June 2012

Heterogeneity issues in the meta-analysis of cluster randomization trials.

Shun Fu Chen
The University of Western Ontario

Supervisor
Drs Allan Donner
The University of Western Ontario

Joint Supervisor
Neil Klar
The University of Western Ontario

Graduate Program in Epidemiology and Biostatistics

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

HETEROGENEITY ISSUES IN THE META-ANALYSIS OF CLUSTER RANDOMIZATION TRIALS

Thesis format: Monograph

by

Shun Fu Chen

Graduate Program in Epidemiology & Biostatistics

**A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy**

**The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada**

CERTIFICATE OF EXAMINATION

THE UNIVERSITY OF WESTERN ONTARIO
THE SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES

Joint-Supervisor

Dr. Allan Donner

Joint-Supervisor

Dr. Neil Klar

Examiners

Dr. Shelley Bull

Dr. Yun-Hee Choi

Dr. John Koval

Dr. Serge Provost

The thesis by

Shun Fu Chen

entitled

HETEROGENEITY ISSUES IN THE META-ANALYSIS OF CLUSTER RANDOMIZATION TRIALS

is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Date _____

Chair of the Examination Board

ABSTRACT

An increasing number of systematic reviews summarize results from cluster randomization trials. Applying existing meta-analysis methods to such trials is problematic because responses of subjects within clusters are likely correlated. The aim of this thesis is to evaluate heterogeneity in the context of fixed effects models providing guidance for conducting a meta-analysis of such trials. The approaches include the adjusted Q statistic, adjusted heterogeneity variance (τ_c^2) estimators and their corresponding confidence intervals and adjusted measures of heterogeneity (H_a^2, R_a^2, I_a^2) and their corresponding confidence intervals. Attention is limited to meta-analyses of completely randomized trials having a binary outcome. An analytic expression for power of Q test is derived, which may be useful in planning a meta-analysis. The Type I error and power for the Q statistic, bias and mean square errors for the estimators and the coverage, tail errors and interval width for the confidence interval methods are investigated using Monte Carlo simulation.

Simulation results show that the adjusted Q statistic has a Type I error close to the nominal level of 0.05 as compared to the unadjusted Q statistic which has a highly inflated Type I error. Power estimated using the algebraic formula had similar results to empirical power. For τ_c^2 estimators, the iterative REML estimator consistently had little bias. However, the noniterative MVVC and DLVC estimators with relatively low bias may also be recommended for small and large heterogeneity, respectively. The Q profile confidence interval approach for τ_c^2 had generally nominal coverage for large heterogeneity. The measures of heterogeneity had generally low bias for large number of trials. For confidence interval approaches, the MOVER consistently maintained nominal coverage for 'low' to 'moderate' heterogeneity. For the absence of heterogeneity, the approach based on the Q statistic is preferred. Data from four cluster randomization trials are used to illustrate methods of analysis.

Keywords: cluster randomization; meta-analysis; heterogeneity; binary outcome; Q statistic; power; confidence intervals

ACKNOWLEDGMENTS

First, I would like to express my deepest appreciation to my supervisors, Drs Allan Donner and Neil Klar, for offering the opportunity to pursue my PH.D studies in Biostatistics at the Department of Biostatistics and Epidemiology of Univerisity of Western Ontario. I would also like to thank them for their persistent patience, warm encouragement and effective guidance while working on the thesis. This thesis would not have been completed without their continuous help.

I would thank the authors of the four papers (Moher et al., 2001; Woodcock et al., 1999; Montgomery et al., 2000; Jolly et al., 1999) for use of their data and Dr. Martin Gulliford for facilitating their use.

With all my affection, I thank my parents, my husband and my son, for their love and support.

Finally, I thank God to provide all those people in my life who have helped me and prayed for me to complete this thesis which seems to be impossible years ago.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Cluster randomization trials

Randomized controlled trials are often deemed the gold standard to assess the effectiveness of an intervention in health research (Wade, 1999). A key benefit of randomization is the potential for elimination of bias due to confounding. The units of randomization in randomized trials are usually the individual.

Over the past two decades, randomized trials in which the unit of randomization is at the cluster level have been more frequently adopted in the evaluation of health care interventions, screening and educational programs (Bland, 2004). Such trials are characterized by random assignment of intact social units (e.g., worksites, clinical practices, schools or entire communities) instead of individual study subjects (Donner and Klar, 2000).

For example, a study evaluating the effect of vitamin A supplementation on childhood mortality (Sommer et al., 1986) adopted cluster randomization because it was not politically feasible to randomize individuals. Hence, the units of randomization for this study were villages instead of individuals within a village. Contamination could have arisen using individual random assignment. For instance, contamination would occur if

individuals from the same village who were assigned to different interventions shared their vitamin A supplement. Cluster randomization trials are preferred in situations where the ethical issues, the desire to control costs or the attempt to minimize experimental contamination are major concerns.

However, the cluster randomization design is statistically less efficient compared to individual randomization because responses of individuals in a cluster tend to be more similar to each other than to responses of individuals in different clusters. The degree of similarity is measured using the intracluster correlation coefficient denoted by ρ (Donner and Klar, 2000, p2) which takes a value between 0 and 1. In order to adjust for clustering, the variance of the estimated intervention effect is multiplied by a variance inflation factor (or design effect), $IF = 1 + (\bar{m} + 1)\rho$ where $\bar{m}$ is the average cluster size. Intracluster correlation coefficients may be quite small particularly for community intervention trials where they rarely take on values above 0.1 (Murray et al., 2000). However even then design effects may be quite large since such trials typically recruit hundreds of subjects per cluster. Subsequently, ignoring clustering effects could result in spurious statistical significance where the variance estimators will tend to be underestimated.

1.2 Meta-analysis of individually randomized trials

Since the 1980s there have been a growing number of published systematic reviews summarizing results from clinical trials (Whitehead, 2002, page xiii). At the same time, there has also been increasing methodological research dealing with meta-analytic methods. A key challenge in combining study results is possible heterogeneity in the estimated intervention effect. Heterogeneity may reflect systematic differences in study design or in characteristics of participating subjects or may be a consequence of random variation (Whitehead, 2002).

1.2.1 Fixed versus random effects modeling of heterogeneity

Fixed effects and random effects models depend on different assumptions about heterogeneity of the intervention effect. The fixed effects model assumes that there is a common fixed effect (at least for interval estimation) and a random component (sampling error) that is responsible for differences among trial results. Often, however, there may be some heterogeneity of intervention effects across trials. A test of heterogeneity is frequently used to evaluate the assumption of a common fixed effect. On the other hand, the random effects model assumes that the observed trials are a random sample from a hypothetical population of trials. To account for the variation among trial results, an additional random term is added to the model. This added term, recognized as the heterogeneity variance parameter, is denoted by τ^2. Consequently, the random effects model generally yields more conservative inferences about the intervention effect as compared to the fixed effects model (Schulze, 2007; Villar et al., 2001).

1.2.2 Tests of heterogeneity: Q statistic

The Q statistic (Cochran, 1954) tests the null hypothesis: $H_o : \theta_1 = \theta_2 = \ldots = \theta_k = \theta$ versus the alternative H_A: at least one trial had a truly different intervention effect as compared to the other trials, where θ_j denotes the intervention effect for trial j, $j = 1, \ldots, k$. Mathematically, the Q statistic is defined as a weighted sum of squares of the deviations of individual study estimates $\hat{\theta}_j$, from the overall estimate $\hat{\theta}$. The Q statistic when H_o is true, is approximately a chi-square random variable with $k - 1$ degrees of freedom. If the null hypothesis H_o is rejected, one concludes that there is at least one study which truly differs from other studies in terms of the intervention effect. Further analyses are then usually recommended to identify covariates that stratify studies into homogeneous populations.

Several other test statistics are available to test for heterogeneity (e.g. likelihood ratio test, score test). In a simulation study generating continuous outcome data, Viechtbauer (2007*b*) showed that the Q statistic as compared to other test statistics kept the tightest control of the Type I error rate for meta-analyses based on studies having at least moderately large sample size, such that the number of trials is from 5 to 80 and the average sample size per trial is 20 to 640. He also suggested that if the amount of heterogeneity was small, sample sizes exceeding 100 observations within each study would be required to detect it. As for binary outcome data in large sample sizes, the Q statistic based on the Woolf estimator is conservative in general and least powerful for severely unbalanced and within-strata unbalanced designs (Paul and Donner, 1989). However, for balanced and mildly unbalanced designs, the Q statistic, which is easy to calculate, is recommended.

According to Hardy and Thompson (1998), the Q statistic may detect clinically unimportant heterogeneity when there are many studies but is unable to detect clinically significant heterogeneity when there are few studies. Therefore, power calculations for the Q statistic prior to conducting a meta-analysis may prove helpful in assessing statistical power (Hedges and Pigott, 2001; Valentine et al., 2010). Power for the Q statistic is a function of the selected effect measure, the specified Type I error, number of trials and sample size per trial.

1.2.3 Heterogeneity variance estimators τ^2

Point estimation

Heterogeneity variance estimators are also useful in assessing heterogeneity in meta-analysis. Their advantage is that they do not depend on the number, or size of trials in a meta-analysis like the Q statistic. A disadvantage they share with the Q statistic is that comparisons across meta-analyses must be limited to trials with the same effect

measures (e.g. odds ratio, risk ratio and hazard ratio) (R*ü*cker et al., 2008). Seven methods of estimating the parameter τ^2 were compared in terms of bias and mean square error under a random effects model for a binary outcome in a simulation study Sidik and Jonkman (2007). Four of these estimators are simple to compute while the remaining three approaches require relatively extensive computation.

Hedges (1983) originally developed a method of moments estimator obtained by setting the usual sample variance equal to its expected value and solving for τ^2, known as the variance component type estimator. Another method of moments estimator proposed by DerSimonian and Laird (1986) using the expectation of the Q statistic (Cochran, 1954) is commonly used in random effects meta-analysis (Brockwell and Gordon, 2001; Thompson and Sharp, 1999). Furthermore, given that the DerSimonian and Laird estimator often underestimates the true value (e.g., Bohning et al., 2002; DerSimonian and Kacker, 2007; DerSimonian and Laird, 1986; Sidik and Jonkman, 2007), DerSimonian and Kacker (2007) proposed a two-step method to avoid using iterative methods (e.g. likelihood approaches). Besides the method of moments estimators, Sidik and Jonkman (2005) proposed an estimator based on the unbiased estimation of the error variance in a linear model, called a model error variance type estimator. The model error variance type of estimator requires an initial estimate of τ^2. The simplest is to use the empirical variance estimate as an initial estimator for τ^2. Later, Sidik and Jonkman (2007) suggested an improved version of this approach by using the variance component type estimate as an initial estimate of τ^2. The common iterative approaches to estimating τ^2 are maximum likelihood estimation and restricted maximum likelihood estimation (Hardy and Thompson, 1996; Harville, 1977; Raudenbush and Bryk, 1985). Another iterative approach is obtaining by using the empirical Bayes estimator (Morris, 1983).

F

Sidik and Jonkman's (2007) simulation results showed that the improved variance com-

Table 1.1: Summary of the heterogeneity variance estimators

Method	**Description**
Non-Iterative	
Variance Component (VC)	Method of moments
DerSimonian and Laird (DL)	Method of moments
Two-step DL (DLVC)	Empirical variance as an initial estimator
Two-step DL (DL2)	Variance component as an initial estimator
Model Error Variance (MV)	Empirical variance as an initial estimator
Improved (MVVC)	Variance component as an initial estimator
Iterative	
Maximum Likelihood (ML)	Likelihood
Restricted (REML)	Likelihood
Bayes	Bayesian

ponent type estimator and the empirical Bayes estimator provide the most accurate estimation when the heterogeneity is moderate to large (i.e. $\tau^2 \geq 0.5$). The variance component estimator and the model error variance both tend to overestimate the true heterogeneity variance except for meta-analyses with large number of trials and unless the heterogeneity variance is large, respectively. However, the likelihood estimators and DerSimonian and Laird's estimator tend in general to underestimate the true heterogeneity variance. Schlattmann (2009, Chapter 7) found similar results. Table 1.1 provides a summary of the heterogeneity variance estimators mentioned above.

Interval estimation

It is often useful to report a confidence interval in addition to a point estimator of τ^2. Viechtbauer (2007*a*) proposed a new method called the Q profile to construct such intervals and evaluated its performance in terms of nominal coverage as compared with other existing approaches, including Biggerstaff-Tweedie (Biggerstaff and Tweedie, 1997), profile likelihood, Wald-type, Sidik-Jonkman (Sidik and Jonkman, 2005), parametric bootstrap and non-parametric bootstrap using Monte-Carlo simulation.

The Q statistic approximately follows a chi-square distribution with $k - 1$ degrees of free-

dom under the null hypothesis. Alternatively, the 95 percent confidence interval obtained from the Q profile method is constructed based on the 2.5th and 97.5th percentile of this distribution. Based on a similar idea, the Biggerstaff-Tweedie confidence interval can be obtained by approximating the distribution of Q with a gamma distribution (Biggerstaff and Tweedie, 1997). For the profile likelihood method, the confidence intervals can be obtained by profiling the likelihood ratio statistic with the maximum likelihood or restricted maximum likelihood estimates. Then, the inverse of the Fisher information matrix is used to calculate the asymptotic sampling variances of the maximum likelihood and restricted maximum likelihood estimates of the heterogeneity estimator in order to construct Wald-type confidence intervals. The confidence interval for the model error variance is based on the assumption that its estimator approximately follows a chi-square distribution with $k-1$ degrees of freedom (Sidik and Jonkman, 2005). Last is the bootstrap confidence interval constructed by taking the 2.5th and 97.5th empirical percentiles of the heterogeneity estimate based on the bootstrap sample after repeating the same process up to 1000 times.

According to simulation results (Viechtbauer, 2007*a*), the profile likelihood method with the Q statistic yields the most accurate coverage, closely followed by Biggerstaff and Tweedie's method. The performance of other methods was poor in general with a coverage probability either too low or too high.

1.2.4 Measures of heterogeneity

Higgins and Thompson (2002) proposed three statistics which measure the impact of heterogeneity on a meta-analysis: H, R, and I^2. The advantage of these measures as compared to the heterogeneity variance is that they allow heterogeneity of the intervention effect to be compared across meta-analyses including different numbers of studies and different outcome measures.

The H statistic is given by the square root of the Q statistic divided by its degrees of freedom. Since the expectation of Q is equal to $k-1$ under H_o, $H=1$ indicates that the intervention effects are homogeneous across trials. Values of H exceeding 1.5 may suggest heterogeneity complicating interpretation of the summary estimates of the intervention effect. The R statistic is the ratio of the standard error of a random effects meta-analytic summary estimate to the standard error of a fixed effects meta-analytic summary estimate. It describes the inflation in the confidence interval for a summary intervention effect estimate under a random effects model compared with a fixed effects model. When the value is 1, it indicates that the two models yield identical inferences and the fixed effects model is sufficient. The I^2 statistic is interpreted as the proportion of total variation in the estimate of an intervention effect that is due to heterogeneity between studies. When the I^2 statistic is 0 percent, the variation is considered only due to sampling error and not due to heterogeneity. Similarly, an I^2 statistic of 20 percent indicates that 20 percent of variability in the trials may be attributed to between-study variation. Several investigators (e.g. Higgins and Green, 2008; Higgins and Thompson, 2002; Higgins et al., 2002*a*) recommend including the H or I^2 statistics when reporting meta-analyses.

