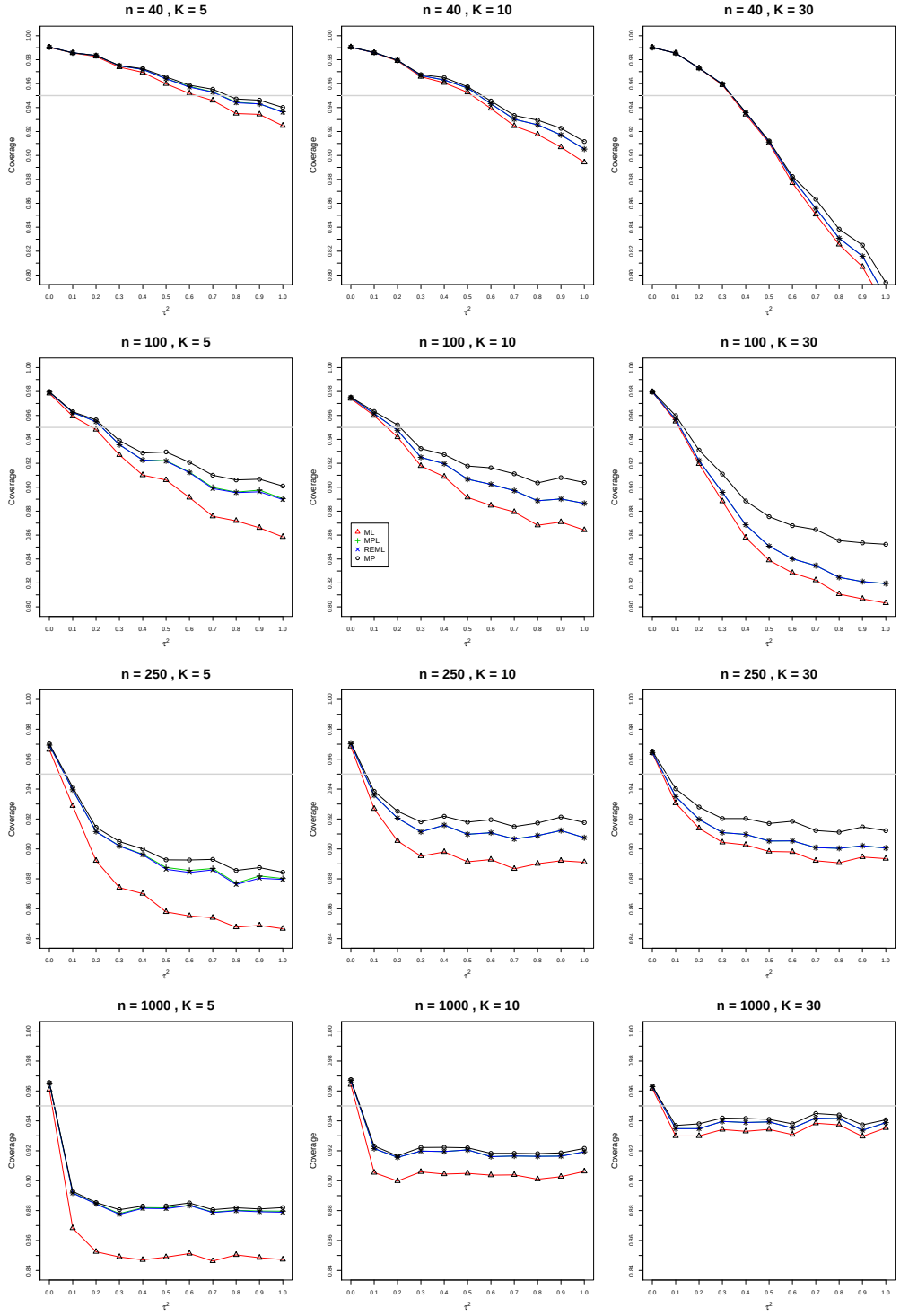


Figure C.36: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_w$ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.1$, $\theta_w = 0$ and $0 \leq \tau^2 \leq 1$. The inverse-variance weights use the following estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

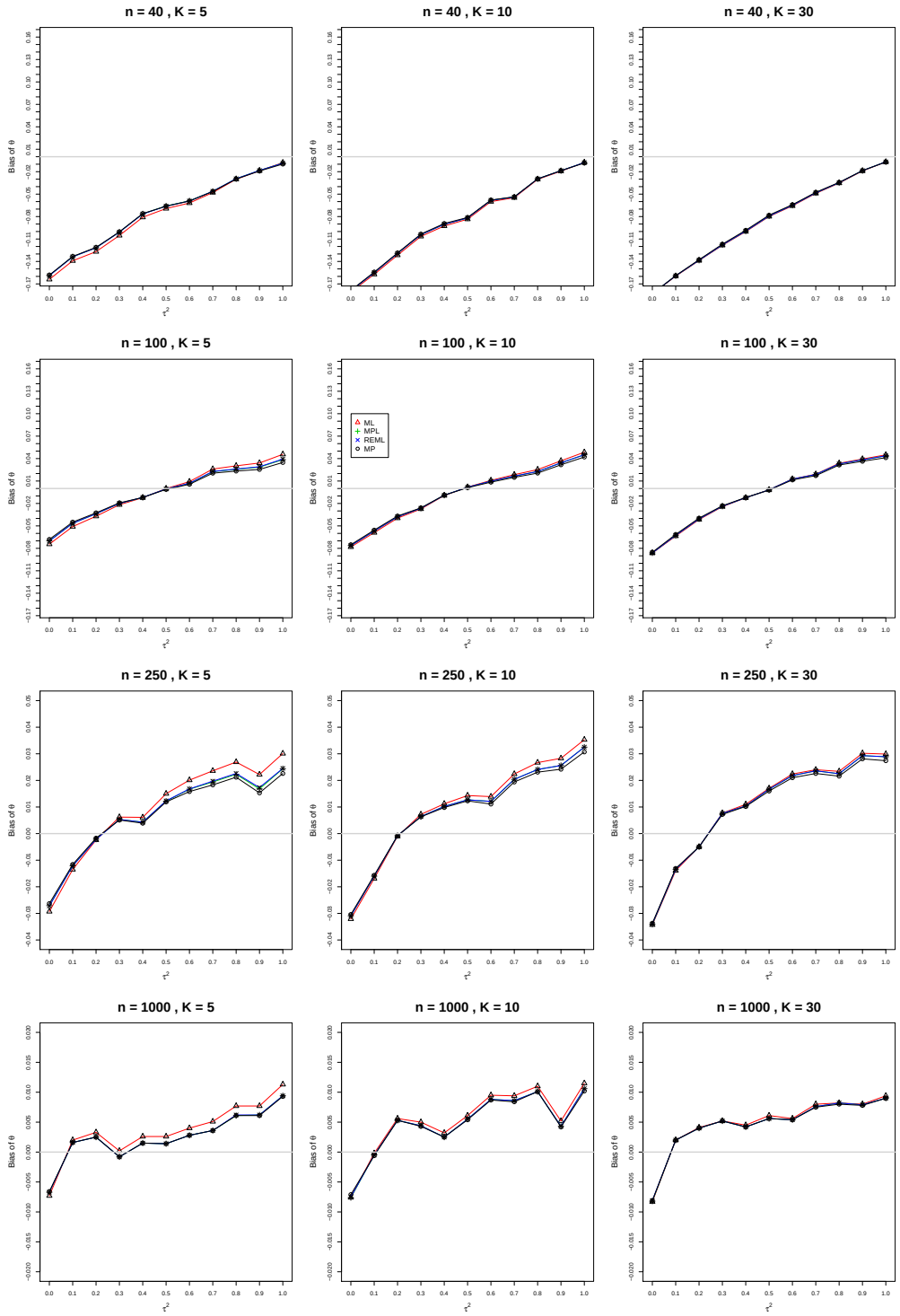


Figure C.37: Bias of the estimated overall effect measure $\hat{\theta}_w$ obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 1$ and $0 \leq \tau^2 \leq 1$. The estimators of $\hat{\theta}_w$ include the estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

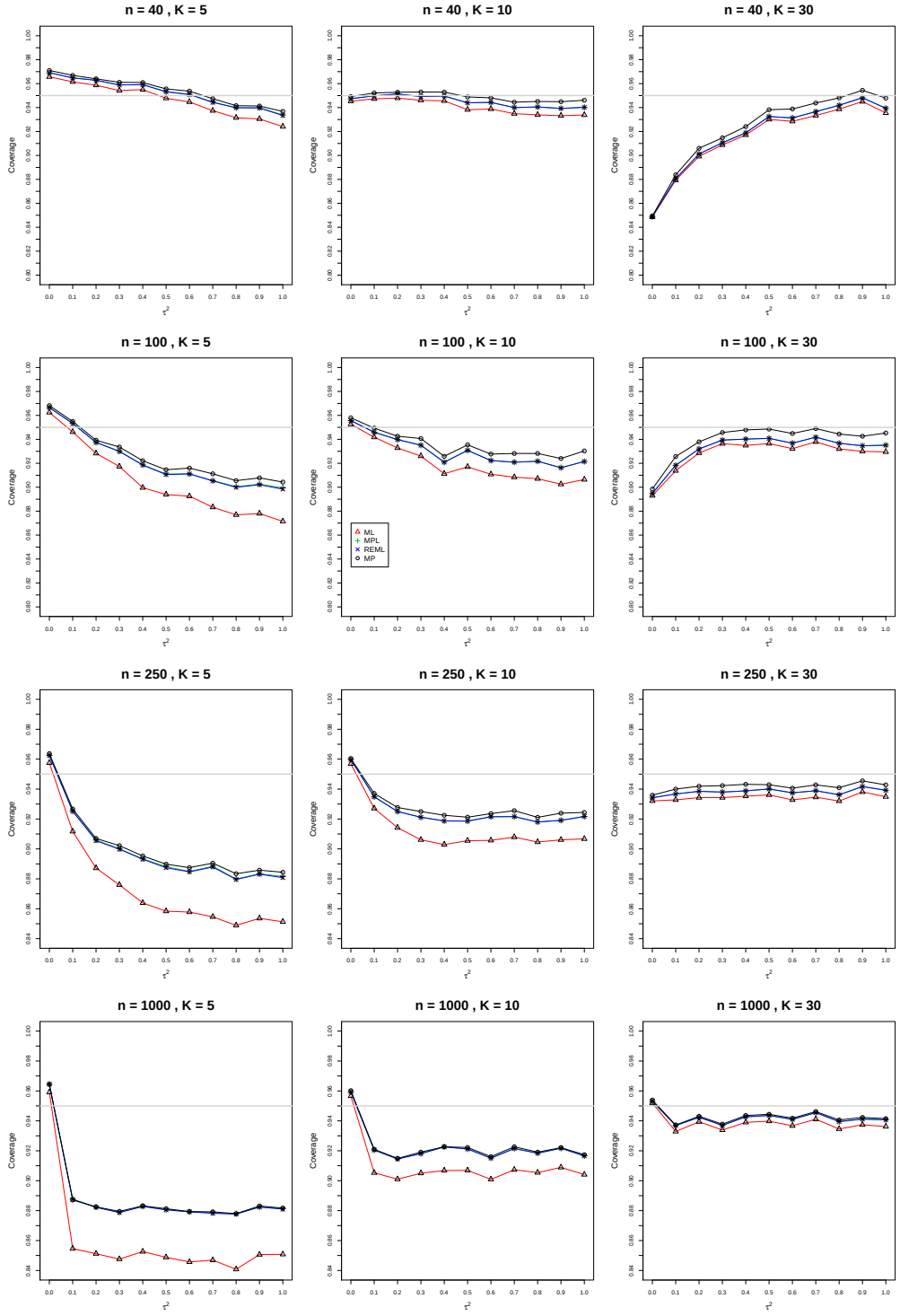


Figure C.38: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_w$ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.1$, $\theta_w = 1$ and $0 \leq \tau^2 \leq 1$. The inverse-variance weights use the following estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

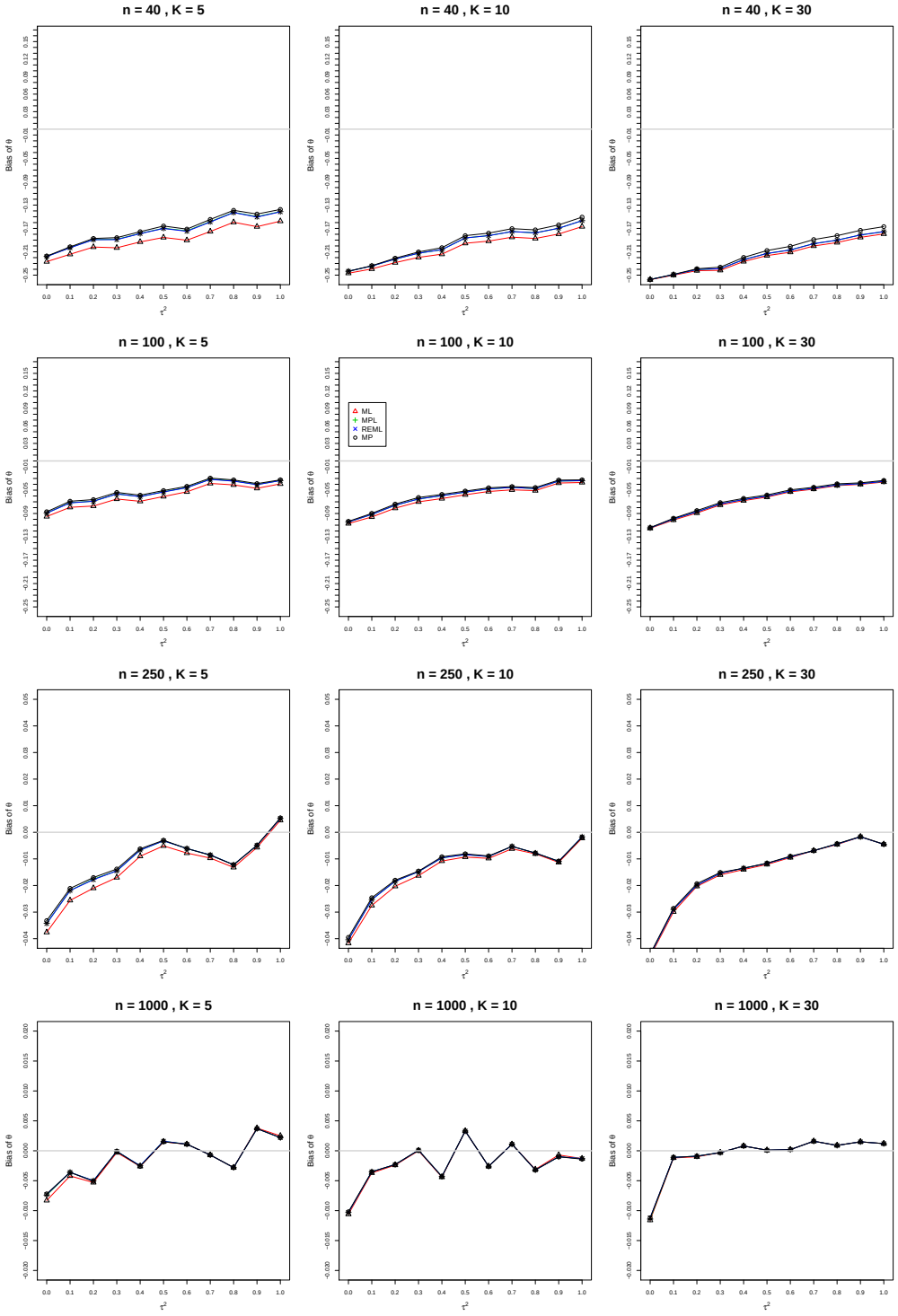


Figure C.39: Bias of the estimated overall effect measure $\hat{\theta}_w$ obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 1$. The estimators of $\hat{\theta}_w$ include the estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

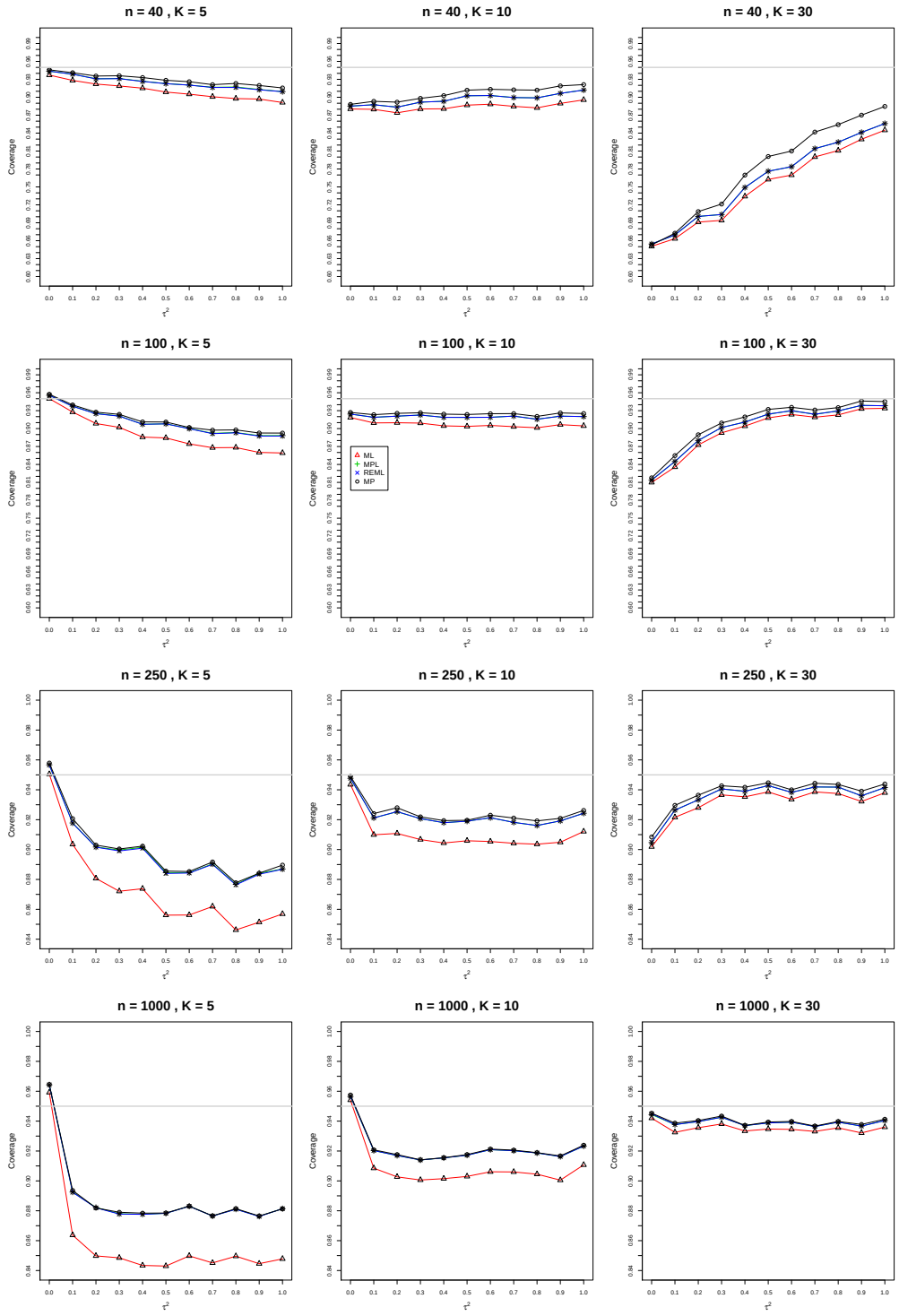


Figure C.40: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_w$ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.1$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 1$. The inverse-variance weights use the following estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

Results of simulation study for bias and coverage of τ^2 in case of moderate-large heterogeneity

Similarly as for the method of simulation by Viechtbauer (2007), we performed simulation study for the method of simulation by Kosmidis et al. (2017) with moderate-large values of heterogeneity across studies. Figures C.41, C.43, C.45 and C.42, C.44, C.46 show the bias and coverage of τ^2 from point and interval estimator of τ^2 for moderate-large heterogeneity $0 \leq \tau^2 \leq 10$. The bias of τ^2 for moderate-large heterogeneity $0 \leq \tau^2 \leq 10$ is similar to small-moderate heterogeneity $0 \leq \tau^2 \leq 1$. Again, Mandel-Paule method outperforms likelihood based methods. Overall for $p_{2j} = 0.1$, the bias of Mandel-Paule estimator varies between 0.1-80%. Whereas, the bias of maximum likelihood and penalized likelihood estimators vary between 4.58-86.22% and 0.05-83.25%. Again $\hat{\tau}_{MPL}$ reduces the bias of $\hat{\tau}_{MPL}$. However, the reductions in bias is similar to restricted maximum likelihood. When $p_{2j} = 0.1$ and $\theta = 0$, the bias of maximum likelihood and penalized likelihood estimators vary between 9.42-86.22% ans 3.27-83.25% respectively. The biases are large when both arms have low probabilities in comparison to when only control arm has a low probability. The bias of both likelihood based estimators is large in comparison to simulation results with the method of simulation by Viechtbauer (2007). We would expect the smaller bias when $p_{2j} = 0.1$ with $\theta = 1$ and $\theta = 2$, since only control arm has a lower probabilities across K studies.