Several simulation studies (Huedo-Medina et al., 2006; Mittlböck and Heinzl, 2006) examined the properties of H and I^2 as a function of the ratio of between and within study variances. They concluded that I^2 but not H may depend on the number of trials when that number is small (i.e. $k \leq 10$). In addition, the values of I^2 may be affected by the ratio of between and within study variances rather than the between study variance alone.

Possible approaches for constructing confidence intervals for each measure were also summarized in the appendix of the Higgins and Thompson's article (2002): i) based on the distribution of Q, ii) based on the statistical significance of Q (test-based method), iii) based on the estimation of a heterogeneity estimator, and iv) using a non-parametric

bootstrap procedure. In addition, the confidence interval approach known as the method of variance estimates recovery (MOVER) originally proposed by Zou (2008) may be used to construct a confidence interval for H by treating it as a ratio of within study and between study variances (Donner and Zou, 2010).

According to the simulation results presented in Table A1 for the H statistic (Higgins and Thompson, 2002), it appears that a confidence interval constructed based on the distribution of Q has coverage close to 100 per cent even with large number of trials k (i.e. $k = 30$) except for large heterogeneity. The maximum likelihood, restricted likelihood and bootstrap confidence intervals have inadequate coverage above nominal for small heterogeneity and below nominal for large heterogeneity. The coverage of the test-based confidence interval appears to be conservative in most of the situations except when significant heterogeneity is present or the number of studies is large. The Pearson type III confidence interval constructed for the heterogeneity estimator provides good coverage in all situations, but is complicated to calculate. Finally, the accuracy of MOVER will depend heavily on the performance of confidence intervals for numerator and denominator of the given ratio (Schuster and Metzger, 2010, Chapter 11).

In summary, despite the different assumptions and methods regarding the assessment of heterogeneity among studies, the fixed and random effects approaches in principle both use weighted averages with only a change in the weights to calculate the overall mean effect size. When the two modeling approaches yield similar results, the conclusions based on these results gain credibility. When the intervention effects are considered homogeneous, the results from both models are identical with the heterogeneity variance equal to zero.

1.3 Meta-analysis of cluster randomization trials

In response to the frequent use of cluster randomization designs in the health research field, the need to conduct meta-analyses for such trials becomes increasingly evident and necessary. The challenge in planning and conducting such a meta-analysis involves the need for accounting for clustering effects. Not recognizing that the unit of randomization for cluster randomization trials is at the cluster level with outcome measures collected and analyzed at the individual level will generally lead to underestimating the variance due to lack of independence between individuals.

Heterogeneity is recognized as another important analytic issue in performing a meta-analysis by investigators who have performed separate meta-analyses on trials that involve very different randomization units. For example, a study was conducted by Fawzi et al. (1993) to investigate the effect of vitamin A supplementation on child mortality. The participants of this study were taken from studies of hospitalized children with measles, as well as other studies involving healthy children participating in community-based trials. Individual children were assigned to intervention in the four hospital-based trials, while allocation was by village, district or household in the eight community-based trials. Therefore, the meta-analysis was performed separately for the hospital-based trials and the community studies. When the results agree, an important advantage is the confidence gained that the intervention tested is effective (or ineffective) in more than one setting. Otherwise, the investigator can further study the impact of different choices of randomization unit as part of a sensitivity analysis.

One approach to testing heterogeneity in the meta-analysis of cluster randomized trials is to use the Q statistic adjusted for clustering discussed in Donner et al. (2001). In principle, the idea is similar to the Q statistic used to test for heterogeneity in meta-analysis of individually randomized trials, except its weights are modified to account for clustering to ensure test validity. A similar method of adjusting tests of heterogeneity for clustering was

described independently by Song (2004), suggesting that the adjusted tests maintained the nominal significance level in a stimulation study.

Methodological researches on meta-analytic methods involving cluster randomized trials have mainly focused on fixed effects models. By assuming there is no variation between studies, there are several statistical approaches that can be applied to a meta-analysis of cluster randomization trials with a binary endpoint. Statistical methods include the adjusted Mantel-Haenszel procedures, the ratio estimator approach, the general inverse variance approach, Woolf procedures and generalized estimating equations (GEE) using robust variance estimation (Donner et al., 2001).

The adjusted Mantel-Haenszel test statistic (Donner and Klar, 2000) is slightly modified from the standard Mantel-Haenszel test statistic to account for clustering effects. The null hypothesis is that the overall odds ratio of all 2x2 tables is equal to one and the test statistic follows approximately a chi-squared distribution with one degree of freedom. The ratio estimator approach is based on an adjustment of the Mantel-Haenszel chi-square statistic in which the event rate is regarded as a ratio rather than as a proportion. It was developed by Rao and Scott (1992) and involves dividing the observed sample frequencies (counts) in a given study by the estimated design effect. The general inverse variance approach (GIV) is obtained by combining study estimates in a meta-analysis using a weighted average of estimated effect measures that are calculated separately for each trial. This approach is recommended in the guidance provided by the Cochrane Collaboration. The Woolf procedure, which is best applied with a small number of clusters each of fairly large size, transforms the intervention odds ratio of each trial to the logarithmic scale in order to obtain a distribution which is more likely to be normally distributed. Then, the average of the transformed odds ratios is computed using a weighting scheme originally described by Woolf (1955) and modified for cluster randomization trials by Donner and Donald (1987a).

Furthermore, a simulation study (Darlington and Donner, 2007) was performed to compare the unadjusted Mantel-Haenszel method, the adjusted Mantel-Haenszel methods, the ratio procedure, the general inverse variance, and the Woolf procedure. This simulation study had two important results. First, the simulation results clearly showed that it is inappropriate to use the unadjusted Mantel-Haenszel method due to elevated Type I error rate. Second, the adjusted Mantel-Haenszel method had the greatest power and slightly outperformed the general inverse variance method since it uses information on the cluster sizes and intracluster coefficient ρ for each trial, while the general inverse method is a generic procedure.

1.4 Scope of thesis

Most meta-analytic methods focus on combining study results of individually randomized trials where observations are independent. Meta-analytic methods for cluster randomization trials are largely extensions of meta-analytic methods for individually randomized trials. However, applying existing meta-analytic methods to handle heterogeneity of cluster randomized trials is problematic with correlated observations. The rationale for limiting attention to heterogeneity among studies is that this is a substantial issue for meta-analysis, since when present it complicates discussion of an overall intervention effect.

This research focuses mainly on binary outcomes because such outcomes have been most frequently used in cluster randomization trials (Laopaiboon, 2003). There are three frequently used designs in cluster randomization trials: completely randomized, matched-pair and stratified. The completely randomized design is best suited to trials that have a fairly large numbers of clusters, whereas matching or stratification is more effective in small studies (Donner and Klar, 2000). The challenge of extending all methods to stratified and pair-matched designs is an area for future research and will not be further discussed. For

simplicity, the discussion will thus be focused on designs where there is a single binary, cluster-level covariate, i.e., trials where there is one experimental group and one control group.

In summary, my thesis will focus on exploring and evaluating heterogeneity in the context of fixed effects models with the aim of providing general guidance in conducting a meta-analysis of cluster randomization trials. Attention will also be limited to meta-analyses of community intervention trials which typically enroll a small number of large clusters. This focus reflects the relatively greater methodological challenge of statistical inferences when estimates of variance inflation are less precisely estimated. Intervention effects for binary outcomes will be measured using odds ratio estimators comparing an experimental group to a control intervention.

1.5 Thesis objectives

The primary objectives of this research are:

1. Analytics

   (a) To extend the Q statistic, as commonly applied to test for heterogeneity in meta-analyses of individually randomized trials, to the meta-analysis of cluster randomization trials by specifying a weight accounting for clustering.

   (b) To obtain an analytic expression for the power curve of the adjusted Q statistic.

   (c) To derive heterogeneity variance estimators and their confidence intervals accounting for clustering.

   (d) To derive measures of heterogeneity and their confidence intervals accounting for clustering.

2. Simulation

   (a) To evaluate the performance of the adjusted Q statistic in terms of Type I error and statistical power and to compare its power with the proposed formula.

   (b) To assess the bias and mean square error for the adjusted heterogeneity variance estimators and to evaluate the coverage, tail errors and interval width of the proposed confidence interval methods.

   (c) To assess the bias and mean square error for the adjusted measures of heterogeneity and to evaluate the coverage, tail errors and interval width of the proposed confidence interval methods.

3. Example

   (a) To illustrate the application of results, both in fixed and random effects models, using data from four cluster randomization trials.

1.6 Organization of the thesis

This thesis includes eight chapters. Chapter 2 extends the Q statistic, as commonly applied to test for heterogeneity in meta-analyses of individually randomized trials, to the meta-analysis of cluster randomization trials. An analytic expression for the power of the Q statistic is derived. The effect on the power of cluster size, number of clusters, degree of heterogeneity, and magnitude of intracluster correlation is explored. Chapter 3 presents analytic expressions for the heterogeneity variance estimators adjusted for clustering and describes approaches for constructing confidence intervals. Chapter 4 presents analytic expressions for the measures of heterogeneity adjusted for clustering and approaches for constructing confidence intervals.

Chapter 5 describes the design of a simulation study used to assess the procedures and to validate analytical findings. Performance is evaluated in terms of Type I error and statistical power for the Q statistic, bias and mean square error for both heterogeneity variance estimators and measures of heterogeneity, and coverage for the confidence intervals approaches. Results of the simulation study are described in Chapter 6. Chapter 7 presents a meta-analysis of 4 cluster randomization trials to illustrate the application of the proposed methods. Finally, Chapter 8 summarizes the main results, the recommendations based on the main results, the limitations of this thesis, and directions for future research.

Chapter 2

Approximate power of the adjusted Q statistic

2.1 Introduction

The Q statistic was introduced in Chapter 1 for meta-analysis of individually randomized trials. An extension of this statistic was also described for cluster randomization trials. Applying the Q statistic to a meta-analysis of cluster randomized trials without adjusting for clustering is problematic because the unadjusted Q statistic tends to have inflated Type I error rates. Therefore, we will derive the adjusted Q statistic to account for clustering as well as derive a formula for its power that may be useful in planning a meta-analysis of such trials. It can be quite time consuming to review randomized trials and combine their results for meta-analyses. Thus, performing power calculations prior to conducting a meta-analysis may prevent wasting time, money and energy in the searching and collection of representative trials when there is scant likelihood of detecting clinically relevant amounts of heterogeneity (Donner et al., 2003).

Approaches to computing the power of the Q statistic as applied to the meta-analyses of individually randomized trials have been frequently discussed (e.g. Biggerstaff and

Jackson (2008); Hardy and Thompson (1998); Hedges and Pigott (2001); Jackson (2006); Valentine et al. (2010)). However, relatively little attention has been given to considering the power of the Q statistic in planning a meta-analysis of cluster randomized trials. Therefore, the aim of this chapter is to extend existing approaches to approximating power of the adjusted Q statistic (i.e. adjusted for clustering). Specifically, interest focuses on investigating the power of the adjusted Q statistic as a function of number of trials, number of clusters, cluster size, disease risk rates, intracluster correlation coefficient and degree of odds ratio heterogeneity across trials.

Section 2.2 provides the notation used throughout this thesis. An analytic expression for the adjusted Q statistic is derived in Section 2.3.1, followed by a power formula approximating the power of the adjusted Q statistic in Section 2.3.2. Summary comments are provided in Section 2.4.

2.2 Notation

The data layout for a meta-analysis of k cluster randomized trials is provided in Table 2.1, where the notation used is defined in Table 2.2.

Suppose $\theta_1, \ldots, \theta_k$ are the intervention effects of k trials, each measured as a log odds ratio. Then the estimated intervention effect of θ_j of trial j is denoted

$$\hat{\theta}_j = ln\left[\frac{\hat{P}_{1j}(1-\hat{P}_{2j})}{\hat{P}_{2j}(1-\hat{P}_{1j})}\right].$$

Table 2.1: Data layout for a meta-analysis of k cluster randomized trials.

Trial	Intervention	Number of clusters	Number of events	Subjects per cluster	Number of subjects
1	Experimental	n_{11}	A_{11}	m_{11l}	M_{11}
	Control	n_{21}	A_{21}	m_{21l}	M_{21}
	Total	N_1	A_1		M_1
2	Experimental	n_{12}	A_{12}	m_{12l}	M_{12}
	Control	n_{22}	A_{22}	m_{22l}	M_{22}
	Total	N_2	A_2		M_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	Experimental	n_{1k}	A_{1k}	m_{1kl}	M_{1k}
	Control	n_{2k}	A_{2k}	m_{2kl}	M_{2k}
	Total	N_k	A_k		M_k

Table 2.2: Notation used in Table 2.1 given that intervention groups $i = 1, 2$, cluster $l = 1, \ldots, n_{ij}$ and trial $j = 1, \ldots, k$.

Symbol	Description
m_{ijl}	size of the i^{th} group in cluster l of trial j
n_{ij}	total number of clusters in group i of trial j
$N_j = \sum_{i=1}^{2} n_{ij}$	total number of clusters in trial j
$M_{ij} = \sum_{l=1}^{n_{ij}} mijl$	total number of subjects in group i of trial j
$M_j = \sum_{i=1}^{2} M_{ij}$	total number of subjects in trial j
A_{ijl}	number of events of the i^{th} group in cluster l of trial j
$A_{ij} = \sum_{l=1}^{n_{ij}} A_{ijl}$	number of events in group i of trial j
$A_j = \sum_{i=1}^{2} A_{ij}$	total number of events in trial j
$\hat{P}_{ijl} = A_{ijl}/m_{ijl}$	proportion of events of the i^{th} group in cluster l of trial j
$\hat{P}_{ij} = A_{ij}/M_{ij}$	total event rate in group i of trial j

2.3 Fixed effects model

A fixed effects model assumes that there is a common effect measure and a random component (within study sampling error), which is responsible for observed between study heterogeneity in a meta-analysis. The fixed effects model for the observed study-specific intervention effect $\hat{\theta}_j$ is given (Whitehead, 2002) by

$$\hat{\theta}_j = \theta + \epsilon_j, \tag{2.1}$$

where the sampling error ϵ_j is assumed to be approximately independently and normally distributed with mean 0 and within study variance σ_j^2 and with overall mean effect size θ, respectively, for trial j, $j = 1, \ldots, k$.

2.3.1 Adjusted Q statistic

Individually randomized trials

The null hypothesis H_o for the Q statistic is given by $\theta_1 = \theta_2 = \ldots = \theta_k = \theta$ and the alternative hypothesis H_A is $\theta_i \neq \theta_j$ for some $i \neq j$. The mathematical expression for the Q statistic is defined as a weighted sum of squares of the deviations of individual study estimates from the overall mean effect size, given by

$$Q = \sum_{j=1}^{k} \hat{w}_j(\hat{\theta}_j - \hat{\theta})^2 \tag{2.2}$$

where the estimated overall mean effect size is given by

$$\hat{\theta} = \frac{\sum_{j=1}^{k} \hat{w}_j \hat{\theta}_j}{\sum_{j=1}^{k} \hat{w}_j}.$$

The estimated weights are the reciprocals of the estimated within study variances, given by $\hat{w}_j = 1/\hat{\sigma}_j^2$. These particular weights are chosen to provide the most precise estimate of θ by minimizing the variance of θ (Hardy and Thompson, 1996). Under H_o, the Q statistic is distributed as a chi square random variable with $k-1$ degrees of freedom. The derivation of the Q statistic is provided in Appendix A.