Coming to coverages, overall Q-profile and Profile-likelihood based confidence intervals performs worse than they performed for small-moderate heterogeneity. Particularly, the coverages of Profile-likelihood based confidence interval deteriorate when number of studies increase from $K = 5$ to $K = 10$ and $K = 30$ (see Figures C.42, C.44, C.46).

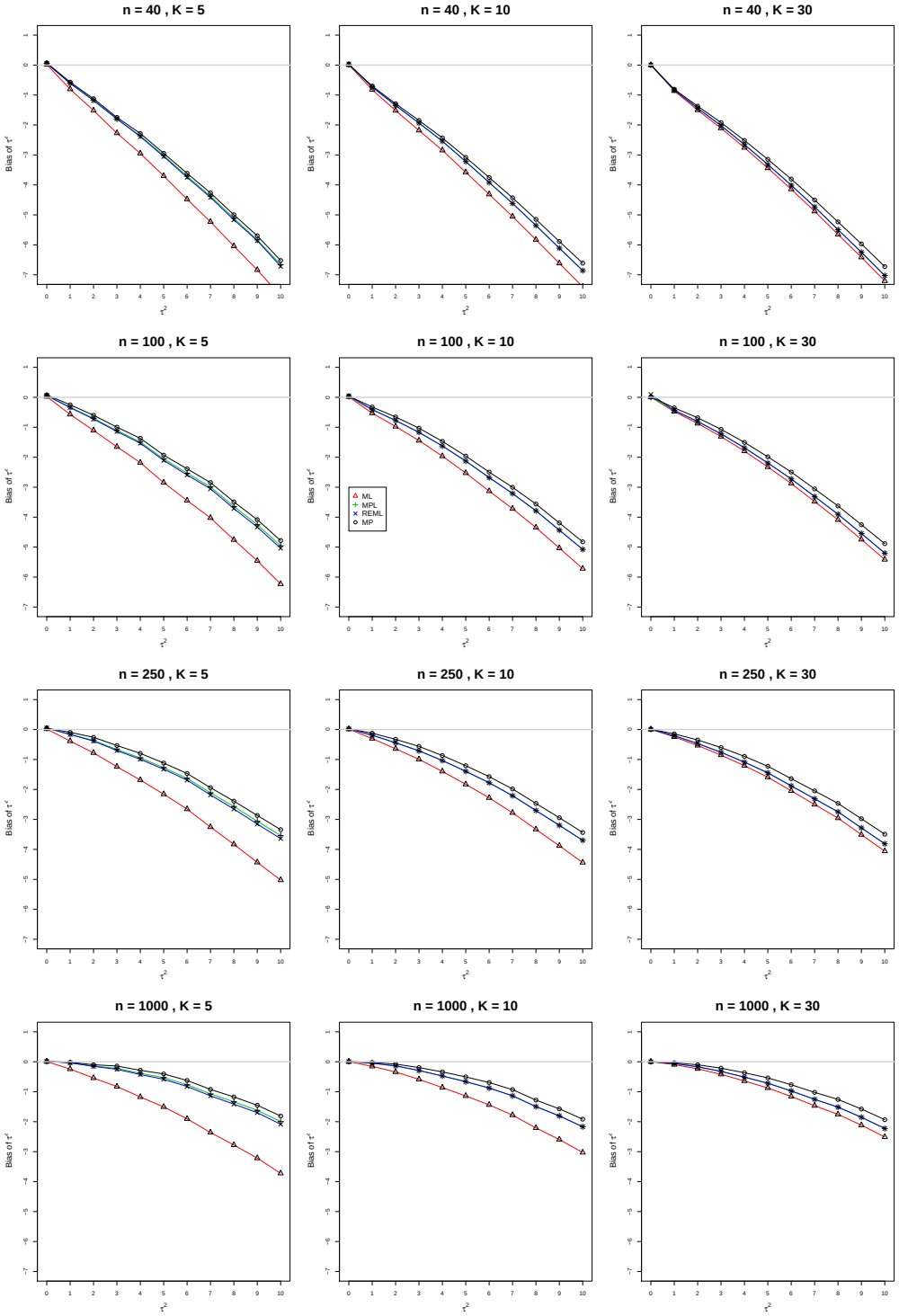


Figure C.41: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 0$ and $0 \leq \tau^2 \leq 10$. The estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

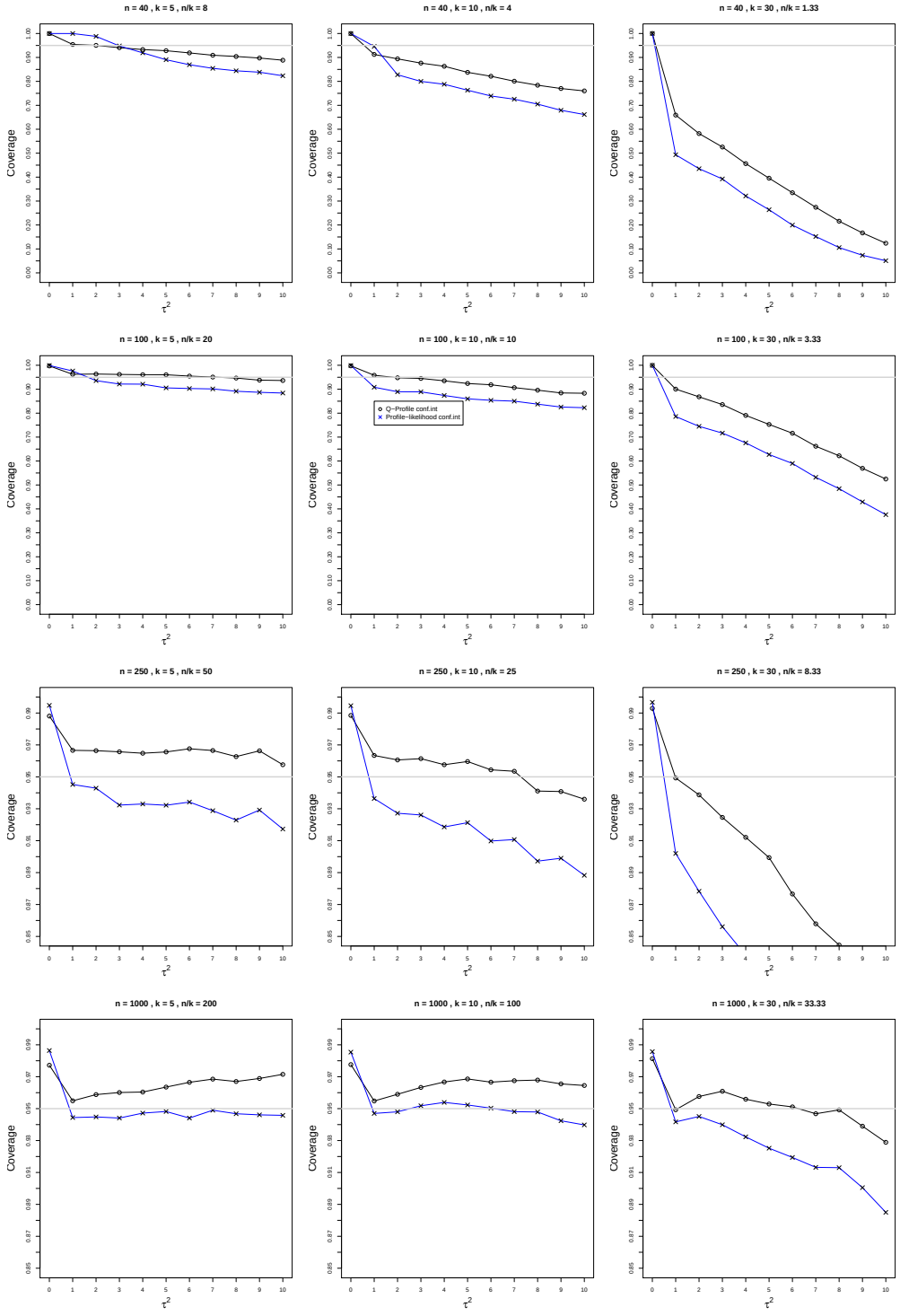


Figure C.42: Coverage at the nominal confidence level of 0.95 of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 0$ and $0 \leq \tau^2 \leq 10$. Interval estimation methods: black circles – Q-profile confidence interval for τ^2 based on χ^2 distribution and dark blue crosses – Profile likelihood confidence intervals. Light grey line at 0.95.

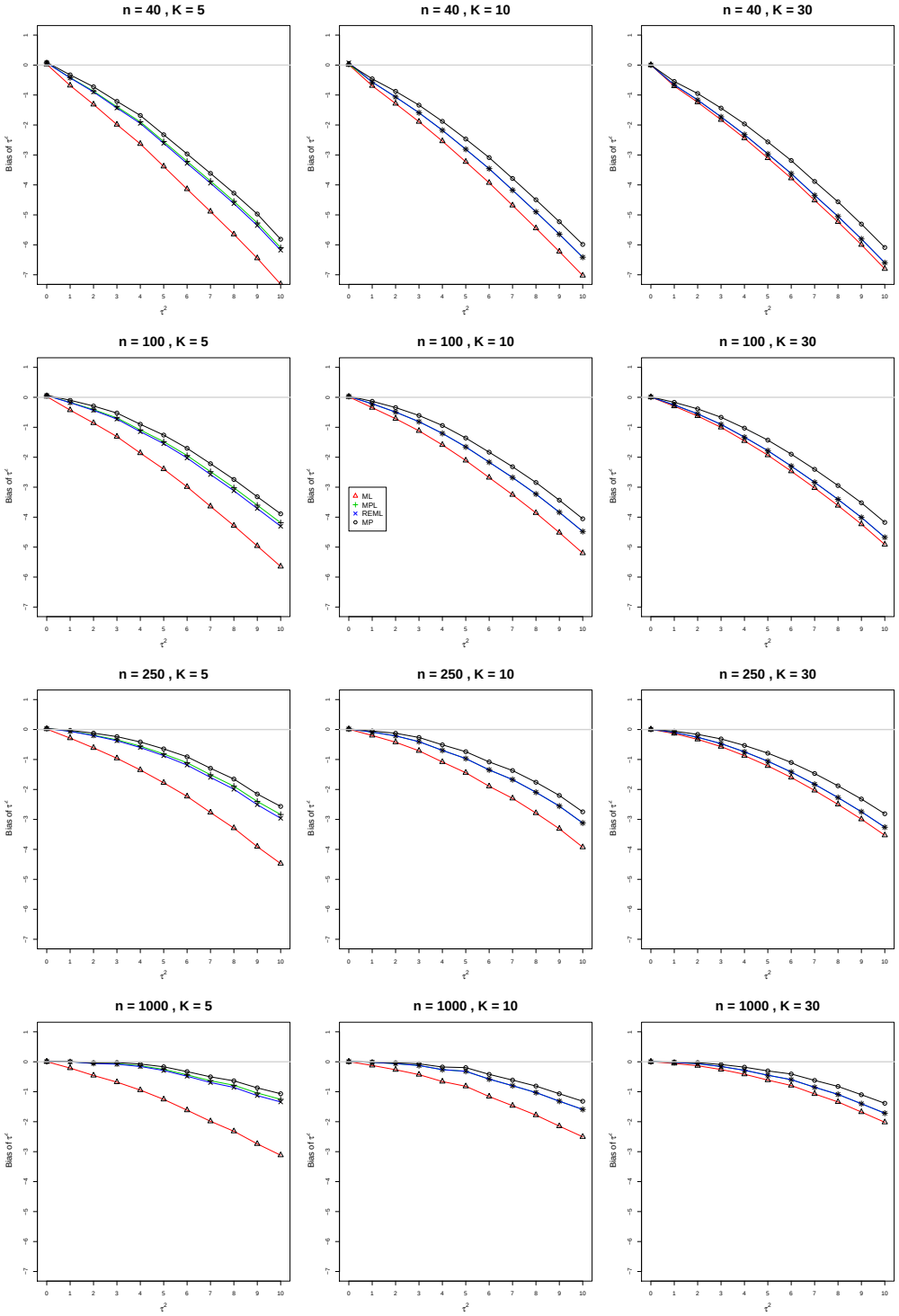


Figure C.43: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 1$ and $0 \leq \tau^2 \leq 10$. The estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

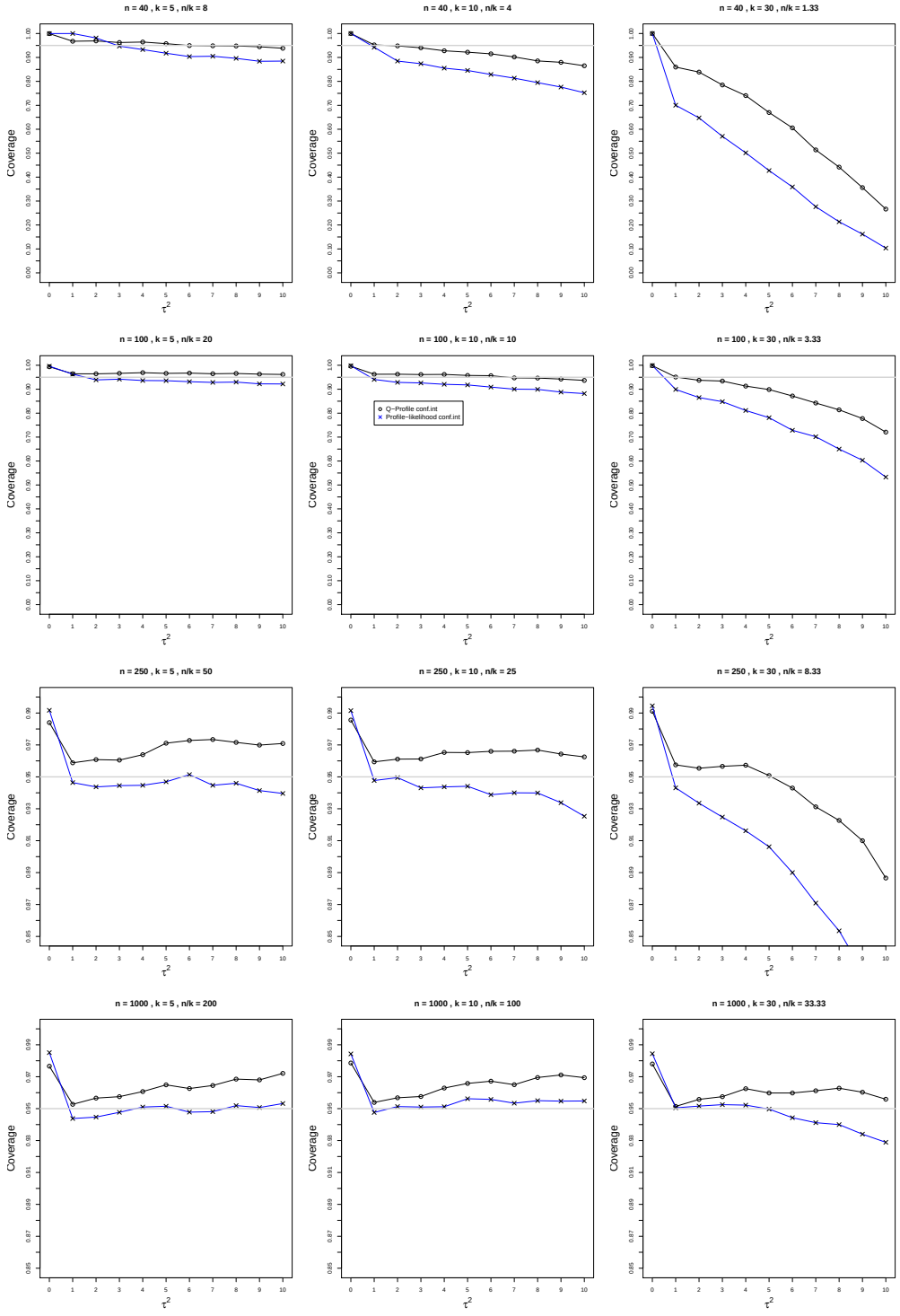


Figure C.44: Coverage at the nominal confidence level of 0.95 of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 1$ and $0 \leq \tau^2 \leq 10$. Interval estimation methods: black circles – Q-profile confidence interval for τ^2 based on χ^2 distribution and dark blue crosses – Profile likelihood confidence intervals. Light grey line at 0.95.

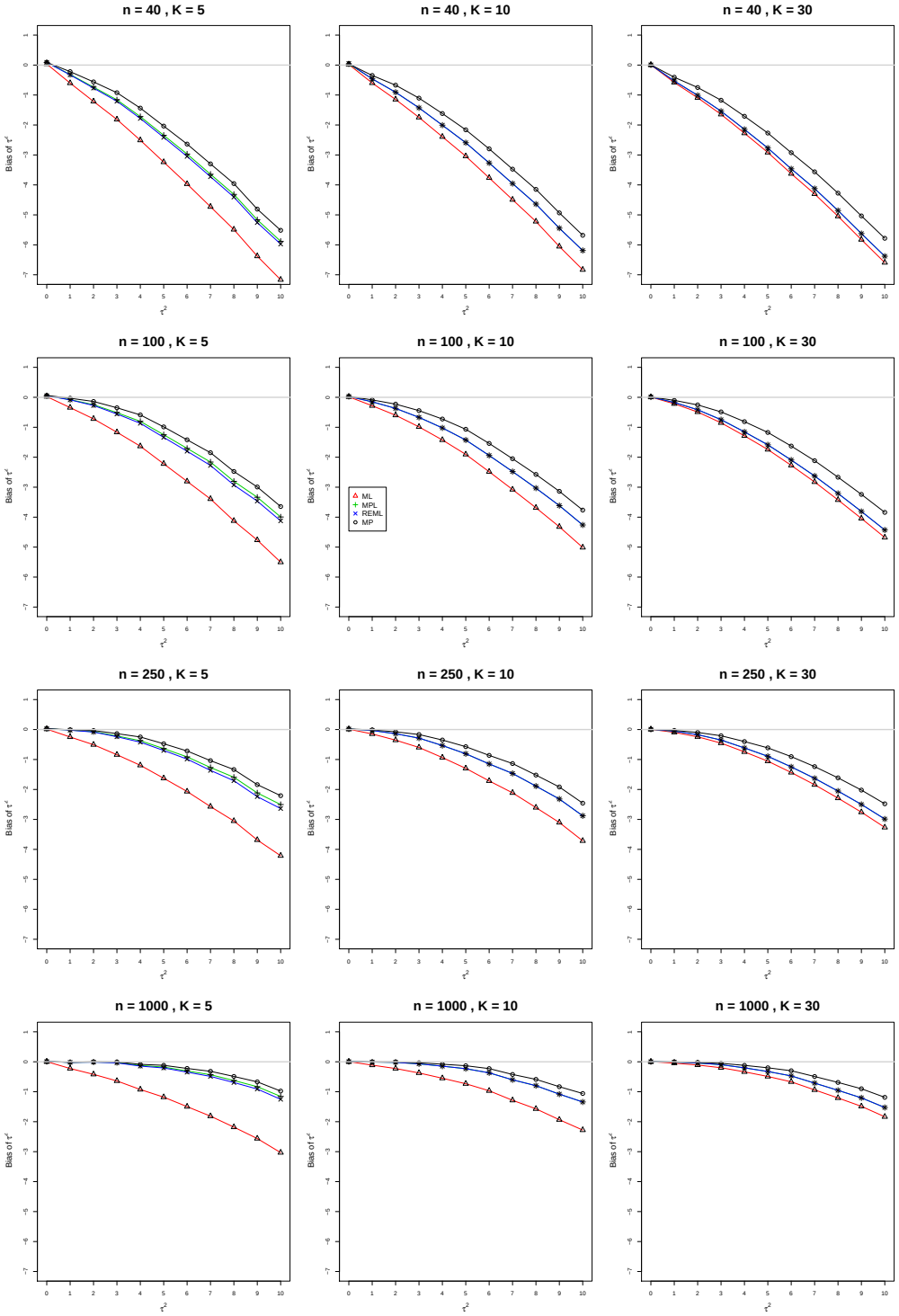


Figure C.45: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 10$. The estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

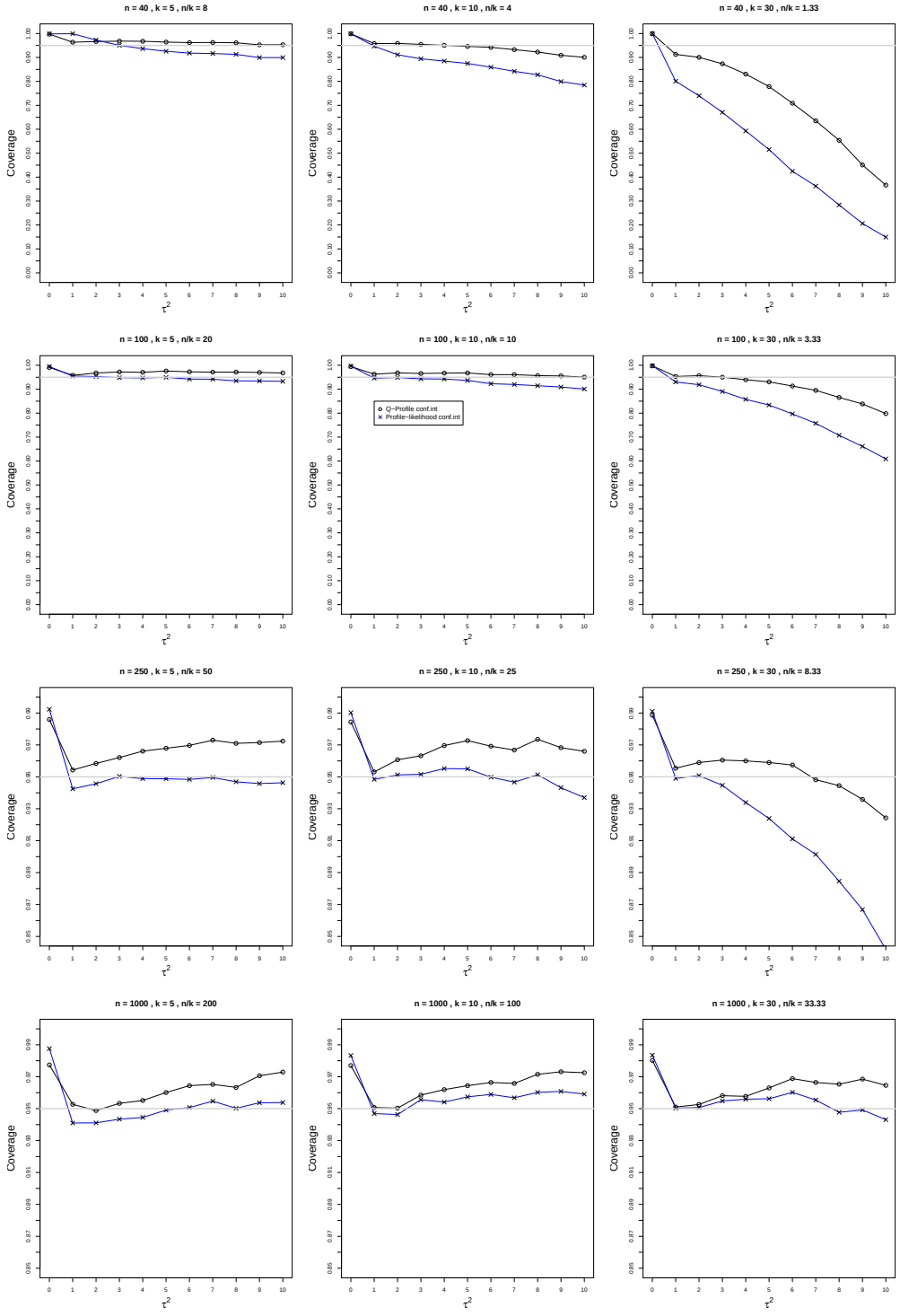


Figure C.46: Coverage at the nominal confidence level of 0.95 of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.1$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 10$. Interval estimation methods: black circles – Q-profile confidence interval for τ^2 based on χ^2 distribution dark blue crosses – Profile likelihood confidence intervals. Light grey line at 0.95.

Results of simulation study for bias and coverage of θ in case of moderate-large heterogeneity

Similarly to results of simulations with the method of simulation by Viechtbauer (2007), for moderate-large heterogeneity ($0 \leq \tau^2 \leq 10$), the bias of estimated overall log-odds ratio $\hat{\theta}_w$ was practically the same regardless of the method used. Figures C.47 show the bias of $\hat{\theta}_{MPL}$ for $\theta = 0$, 1 and 2. We can clearly see that, the bias of $\hat{\theta}_{MPL}$ reduces with increasing θ. Asymptotically, the biases of $\hat{\theta}_{MPL}$ reduces with the sample size N.

Figures C.48, C.49, C.50, show the coverages of estimated overall log-odds-ratio $\hat{\theta}$ from four methods. The four methods are Mandel-Paule estimator, Maximum-likelihood estimator, Penalized Maximum-likelihood estimator and Restricted Maximum-likelihood estimator. Overall from Figures C.48, C.49, C.50, the coverages of $\hat{\theta}_{RE}$ from all the methods do not reach the 95% significance level. Among all the methods, the Mandel-Paule method remain the best method with least biased estimate of τ^2 and better coverage than likelihood based methods. The coverage from of penalized maximum likelihood estimator $\hat{\theta}_{MPL}$ are slightly better than coverages from maximum likelihood estimator $\hat{\theta}_{ML}$. However, for $p_{2j} = 0.1$ and $\theta = 0$, they are still below 90% for $K = 5$ and $K = 10$ with $N \leq 100$. With increasing the sample size N, the coverage of penalized maximum likelihood and restricted maximum likelihood improve from 80-85% to around 92%. Coverage of penalized maximum likelihood and restricted maximum likelihood are practically identical. The coverages from all methods deteriorate when $K = 30$ and $N \leq 250$. This is due to large negative bias of between study variance when $K = 30$ and $N \leq 250$. Thus, the performance of methods for small-moderate heterogeneity is not the same as for moderate-large heterogeneity.

For $p_{2j} = 0.1$ with $\theta = 1$ and $\theta = 2$, all the coverage go up with increasing K and N (see Figures C.49 and C.50). There is no dramatic deterioratio for coverages when $p_{2j} = 0.1$ with $\theta = 1$ and $\theta = 2$ in comparison to the case when $p_{2j} = 0.1$ and $\theta = 0$. Thus, when heterogeneity is large and data is sparse in both arms ($p_{2j} = 0.1$ and $\theta = 0$), the confidence intervals tend to become shorter with increasing K from $K = 5$ to $K = 30$ for $N \leq 250$. This is probably because the bias of τ^2 increases with increase in K from $K = 5$ to $K = 30$. For example the bias of $\hat{\tau}^2_{MPL}$ for $K = 5$ and $\theta = 0$ varies between $3.27 - 66.612\%$. Whereas, the bias of $\hat{\tau}^2_{MPL}$ for $K = 30$ and $\theta = 0$ varies between $5.95 - 83.25\%$. Thus we get worse coverages for $K = 30$ and then for $K = 5$ when $N \leq 250$. Overall the bias of τ^2 depend on the

combination of K and N. The main conclusion is the performance of method of simulation similarly to Viechtbauer (2007) and similarly to Kosmidis et al. (2017) are not the same. This might be because of different data generation structure. In method of simulation similarly to Viechtbauer (2007), we generate the effect measures directly from the normal distribution. Whereas, in the method of simulation similarly to Kosmidis et al. (2017), we generate the data from logistic regression and obtain the estimates of effect measure as the coefficients of covariates. The generation of sample sizes are also different in two methods of simulation. In the method of simulation similarly to Viechtbauer (2007), the sample sizes are fixed. Whereas in the method of simulation similarly to Kosmidis et al. (2017), the sample sizes are variable. One more difference these two methods of simulation is that in the method of simulation by Kosmidis et al. (2017) we have two random effects that are added to coefficients for slope and intercept.

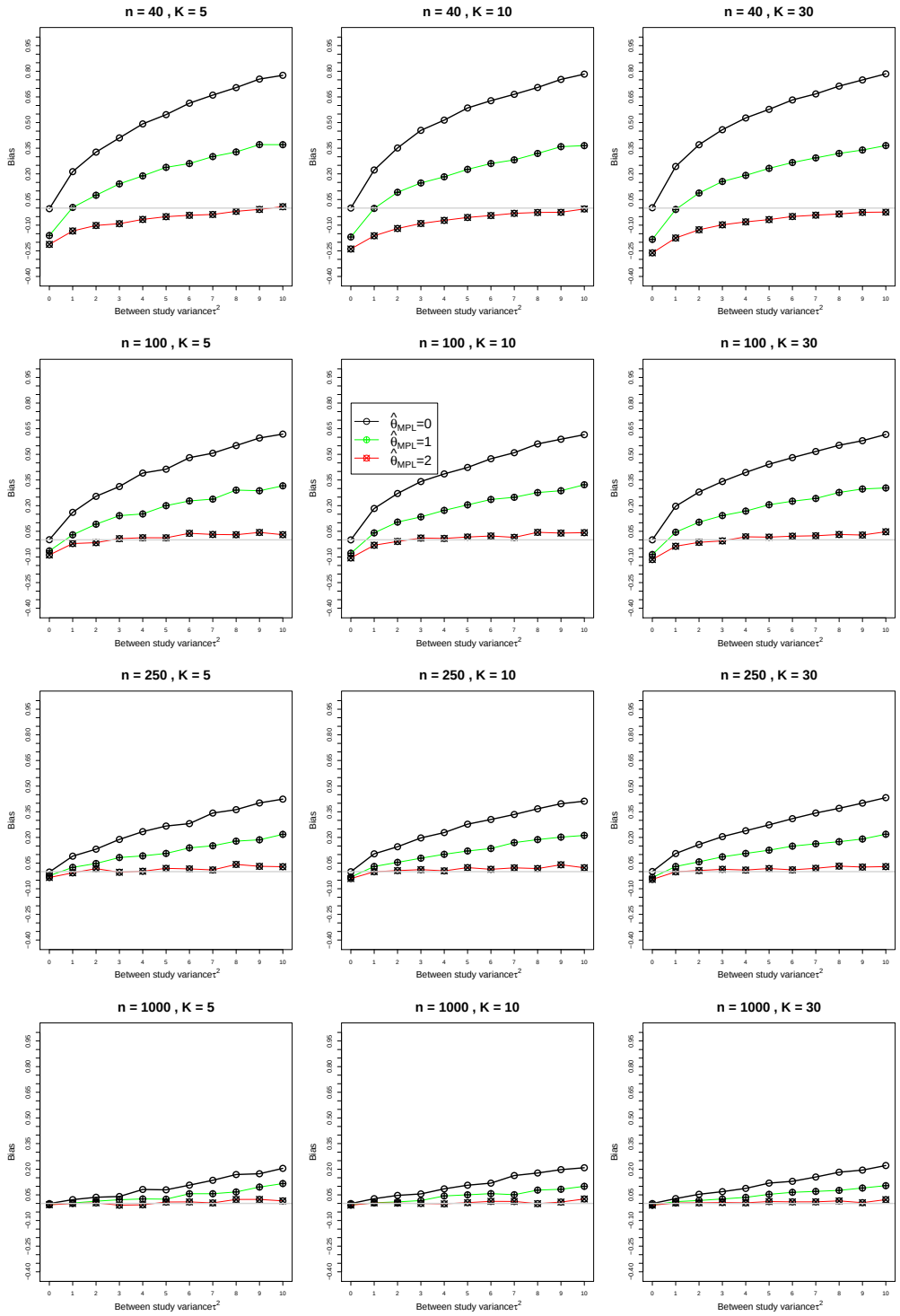


Figure C.47: Bias of overall odds ratio θ_{IV} obtained from K studies by the inverse-variance method with the moment estimator $\hat{\tau}^2_{MPL}$ in the weights, for $p_{2j} = 0.1$, and $0 \leq \tau^2 \leq 10$. The biases are given for $\theta = 0$ (circles), $\theta = 1$ (circle plus), and $\theta = 2$ (circle cross). Light grey line at 0.

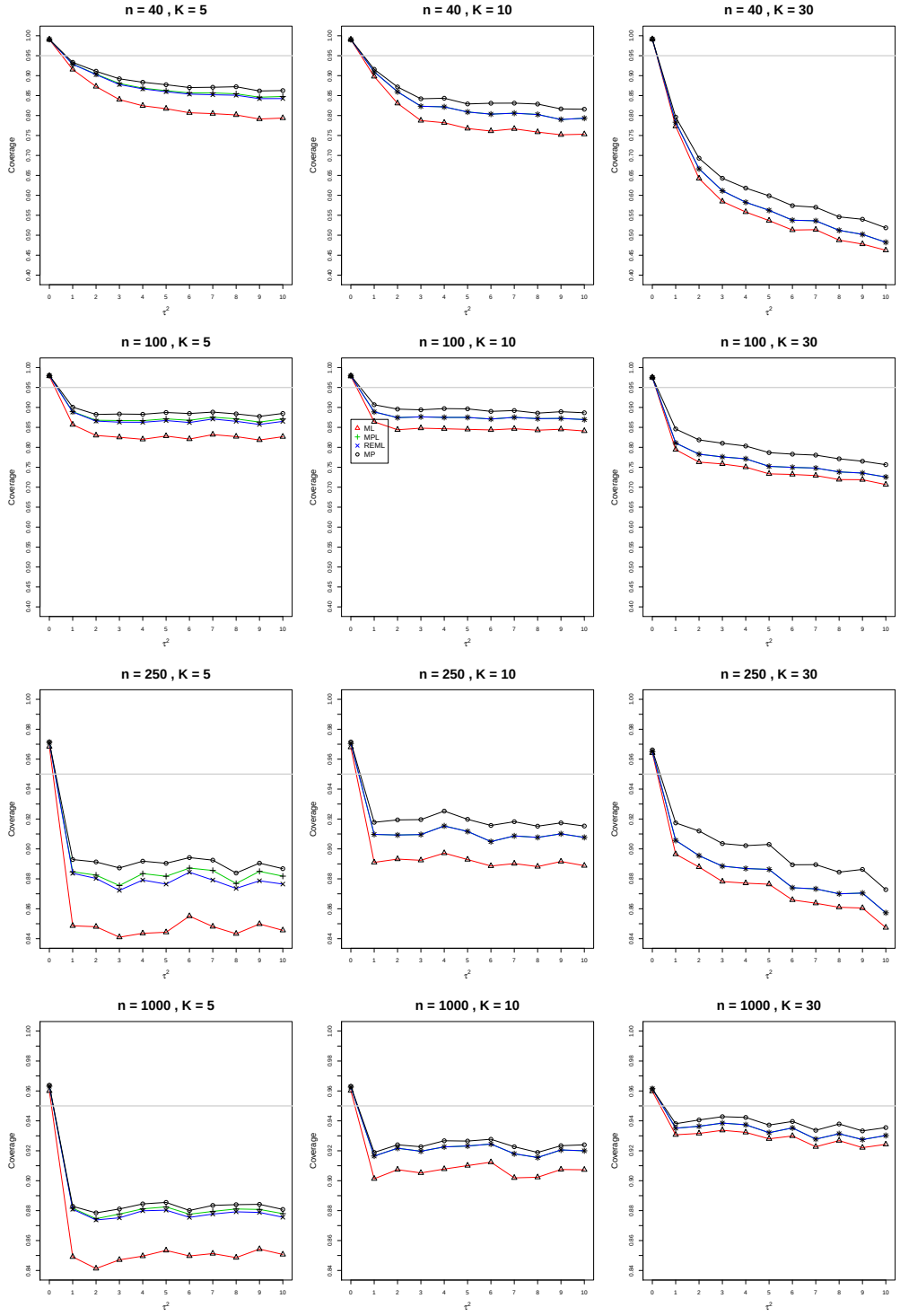


Figure C.48: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_w$ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.1$, $\theta_w = 0$ and $0 \leq \tau^2 \leq 10$. The inverse-variance weights use the following estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

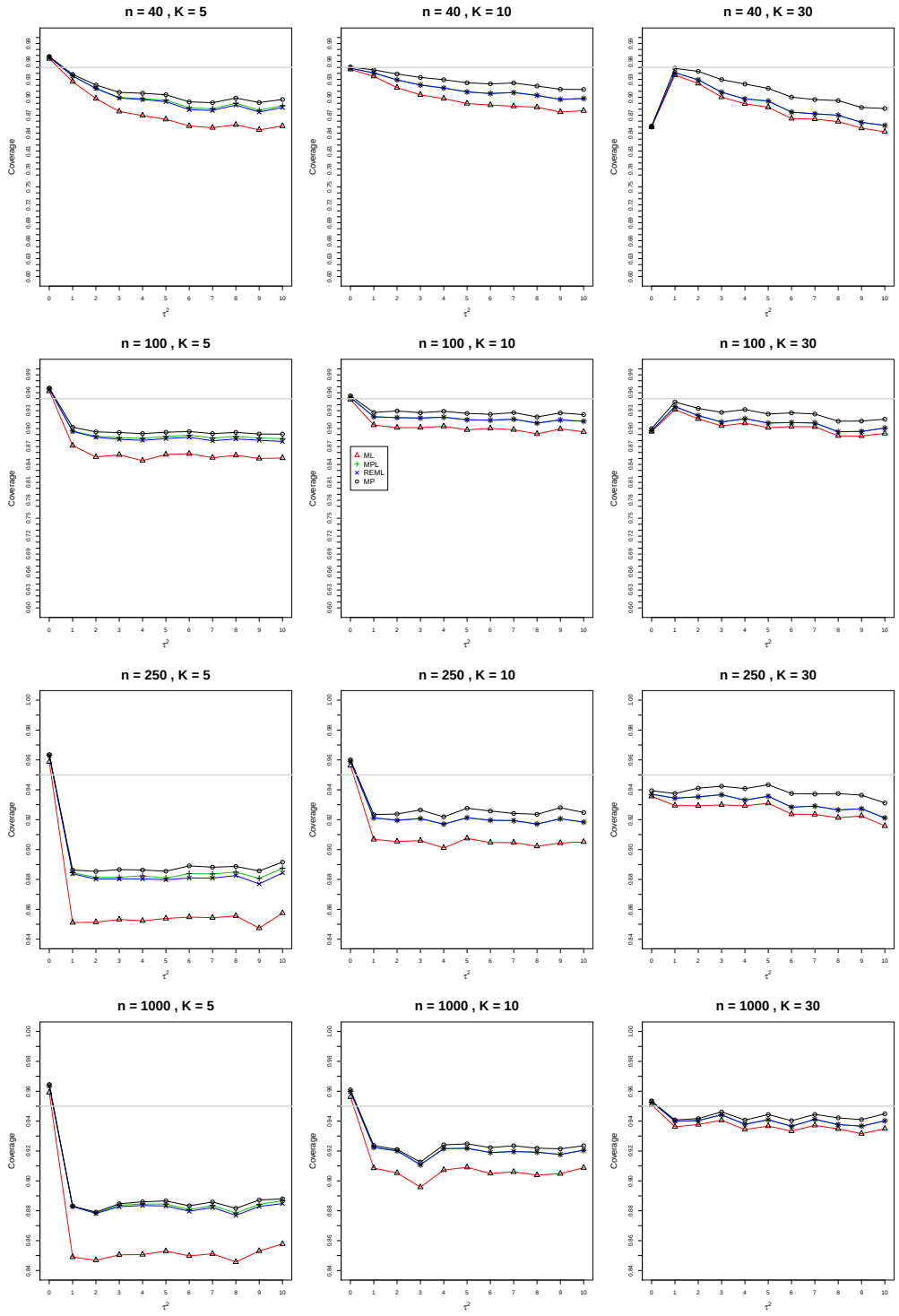


Figure C.49: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_w$ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.1$, $\theta_w = 1$ and $0 \leq \tau^2 \leq 10$. The inverse-variance weights use the following estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