In practice, the within study variance σ_j^2 is estimated using data from jth trial, $j = 1, \ldots, k$. Given the responses for the jth trial in Table 2.3,

Table 2.3: Responses for the jth trial

	Positive	Negative	Total	P(positive)
Experimental	A_{1j}	$M_{1j} - A_{1j}$	M_{1j}	$P_{1j} = A_{1j}/M_{1j}$
Control	A_{2j}	$M_{2j} - A_{2j}$	M_{2j}	$P_{2j} = A_{2j}/M_{2j}$

and applying Woolf (1955)'s approach to estimate the within study variance for the estimated log odds ratio, denoted as $\hat{\theta}_j$:

$$\begin{aligned} \hat{w}_j = (\hat{\sigma}_j^2)^{-1} &= \left[\frac{1}{A_{1j}} + \frac{1}{M_{1j} - A_{1j}} + \frac{1}{A_{2j}} + \frac{1}{M_{2j} - A_{2j}}\right]^{-1} \\ &= \left[\frac{1}{M_{1j}\hat{P}_{1j}(1-\hat{P}_{1j})} + \frac{1}{M_{2j}\hat{P}_{2j}(1-\hat{P}_{2j})}\right]^{-1}. \end{aligned}$$

Cluster randomized trials

In the case of cluster randomization trials where individuals of the same cluster are correlated with a positive intracluster correlation coefficient ρ, Donner and Donald (1987b) suggested an adjustment using the variance inflation factor for each intervention group $i = 1, 2$ defined as

$$C_{ij} = \sum_{l=1}^{n_{ij}} m_{ijl}[1 + (m_{ijl} - 1)\hat{\rho}_j]/M_{ij}. \tag{2.3}$$

The intracluster correlation coefficient ρ_j may be obtained by using the 'analysis of variance' (ANOVA) estimator proposed by Snedecor and Cochran (1980) (see Appendix B for details). Consequently, the weights adjusting for clustering w_{jc} become

$$\hat{w}_{jc} = (\hat{\sigma}^2_{jc})^{-1} = \left[\frac{C_{1j}}{M_{1j}\hat{P}_{1j}(1-\hat{P}_{1j})} + \frac{C_{2j}}{M_{2j}\hat{P}_{2j}(1-\hat{P}_{2j})}\right]^{-1}. \tag{2.4}$$

Accordingly, replacing $\hat{w}_j$ in (2.2) by $\hat{w}_{jc}$, the adjusted Q statistic is obtained by

$$Q_a = \sum_{j=1}^{k} \hat{w}_{jc}(\hat{\theta}_j - \hat{\theta}_c)^2, \tag{2.5}$$

where the estimated adjusted overall mean effect size is given by

$$\hat{\theta}_c = \frac{\sum_{j=1}^{k} \hat{w}_{jc}\hat{\theta}_j}{\sum_{j=1}^{k} \hat{w}_{jc}}. \tag{2.6}$$

The adjusted Q statistic asymptotically follows a chi square distribution with $k-1$ degrees of freedom under H_o (Song, 2004). Note that if $\hat{\rho}_j = 0$ or $C_{ij} = 1$ for $i = 1, 2$ and $j = 1, \ldots, k$, indicating there is no clustering, $\hat{w}_{jc}$ reduces to $\hat{w}_j$ and Q_a equals Q.

Now, the adjusted Q statistic from equation (2.5) may be rewritten as

$$Q_a = \sum_{j=1}^{k} \hat{w}_{jc}(\hat{\theta}_j - \theta)^2 - \sum_{j=1}^{k} \hat{w}_{jc}(\hat{\theta}_c - \theta)^2 \tag{2.7}$$

and the variance of the overall mean effect size is computed as

$$var(\hat{\theta}_c) = \frac{\sum_{j=1}^{k} \hat{w}_{jc}^2 var(\hat{\theta}_j)}{\left(\sum_{j=1}^{k} \hat{w}_{jc}\right)^2} = \frac{\sum_{j=1}^{k} \hat{w}_{jc}^2 \hat{w}_{jc}^{-1}}{\left(\sum_{j=1}^{k} \hat{w}_{jc}\right)^2} = \frac{1}{\sum_{j=1}^{k} \hat{w}_{jc}}. \quad (2.8)$$

Assuming within study variances are known, the expectation of the adjusted Q statistic under the null hypothesis based on equations (2.7) and (2.8) is given by

$$\begin{aligned} E[Q_a|H_o] &= \sum_{j=1}^{k} w_{jc} E(\hat{\theta}_j - \theta)^2 - \sum_{j=1}^{k} w_{jc} E(\hat{\theta}_c - \theta)^2 \\ &= \sum_{j=1}^{k} w_{jc} var(\hat{\theta}_j) - \sum_{j=1}^{k} w_{jc} var(\hat{\theta}_c) \\ &= \sum_{j=1}^{k} w_{jc} w_{jc}^{-1} - \sum_{j=1}^{k} w_{jc} \left(\sum_{j=1}^{k} w_{jc}\right)^{-1} = k - 1. \end{aligned}$$

2.3.2 Approximate power

Let δ_j denote the deviation of the intervention effect θ_j from the overall mean effect size θ for trial j, $j = 1, \ldots, k$, under the alternative hypothesis such that $\theta_i \neq \theta_j$ for at least one pair (i, j) or equivalently $\delta_j \neq 0$ for at least one j (Montgomery, 2000, p.64). This implies that the model in equation (2.1) for the observed intervention effect $\hat{\theta}_j$ becomes

$$\begin{aligned} \hat{\theta}_j &= \theta + \delta_j + \epsilon_j \quad (2.9) \\ &= \theta_j + \epsilon_j, \end{aligned}$$

since $\delta_j = \theta_j - \theta$, where θ is the overall mean effect size. The sampling error ϵ_j is approximately normally distributed with mean 0 and within study variance σ_{jc}^2 for trial j, $j = 1, \ldots, k$. The fixed effects δ_j has a constraint such that the sum of weighted δ_j equals zero to satisfy the condition that $\hat{\theta}_c$ remains an unbiased estimator of θ with variance of $1/\sum_{j=1}^{k} w_{jc}$. Moreover, the expectations of all cross-products with ϵ_j are set to zero

because the expectation of ϵ_j is equal to zero (i.e. $E(\epsilon_j) = 0$).

Next, replacing $\hat{\theta}_j$ defined in (2.7) by (2.9), the expectation of the adjusted Q under the alternative hypothesis assuming the within study variance being known is given by

$$\begin{aligned} E[Q_a|H_A] &= \sum_{j=1}^{k} w_{jc} E(\theta + \delta_j + \epsilon_j - \theta)^2 - \sum_{j=1}^{k} w_{jc} var(\hat{\theta}_c) \\ &= \sum_{j=1}^{k} w_{jc} E(\delta_j^2) + \sum_{j=1}^{k} w_{jc} E(\epsilon_j^2) - 1 \\ &= \sum_{j=1}^{k} w_{jc} (\theta_j - \theta)^2 + k - 1, \end{aligned}$$

where the adjusted Q statistic is distributed as a noncentral chi square distribution with $k-1$ degrees of freedom and noncentrality parameter NC defined as

$$NC = \sum_{j=1}^{k} w_{jc} (\theta_j - \theta)^2. \tag{2.10}$$

The overall mean effect size θ may be estimated using $\hat{\theta}_c$ in equation (2.6). It follows that the power of the adjusted Q statistic at significance level α is defined as

$$\begin{aligned} power &= 1 - P(Accept\ H_o|H_A) \\ &= 1 - P(Q_a \leq \chi^2_{k-1}|H_A) \\ &= 1 - F(c_\alpha|k-1, NC), \end{aligned} \tag{2.11}$$

where $F(c_\alpha|k-1; NC)$ is the cumulative distribution function of the noncentral chi-square with $k-1$ degrees of freedom and noncentrality parameter NC given in equation (2.10) and c_α is the $100(1-\alpha)$ percent point of the chi-square distribution (Hedges and Pigott,

2001).

In practice, parameters used for calculating statistical power are rarely available; therefore, it is common to make some *a priori* assumptions. Based on these assumptions, we investigate the approximate power of the adjusted Q statistic as a function of the number of clusters, cluster size, disease risk rates, intracluster correlation coefficient and degree of heterogeneity.

First, we focus on the case of an equal number of clusters n per intervention group, where each cluster has a constant cluster size of m. Equal allocation is considered to be statistically efficient as compared to unequal allocation, which requires more clusters to obtain the same statistical power. In the case of unequal cluster sizes, we may replace m by average cluster size $\bar{m}$. The slight underestimation of the actual sample size can be negligible, providing that the variation in cluster size is not substantial. If m is replaced by m_{max}, the statistical power calculated will be more conservative (Donner and Klar, 2000, p.57).

Second, for simplicity, we assume a constant within study variance across trials (i.e. $\sigma_{jc}^2 = \sigma_c^2$ for $j = 1, \ldots, k$). However, in the case where the within study variance varies, the noncentrality parameter assuming a constant within study variance tends to be overestimated. Thus, the statistical power of the test obtained based on a constant within study variance will be overestimated.

Third, Donner and Klar (2000, p.56) noted that the study design has an impact on the estimates of intracluster correlation coefficient. Since each of the trials in the meta-analysis is assumed to be completely randomized, we will assume the intracluster correlation coefficient ρ is constant across trials.

Following these assumptions, the variance inflation factors defined in equation (2.3) are reduced to $1 + (m-1)\rho$ and the within study variance in (2.4) is simplified to a common variance denoted by σ_c^2, given by

$$\sigma_c^2 = \frac{1 + (m-1)\rho}{nmP_1(1-P_1)} + \frac{1 + (m-1)\rho}{nmP_2(1-P_2)}. \tag{2.12}$$

Furthermore, the between study variance denoted by τ_c^2 (also referred to as the heterogeneity variance) can be estimated using the sample variance, given by $\sum_{j=1}^{k}(\theta_j - \bar{\theta})^2/(k-1)$, where $\bar{\theta} = \sum_{j=1}^{k}\theta_j/k$. When the weights are assumed constant across trials, $\bar{\theta}$ is equivalent to the overall mean effect size θ_c. Subsequently, $\sum_{j=1}^{k}(\theta_j - \theta_c)^2$ may be approximated by $(k-1)\tau_c^2$ (Hedges and Pigott, 2001). Therefore, the noncentrality parameter in equation (2.10) is approximated by

$$NC \quad = \quad (k-1)\tau_c^2/\sigma_c^2. \tag{2.13}$$

Note that the ratio τ_c^2/σ_c^2 is a measure of the degree of heterogeneity (see section 5.3). However, for plotting purposes, τ_c^2 and σ_c^2 are considered as two separate quantities. Therefore, without loss of generality, let the effect size (ES) be defined as $d = 2|arcsin(P_1)^{(1/2)} - arcsin(P_2)^{(1/2)}|$ (Cohen, 1992), the values of disease rates (P_1, P_2) corresponding to $d = 0.20$ (small effect size) and 0.50 (medium effect size) are $(0.1, 0.168)$ and $(0.1, 0.293)$, respectively. The effect sizes are defined in term of the disease rates in order to plot the power of the adjusted Q statistic in function of the between study variance τ_c^2 or the intracluster correlation coefficient ρ. Given the number of trials k, number of clusters n, cluster size m, disease rates (P_1, P_2), between study variance τ_c^2, and intracluster correlation coefficient ρ, the power of the adjusted Q statistic was calculated for the following parameter values:

$$
\begin{array}{rcc}
(P_1, P_2) & = & (0.1,\ 0.168),\ (0.1,\ 0.293) \\
(n, m) & = & (5,\ 50),\ (5,\ 100),\ (10,\ 50) \\
k & = & 5,\ 10,\ 20 \\
\tau_c^2 & = & 0 \text{ to } 0.5 \text{ in steps of } 0.1 \\
\rho & = & 0 \text{ to } 0.05 \text{ in steps of } 0.01
\end{array}
$$

The values for the number of trials k and between study variance τ_c^2, are taken from Hardy and Thompson (1998). Also, the values of the effect sizes and intracluster coefficients in community intervention trials are frequently small (Donner and Klar, 1996). For example, the intracluster correlation coefficients for four cluster randomization trials (Jolly et al., 1999; Moher et al., 2001; Montgomery et al., 2000; Woodcock et al., 1999) performed to compare two or more interventions in primary care for cardiovascular heart disease (CHD), which will be used as an example for this research, were in the range of 0 to 0.0125. Also, when the approximate power of the adjusted Q statistic plotted against the between study variance τ_c^2, the intracluster correlation coefficient ρ was set to 0.01. But when the approximate power of the adjusted Q statistic plotted against the intracluster correlation coefficient ρ, the between study variance τ_c^2 was then set to 0.1.

Figure 2.1(a)-(b) shows that the approximate power of the adjusted Q statistic increases as the number of trials increases, while holding other variables constant. Similarly, in Figure 2.1(c)-(d), the approximate power of the adjusted Q statistic increases as the total sample size increases while holding other variables constant. In addition, for a given total sample size ($nm = 500$), the adjusted Q statistic with $n = 10$ has greater power than with $n = 5$. In Figure 2.1(e)-(f), the approximate power increases as the effect size becomes larger while holding other variables constant because larger effect size with larger $P_2(1 - P_2)$ in equation (2.12) given P_1 fixed results in smaller constant within study variance. From the plots of power against τ_c^2 (first column), it is seen that the approximate power increases as τ_c^2 increases. For instance, in Figure 2.1(a), the approximate power of the adjusted Q statistic is approximately 60% for $\tau_c^2 = 0.2$ and 80% for $\tau_c^2 = 0.4$ while fixing $k = 5$, $(n, m) = (5, 50)$, $\rho = 0.01$, and $ES = small$. On the contrary, for the plots

of power against ρ (second column), it is seen that the approximate power decreases as ρ increases. For instance, in Figure 2.1(b), the approximate power of the adjusted Q statistic is approximately 80% for $\rho = 0$ and 40% for $\rho = 0.02$ while fixing $k = 10$, $(n, m) = (5, 50)$, $\tau_c^2 = 0.1$, and $ES = small$.

2.4 Summary

In summary, since the validity of the unadjusted Q statistic in the presence of clustering becomes questionable with inflated Type I error rates, the adjusted Q statistic has been introduced. In addition, the approximate power of the adjusted Q statistic was derived from a noncentral chi square distribution with $k - 1$ degrees of freedom and a specified noncentrality parameter.

We have also investigated the power in terms of parameters including the number of trials, number of clusters, cluster size, disease rates between intervention groups (i.e. effect size), between study variance and intracluster correlation coefficient. It appears that the power of the adjusted Q statistic increases by increasing any of the following parameters: number of trials, overall sample size per trial (i.e. $n \times m$), effect size (disease rates between intervention groups) and between study variance. In contrast, the power decreases as the intracluster correlation coefficient increases. Moreover, for a fixed sample size, it is seen that the power of the adjusted Q statistic is greater for a large number of small clusters than for a small number of large clusters.