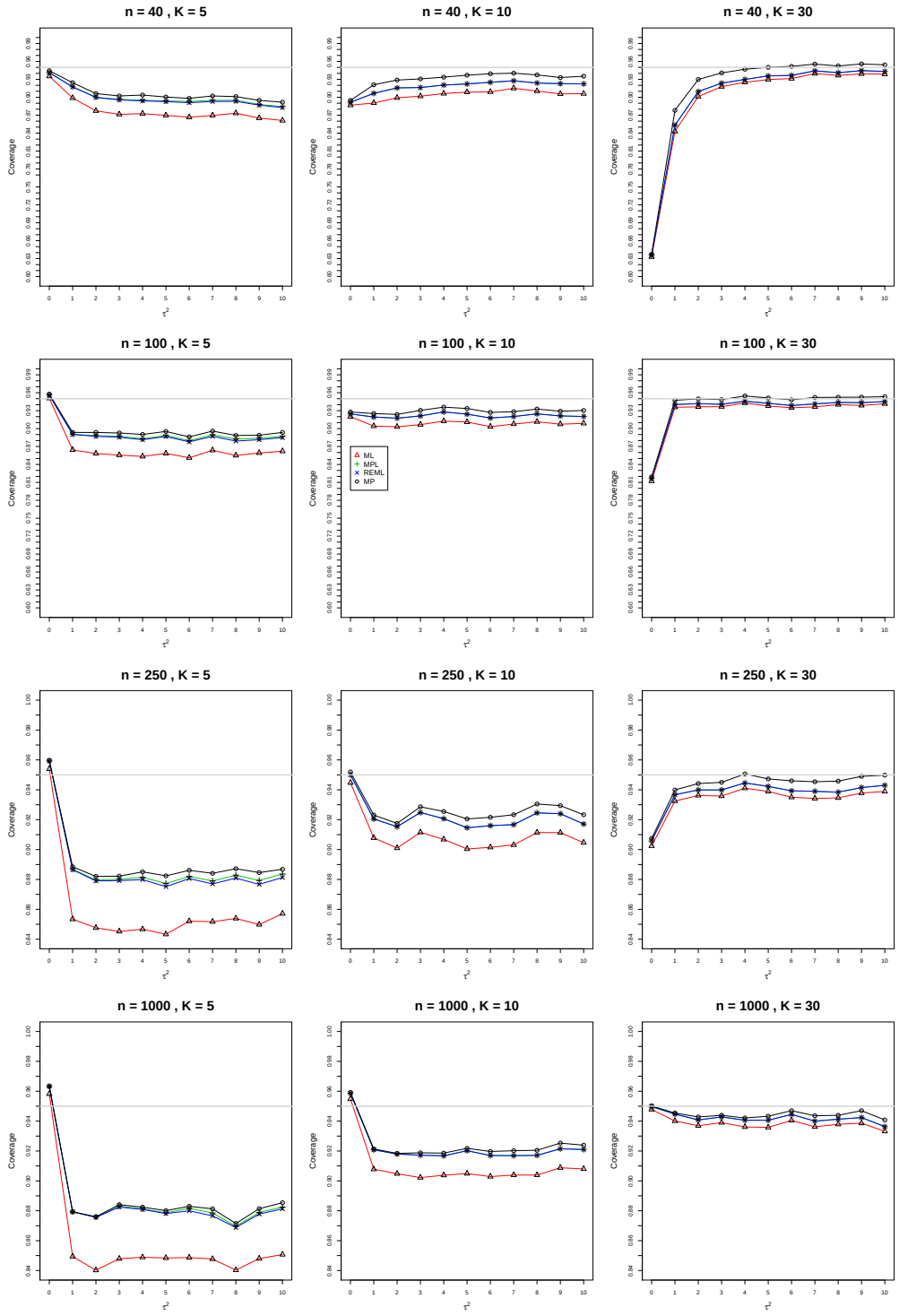


Figure C.50: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio $\hat{\theta}_w$ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.1$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 10$. The inverse-variance weights use the following estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

Simulation results for bias and coverage of τ^2 for small-moderate heterogeneity when $p_{2j} = 0.2$ and $p_{2j} = 0.4$

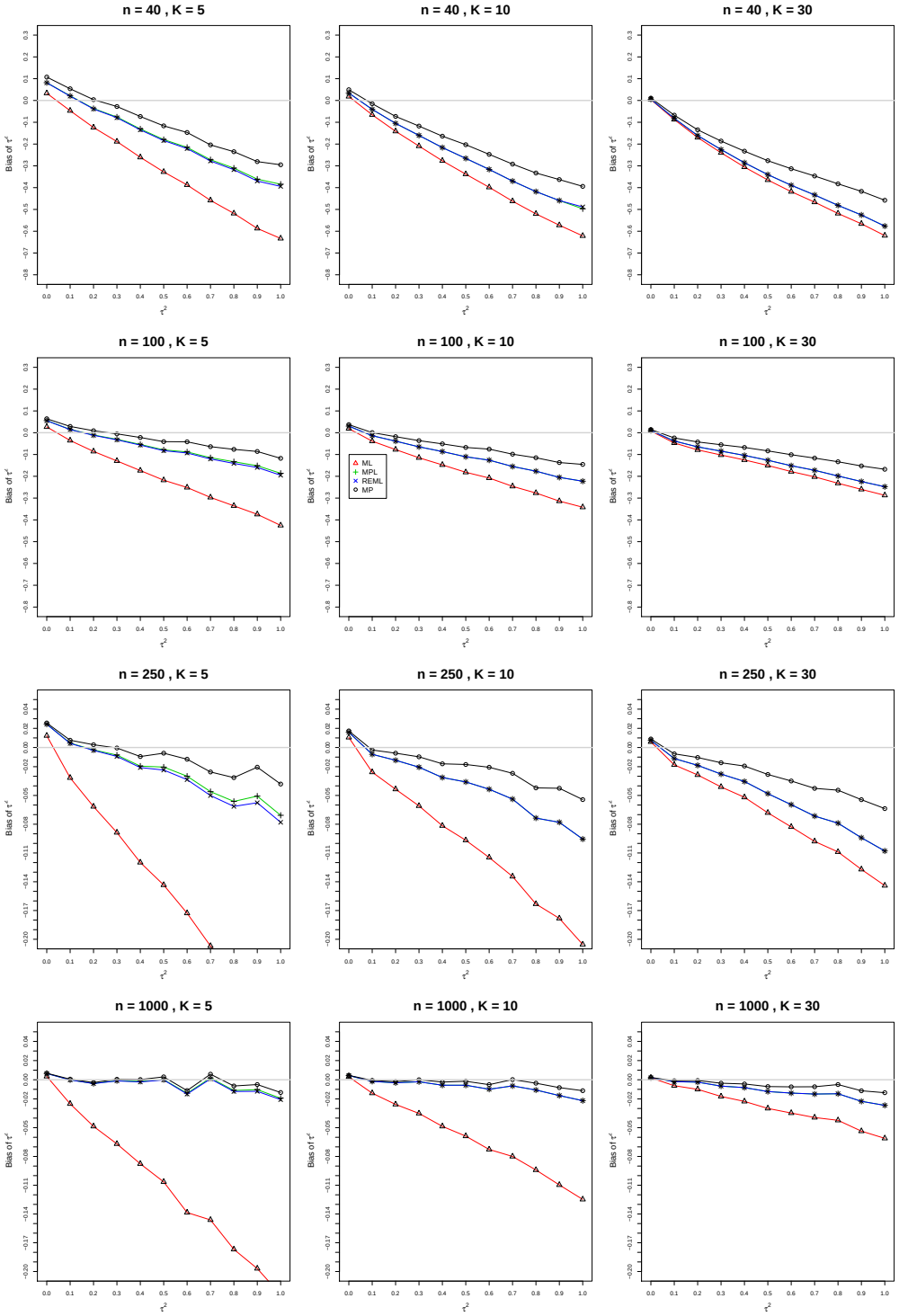


Figure C.51: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.2$, $\theta_w = 0$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

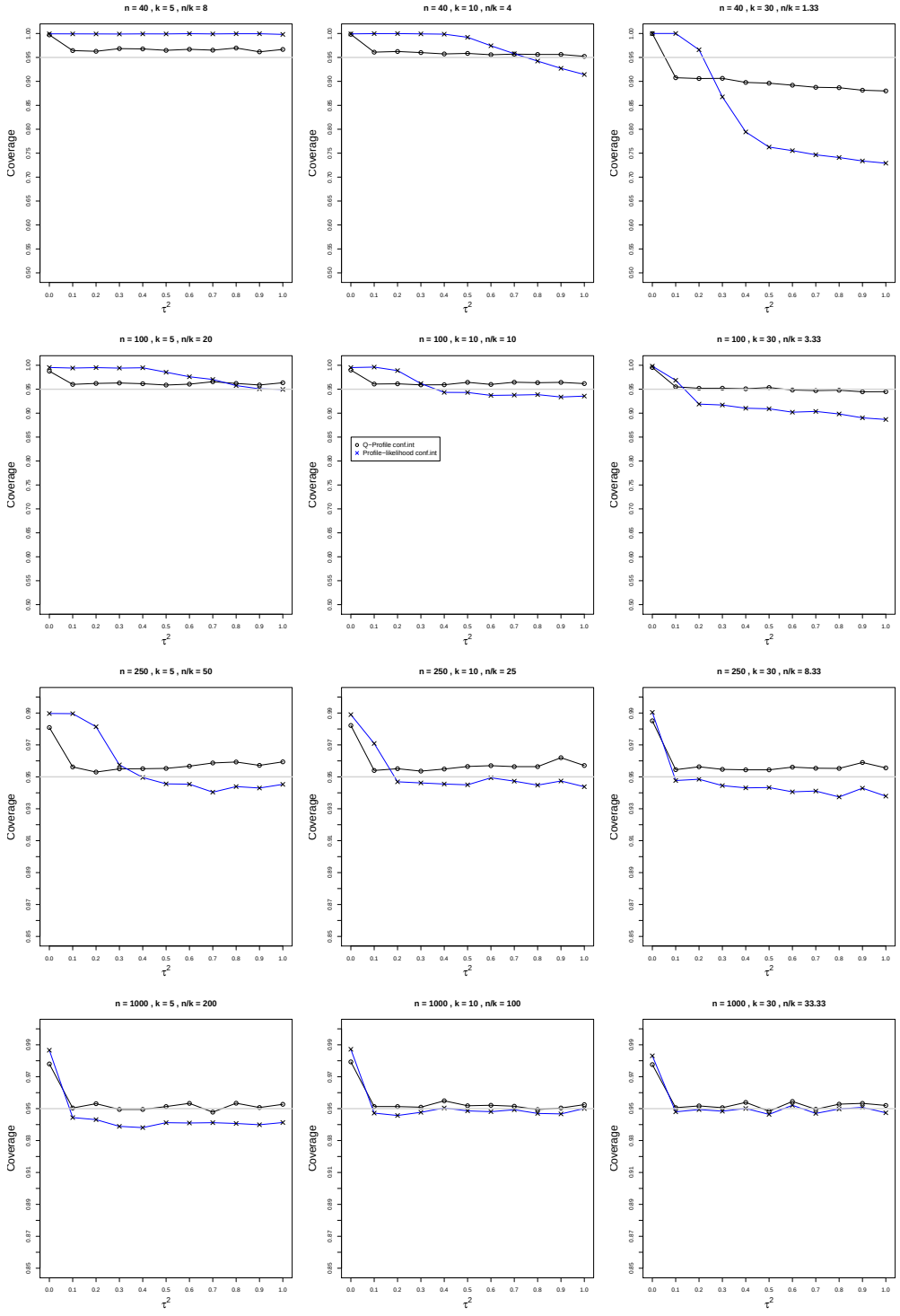


Figure C.52: Coverage at the nominal confidence level of 0.95 of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.2$, $\theta_w = 0$ and $0 \leq \tau^2 \leq 1$. Interval estimation methods: black circles – Q-profile confidence interval for τ^2 based on χ^2 distribution and dark blue crosses – Profile likelihood confidence intervals. Light grey line at 0.95.

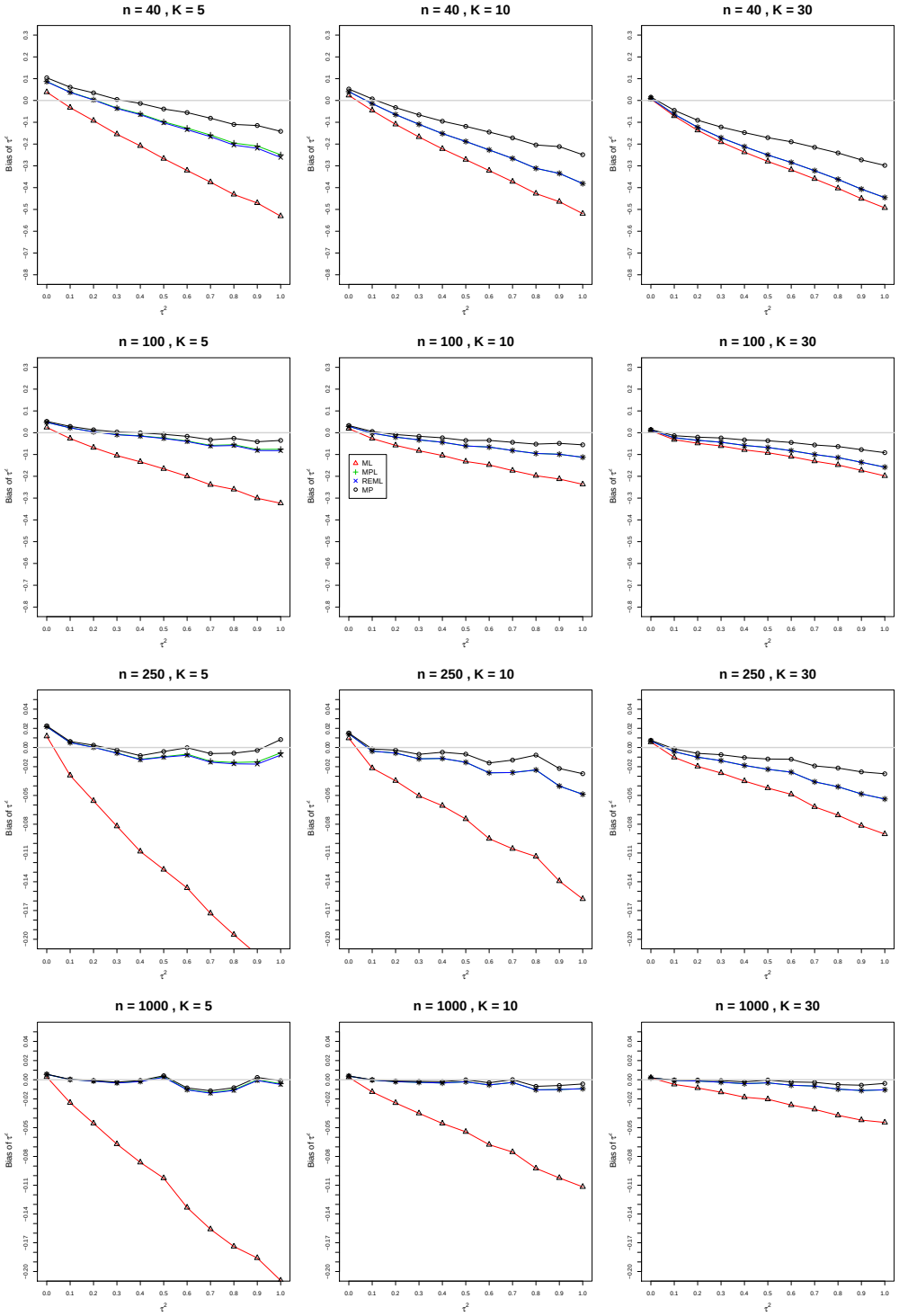


Figure C.53: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.2$, $\theta_w = 1$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

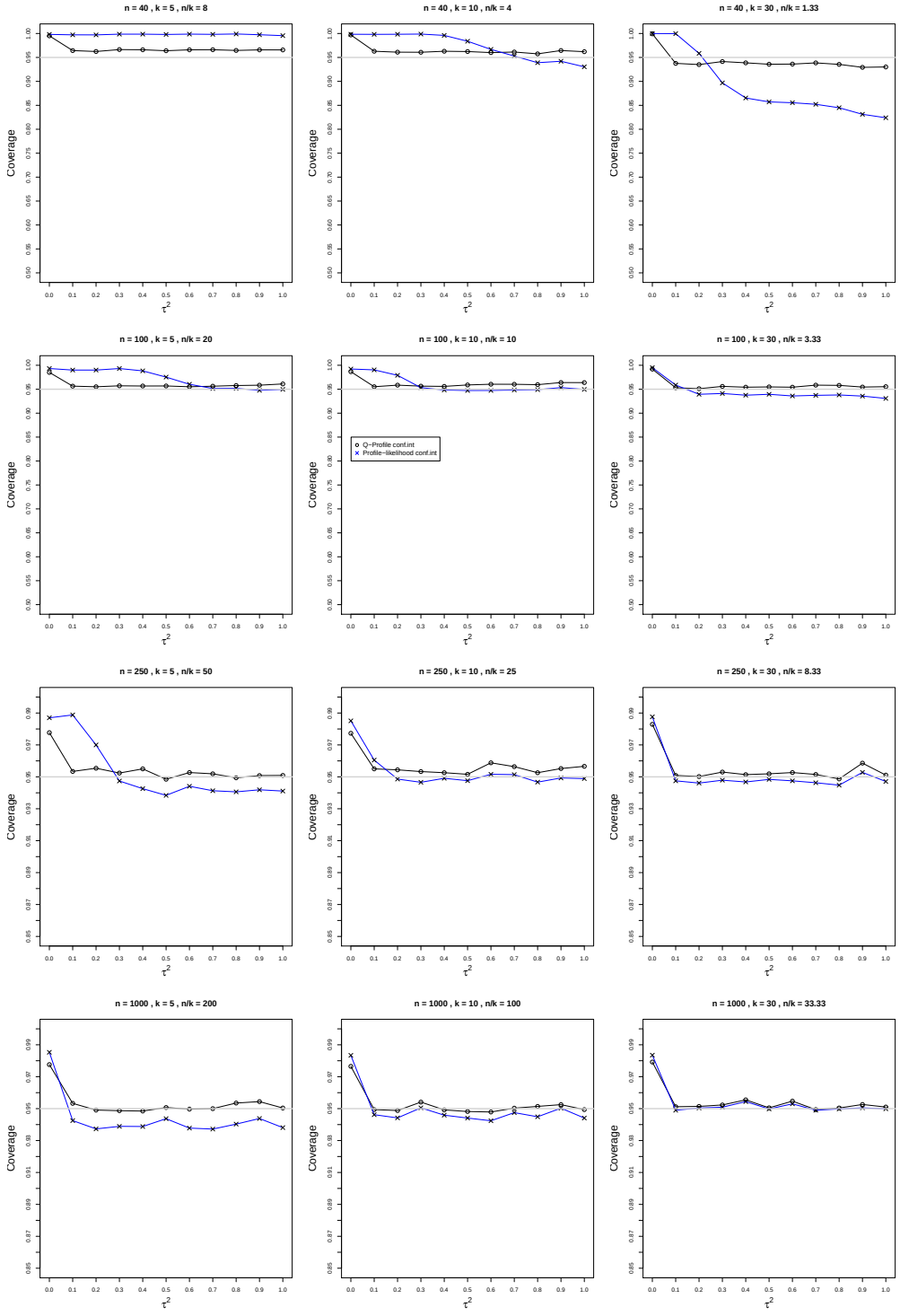


Figure C.54: Coverage at the nominal confidence level of 0.95 of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.2$, $\theta_w = 1$ and $0 \leq \tau^2 \leq 1$. Interval estimation methods: black circles – Q-profile confidence interval for τ^2 based on χ^2 distribution dark blue crosses – Profile likelihood confidence intervals. Light grey line at 0.95.