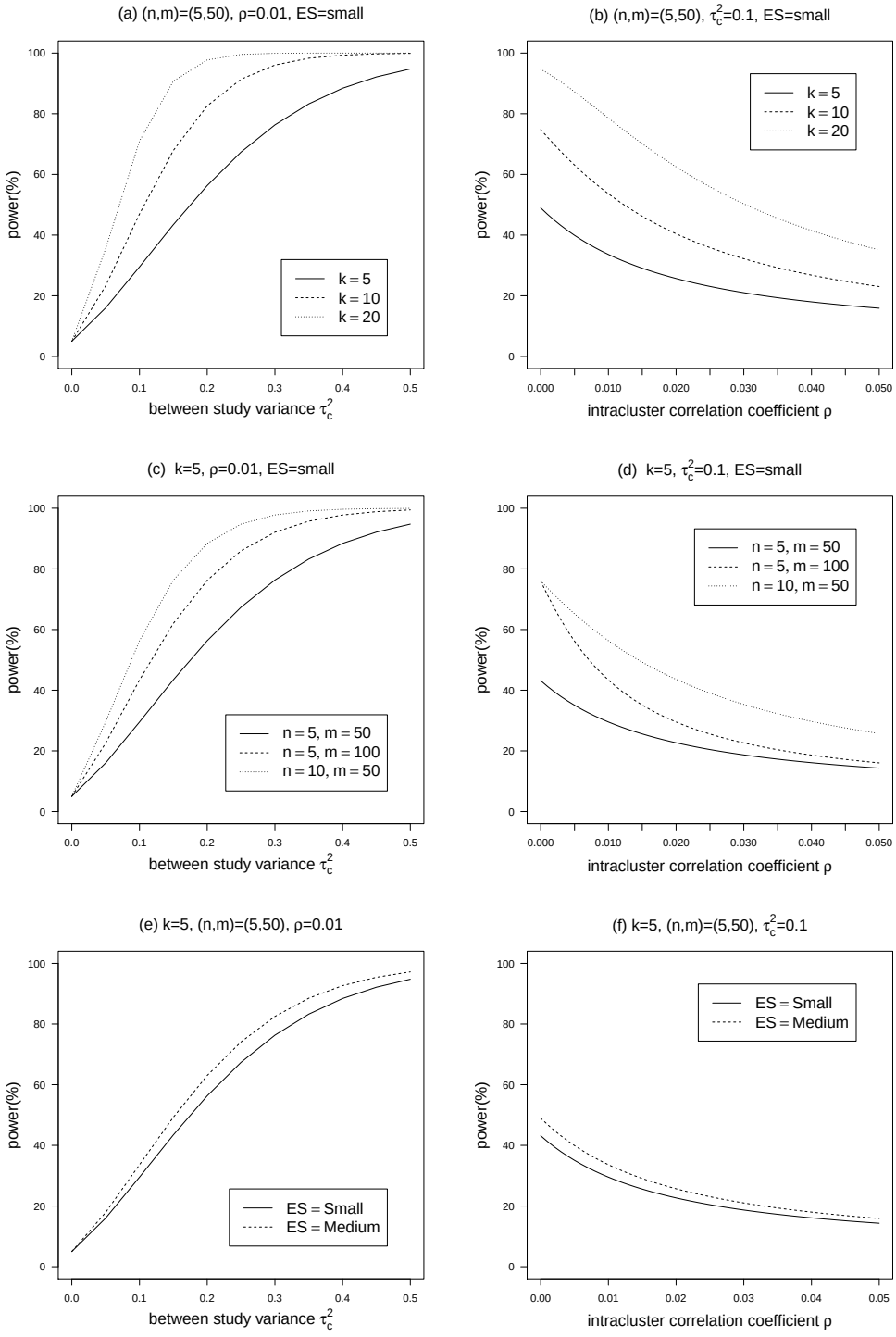

Figure 2.1: Approximate power of Q_a plotted against τ_c^2 (first column) and ρ (second column). (a)-(b) varying numbers of trials k ($k = 5, 10, 20$); (c)-(d) varying number of clusters per trial n and cluster size m ($(n, m) = (5, 50), (5, 100), (10, 50)$); (e)-(f) varying effect size ES ($(P_1, P_2) = (0.1, 0.168), (0.1, 0.293)$).

Chapter 3

Heterogeneity variance estimation

3.1 Introduction

The fixed effects model described in Chapter 2 assumes homogeneity of intervention effects across the k trials. In contrast, the random effects model assumes that the observed trials are a random sample from a hypothetical population of trials. In order to account for the variation among trials, a random term known as heterogeneity variance is added to compute the weights in the random effects model; this tends to equalize the weights assigned to small and large trials. Subsequently, the random effects model may lead to wider confidence intervals for the overall intervention effect.

Heterogeneity variance is also used as a measure of heterogeneity in meta-analysis. Although heterogeneity variance may be solely limited to trials with the same effect measures (e.g. odds ratio, risk ratio and hazard ratio), its value does not depend on the number, or size of trials in a meta-analysis unlike the other measures such as the Q statistic (R*ü*cker et al., 2008).

The aim of this chapter is to extend existing approaches for estimating the heterogeneity variance of meta-analysis of individually randomized trials to meta-analysis of cluster

randomization trials. We begin by considering eight methods for estimating the heterogeneity variance. In addition to a point estimate, confidence intervals for the heterogeneity variance estimate may be useful, as they indicate its precision while also conveying all the information contained in the corresponding test of heterogeneity (Hardy and Thompson, 1996; Viechtbauer, 2007*a*). Moreover, such confidence intervals may be also used to construct confidence intervals for measures of heterogeneity (Higgins and Thompson, 2002), which will be discussed in Chapter 4.

The random effects model for cluster randomization trials is briefly described in Section 3.2. The eight approaches for estimating the heterogeneity variance adjusted for clustering and the six methods for constructing confidence intervals, which are introduced in Section 1.2.3, are discussed with corresponding mathematical expressions presented in Section 3.3 and 3.4, respectively. Furthermore, a simulation study conducted in order to assess the bias and mean square error of the adjusted heterogeneity estimators and the coverage probabilities of the confidence intervals appears in Chapter 5.

3.2 Random effects model

Let $\hat{\theta}_j$ denote the estimated intervention effect (experimental vs. control) on the log odds ratio of the study outcome for the jth trial, $j = 1, \ldots, k$. The random effects meta-analysis model (Whitehead, 2002, p.88) is given by

$$\hat{\theta}_j = \theta + \nu_j + \epsilon_j,$$

where θ is the true overall mean effect size. Also, two independent random effects included in the model are the random study effects and the error terms, denoted by ν_j and ϵ_j, respectively. Random study effects are assumed to be independently and normally distributed with mean 0 and variance τ_c^2 (i.e. $\nu_j \sim N(0, \tau_c^2)$) and similarly, the error terms

are assumed to be independently and normally distributed with mean 0 and variance σ_{jc}^2, (i.e. $\epsilon_j \sim N(0, \sigma_{jc}^2)$), where τ_c^2 is the between study component of variance also known as the heterogeneity variance and σ_{jc}^2 is the within study component.

In practice, the within study variance is estimated using equation (2.4) for cluster randomization trials ignoring the sampling errors within the trial. This practice is often used because the within study variance tends to be relatively small as compared to the between study variance. Therefore, the errors can be negligible. However, caution must be taken using the estimated within study variance as the true variance for trials with small overall sizes, where the large sample approximation may be questionable (Bohning et al., 2002; Brockwell and Gordon, 2001; Sidik and Jonkman, 2006). In this chapter, the focus will be restricted to estimating heterogeneity variance τ_c^2, with σ_{jc}^2 being assumed known.

3.3 Adjusted heterogeneity variance estimators τ_c^2

3.3.1 Variance component estimator (VC)

Hedges and Olkin (1985) proposed a simple approach to estimate the heterogeneity variance using a method similar to that for estimating the variance components in a random effects analysis of variance. Given the unweighted mean $\bar{\theta} = \sum_{j=1}^{k} \hat{\theta}_j / k$, the usual sample variance of $\hat{\theta}_j$ may be expressed as

$$S_\theta^2 = \frac{1}{k-1} \sum_{j=1}^{k} (\hat{\theta}_j - \bar{\theta})^2.$$

Then the expected value of S_θ^2 in terms of variance components is

$$E[S_\theta^2] = \tau_c^2 + \frac{1}{k} \sum_{j=1}^{k} \hat{\sigma}_{jc}^2,$$

since as noted in Section 3.2, σ_{jc}^2 is assumed known. The variance component estimate of τ_c^2 is obtained as

$$\hat{\tau}_{c.VC}^2 = \frac{1}{k-1}\sum_{j=1}^{k}(\hat{\theta}_j - \bar{\theta})^2 - \frac{1}{k}\sum_{j=1}^{k}\hat{\sigma}_{jc}^2. \tag{3.1}$$

Negative values of $\hat{\tau}_{c.VC}^2$ will be truncated (i.e. $\max\{0, \hat{\tau}_{c.VC}^2\}$). The variance component estimator is a method of moments estimator.

3.3.2 DerSimonian and Laird estimator (DL)

Another method of moments estimator is the DerSimonian and Laird type estimator. This estimator is also implemented in RevMan (DerSimonian and Laird, 1986) software, which is the software recommended by The Cochrane Collaboration. Let w_{jc} denote the adjusted weights. The expectation of the adjusted Q statistic in equation (2.7) is then calculated (the derivation is given by Whitehead (2002, p.90)) as

$$E[Q_a] = \tau_c^2\left(\sum_{j=1}^{k} w_{jc} - \frac{\sum_{j=1}^{k} w_{jc}^2}{\sum_{j=1}^{k} w_{jc}}\right) + \left(\sum_{j=1}^{k} w_{jc}\sigma_{jc}^2 - \frac{\sum_{j=1}^{k} w_{jc}^2\sigma_{jc}^2}{\sum_{j=1}^{k} w_{jc}}\right). \tag{3.2}$$

By equating the expression $\sum_{j=1}^{k} w_{jc}(\hat{\theta}_j - \hat{\theta}_c)^2$ to its expected value given by equation (3.2), solving for τ_c^2 and then substituting $\hat{\sigma}_{jc}$ for σ_{jc}, $j = 1, \ldots, k$, a general method of moment estimator for τ_c^2 without any particular weights assigned to the trials is as follows:

$$\hat{\tau}_c^2 = \frac{\sum_{j=1}^{k} \hat{w}_{jc}(\hat{\theta}_j - \hat{\theta}_c)^2 - \left(\sum_{j=1}^{k} \hat{w}_{jc}\hat{\sigma}_{jc}^2 - \sum_{j=1}^{k} \hat{w}_{jc}^2\hat{\sigma}_{jc}^2 / \sum_{j=1}^{k} \hat{w}_{jc}\right)}{\sum_{j=1}^{k} \hat{w}_{jc} - \sum_{j=1}^{k} \hat{w}_{jc}^2 / \sum_{j=1}^{k} \hat{w}_{jc}}. \tag{3.3}$$

A negative estimate of $\hat{\tau}_c^2$ is set to zero.

It is noted that the variance component estimator described in Section 3.3.1 and the DerSimonian and Laird estimator for τ_c^2 are special cases of the general method of moments estimator presented in equation (3.3) differing only in the choices of $\hat{w}_{jc}$ (DerSimonian and Kacker, 2007). For the variance component estimator, $\hat{w}_{jc} = 1/k$ where k is the number of trials. In this case, equation (3.3) is simply equal to $\hat{\tau}_{c.VC}^2$ of equation (3.1). On the other hand, the DerSimonian and Laird estimator may be obtained by assigning $\hat{w}_{jc} = 1/\hat{\sigma}_{jc}^2$ to equation (3.3), given by

$$\hat{\tau}_{c.DL}^2 = \frac{Q_a - (k-1)}{\sum_{j=1}^k \hat{w}_{jc} - \sum_{j=1}^k \hat{w}_{jc}^2 / \sum_{j=1}^k \hat{w}_{jc}}. \tag{3.4}$$

Furthermore, to improve performance, two two-step estimators to the one-step non-iterative procedures may be derived based on the variance component estimate $\hat{\tau}_{c.VC}^2$ in equation (3.1) and the DerSimonian and Laird estimate $\hat{\tau}_{c.DL}^2$ in equation (3.4) (DerSimonian and Kacker, 2007). Specifically, the first two-step estimate (DLVC) is obtained by assigning $\hat{w}_{jc.VC} = 1/(\hat{\sigma}_{jc}^2 + \hat{\tau}_{c.VC}^2)$ to $\hat{w}_{jc}$ in equation (3.3), given by

$$\begin{aligned} \hat{\tau}_{c.DLVC}^2 &= \frac{\sum_{j=1}^k \hat{w}_{jc.VC}(\hat{\theta}_j - \hat{\theta}_{c.VC})^2}{\sum_{j=1}^k \hat{w}_{jc.VC} - \sum_{j=1}^k \hat{w}_{jc.VC}^2 / \sum_{j=1}^k \hat{w}_{jc.VC}} \\ &- \frac{\sum_{j=1}^k \hat{w}_{jc.VC}\hat{\sigma}_{jc}^2 - \sum_{j=1}^k \hat{w}_{jc.VC}^2 \hat{\sigma}_{jc}^2 / \sum_{j=1}^k \hat{w}_{jc.VC}}{\sum_{j=1}^k \hat{w}_{jc.VC} - \sum_{j=1}^k \hat{w}_{jc.VC}^2 / \sum_{j=1}^k \hat{w}_{jc.VC}}, \end{aligned} \tag{3.5}$$

where $\hat{\theta}_{c.VC} = \sum_{j=1}^k \hat{w}_{jc.VC}\hat{\theta}_j / \sum_{j=1}^k \hat{w}_{jc.VC}$. Alternatively, the second two-step estimate (DL2) is obtained by assigning $\hat{w}_{jc.DL} = 1/(\hat{\sigma}_{jc}^2 + \hat{\tau}_{c.DL}^2)$ to w_{jc} in equation (3.3), given by

$$\hat{\tau}_{c.DL2}^2 = \frac{\sum_{j=1}^k \hat{w}_{jc.DL}(\hat{\theta}_j - \hat{\theta}_{c.DL})^2}{\sum_{j=1}^k \hat{w}_{jc.DL} - \sum_{j=1}^k \hat{w}_{jc.DL}^2 / \sum_{j=1}^k \hat{w}_{jc.DL}}$$

$$- \quad \frac{\sum_{j=1}^{k} \hat{w}_{jc.DL}\hat{\sigma}^2_{jc} - \sum_{j=1}^{k} \hat{w}^2_{jc.DL}\hat{\sigma}^2_{jc} / \sum_{j=1}^{k} \hat{w}_{jc.DL}}{\sum_{j=1}^{k} \hat{w}_{jc.DL} - \sum_{j=1}^{k} \hat{w}^2_{jc.DL} / \sum_{j=1}^{k} \hat{w}_{jc.DL}}, \tag{3.6}$$

where $\hat{\theta}_{c.DL} = \sum_{j=1}^{k} \hat{w}_{jc.DL}\hat{\theta}_j / \sum_{j=1}^{k} \hat{w}_{jc.DL}$.

3.3.3 Model error variance estimator (MV)

Consider the random effects meta-analysis model presented in Section 3.2 as a linear regression model with no covariates. To obtain the model error variance estimator of τ_c^2, we reparameterize the total variance of $\hat{\theta}_j$ (Sidik and Jonkman, 2005), such that the total variance $\hat{\sigma}^2_{jc}+\tau^2_c$ becomes $\tau^2_c(\hat{r}_{jc}+1)$, where $\hat{r}_{jc}$ denotes the ratio of $\hat{\sigma}^2_{jc}$ to τ^2_c with $\tau^2_c \neq 0$.

Now, letting $\hat{\boldsymbol{\theta}}$ be a vector of the elements $\hat{\theta}_1, \ldots, \hat{\theta}_k$, the expectation and the variance of $\hat{\boldsymbol{\theta}}$ are expressed in terms of matrices as $E(\hat{\boldsymbol{\theta}}) = \mathbf{X}\theta$ and $Var(\hat{\boldsymbol{\theta}}) = \tau^2_c\mathbf{V}$, respectively, where $\mathbf{X}$ is a vector of all ones and $\mathbf{V}$ is a diagonal matrix with $\hat{r}_{1c}+1, \ldots, \hat{r}_{kc}+1$. Therefore, the best linear unbiased estimator for the overall mean effect size θ is the weighted least squares estimator

$$\hat{\theta}_{vc} \quad = \quad (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\hat{\boldsymbol{\theta}} = \sum_{j=1}^{k} \hat{v}_{jc}^{-1}\hat{\theta}_j / \sum_{j=1}^{k} \hat{v}_{jc}^{-1}, \tag{3.7}$$

where $\hat{v}_{jc} = \hat{r}_{jc} + 1$. The estimated variance of $\hat{\theta}_{vc}$ is given by

$$\widehat{var(\hat{\theta}_{vc})} \quad = \quad (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\hat{\tau}^2_c = \hat{\tau}^2_c / \sum_{j=1}^{k} \hat{v}_{jc}^{-1}. \tag{3.8}$$

Analogous to the usual weighted least squares estimator (Dobson, 2002), an estimate of τ^2_c may be obtained as

$$\hat{\tau}^2_{c.MV} = \frac{(\hat{\boldsymbol{\theta}} - \mathbf{X}\hat{\theta}_{vc})^T \mathbf{V}^{-1} (\hat{\boldsymbol{\theta}} - \mathbf{X}\hat{\theta}_{vc})}{k-1} = \frac{1}{k-1} \sum_{j=1}^{k} \hat{v}_{jc}^{-1} (\hat{\theta}_j - \hat{\theta}_{vc})^2. \quad (3.9)$$

In order to compute $\hat{r}_{jc}$, an initial estimate of τ_c^2 is required, denoted by $\hat{\tau}_o^2$. The commonly used one is the empirical variance estimate given by $\hat{\tau}_o^2 = \sum_{j=1}^{k} (\hat{\theta}_j - \bar{\theta})^2/k$, where the unweighted overall mean effect size is $\bar{\theta} = \sum_{j=1}^{k} \hat{\theta}_j/k$. According to Sidik and Jonkman (2005), this initial estimate works reasonably well for estimating moderate and large values of the heterogeneity variance. Unlike the other estimators, $\hat{\tau}^2_{c.MV}$ will always yield a nonnegative value.

In addition, Sidik and Jonkman (2007) proposed an improved version of the model error variance estimator obtained by replacing $\hat{\tau}_o^2$ with $\hat{\tau}^2_{c.VC}$ in (3.1). This improved model error variance estimate (MVVC) of τ_c^2 is referred to as $\hat{\tau}^2_{c.MVVC}$.

3.3.4 Maximum likelihood estimator (ML)

The maximum likelihood approach requiring an iterative numerical solution may also be used to estimate the adjusted heterogeneity variance (DerSimonian and Laird, 1986; Hardy and Thompson, 1996; Harville, 1977). Given that the marginal distribution of $\hat{\theta}_j$ is assumed to be normally distributed with mean θ and variance $\hat{\sigma}_{jc}^2 + \tau_c^2$, the log likelihood function is given by

$$lnL(\theta, \tau_c^2) = -\frac{k}{2} ln 2\pi + \frac{1}{2} \sum_{j=1}^{k} ln(\hat{w}_{jc}^*) - \frac{1}{2} \sum_{j=1}^{k} \hat{w}_{jc}^* (\hat{\theta}_j - \theta)^2, \quad (3.10)$$

where $\hat{w}_{jc}^* = 1/(\hat{\sigma}_{jc}^2 + \tau_c^2)$. By setting the first derivative of $lnL(\theta, \tau_c^2)$ with respect to τ_c^2 equal to zero, the maximum likelihood estimate of τ_c^2 (for details see Appendix D) is

given by

$$\hat{\tau}^2_{c.ML} = \frac{\sum_{j=1}^{k} \hat{w}^2_{jc.ML}\{(\hat{\theta}_j - \hat{\theta}_{c.ML})^2 - \hat{\sigma}^2_{jc}\}}{\sum_{j=1}^{k} \hat{w}^2_{jc.ML}}, \tag{3.11}$$

where $\hat{w}_{jc.ML} = 1/(\hat{\sigma}^2_{jc} + \hat{\tau}^2_{c.ML})$ and $\hat{\theta}_{c.ML} = \sum_{j=1}^{k} \hat{w}_{jc.ML}\hat{\theta}_j / \sum_{j=1}^{k} \hat{w}_{jc.ML}$. In equation (3.11), $\hat{\tau}^2_{c.ML}$ may be obtained iteratively with an initial value of $\hat{\tau}^2_{c.ML} = 0$. At each iteration, a positive value of $\hat{\tau}^2_{c.ML}$ is assured by setting the negative values to zero until convergence is reached.

3.3.5 Restricted maximum likelihood estimator (REML)

The restricted maximum likelihood estimator is often recommended over the maximum likelihood estimator, which tends to underestimate the variances (Harville, 1977). By modifying the log likelihood function for the ML estimator in (3.10), the log likelihood function for the REML estimator becomes

$$lnL_R(\theta) = -\frac{k}{2} ln 2\pi + \frac{1}{2}\sum_{j=1}^{k} ln(\hat{w}^*_{jc}) - \frac{1}{2}\sum_{j=1}^{k} \hat{w}^*_{jc}(\hat{\theta}_j - \theta)^2 - \frac{1}{2} ln \sum_{j=1}^{k} \hat{w}^*_{jc}, \tag{3.12}$$

where $\hat{w}^*_{jc} = 1/(\hat{\sigma}^2_{jc} + \tau^2_c)$. Similar to the ML estimator, by setting the first derivative of $lnL_R(\theta)$ with respect to τ^2_c equal to zero, the REML estimate of τ^2_c (for details see Appendix E) is

$$\hat{\tau}^2_{c.RE} = \frac{\sum_{j=1}^{k} \hat{w}^2_{jc.RE}\{(\hat{\theta}_j - \hat{\theta}_{c.RE})^2 + (1/\sum_{j=1}^{k} \hat{w}_{jc.RE}) - \hat{\sigma}^2_{jc}\}}{\sum_{j=1}^{k} \hat{w}^2_{jc.RE}}, \tag{3.13}$$

where $\hat{w}_{jc.RE} = 1/(\hat{\sigma}^2_{jc} + \hat{\tau}^2_{c.RE})$ and $\hat{\theta}_{c.RE} = \sum_{j=1}^{k} \hat{w}_{jc.RE}\hat{\theta}_j / \sum_{j=1}^{k} \hat{w}_{jc.RE}$. Similarly, with

an initial value of $\hat{\tau}^2_{c.RE} = 0$, $\hat{\tau}^2_{c.RE}$ in (3.13) may be obtained iteratively. For each iteration, negative values are truncated until convergence.

3.4 Confidence intervals for τ_c^2

3.4.1 Q profile confidence intervals

Q profile confidence intervals for τ_c^2 are constructed based on the distribution of the adjusted Q statistic, followed by $P(\chi^2_{k-1,0.025} \leq Q(\tau_c^2) \leq \chi^2_{k-1,0.975})$ where $\chi^2_{k-1,0.025}$ and $\chi^2_{k-1,0.975}$ denote the 2.5th and 97.5th percentiles of a χ^2 distribution with $k-1$ degrees of freedom, respectively. Thus, the lower and upper bounds of a 95 percent confidence interval for $\hat{\tau}_c^2$ is determined by solving for $\tilde{\tau}_c^2$ from

$$\left(Q(\tilde{\tau}_c^2) = \chi^2_{k-1,0.975}, \quad Q(\tilde{\tau}_c^2) = \chi^2_{k-1,0.025}\right). \tag{3.14}$$

The iterative procedure is used by repeatedly computing $Q(\tilde{\tau}_c^2)$ with increasing values of $\hat{\tau}_c^2$ until the critical values of χ^2 distribution are reached. A lower bound with negative values, which is outside of the parameter space, is truncated to zero in order to ensure a positive confidence interval. When $Q(\tau_c^2 = 0) \leq \chi^2_{k-1,0.025}$, the upper bound is set equal to the null set.

3.4.2 Biggerstaff-Tweedie confidence intervals

The Biggerstaff-Tweedie confidence intervals are constructed based on an approximation of the distribution of the adjusted Q statistic, the gamma distribution with shape γ and scale parameter ϕ. More specifically, the shape and scale parameters are defined as $\gamma(\tau_c^2) = E(Q_a)^2/var(Q_a)$ and $\phi(\tau_c^2) = var(Q_a)/E(Q_a)$, respectively as functions of the

expected value and variance of the adjusted Q statistic, which are expressed as

$$E(Q_a) \quad = \quad (k-1) + \left(s_1 - \frac{s_2}{s_1}\right)\tau_c^2, \tag{3.15}$$

and

$$var(Q_a) \quad = \quad 2(k-1) + 4\left(s_1 - \frac{s_2}{s_1}\right)\tau_c^2 + 2\left(s_2 - 2\frac{s_3}{s_1} + \frac{s_2^2}{s_1^2}\right)\tau_c^4, \tag{3.16}$$

respectively, where $s_t = \sum_{j=1}^{k} w_{jc}^t$ (the proof is given by Biggerstaff and Tweedie (1997)). The lower and upper bounds of a 95 percent confidence interval for τ_c^2 can be obtained by finding those two values of $\tilde{\tau}_c^2$ such that

$$\int_{Q/\phi(\tilde{\tau}_c^2)}^{\infty} f(x/\gamma(\tilde{\tau}_c^2))dx \quad = \quad 0.025 \tag{3.17}$$

and

$$\int_{0}^{Q/\phi(\tilde{\tau}_c^2)} f(x/\gamma(\tilde{\tau}_c^2))dx \quad = \quad 0.025, \tag{3.18}$$

where $f(x/\gamma(\tilde{\tau}_c^2))$ denotes the density function of a gamma distribution with shape parameter $\gamma(\hat{\tau}_c^2)$ and scale parameter 1. This approach yields non-negative values. Similar to the Q profile approach, it requires iteratively inputting monotonic increasing values of τ_c^2 until conditions (3.17) and (3.18) are satisfied for the lower and upper bounds, respectively. Also, when the negative upper bound is obtained, where the second integral (3.18) is smaller than 0.025, the interval is set to be null. However, the accuracy of this approach relies on the approximation of the gamma distribution to the true distribution of the Q_a statistic which follows a chi square with $k-1$ degrees of freedom when $\tau_c^2 = 0$

but a noncentral chi square with $k-1$ degrees of freedom when $\tau_c^2 \neq 0$.

3.4.3 Profile likelihood confidence intervals

The profile likelihood approach is mainly used for constructing confidence intervals for the ML estimator in (3.11) and the REML estimator in equation (3.13). Given the log-likelihood function of θ and τ_c^2, $lnL(\theta, \tau_c^2)$ in equation (3.10), and the restricted log-likelihood function of τ_c^2, $lnL_R(\theta)$ in equation (3.12), the 95 percent confidence intervals for $\hat{\tau}_{c.ML}^2$ in equation (3.11) and $\hat{\tau}_{c.RE}^2$ in equation (3.13) are given by a set of $\hat{\tau_c}^2$ values satisfying the following conditions

$$lnL(\theta, \tau_c^2) \geq lnL(\hat{\theta}_{c.ML}, \hat{\tau}_{c.ML}^2) - 3.84/2 \tag{3.19}$$

$$lnL_R(\tau_c^2) \geq lnL_R(\hat{\tau}_{c.RE}^2) - 3.84/2 \tag{3.20}$$

The value of 3.84 is the 5% point of the chi-square distribution with one degree of freedom. The lower and upper bounds for $\tau_{c.ML}^2$ and $\tau_{c.RE}^2$ may be found iteratively by substituting values into equations (3.19) and (3.20), respectively, until convergence is reached. Since $\tau_{c.ML}^2$ and $\tau_{c.RE}^2$ always yield nonnegative values with truncation, the lower bound of the profile likelihood confidence intervals remains nonnegative, followed by the positive upper bound. Unlike the likelihood functions, the profile likelihood function takes into consideration that τ_c^2 is estimated while varying values of the overall mean $\hat{\theta}_c$.

3.4.4 Wald-type confidence intervals

Wald-type confidence intervals are another option for the maximum likelihood estimators. Such confidence intervals are constructed by taking inverse elements of the Fisher information matrix as the asymptotic sampling variances of the ML and REML estimates.

The 95 percent Wald-type confidence intervals for $\hat{\tau}^2_{c.ML}$ in equation (3.11) and $\hat{\tau}^2_{c.RE}$ in equation (3.13) are then given by

$$\begin{aligned} \hat{\tau}^2_{c.ML} &\pm 1.96\sqrt{\hat{var}(\hat{\tau}^2_{c.ML})} \\ \hat{\tau}^2_{c.RE} &\pm 1.96\sqrt{\hat{var}(\hat{\tau}^2_{c.RE})} \end{aligned}$$

where

$$\begin{aligned} \hat{var}(\hat{\tau}^2_{c.ML}) &= 2\left(\sum_{j=1}^{k} \hat{w}^2_{jc.ML}\right)^{-1} \\ \hat{var}(\hat{\tau}^2_{c.RE}) &= 2\left(\sum_{j=1}^{k} \hat{w}^2_{jc.RE} - 2\frac{\sum_{j=1}^{k} \hat{w}^3_{jc.RE}}{\sum_{j=1}^{k} \hat{w}_{jc.RE}} + \left(\frac{\sum_{j=1}^{k} \hat{w}^2_{jc.RE}}{\sum_{j=1}^{k} \hat{w}_{jc.RE}}\right)^2\right)^{-1} \end{aligned}$$

with $\hat{w}_{jc.ML} = 1/(\hat{\sigma}^2_{jc} + \hat{\tau}^2_{c.ML})$ and $\hat{w}_{jc.RE} = 1/(\hat{\sigma}^2_{jc} + \hat{\tau}^2_{c.RE})$, respectively. Details on how the sampling variances are obtained can be found in Appendix A and Appendix B, respectively. The negative lower bound are suggested not be truncated to zero to preserve the precision of the τ^2_c estimate. The upper bound will always be positive for the confidence interval are constructed around a non-negative τ^2_c.

3.4.5 Sidik-Jonkman confidence intervals

Sidik-Jonkman confidence intervals may be constructed for the model error variance heterogeneity estimator. Given that $(k-1)\hat{\tau}^2_{c.MV}/(\sum_{j=1}^{k}(\theta_j - \bar{\theta})^2/k)$ approximately follows a χ^2 distribution with $k-1$ degrees of freedom, a 95 percent confidence interval for $\hat{\tau}^2_{c.MV}$ in equation (3.9) can be obtained as

$$\left(\frac{(k-1)\hat{\tau}^2_{c.MV}}{\chi^2_{k-1,0.975}}, \quad \frac{(k-1)\hat{\tau}^2_{c.MV}}{\chi^2_{k-1,0.025}} \right)$$

This approach is straightforward and does not require iterative solutions. Moreover, the estimates of τ_c^2 obtained by using Sidik-Jonkman are always greater than zero. Thus, the lower and upper bound of Sidik-Jonkman confidence intervals are also greater than zero.