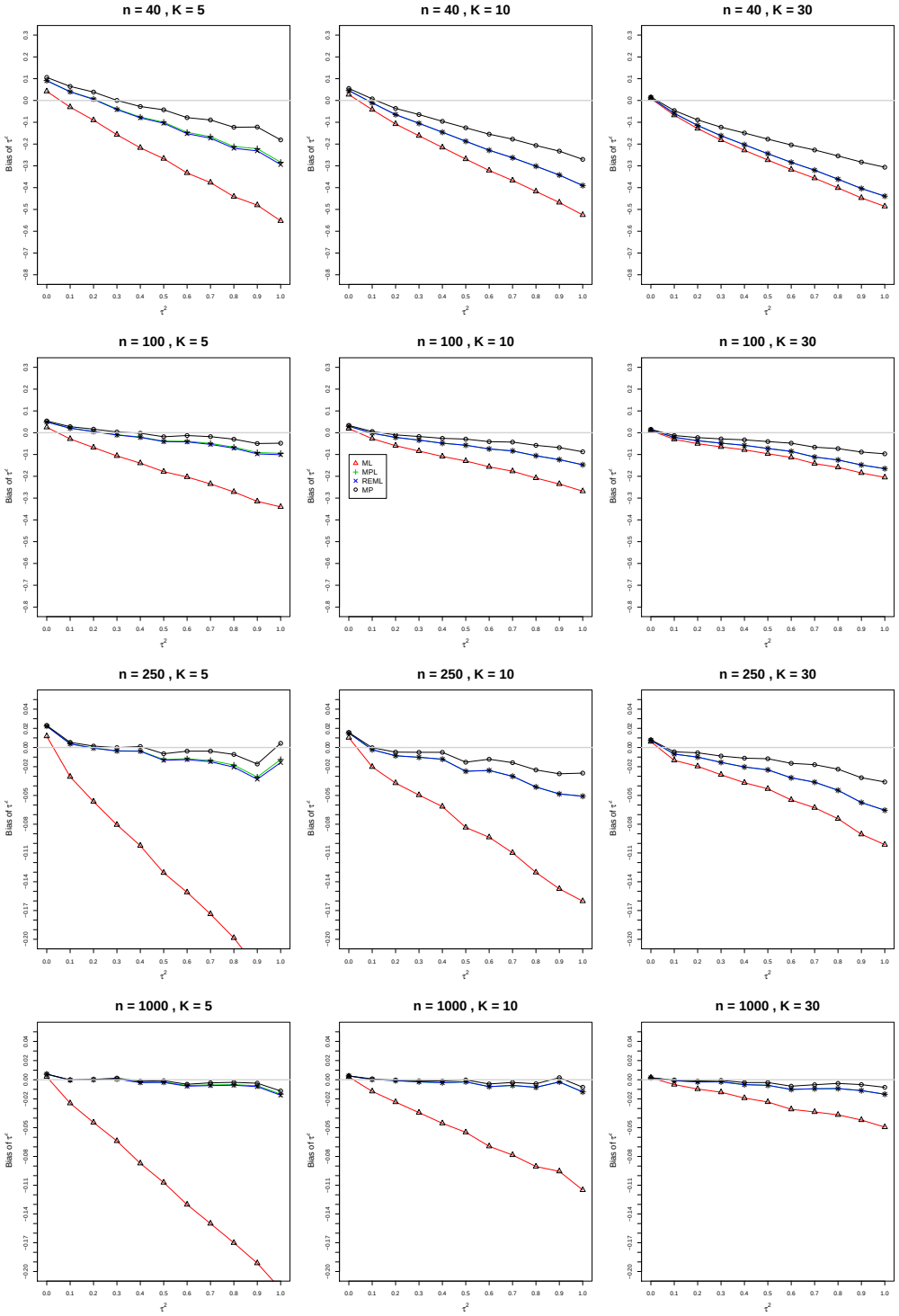


Figure C.55: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.2$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

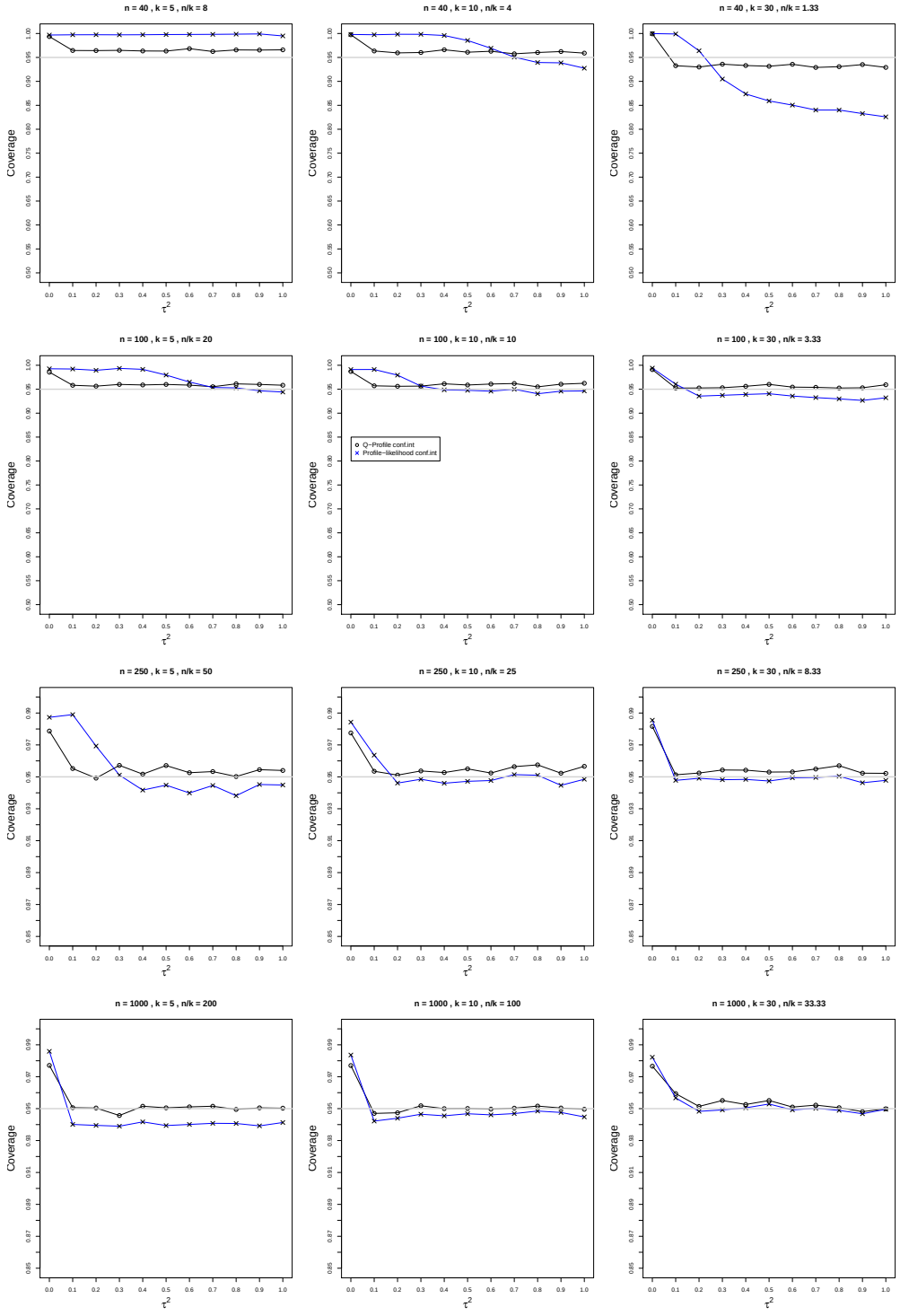


Figure C.56: Coverage at the nominal confidence level of 0.95 of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.2$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 1$. Interval estimation methods: black circles – Q-profile confidence interval for τ^2 based on χ^2 distribution and dark blue crosses – Profile likelihood confidence intervals. Light grey line at 0.95.

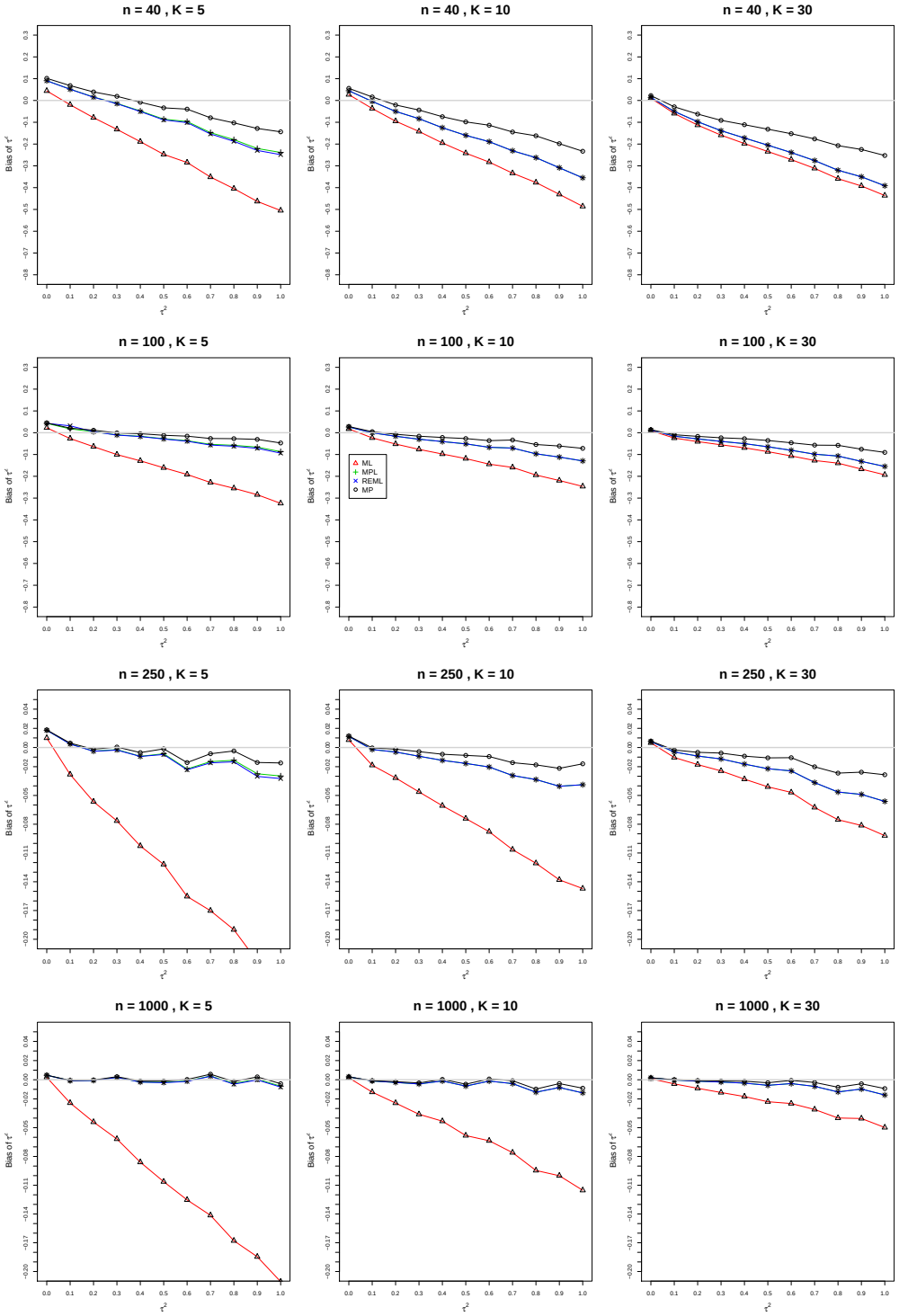


Figure C.57: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.4$, $\theta_w = 0$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

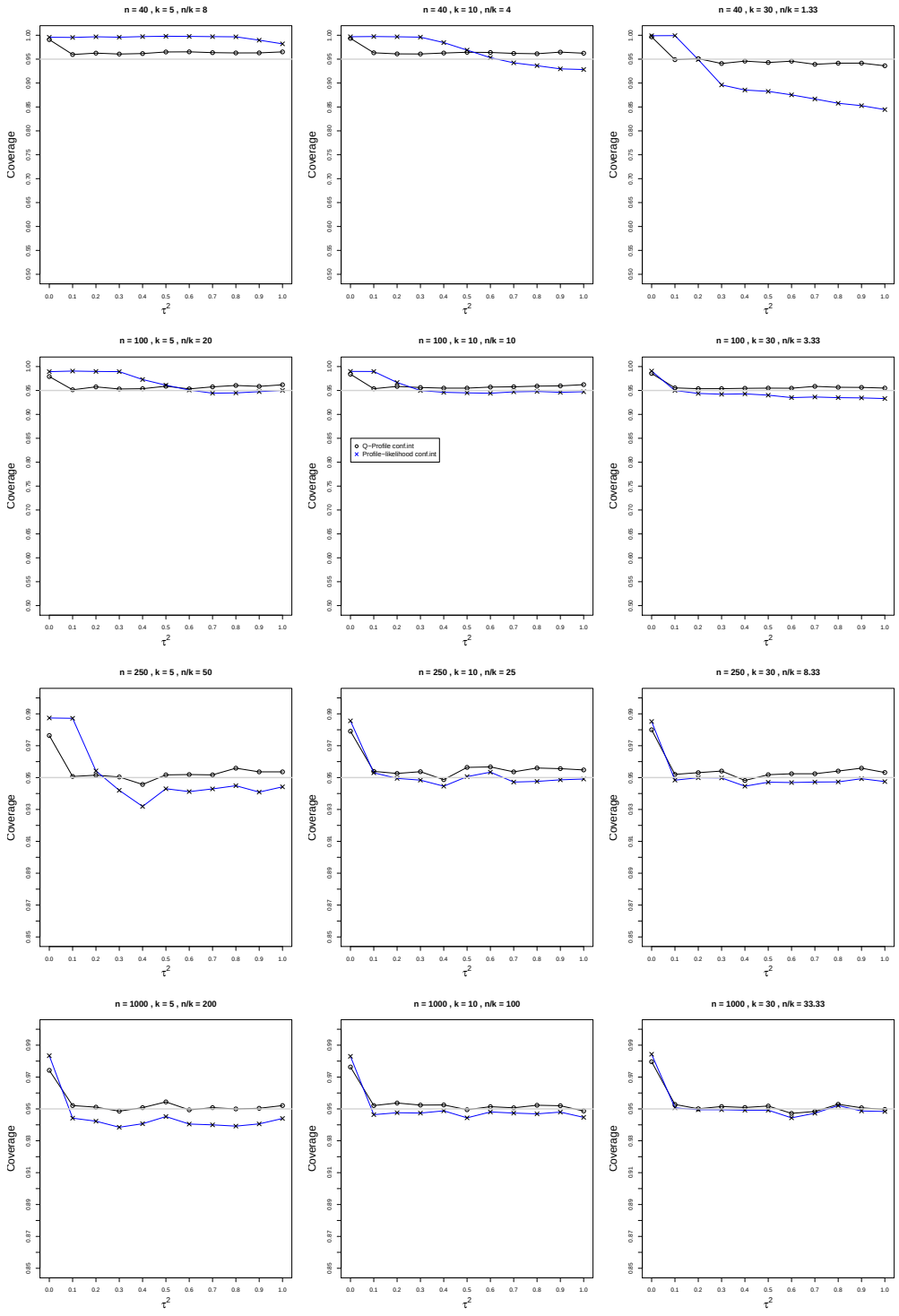


Figure C.58: Coverage at the nominal confidence level of 0.95 of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.4$, $\theta_w = 0$ and $0 \leq \tau^2 \leq 1$. Interval estimation methods: black circles – Q-profile confidence interval for τ^2 based on χ^2 distribution and dark blue crosses – Profile likelihood confidence intervals. Light grey line at 0.95.

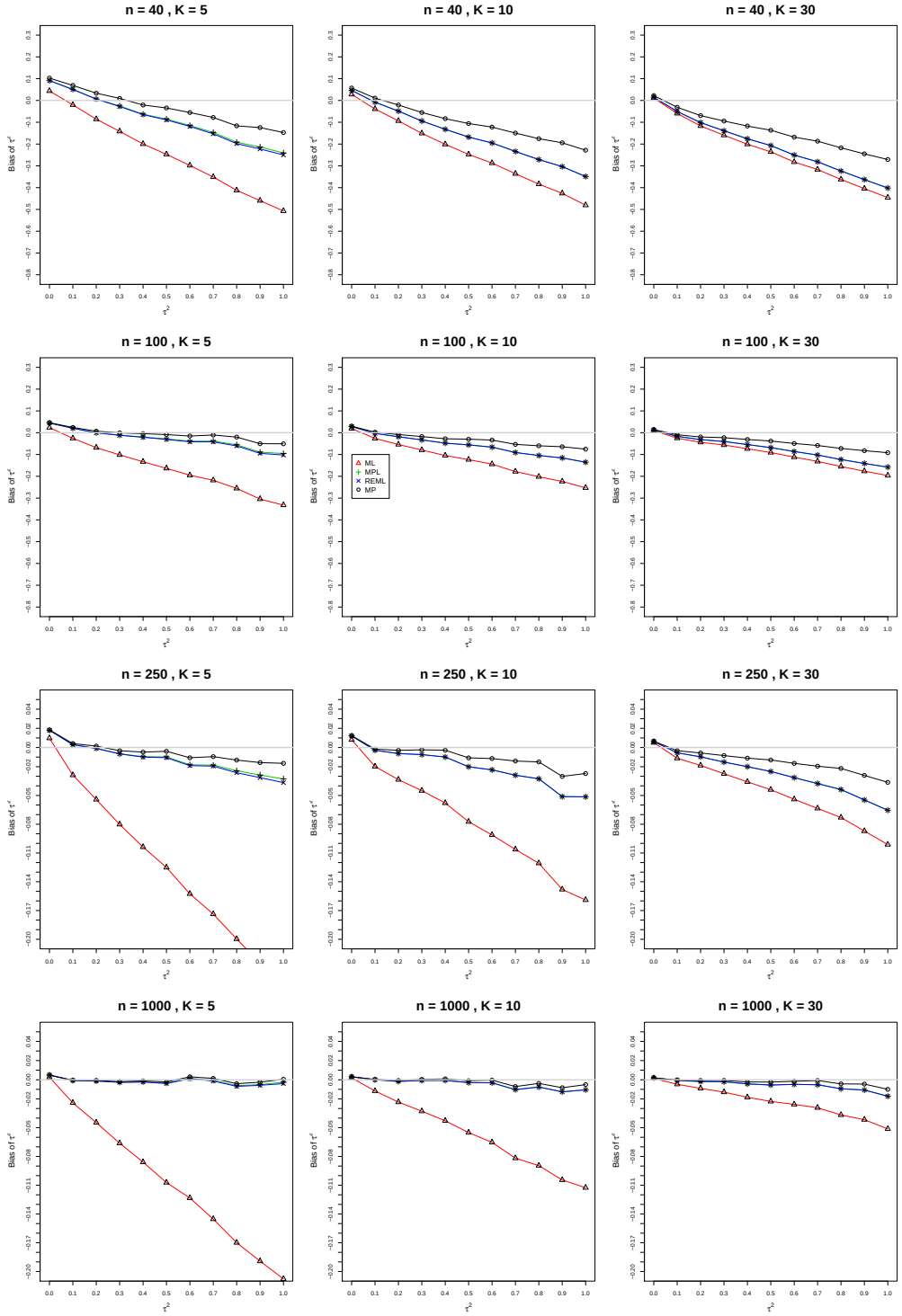


Figure C.59: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.4$, $\theta_w = 1$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

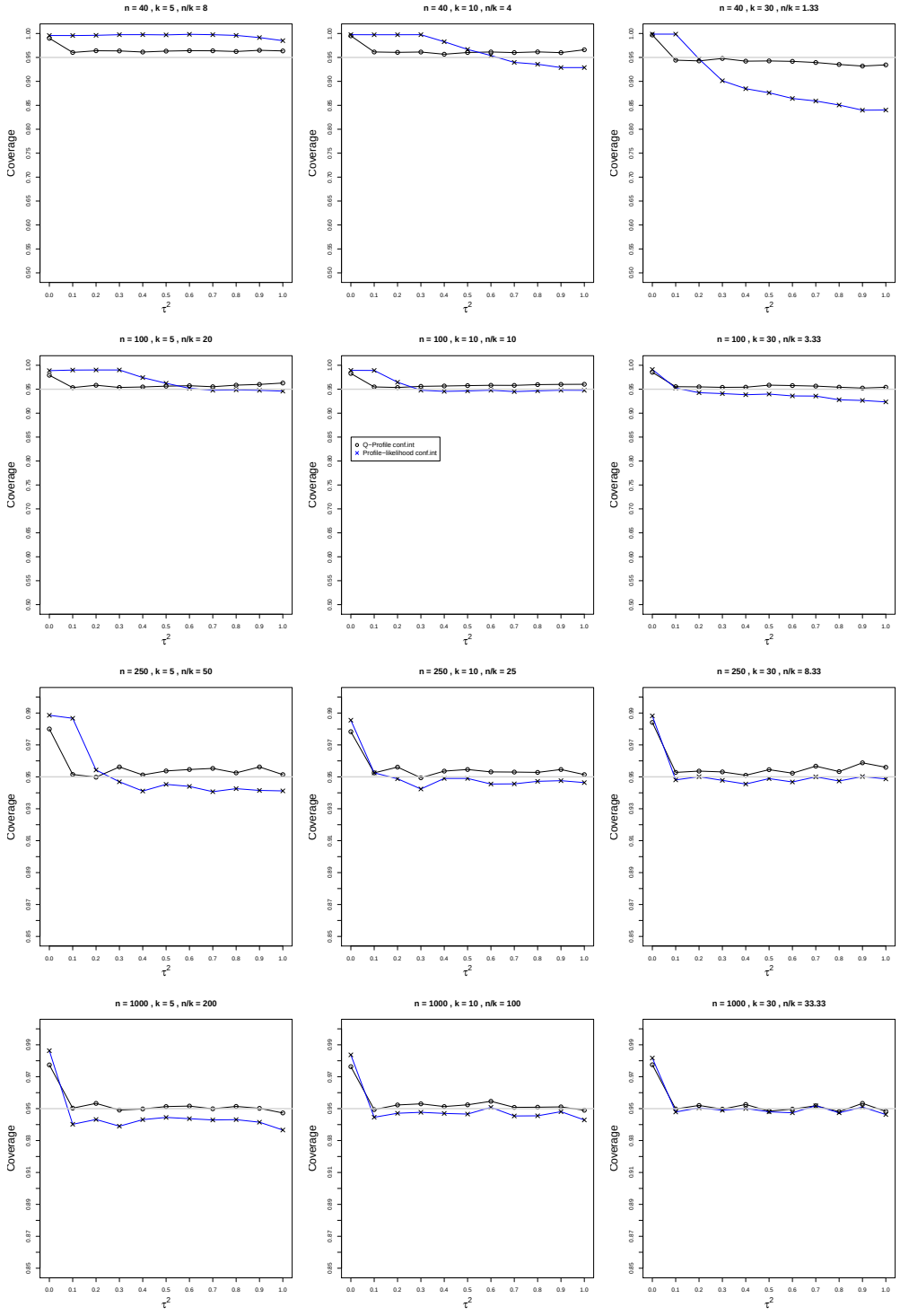


Figure C.60: Coverage at the nominal confidence level of 0.95 of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.4$, $\theta_w = 1$ and $0 \leq \tau^2 \leq 1$. Interval estimation methods: black circles – Q-profile confidence interval for τ^2 based on χ^2 distribution dark blue crosses – Profile likelihood confidence intervals. Light grey line at 0.95.