3.4.6 Nonparametric bootstraps confidence intervals

The advantage of the nonparametric bootstraps approach is that it does not require any distributional assumptions for the estimators. In addition, it is relatively simple to implement. A set of 1000 nonparametric bootstraps samples can be obtained by sampling with replacement the same number of observations as in the original dataset consisting of θ_j and the corresponding $\hat{\sigma}^2_{jc}$. For each bootstraps sample, τ_c^2 can be estimated using the DerSimonian and Laird method, where we denote the resulting estimate as $\hat{\tau}^2_b$. Repeating this process 1000 times, a 95 percent confidence interval for τ_c^2 is then given by the 2.5th and 97.5th empirical percentiles of the 1000 $\hat{\tau}^2_b$ values. Since the DerSimonian and Laird method is applied, which has the negative values truncated to zero, the confidence interval will be positive.

3.5 Summary

The approaches for point and interval estimation of τ_c^2 are summarized in Table 3.1. Overall, eight heterogeneity variance estimators have been described, including the four noniterative estimators VC, DL, MV and MVVC, the two two-step estimators DLVC and DL2 and finally the two iterative estimators ML and REML. Most of the estimators will

yield nonnegative estimates with truncation at zero, while the adjusted MV and improved MV will always give positive τ_c^2 estimates without truncation.

In addition to the point estimation, a total of six approaches for constructing confidence intervals are provided including the Q profile, Biggerstaff-Tweedie, profile likelihood, Wald-type, Sidik and Jonkman and nonparametric bootstraps confidence intervals. The iterative approaches are Q profile, Biggerstaff-Tweedie, profile likelihood, and nonparametric, while the rest of confidence interval approaches are noniterative. The Q profile and Biggerstaff-Tweedie approaches are mainly used to construct confidence intervals for the method of moments estimators. The profile likelihood and Wald-type approaches are used to constructed confidence intervals for the ML and REML estimators. Finally, the Sidik-Jonkman approach is uniquely designed to construct confidence intervals for the Sidik-Jonkman estimator. Most of the confidence intervals require truncation at zero, except that the Wald-type confidence interval allows a negative lower bound to enhance the precision of the τ_c^2 estimate and the Sidik and Jonkman confidence interval always has a lower bound greater than zero.

The performance of the proposed methods has not yet been compared for meta-analysis of cluster randomized trials. Therefore, we will conduct a simulation study evaluating the adjusted heterogeneity variance estimators in terms of bias and mean square errors and the confidence interval approaches in terms of coverage, tail errors and interval width in Chapter 5.

Table 3.1: Summary of the adjusted heterogeneity variance estimators with the methods of constructing confidence intervals

Estimator			**Confidence Intervals**
Non-Iterative			
Variance Component (VC)	$\hat{\tau}^2_{c.VC}$	(3.1)	Q Profile, Biggerstaff-Tweedie
DerSimonian and Laird (DL)	$\hat{\tau}^2_{c.DL}$	(3.4)	Q Profile, Biggerstaff-Tweedie
Two-step DL based on VC (DLVC)	$\hat{\tau}^2_{c.DLVC}$	(3.5)	Q Profile, Biggerstaff-Tweedie
Two-step DL based on DL (DL2)	$\hat{\tau}^2_{c.DL2}$	(3.6)	Q Profile, Biggerstaff-Tweedie
Model Error Variance (MV)	$\hat{\tau}^2_{c.MV}$	(3.9)	Sidik-Jonkman
Improved based on VC (MVVC)	$\hat{\tau}^2_{c.MVVC}$	(3.9)	Sidik-Jonkman
Iterative			
Maximum Likelihood (ML)	$\hat{\tau}^2_{c.ML}$	(3.11)	Profile likelihood, Wald-type
Restricted (REML)	$\hat{\tau}^2_{c.RE}$	(3.13)	Profile likelihood, Wald-type

Chapter 4

Measures of heterogeneity

4.1 Introduction

Heterogeneity of the intervention effect in a meta-analysis of cluster randomized trials may be assessed using the adjusted Q statistic or by estimating the adjusted heterogeneity variance. However, the power of the adjusted Q statistic depends on the number of trials included in the meta-analysis, while estimation of the heterogeneity variance is limited to trials using the same intervention effect measures.

Higgins and Thompson (2002) developed three statistics (i.e. H, R and I^2), known as measures of heterogeneity, that avoid these two above-mentioned shortcomings. These three statistics, described earlier in Section 1.2.4, provide intuitive interpretations allowing comparisons across meta-analyses regardless of the number of trials, the type of outcome data (e.g. dichotomous, quantitative, or time to event) and the choice of intervention effect measure (e.g. odds ratio or hazard ratio). Consequently, these three statistics have been adopted by many researchers for quantifying heterogeneity across trials. Due to this wide usage, the I^2 statistic is now recommended in the guidelines for conducting meta-analysis provided by The Cochrane Collaboration, an international network for maintaining and ensuring the accessibility of systematic reviews in health care.

In addition to the estimated statistics, the confidence intervals are informative in summarizing precision by providing a range of values that reflect the degree of uncertainty in the estimation procedure. Moreover, the use of confidence intervals in presenting research results is usually recommended in reporting guidelines (e.g. the CONSORT statement (Altman et al., 2001)).

The objectives of this chapter are to adapt H, R and I^2 to the meta-analysis of cluster randomization trials and to modify existing approaches for constructing confidence intervals for the adjusted statistics. Specifically, the formula, interpretation and properties of the adjusted H, R and I^2 statistics are explicitly described in Section 4.2. Confidence intervals for the adjusted statistics are presented in Section 4.3. Key results are summarized in Section 4.4.

Small sample properties and confidence intervals for adjusted H, R and I^2 will be evaluated by simulation in Chapter 5.

4.2 Measures of Heterogeneity

4.2.1 Quantifying Heterogeneity

The measures of heterogeneity developed for individual randomized trials were derived based on three criteria (Higgins and Thompson, 2002). The first criterion is that the measure depends on the extent of heterogeneity (i.e. τ_c^2). The second criterion requires the measure to be scale invariant, allowing comparisons across meta-analyses involving different outcome data (e.g. lbs vs. kg or odds ratio vs. hazard ratio). The third criterion requires the measure to be invariant to the number of trials.

The first and second criteria suggest that the measures should be monotonically increasing with the between study variance τ_c^2 and depend on the within study variance σ_c^2, but not depend on the choice of intervention effect and the number of trials (according to the second and third criteria, respectively). Therefore, the measures proposed by Higgins and Thompson (2002) must be a monotonic increasing function of $\gamma = \tau_c^2/\sigma_c^2$.

The assumption of equal within study variances across trials is often applied to obtain a common within study variance. However, this assumption may not hold when binary outcomes are considered. The assumption of equal within study variances, equivalent to assuming the same disease risk, may not be realistic because the trials with the same disease risk are mostly likely to be replicates (i.e. the disease rates are for each trial). Alternatively, the individual within study variances may be summarized to obtain an overall within study variance referred to as a typical within study variance $\hat{\sigma}_{wc}^2$ by Higgins and Thompson (2002). There have been two suggestions for estimating a typical within study variance.

One is to estimate a typical within study variance using the harmonic mean, i.e., the reciprocal of the arithmetic mean weight (Takkouche et al., 1999), given by

$$\hat{\sigma}_{wc,1}^2 = k/\sum_{i=1}^{k} \hat{w}_{jc}. \tag{4.1}$$

where $\hat{w}_{jc} = 1/\hat{\sigma}_{jc}^2$. The within study variance σ_{jc}^2 may be estimated using equation (2.4). A second possibility discussed by Higgins and Thompson (2002) is given by

$$\hat{\sigma}_{wc,2}^2 = \frac{(k-1)\sum_{i=1}^{k} \hat{w}_{jc}}{\left(\sum_{i=1}^{k} \hat{w}_{jc}\right)^2 - \sum_{i=1}^{k} \hat{w}_{jc}^2}. \tag{4.2}$$

According to Mittlböck and Heinzl's (2006) simulation results for individual randomized trials, $\hat{\sigma}^2_{wc,2}$ is preferable because it is derived from the expectation of Q_a. However, $\hat{\sigma}^2_{wc,2}$ is approximately equal to $\hat{\sigma}^2_{wc,1}$ when there is little variation in the within study variances.

4.2.2 Adjusted H statistic

One approach that meets the three proposed criteria is to calculate the adjusted H statistic, given by

$$H_a = \sqrt{\frac{\hat{\tau}^2_c + \hat{\sigma}^2_{wc}}{\hat{\sigma}^2_{wc}}} \tag{4.3}$$

When the DerSimonian and Laird estimator $\hat{\tau}^2_{c.DL}$ in equation (3.4) and the typical within study variance $\hat{\sigma}^2_{wc,2}$ in equation (4.2) are used to estimate $\hat{\tau}^2_c$ and σ^2_{wc} in equation (4.3), respectively, the adjusted H is expressed as

$$H_a = \sqrt{\frac{\hat{\tau}^2_{c.DL} + \hat{\sigma}^2_{wc.2}}{\hat{\sigma}^2_{wc.2}}} = \sqrt{\frac{Q_a}{k-1}} \tag{4.4}$$

The adjusted H statistic describes the square root of the relative excess in Q_a over its degrees of freedom. Specifically, the H_a values reflect the relation of between to within study variance. For instance, when the between study variance is equal zero or in the case of homogeneity, H_a has a value of 1. A rough guideline will be that values exceeding 1.5 may suggest heterogeneity among trials and values below 1.2 may suggest little heterogeneity.

The next step is to investigate the values of H_a as a function of parameters including the number of trials, number of clusters per group, cluster size, intracluster correlation

coefficient and degree of heterogeneity. For simplicity, the focus will be restricted to the case of an equal number of clusters per group (i.e. $N_j = N_{1j} = N_{2j}$) having constant cluster size m. The measure of degree of heterogeneity is defined in Section 5.3.

The simulation results are illustrated in Figure 4.1 where values of H_a are plotted against the degree of heterogeneity, while varying the intracluster correlation coefficient, number of trials, number of clusters per trial, and cluster size. The plotted lines indicate the values of $H_a = 1.2$ and $H_a = 1.5$ corresponding to the boundaries for 'low' to 'moderate', and 'moderate' to 'high' heterogeneity, respectively. Unlike the adjusted Q statistic, the value of H_a does not intrinsically depend on the number of trials and increases with the degree of heterogeneity. However, the variability of H_a across different values of ρ is large when the number of trials is small and the suggested guideline may not be applicable. Consequently, it becomes difficult to distinguish 'moderate' heterogeneity from chance. The variability of H_a across different ρ is slightly reduced in the case of a large number of small clusters (e.g. $(N_j, m) = (40, 100)$) compared to a small number of large clusters (e.g. $(N_j, m) = (20, 200)$) for a fixed sample size per trial and significantly reduced when the number of trials increases (e.g. to $k = 20$ or $k = 40$).

4.2.3 Adjusted R statistic

The adjusted R statistic is an alternative to equation (4.3). Let the ML estimate in equation (3.4) be used to estimating the heterogeneity variance. It uses the estimated variances of the estimated intervention effect under the fixed and random effects model denoted by $\hat{v}_F$ and $\hat{v}_R$, respectively (Higgins and Thompson, 2002), and given by

$$R_a = \sqrt{\frac{\hat{v}_F}{\hat{v}_R}} = \sqrt{\frac{\sum_{j=1}^{k} \hat{w}_{jc}}{\sum_{j=1}^{k} \hat{w}^*_{jc}}}, \tag{4.5}$$

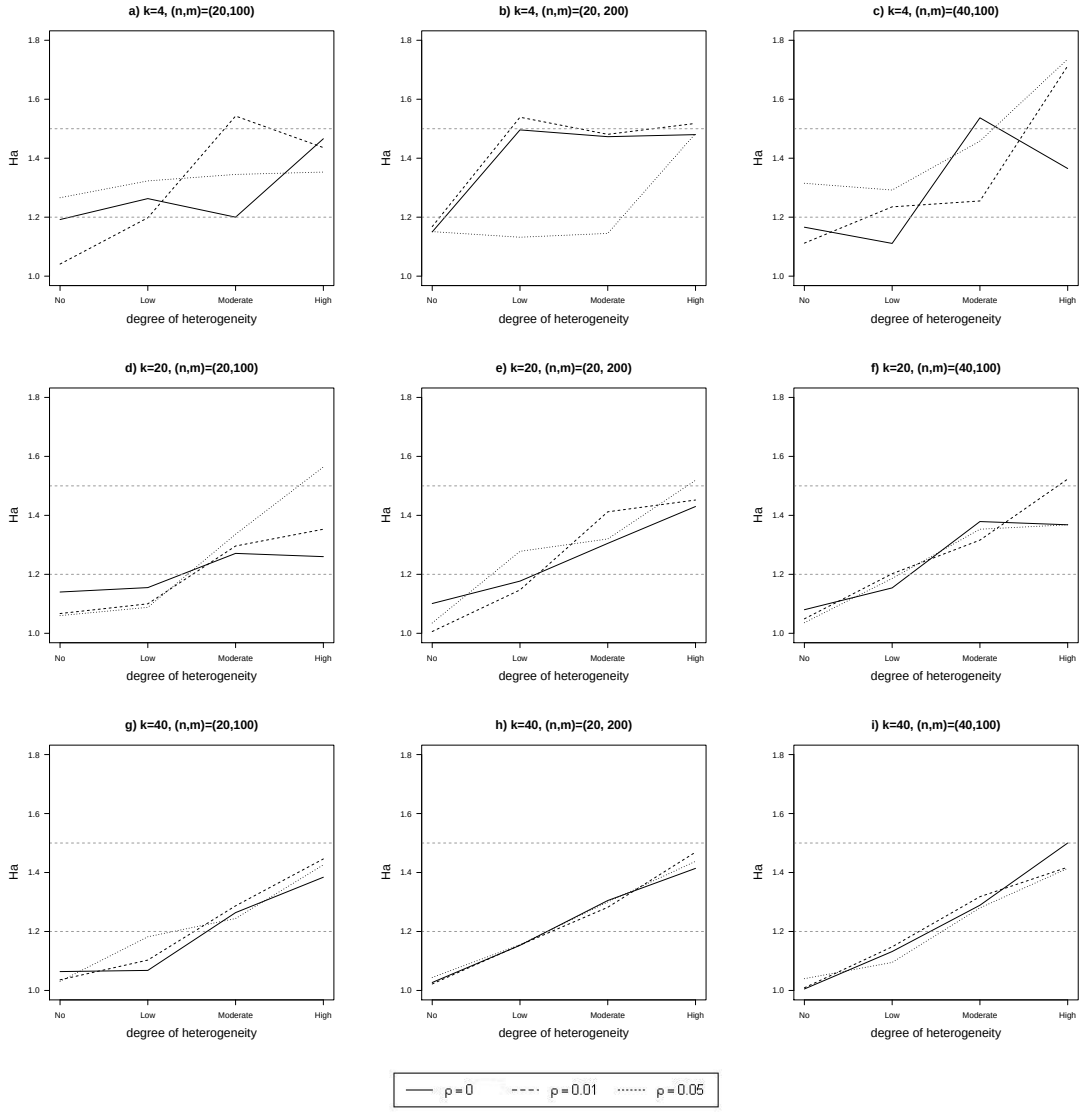

Figure 4.1: Estimated H_a plotted against degree of heterogeneity ('no', 'low', 'moderate', 'high') varying numbers of trials k ($k = 4, 20, 40$), number of clusters per group N_j and cluster size m ($(N_j, m) = (20, 100), (20, 200), (40, 100)$) and intracluster correlation coefficient ρ ($\rho = 0, 0.01, 0.05$). Plotted lines indicate the values of $H_a = 1.2$ and $H_a = 1.5$.

where $\hat{w}_{jc} = 1/\hat{\sigma}^2_{jc}$ and $\hat{w}^*_{jc} = 1/(\hat{\tau}^2_{c.DL} + \hat{\sigma}^2_{jc})$. The adjusted R statistic describes the inflation in the confidence interval for a single summary estimate under a random effects model compared with a fixed effects model. A value of 1 indicates identical inferences under the two models in which the intervention effects are homogeneous. The properties for R_a are similar to those for H_a since they estimate the same measure (i.e. $\gamma + 1$) using different approaches. When both estimates have equal precision, values of H_a are equal to values of R_a.

4.2.4 Adjusted I^2 statistic

The adjusted I^2 statistic may be expressed as

$$I^2_a = \frac{\hat{\tau}^2_{c.DL}}{\hat{\tau}^2_{c.DL} + \hat{\sigma}^2_{wc.2}} = \frac{H^2_a - 1}{H^2_a} = \frac{Q_a - (k-1)}{Q_a} \tag{4.6}$$

in terms of either H_a or Q_a and k. The adjusted I^2 statistic describes the total variation across trials due to heterogeneity. It may also be considered as a measure of inconsistency, since it depends on the extent of overlap in confidence intervals across studies (Higgins, 2008). Values of I^2_a are normally expressed as a percentage with a range from 0 to 100, where a value of zero percent indicates no observed heterogeneity. A general guideline for 'low', 'moderate' and 'high' heterogeneity correspond to 25%, 50% and 75% values of I^2_a, respectively. Alternatively, a value of I^2_a greater than 50% may be considered as substantial heterogeneity (Higgins and Green, 2008; Higgins et al., 2002*a*).

Note that values of I^2_a should be interpreted with caution when the number of subjects in the trials and the number of trials in the meta-analysis are low (Huedo-Medina et al., 2006; Mittlböck and Heinzl, 2006; Rücker et al., 2008).

4.3 Confidence intervals

Methods of constructing confidence intervals for the proposed measures include the method of variance estimates recovery (MOVER), using the distribution of the adjusted Q statistic, statistical significance of the adjusted Q statistic, estimation of an adjusted heterogeneity variance estimator, and a nonparametric bootstraps procedure. Following the approach of several previous investigations (Higgins and Thompson, 2002), all approaches except MOVER proceed as if the within study variance is known and the between study variance is unknown. Although the discussion is restricted to constructing a confidence interval for H_a, a confidence interval for I_a^2 can be easily computed using equation (4.6). Note that the approaches based the adjusted Q statistic may not be applicable to construct confidence intervals for the adjusted R statistic.

Assuming the adjusted weights are known, R_a is considered to be a function of τ_c^2; thus, it may be calculated based on the estimators listed in Chapter 3. As a result, the confidence intervals for R_a may be constructed using approaches similar to that of constructing a confidence interval for τ_c^2, as described in Section 3.4.

4.3.1 Intervals based on MOVER

The confidence intervals may be constructed for H_a using the method of variance estimates recovery (MOVER) (Zou, 2008) for a ratio. By rearranging equation (4.3), the adjusted H_a may be considered as a ratio as

$$H_a^2 - 1 \;=\; \frac{\hat{\tau}_{c.DL}^2}{\hat{\sigma}_{wc}^2}$$

The equations for constructing the confidence intervals (R_l, R_u) for a ratio (Donner and Zou, 2010) are given by

$$R_l = \frac{\hat{\theta}_1\hat{\theta}_2 - \sqrt{(\hat{\theta}_1\hat{\theta}_2)^2 - l_1u_2(2\hat{\theta}_1 - l_1)(2\hat{\theta}_2 - u_2)}}{u_2(2\hat{\theta}_2 - u_2)} \tag{4.7}$$

$$R_u = \frac{\hat{\theta}_1\hat{\theta}_2 + \sqrt{(\hat{\theta}_1\hat{\theta}_2)^2 - u_1l_2(2\hat{\theta}_1 - u_1)(2\hat{\theta}_2 - l_2)}}{l_2(2\hat{\theta}_2 - l_2)}. \tag{4.8}$$

Let $\hat{\theta}_1 = \hat{\tau}^2_{c.DL}$ of equation (3.11) and $\hat{\theta}_2 = \hat{\sigma}^2_{wc.1}$ of equation (4.2). Assuming there is relatively little variation in the within study variances, the typical within study variance $\hat{\sigma}^2_{wc.1}$ is used instead of $\hat{\sigma}^2_{wc.2}$ to avoid complex computations.

The confidence interval $[l_1, u_1]$ for $\hat{\tau}^2_{c.DL}$ may be obtained using the Q profile confidence intervals described in section 3.4.1. As for the confidence interval $[l_2, u_2]$ for $\hat{\sigma}^2_{wc.1}$, we will first define the large sample confidence interval for $\hat{w}_{jc} = 1/\hat{\sigma}_{jc}$ as

$$l_{2jc} = w_{jc}\chi_{\alpha/2,M_j-1}/(M_j - 1), \quad u_{2jc} = w_{jc}\chi_{1-\alpha/2,M_j-1}/(M_j - 1),$$

where M_j is the total number of subjects in trial j. Next, the confidence interval for $\sum_{j=1}^{k}\hat{w}_{jc}$ (Zou et al., 2009) is obtained as

$$l_2' = \sum_{j=1}^{k}\hat{w}_{jc} - \sqrt{\sum_{j=1}^{k}[\hat{w}_{jc} - l_{2jc}]^2}, \quad u_2' = \sum_{j=1}^{k}\hat{w}_{jc} - \sqrt{\sum_{j=1}^{k}[\hat{w}_{jc} - u_{2jc}]^2}.$$

Finally, the confidence interval for $\hat{\sigma}^2_{wc.1}$ is simply given by

$$l_2 = k/u_2', \quad u_2 = k/l_2'.$$

The confidence interval for $H_a - 1$ may be obtained by applying equations (4.7) and (4.8).

4.3.2 Intervals based on the distribution of Q_a

In Chapter 2, the adjusted Q statistic was shown to follow a noncentral chi square distribution under the alternative hypothesis that not all intervention effects are the same, with the variance of Q_a given by $2(k-1+2NC)$. The noncentrality parameter NC may be calculated using equation (2.10). As a result, a 95 percent confidence interval for H_a may be obtained as

$$\sqrt{\frac{1}{k-1}\left(Q_a \pm 1.96\sqrt{\hat{var}(Q_a)}\right)}, \tag{4.9}$$

which also can be recognized as a symmetric Wald-type confidence interval (Higgins and Thompson, 2002). Alternatively, the distribution of the adjusted Q statistic may be approximated by a gamma distribution where the variance of Q_a in equation (3.16) is used instead (Biggerstaff and Tweedie, 1997). This approach involves the estimation of quantiles from the cumulative distribution function of the gamma distribution, which requires more complex computation than the one based on a noncentral chi square distribution.

4.3.3 Intervals based on the statistical significance of Q_a

A simple method of constructing a confidence interval for H_a is derived from a test-based standard error for $ln(H_a)$ involving Q_a and k (Abramowitz and Stegun, 1965, formula 26.4.13). Intervals are of the form $exp(lnH_a \pm Z_\alpha \times SE(lnH_a))$ where Z_α is the $(1-\alpha/2)$ quantile of the standard normal distribution and a test-based standard error for $ln(H_a)$ is

$$SE(lnH_a) = \frac{1}{2}\frac{ln(Q_a) - ln(k-1)}{\sqrt{2Q_a} - \sqrt{2k-3}}$$

Whenever $H_a = 1$ or $Q_a \leq (k-1)$, the test-based standard error for H_a may be estimated (Abramowitz and Stegun, 1965, formula 26.4.13) by

$$SE_o(lnH) = \sqrt{\frac{1}{2(k-2)}\left(1 - \frac{1}{3(k-2)^2}\right)}.$$

4.3.4 Intervals based on the estimation of τ_c^2

A confidence interval for H_a may be easily calculated using the approaches of constructing a confidence interval for τ_c^2 illustrated in Chapter 3 by considering H_a as an estimate of

$$\eta_a = \sqrt{\frac{\left(\sum_{j=1}^{k} \hat{w}_{jc} - \sum_{j=1}^{k} \hat{w}_{jc}^2 / \sum_{j=1}^{k} \hat{w}_{jc}\right)\hat{\tau}_c^2}{k-1} + 1} \tag{4.10}$$

where $\hat{w}_{jc} = 1/\hat{\sigma}_{jc}^2$. For instance, if $\hat{\tau}_{c.DL}^2$ is used to replace $\hat{\tau}_c^2$ of equation (4.10), the Q profile approach described in section 3.4.1 may be applied to construct a confidence interval for η_a. The same analogy may be used for constructing a confidence interval for the adjusted R statistic as a function of τ_c^2.

4.3.5 Bootstraps confidence intervals

A nonparametric bootstraps confidence interval for H_a may be obtained by taking samples of size k with replacement from the pairs $(\theta_j, \sigma_{jc}^2)$ and calculating quantiles for the H_a statistic 1000 times. A nonparametric bootstraps approach is selected because it does not require any distributional assumptions and also because it is relatively simple to implement.

4.4 Summary

The adjusted H, R and I^2 statistics may be used to assess the presence of heterogeneity in the meta-analysis of cluster randomization trials in place of the test of heterogeneity. The disadvantage of the latter approach is that it may have low power because of a small number of studies and inappropriately high power with many studies. These measures of heterogeneity are usually recommended in guidelines for conducting meta-analyses because they do not intrinsically depend on the number of trials, at least for relatively large k, and also allow comparisons across meta-analyses with different outcomes. Table 4.4 summarizes the rough guideline for each statistic that corresponds to 'low', 'moderate' and 'high' heterogeneity. Moreover, the confidence intervals for the estimated statistics are more informative since they provide a range of values that reflects the degree of uncertainty in the estimation procedure. The confidence intervals for the estimated statistics may be constructed based on the method of variance estimates recovery (MOVER), the distribution of the adjusted Q statistic, the statistical significance of the adjusted Q statistic, the estimation of an adjusted heterogeneity variance estimator, or a nonparametric bootstraps procedure.

Chapter 5 will also present the evaluation of the performance of the estimated heterogeneity statistics in terms of bias and mean square error and the performance of the corresponding confidence interval approaches in terms of coverage, tail errors and interval width.

Table 4.1: Degree of Heterogeneity

Statistic	Low	Moderate	High
H_a or R_a	< 1.2	$1.2 - 1.5$	> 1.5
I_a^2	$< 25\%$	$25 - 50\%$	$> 50\%$

Chapter 5

Simulation study design

5.1 Introduction

The adjusted Q statistic was derived in Chapter 2 while the adjusted heterogeneity variance estimators and the adjusted measures of heterogeneity were described in Chapter 3 & 4, respectively. The purpose of this chapter is to report on the design of a simulation study evaluating the performances of these statistics.

Objectives for the simulation study are described in Section 5.2. Parameters to be considered are discussed and justified in Section 5.3, while procedures for generating clustered binary data are described in Section 5.4. Finally, the criteria for evaluating the performance of statistical approaches for selected scenarios are provided in Section 5.5. The design of the simulation study follows the guidelines proposed by Burton et al. (2006).

5.2 Objectives

A list of the methods being compared for each type of heterogeneity assessment is provided in Table 5.1. The specific objectives for the simulation study are

1. To evaluate the performance of the adjusted Q statistic in terms of Type I error and statistical power and to compare its power with the proposed formula.

2. To assess the bias and mean square error of the adjusted heterogeneity variance estimators and to evaluate the coverage, tail errors and interval width of the proposed confidence interval methods.

3. To assess the bias and mean square error of the adjusted measures of heterogeneity and to evaluate the coverage, tail errors and interval width of the proposed confidence interval methods.

5.3 Selection of parameters

Given that a fixed effects design is used, the performance of the statistical methods may depend on several factors, including the number of trials, number of clusters, cluster size (mean, variability), disease rates for the control group, intracluster correlation coefficient and degree of heterogeneity in the intervention effects. This section is devoted to justifying the choices of the parameters used in generating the correlated binary data by showing that they reflect practical scenarios encountered in the meta-analysis of cluster randomization trials. The focus is limited to completely randomized designs with a binary outcome.

The focus is further restricted to a meta-analysis of $k = 4, 12, 20, 40$ cluster randomization trials, each with two intervention groups, assuming an equal number of clusters ($N_j = n_{1j} = n_{2j}$) with constant cluster size (m) in each intervention group for each trial. The intervention effect was measured using the log odds ratio which is frequently used

Table 5.1: List of methods being compared for each type of heterogeneity assessment.

Heterogeneity assessment	Method
Heterogeneity variance (τ_c^2)	Variance component (VC) DerSimonian and Laird (DL) Two-step DL (DLVC) Two-step DL (DL2) Model error variance (MV) Improved model error variance (MVVC) Maximum likelihood (ML) Restricted maximum likelihood (REML)
Confidence intervals for τ_c^2	Q profile for DL Biggerstaff-Tweedie for DL Sidik-Jonkman for MV Nonparametric bootstraps for DL Profile likelihood for ML Profile likelihood for REML Wald-type for ML Wald-type for REML
Measures of heterogeneity	H_a R_a I_a^2
Confidence intervals for H_a	MOVER Based on distribution of Q_a Test-based Based on τ_c^2 Nonparametric bootstraps

in medical studies as compared to other types of measures (e.g. relative risks or risk difference) (Bland and Altman, 2000).

Two log odds ratios values were chosen, -0.36 and 0, equivalent to 0.7 and 1 in terms of the odds ratio (ψ), respectively. The results to $\psi = 0.7$ for an experimental group reducing risk are expected to be the same for an experimental group increasing risk with $\psi = 1.4$. Values for the parameters considered in the simulation study are summarized in Table 5.2.

Number of clusters, cluster size

In general, most cluster randomization trials tend to either have a large number of small clusters (e.g. family randomized trials) or a small number of large clusters (e.g. community intervention trials). In this study, we will focus on the latter case. Thus, the selected values for the number of clusters per group and mean cluster size (N_j, m) were (20, 100), (20, 200) and (40, 100) based on the previous simulation studies (Darlington and Donner, 2007; Donner et al., 1990; Donner and Klar, 1996; Eldridge et al., 2004).

Disease rates for the control group

We used two possible sets of disease rates for the control group (r_{1j}) for each k. For instance, the two sets at $k = 4$ were (0.35, 0.45, 0.50, 0.55) and (0.04, 0.07, 0.10, 0.13), which corresponded approximately to the disease rates for the control group used in the simulation study by Darlington and Donner (2007). The sets of disease rates for the control group at $k = 12$, 20 and 40 were the 3-, 5- and 10-fold versions of the two proposed sets of disease rates at $k = 4$, respectively.

Intracluster correlation coefficient

The intracluster correlation coefficient ρ, a measure of the similarity among individuals within the same cluster, tends to be small for larger clusters (e.g. community intervention trials). For example, 220 estimates of ANOVA-based intracluster correlation coefficients from 21 implementation trials limited to hospital and physician randomized trials had a median of 0.048 with a range from 0 to 0.415 (Campbell et al., 2005). In particular, the values of ρ were around 0.05 for primary care trials and were even smaller for binary outcomes (e.g. less than 0.01 for blood pressure $\geq$ 90 mm Hg in a study of hypertension screening and management (Bass et al., 1986)). Based on these findings, the values of ρ

were set to 0, 0.01 and 0.05, similar to those considered by Donner and Klar (1996).