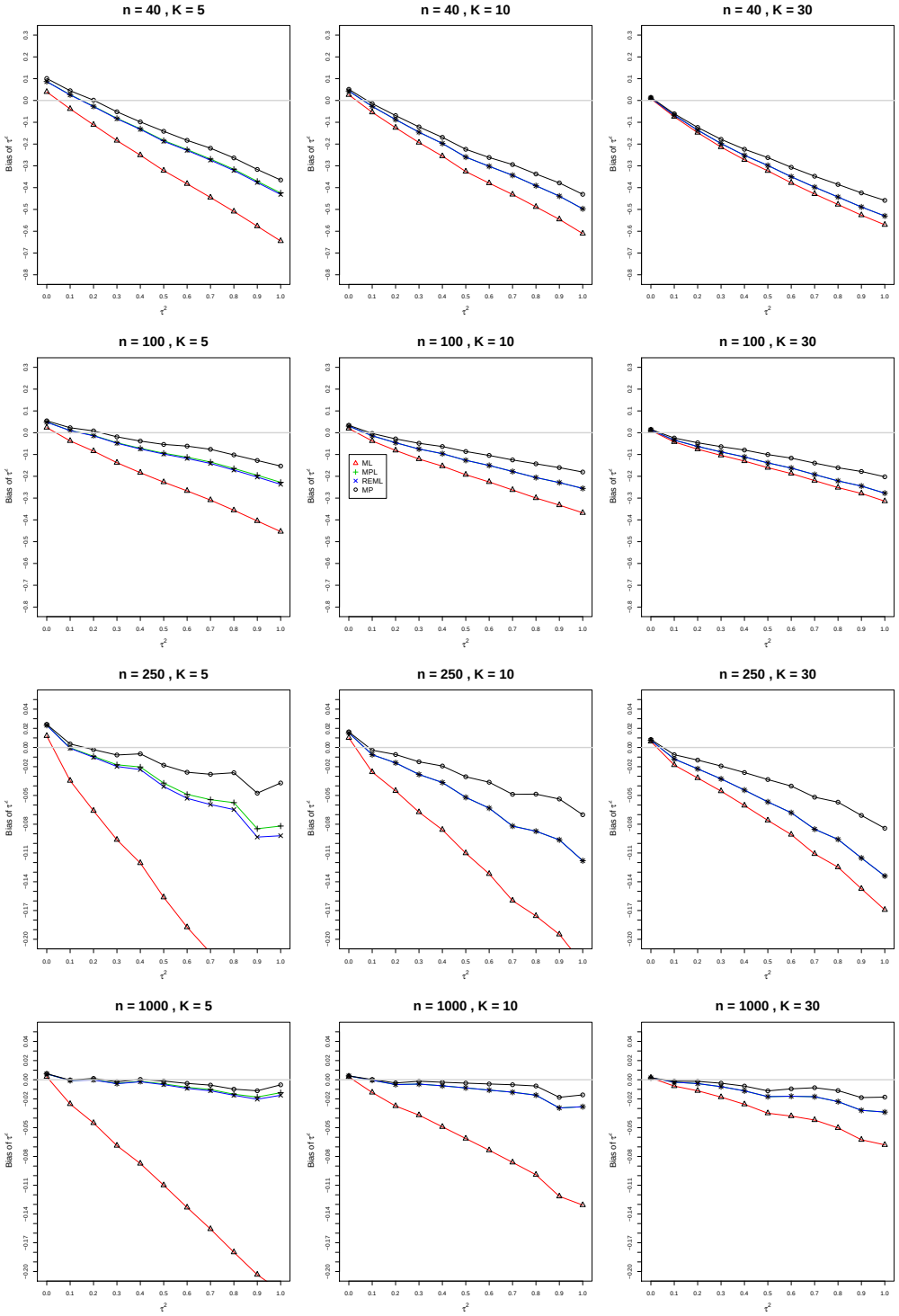


Figure C.61: Bias of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.4$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 1$. The estimators of τ^2: red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$ and black circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

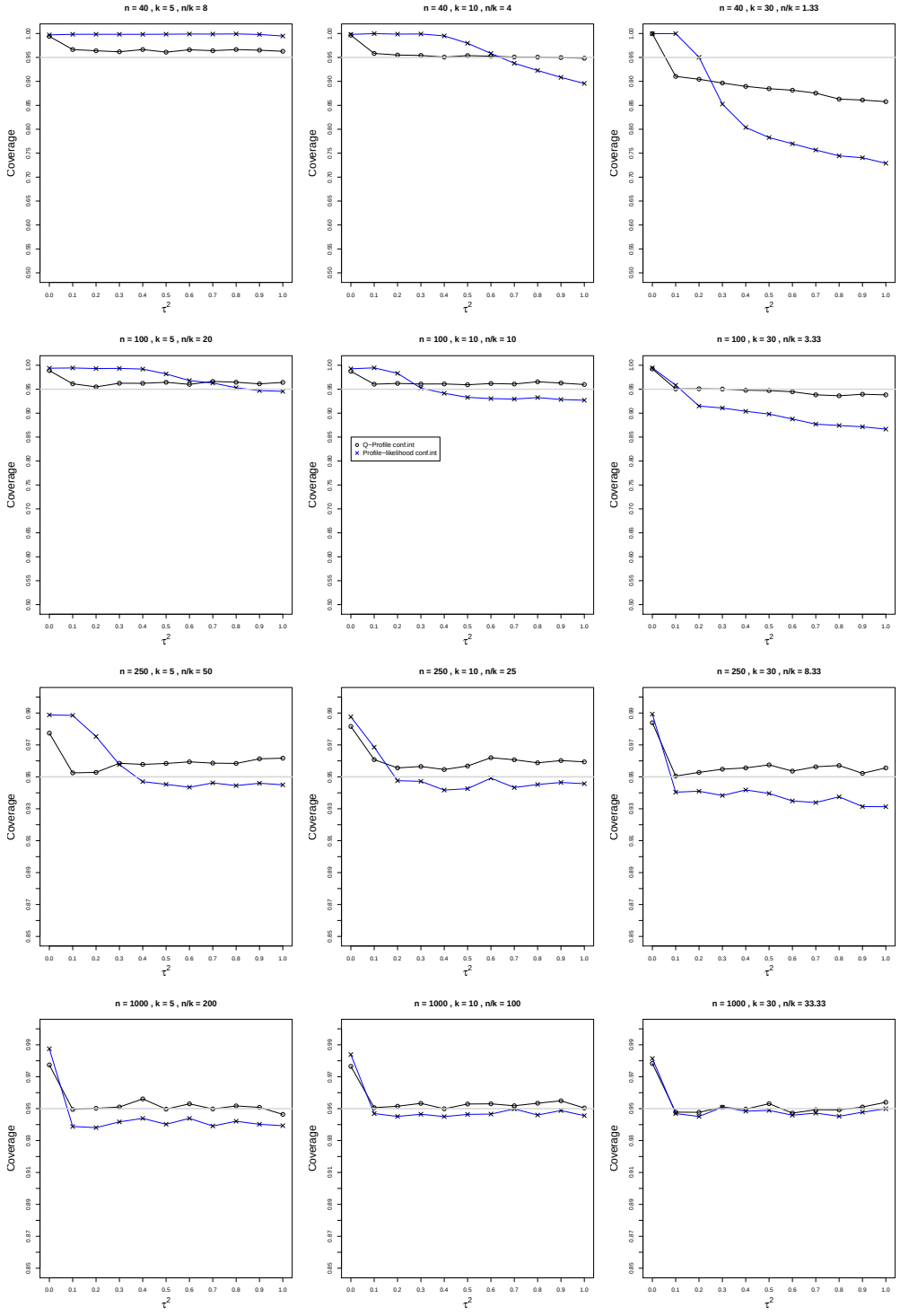


Figure C.62: Coverage at the nominal confidence level of 0.95 of the between study variance τ^2 obtained from K studies for $p_{2j} = 0.4$, $\theta_w = 2$ and $0 \leq \tau^2 \leq 1$. Interval estimation methods: black circles – Q-profile confidence interval for τ^2 based on χ^2 distribution and dark blue crosses – Profile likelihood confidence intervals. Light grey line at 0.95.

C.2 Results of simulation for estimating τ^2 from a model with a pair of beta-binomial distribution when $p_{2j} = 0.2$ and $p_{2j} = 0.4$

C.2.1 Bias and coverage in estimation of between-study variance

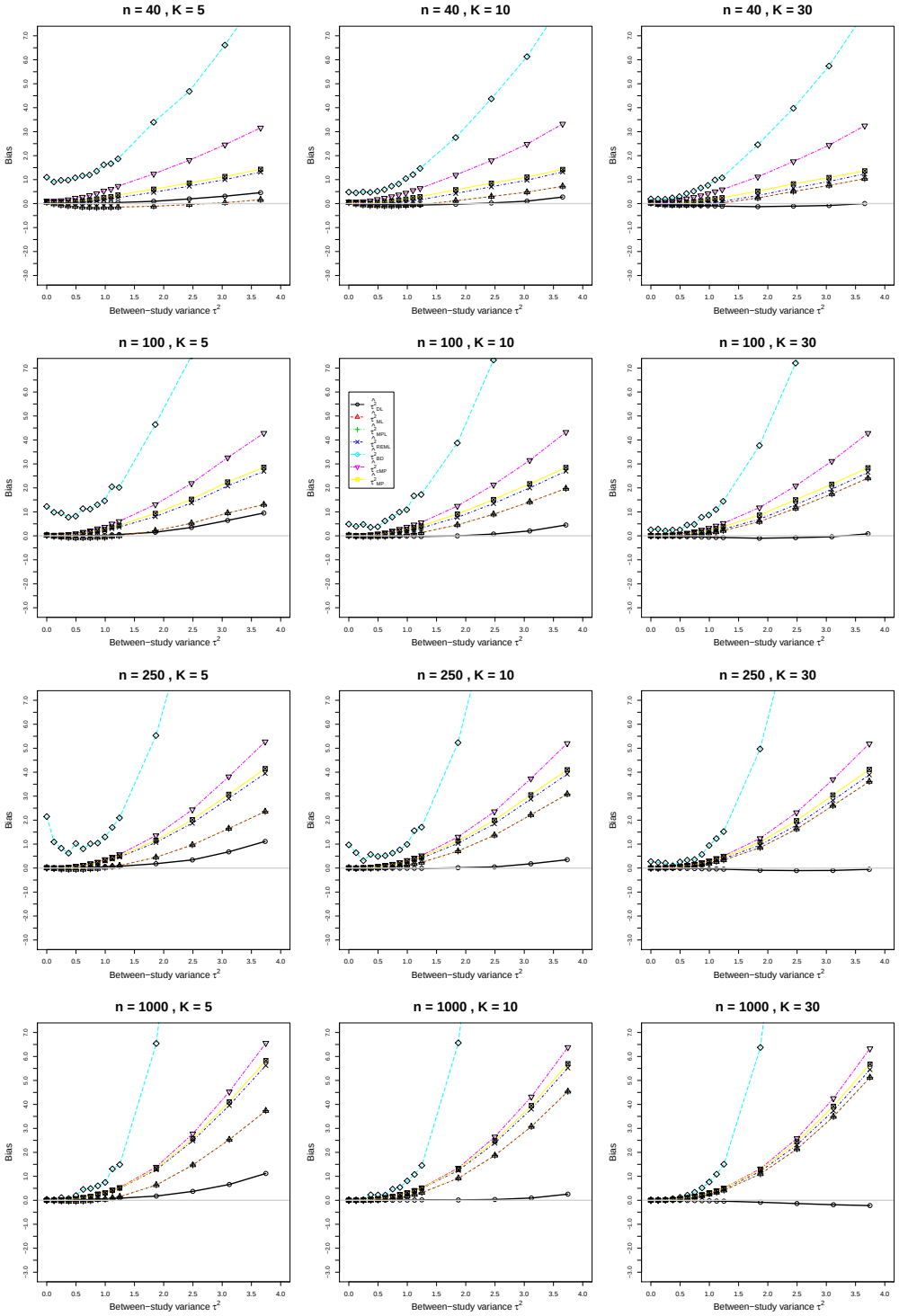


Figure C.63: Bias of the between study variance τ^2 obtained from K studies in beta-binomial model for $p_{2j} = 0.2$,$\theta = 0$ and $0 \leq \rho \leq 0.3$. The estimators of τ^2: black circle – Der-Simonian and Laird estimator $\hat{\tau}^2_{DL}$, red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, dark blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, light blue diamonds – Profiled-Breslow-Day estimator $\hat{\tau}^2_{BD}$,pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and yellow circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

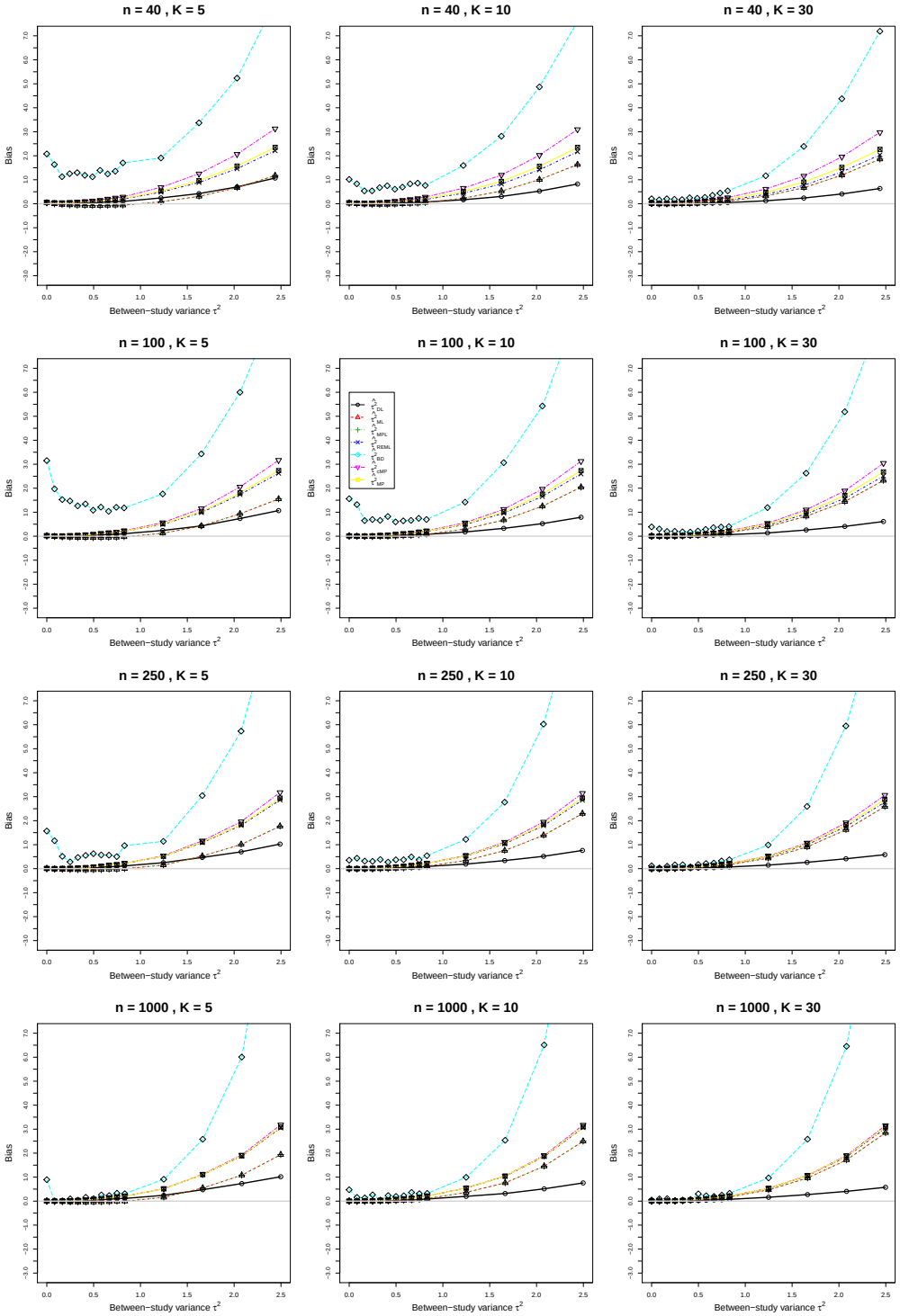


Figure C.64: Bias of the between study variance τ^2 obtained from K studies in beta-binomial model for $p_{2j} = 0.4$,$\theta = 0$ and $0 \leq \rho \leq 0.3$. The estimators of τ^2: black circle – Der-Simonian and Laird estimator $\hat{\tau}^2_{DL}$, red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, dark blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, light blue diamonds – Profiled-Breslow-Day estimator $\hat{\tau}^2_{BD}$,pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and yellow circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$. Light grey line at 0.95.

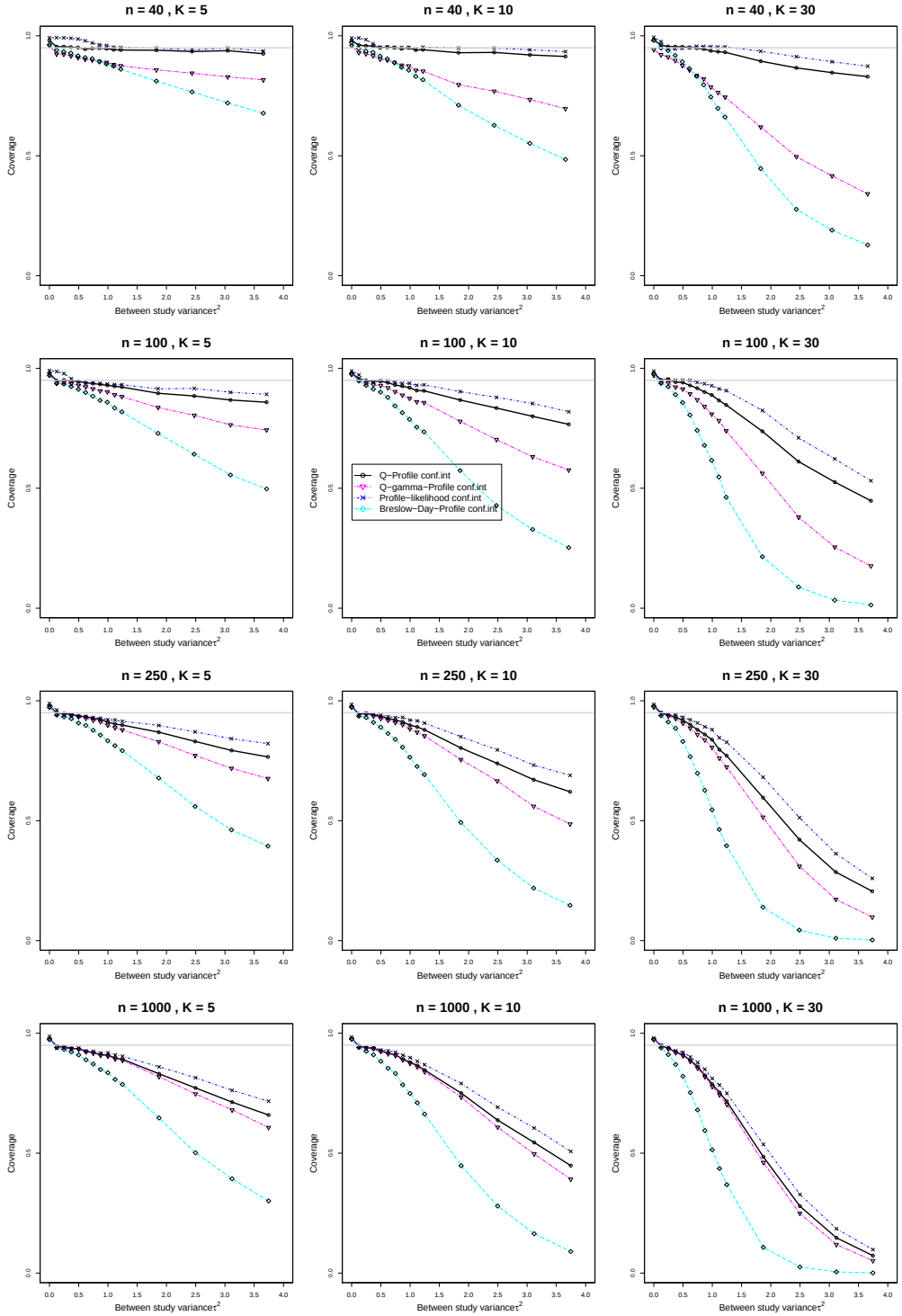


Figure C.65: Coverage at the nominal confidence level of 0.95 of the between-study variance τ^2 estimated from K studies in beta-binomial model for $p_{2j} = 0.2$, $\theta = 0$ and $0 \leq \rho \leq 0.3$. Interval estimation methods: circles – Q-profile confidence interval for τ^2 based on χ^2 distribution, inverse triangles – Q-profile confidence interval for τ^2 based on $\Gamma_{r(\tau^2),\lambda(\tau^2)}$ distribution), crosses – Profile likelihood confidence intervals, diamonds – Breslow-Day-Profile confidence intervals. Light grey line at 0.95.

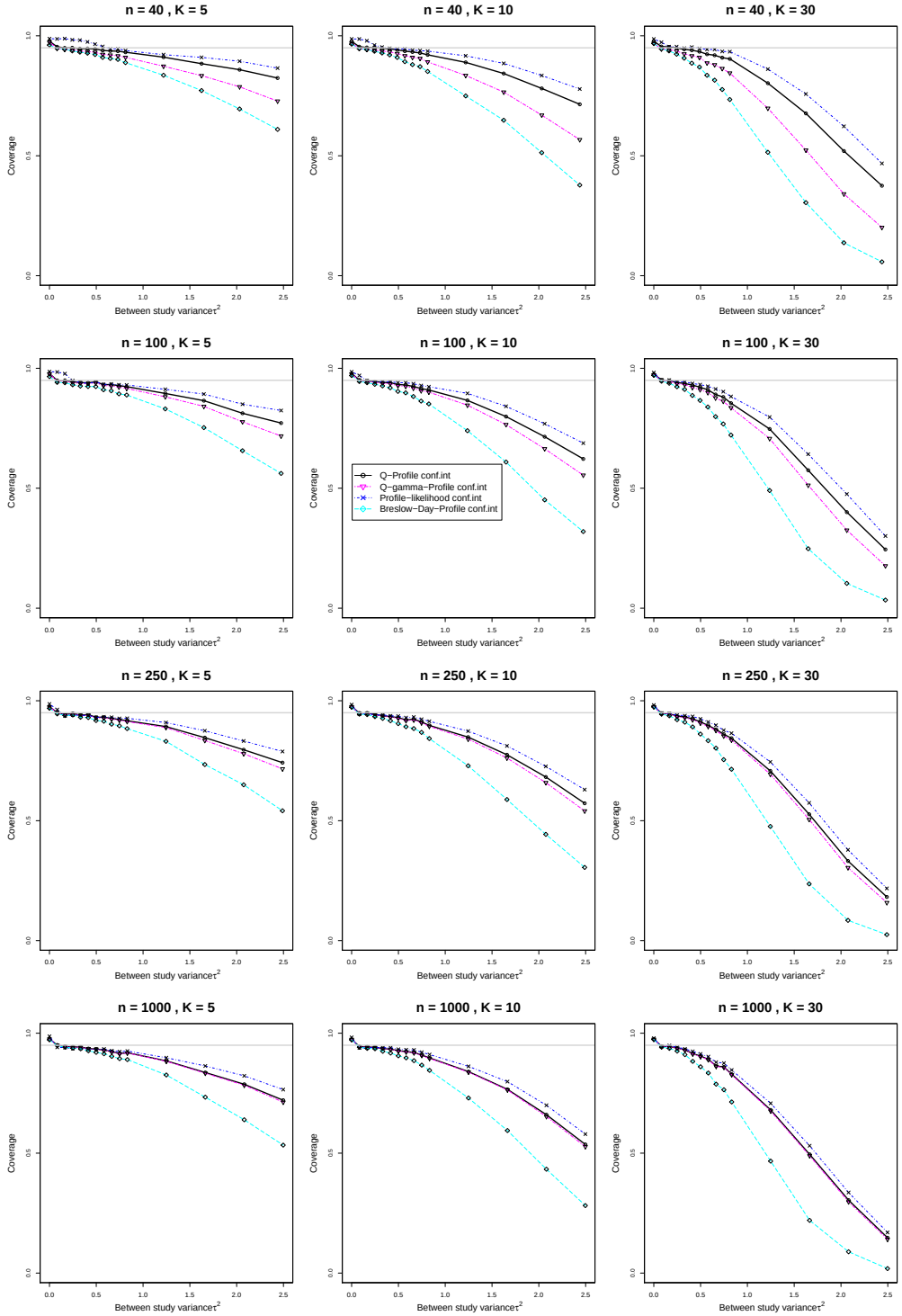


Figure C.66: Coverage at the nominal confidence level of 0.95 of the between-study variance τ^2 estimated from K studies in beta-binomial model for $p_{2j} = 0.4$, $\theta = 0$ and $0 \leq \rho \leq 0.3$. Interval estimation methods: circles – Q-profile confidence interval for τ^2 based on χ^2 distribution, inverse triangles – Q-profile confidence interval for τ^2 based on $\Gamma_{r(\tau^2),\lambda(\tau^2)}$ distribution), crosses – Profile likelihood confidence intervals, diamonds – Breslow-Day-Profile confidence intervals. Light grey line at 0.95.

C.2.2 Bias and coverage in estimation of overall effect measure

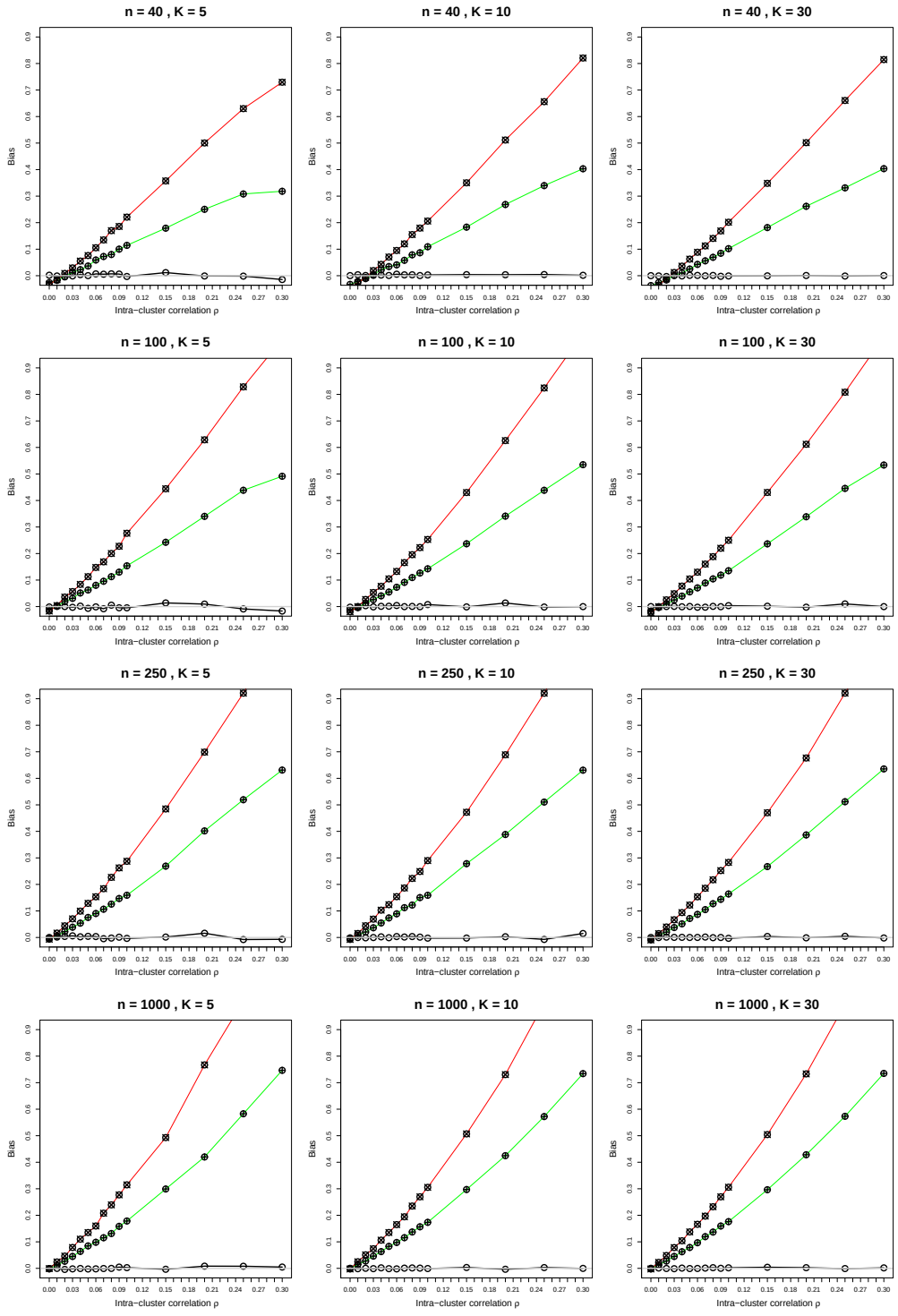


Figure C.67: Bias of overall log odds ratio θ_{IV} obtained from K studies by the inverse-variance method with the moment estimator $\hat{\rho}_{CMP}$ in the weights, for $p_{2j} = 0.2$, and $0 \leq \rho \leq 0.3$. The biases are given for $\theta = 0$ (circles), $\theta = 1$ (circle plus), and $\theta = 2$ (circle cross). Light grey line at 0.

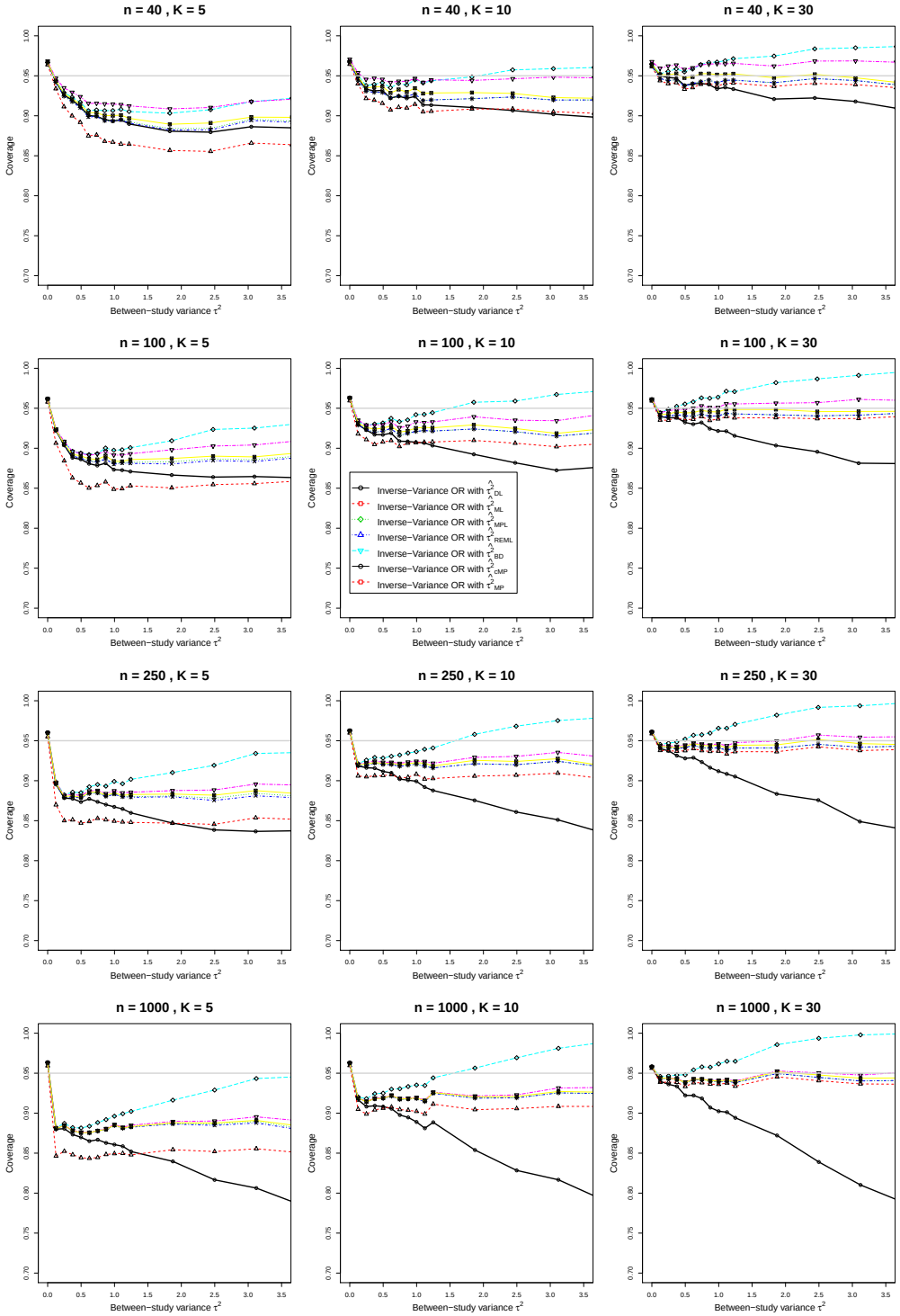


Figure C.68: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio θ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.2$, $\theta = 0$ and $0 \leq \rho \leq 0.3$. The inverse-variance weights use the following estimators of ρ: black circle – Der-Simonian and Laird estimator $\hat{\tau}^2_{DL}$, red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, dark blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, light blue diamonds – Profiled-Breslow-Day estimator $\hat{\tau}^2_{BD}$,pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and yellow circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$.Light grey line at 0.95.

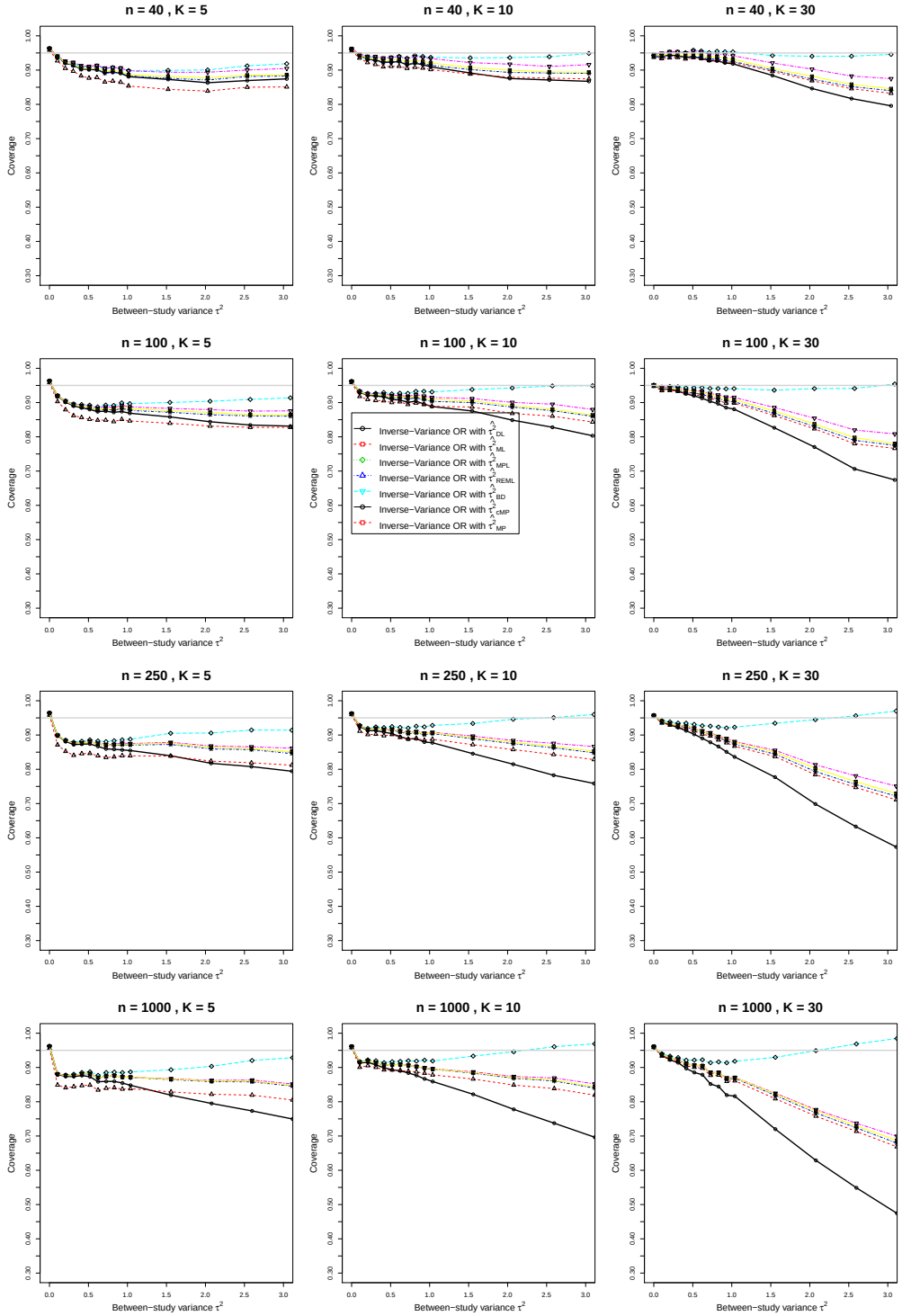


Figure C.69: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio θ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.2$, $\theta = 1$ and $0 \leq \rho \leq 0.3$. The inverse-variance weights use the following estimators of ρ: black circle – Der-Simonian and Laird estimator $\hat{\tau}^2_{DL}$, red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, dark blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, light blue diamonds – Profiled-Breslow-Day estimator $\hat{\tau}^2_{BD}$,pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and yellow circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$.Light grey line at 0.95.

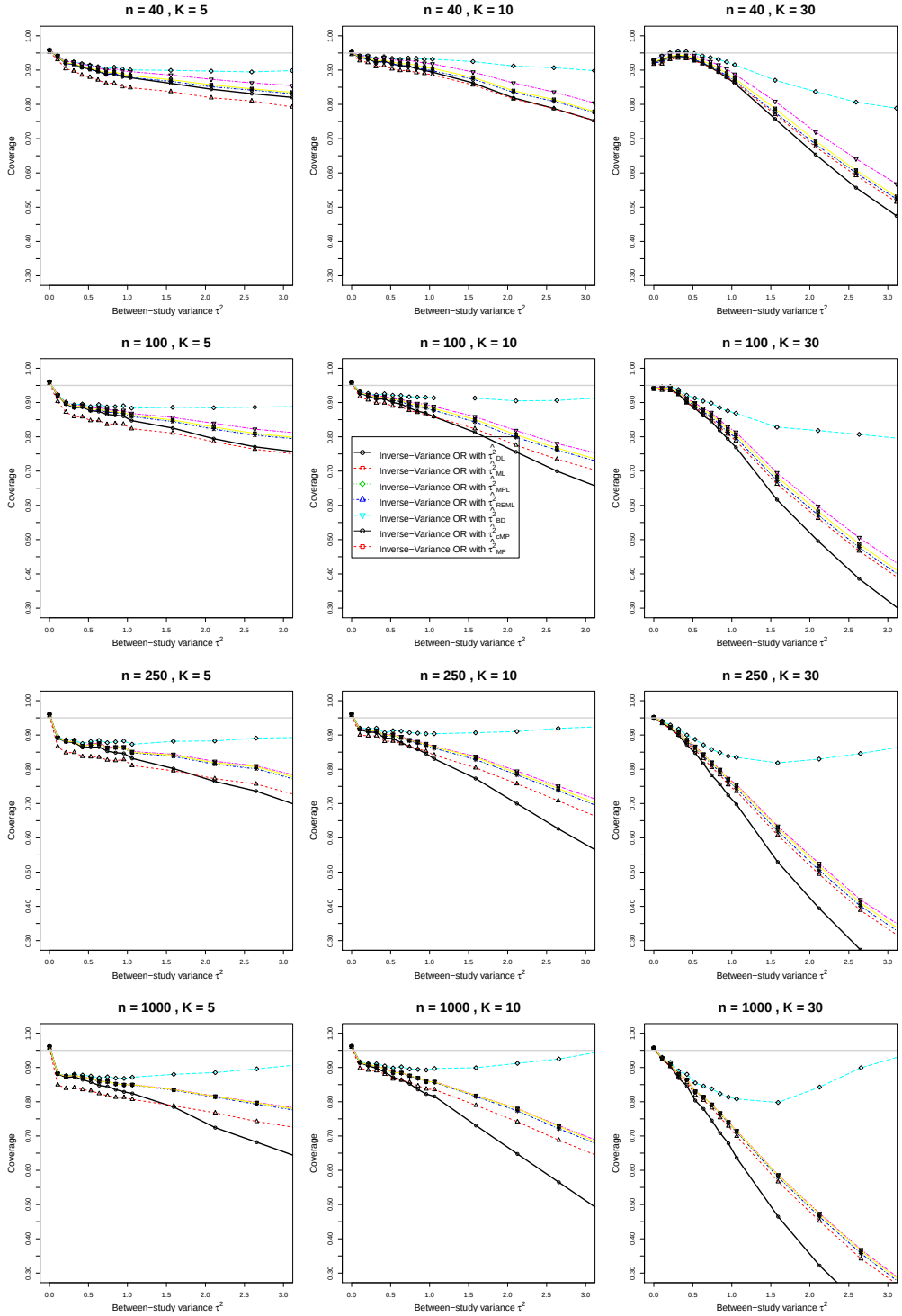


Figure C.70: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio θ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.2$, $\theta = 2$ and $0 \leq \rho \leq 0.3$. The inverse-variance weights use the following estimators of ρ: black circle – Der-Simonian and Laird estimator $\hat{\tau}^2_{DL}$, red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, dark blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, light blue diamonds – Profiled-Breslow-Day estimator $\hat{\tau}^2_{BD}$,pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and yellow circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$.Light grey line at 0.95.

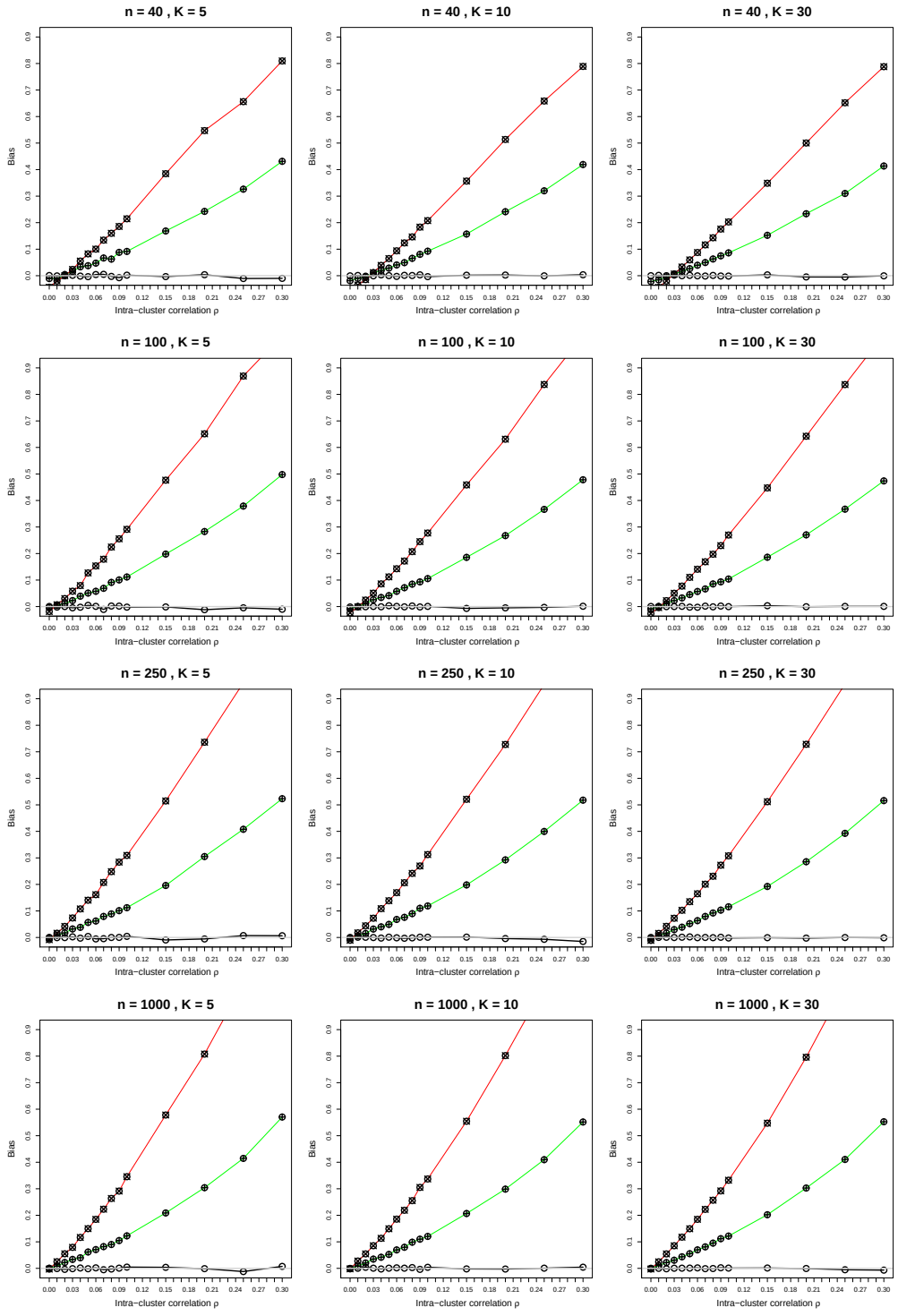


Figure C.71: Bias of overall log odds ratio θ_{IV} obtained from K studies by the inverse-variance method with the moment estimator $\hat{\rho}_{CMP}$ in the weights, for $p_{2j} = 0.4$, and $0 \leq \rho \leq 0.3$. The biases are given for $\theta = 0$ (circles), $\theta = 1$ (circle plus), and $\theta = 2$ (circle cross). Light grey line at 0.

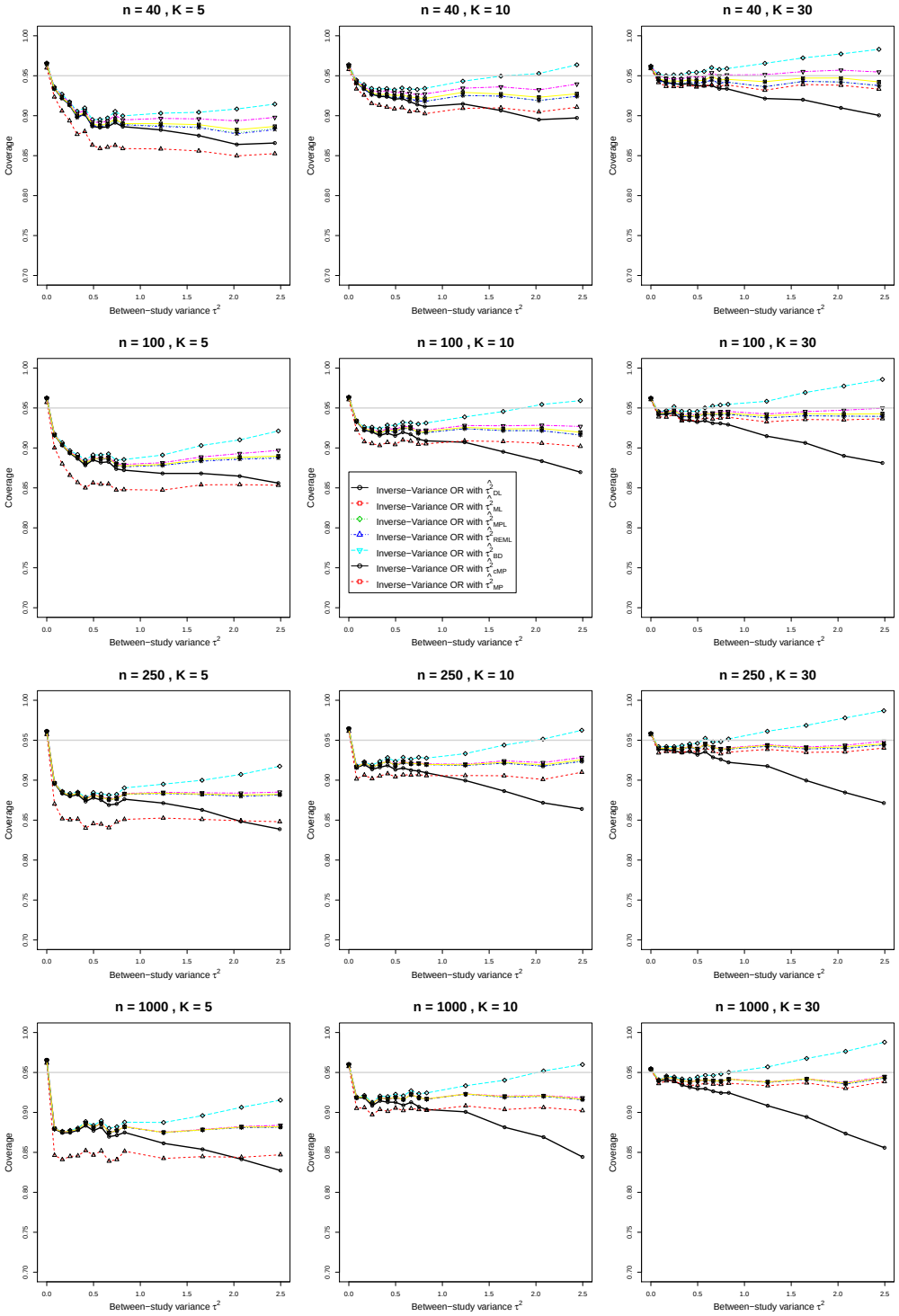


Figure C.72: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio θ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.4$, $\theta = 0$ and $0 \leq \rho \leq 0.3$. The inverse-variance weights use the following estimators of ρ: black circle – Der-Simonian and Laird estimator $\hat{\tau}^2_{DL}$, red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, dark blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, light blue diamonds – Profiled-Breslow-Day estimator $\hat{\tau}^2_{BD}$,pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and yellow circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$.Light grey line at 0.95.

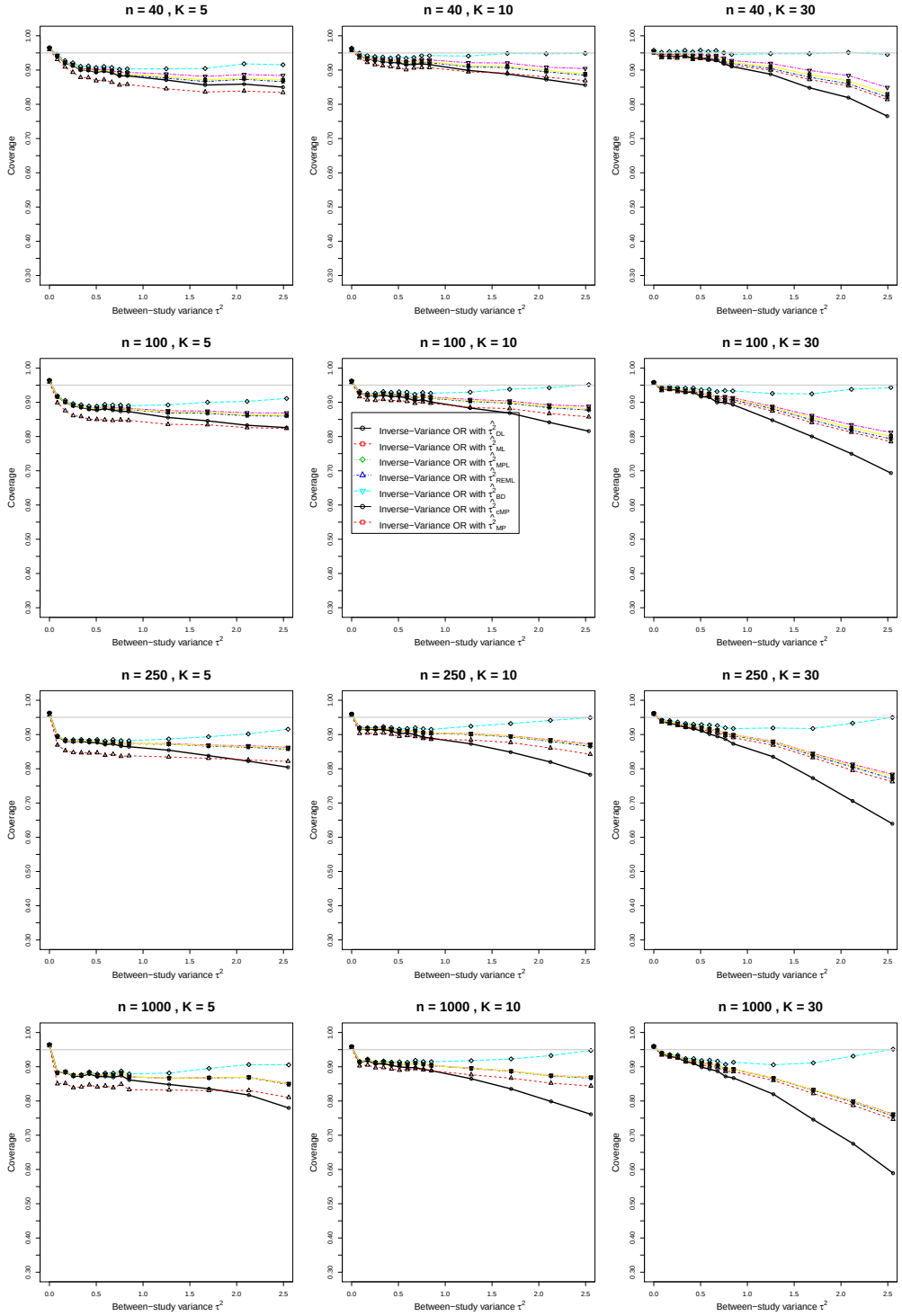


Figure C.73: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio θ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.4$, $\theta = 1$ and $0 \leq \rho \leq 0.3$. The inverse-variance weights use the following estimators of ρ: black circle – Der-Simonian and Laird estimator $\hat{\tau}^2_{DL}$, red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, dark blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, light blue diamonds – Profiled-Breslow-Day estimator $\hat{\tau}^2_{BD}$,pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and yellow circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$.Light grey line at 0.95.

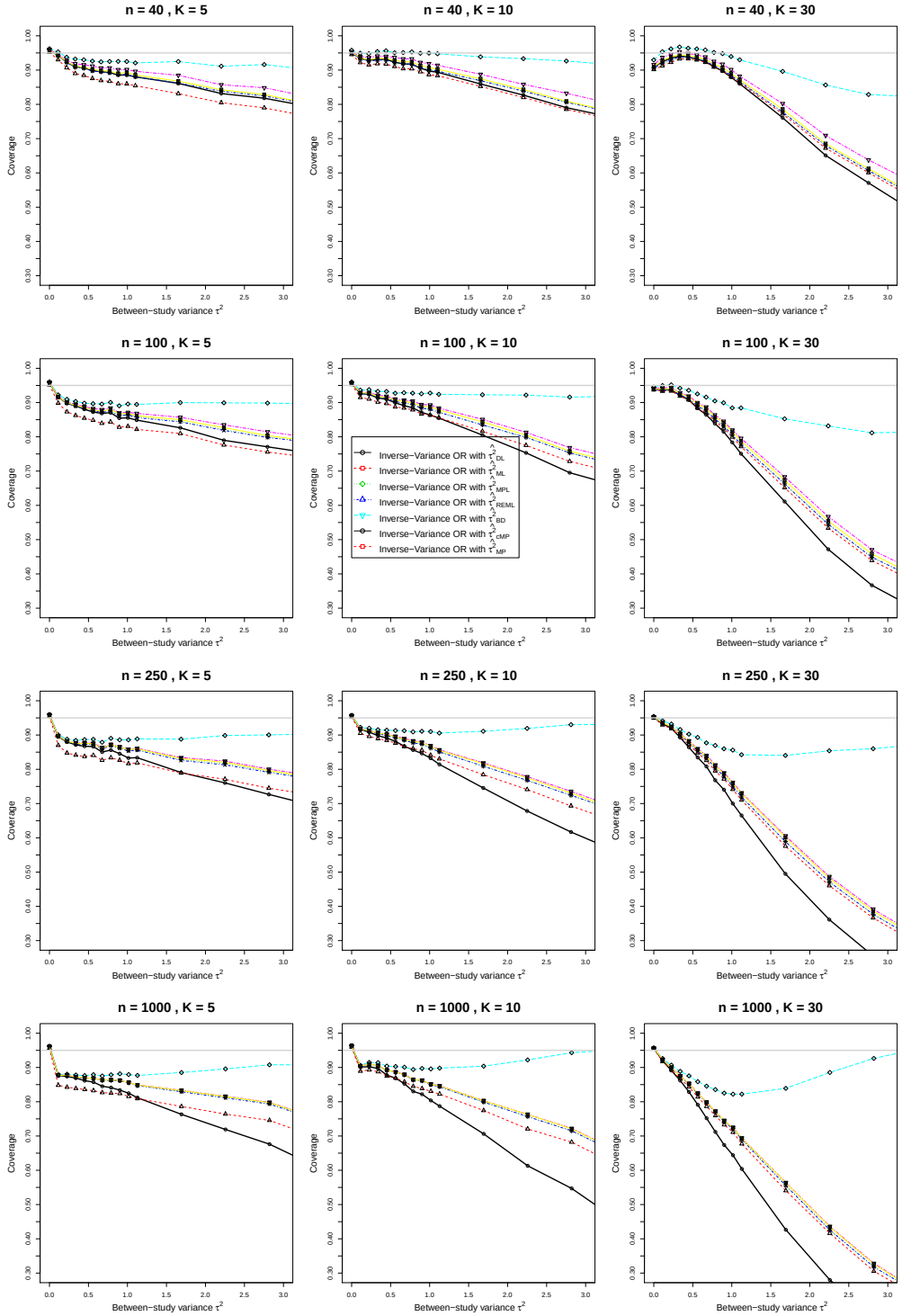


Figure C.74: Coverage at the nominal confidence level of 0.95 of the overall log odds ratio θ obtained from K studies by the inverse-variance method, for $p_{2j} = 0.4$, $\theta = 2$ and $0 \leq \rho \leq 0.3$. The inverse-variance weights use the following estimators of ρ: black circle – Der-Simonian and Laird estimator $\hat{\tau}^2_{DL}$, red triangles – Maximum Likelihood estimator $\hat{\tau}^2_{ML}$, green pluses – Maximum Penalized Likelihood estimator $\hat{\tau}^2_{MPL}$, dark blue crosses – Restricted Maximum Likelihood estimator $\hat{\tau}^2_{REML}$, light blue diamonds – Profiled-Breslow-Day estimator $\hat{\tau}^2_{BD}$,pink reverse triangles – Corrected Mandel-Paule estimator $\hat{\tau}^2_{CMP}$ and yellow circles – standard Mandel-Paule estimator $\hat{\tau}^2_{MP}$.Light grey line at 0.95.