Table 5.2: Simulation parameters for cluster randomization simulation study

Parameter	Values
Number of Studies (k)	4, 12, 20, 40
Odds Ratio ($\theta_j = log(\Psi_j)$)	0.7, 1.0
Number of clusters per group, cluster size (N_j, m)	(20,100), (20,200), (40,100)
Disease rates for the control group at $k = 4$ (r_{1j})	(0.04,0.07,0.10,0.13) (0.35,0.45,0.50,0.55)
Intracluster correlation coefficient (ρ)	0, 0.01, 0.05
Degree of heterogeneity (w) Correspond to	0, 0.33, 0.67, 1 No, Low, Moderate, High

Degree of heterogeneity

The noncentrality parameter NC may be approximated by $(k-1)$ times the ratio of between study variance to the within study variance, given in equation (2.13). In addition, the above ratio denoted by w is typically 0.33 and rarely exceeds one (Schmidt, 1992). Therefore, Hedges and Pigott (2001) suggested to use the convention that the values w =0, 0.33, 0.67 and 1, equivalent to 0%, 25%, 40% and 50% in I_a^2 (refer to Table 4.4), corresponding to 'no', 'low', 'moderate' and 'high' heterogeneity, respectively.

Given the previously mentioned parameters, the odds ratios used to generate the clustered binary data were selected such that the ratio of between study variance to the within study variance corresponded to different degrees of heterogeneity and the deviation of the intervention effect θ_j from the overall mean effect size were summed up to zero under the alternative hypothesis ($\sum_j^k \delta_j = 0$). The list of these odds ratios is given in Table 5.3.

Table 5.3: List of odds ratios to generate clustered binary datasets with odds ratio $\psi = 0.7, 1.0$ as the common overall effect size, disease rates $r_A = (0.04, 0.07, 0.10, 0.13)$, $r_B = (0.35, 0.45, 0.50, 0.55)$.

k	Odds Ratio	Disease rates	Degree of heterogeneity	OR			
4	0.7	r_A	no	0.700	0.700	0.700	0.700
			low	0.642	0.681	0.719	0.764
			medium	0.617	0.672	0.727	0.795
			large	0.599	0.666	0.733	0.819
		r_B	no	0.700	0.700	0.700	0.700
			low	0.670	0.690	0.709	0.730
			medium	0.657	0.685	0.714	0.745
			large	0.648	0.682	0.717	0.756
	1.0	r_A	no	1.000	1.000	1.000	1.000
			low	0.922	0.974	1.025	1.084
			medium	0.890	0.963	1.036	1.124
			large	0.867	0.955	1.044	1.155
		r_B	no	1.000	1.000	1.000	1.000
			low	0.958	0.986	1.013	1.043
			medium	0.940	0.980	1.019	1.063
			large	0.927	0.975	1.024	1.078

For $k = 12, 20, 40$, the odds ratios are 3-, 5- and 10-fold replicates of the odds ratios for $k = 4$.

5.4 Generation of data

Clustered binary data were generated using the method proposed by Lunn and Davies (1998) with an exchangeable correlation ρ. Thus, for cluster l in intervention group i of trial j, m observations were generated as $X_{ijkl} = (1 - U_{ijkl})Y_{ijkl} + U_{ijkl}Z_{ijkl}$ where $i = 1, 2$, $j = 1, \ldots, S$, $l = 1, \ldots, N_j$ and $k = 1, \ldots, m$. Let Y_{ijkl} and Z_{ijkl} be generated independently from the binomial distribution $B(1, r_{ij})$ where r_{ij} is the baseline disease rate for intervention group i in trial j. Let U_{ijkl} be generated from the binomial distribution $B(1, \sqrt{\rho})$. Given that the disease rates r_{1j} for the control group were fixed, the disease rates r_{2j} for the experimental group were then calculated using the following equation: $r_{2j} = r_{1j} \exp(\theta_j)/(1 - r_{1j} + r_{1j} \exp \theta_j)$ to maintain $\theta_j = \text{logit}(r_{2j})\text{-logit}(r_{1j})$, where θ_j

denotes the intervention effect taken as a log odds ratio. The number of events for cluster l in intervention group i of trial j is given by $a_{ijl} = \sum_{k=1}^{m} X_{ijkl}$. Based on a factorial design, the selected values for the different parameters summarized in Table 5.2 were simulated for 576 parameter combinations in total.

There were 1000 randomly generated datasets for each parameter combination. Given that the standard error is calculated as $\sqrt{0.05 \times (1 - 0.05)/1000}$ (Burton et al., 2006), the approximate 95% confidence interval for a five percent rejection rate was (0.036, 0.064) and the approximate 95% confidence interval for coverage probabilities with a 95 percent nominal level was (93.6, 96.3). Therefore, statistical tests which had Type I error rates less than 3.6% were overly conservative, and tests which had Type I error rates greater than 6.4% were overly liberal (e.g. Bradley (1978); Klar and Darlington (2004)). For the purpose of evaluating confidence interval coverage, coverage above 96.3% suggests that the results are conservative. In contrast, coverage below 93.6% indicates that the results are liberal. There are 144 parameter combinations for the case of 'no' heterogeneity ($w = 0$) and 432 parameter combinations for the case of heterogeneity ($w > 0$).

5.5 Evaluation criteria

The evaluation measures used to compare the performance of the proposed approaches include Type I error rate, statistical power, bias, mean square error, confidence interval coverage, tail errors and interval width. The detailed descriptions for each measure are provided as follows:

1. The Type I error rate is calculated as the proportion of simulation samples generated under the null hypothesis which have p-values less than or equal to the nominal 5 percent significance level. Attention is restricted to the 144 parameter combinations generated at $w = 0$.

2. The statistical power is calculated as the proportion of simulation samples generated under the alternative hypothesis which have p-values less than or equal to the nominal 5 percent significance level, given that the corresponding test statistic provides a valid Type I error rate. Attention is restricted to the 432 parameter combinations generated at $w > 0$.

3. Bias is calculated as the difference between the average of 1000 estimates and the true value (i.e. $bias = \hat{\bar{\tau}}_c^2 - \tau_c^2$). The amount of bias considered troublesome varies from $\frac{1}{2}SE(\hat{\tau}_c^2)$ and $2SE(\hat{\tau}_c^2)$.

4. The mean square error (MSE) is calculated as a function of bias and variability, given by $MSE = (\hat{\bar{\tau}}_c^2 - \tau_c^2)^2 + SE(\hat{\tau}_c^2)^2$.

5. The coverage of a confidence interval is calculated as the proportion of simulated confidence intervals including the true estimate. The value for the coverage should be approximately equal to nominal coverage rate, usually 95 percent, consistent with a 5 percent Type I error rate.

6. The left and right tail errors are calculated as the proportion of simulated confidence intervals missing the true estimate from the left and right, respectively. The two-sided left and right tail errors should preferably be approximately 2.5 percent in each tail with a nominal 95 percent confidence interval.

7. The interval width is calculated as the difference between the upper and lower limit averaged over 1000 simulated confidence intervals. An interval estimate with less

width is considered to be more precise.

In the process of calculating the suggested measures from the generated data, we estimated the intracluster correlation coefficients using the ANOVA estimator given in Appendix B. Note that negative estimates of intracluster correlation were truncated at zero.

Furthermore, since the ML and REML heterogeneity variance estimators required iterative solutions, an initial value was pre-specified and the number of simulations for convergence was restricted to 20. For each iteration, a negative estimate was truncated at zero. The convergence criterion adopted from Swallow and Monahan (1984) was given by

$$\frac{|\hat{\tau}^2_{j+1} - \hat{\tau}^2_j|}{1 + \hat{\tau}^2_j} \leq 0.00001,$$

where the value 1 was added in the denominator to prevent singularity and to keep the criterion stringent. When any of the two iterative methods did not reach convergence within 20 iterations, the remaining heterogeneity estimates were estimated under the same number of runs.

All of the computer programs for the simulation study were written in SAS V.9.2 and run on a PC Workstation.

The design of our simulation study is summarized using a flowchart given in Figure 5.1.

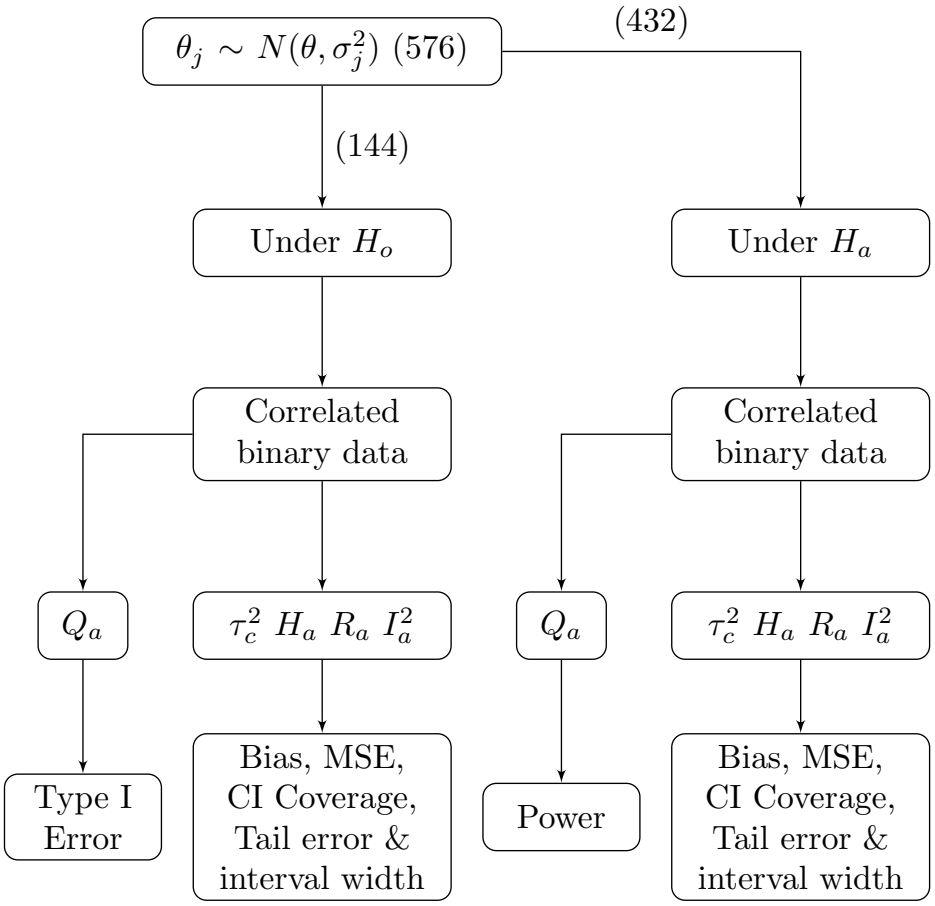

Figure 5.1: Flowchart of the simulation study. Number of parameter combination is noted in parentheses.

Chapter 6

Simulation study results

6.1 Introduction

The results of the simulation study described in Chapter 5 are presented and tabulated in the order of the three study objectives outlined in Section 5.3. In particular, the validity and power of the adjusted Q statistic are summarized in Section 6.2. The performance of the adjusted heterogeneity variance estimators in terms of their bias, mean square error, confidence interval coverage, tail errors and interval width are discussed in Section 6.3. Finally, the performance of the adjusted measures of heterogeneity is summarized in Section 6.4 using the same criteria as for the adjusted heterogeneity estimators.

6.2 Adjusted Q statistic

6.2.1 Type I error

Estimated type I error rates are given in Tables 6.1 and 6.2. The parameters of interest include the number of trials k, number of clusters N_j, cluster size m, disease rates for control group r_{1j} (i.e. r_A and r_B), intracluster correlation coefficient ρ and odds ratio ψ.

Each table displays the results for the unadjusted Q statistic, the adjusted Q statistic with truncated ANOVA-based $\hat{\rho}$ and the adjusted Q statistic omitting truncation for each parameter combination. Type I error rates outside the desired range (3.6%-6.4%) are in bold.

The results in Table 6.1 and Table 6.2 are similar for all parameter combinations. Furthermore, higher disease rates (r_B) had tighter Type I error rates as compared to lower disease rates (r_A), particularly for small values of k.

In Table 6.1, Type I error rates for the unadjusted Q statistic were far greater than nominal when $\rho \geq 0$, with the inflated Type I error rate increasing with ρ. The highest Type I error rate reached up to 100%.

On the other hand, Type I error rates for the adjusted Q statistic with truncated ANOVA-based $\hat{\rho}$ consistently maintained the nominal level for almost all parameter combinations, except at $\rho = 0$. Type I error rates were less than nominal with an average of 3.0% at $\rho = 0$ for both r_A and r_B.

Type I error rates for the adjusted Q statistic omitting truncation fell into the desired range 92% of the time.

6.2.2 Power

Empirical power estimated by simulation is displayed in Tables 6.3 to 6.6, where the power of the unadjusted Q statistic is omitted due to its elevated Type I error rates. Since a negative value of $\hat{\rho}$ is often set equal to zero in practice, the empirical power of the adjusted Q statistic with truncated ANOVA-based $\hat{\rho}$ is given, although the deflated Type I error rates may also be found occasionally when $\rho = 0$. The additional parameter of interest aside from ones considered in evaluating Type I error rates is the degree of

heterogeneity: 'low', 'moderate' and 'high'. Each table displays the empirical power of the adjusted Q statistic computed from data generated using either small disease rates r_A or large disease rates r_B as compared to the power calculated using equation (2.11) (same for both r_A and r_B) for each degree of heterogeneity.

In general, at $\rho = 0$, the empirical power for the adjusted Q statistic omitting truncation was closer to the calculated power than the empirical power for the adjusted Q statistic with truncated ANOVA-based $\hat{\rho}$ (Type I error rates less than nominal for the latter). More specifically, the observed difference between them was approximately 7% on average (Tables 6.5 vs. 6.3; Tables 6.6 vs. 6.4).

The empirical power for $\psi = 1.0$ was similar to that for $\psi = 0.7$ (Tables 6.3 vs. 6.4; Tables 6.5 vs. 6.6). The empirical power generally agrees with the calculated power for all parameter combinations, particularly for large degrees of heterogeneity. For instance, the average differences between the calculated power and the empirical power were 2.3%, 2.2% and 1.2% for 'low', 'moderate', and 'high' heterogeneity at r_B, respectively, in Table 6.3. Also, the empirical power obtained for large disease rates r_B was relatively close to the calculated power as compared to ones obtained for small disease rates r_A. For instance, the average differences between the calculated power and the empirical power at r_A were greater than at r_B, given by 3.5%, 3.5% and 3.1% corresponding to 'low', 'moderate', and 'high' heterogeneity, respectively, in Table 6.3.

The power increases as the degree of heterogeneity increases but decreases as the intra-cluster correlation coefficient increases. Moreover, the increase between the degrees of heterogeneity becomes smaller as ρ increases. For $(N_j, m) = (20, 100)$ at $k \leq 12$, the increase from 'low' to 'moderate' was approximately 20% for $\rho = 0$ as compared to 1% for $\rho = 0.05$. The decrease is dramatic for a small increase in ρ, particularly with large sample size. For $(N_j, m) = (20, 100)$ at $k \leq 40$, the calculated power to detect 'high'

heterogeneity was 84.2% for $\rho = 0$ as compared to 14.2% for $\rho = 0.05$.

The results show that, for a fixed number of subjects, the power is greater for a large number of small clusters than for a small number of large clusters. For instance, at $k = 40$ and $\rho = 0.01$, the calculated power was 84.4% for $(N_j, m) = (40, 100)$ as compared to 60.8% for $(N_j, m) = (20, 200)$ to detect 'high' heterogeneity in Table 6.3.

Overall, at $\rho = 0$, a meta-analysis with at least 12 trials is sufficiently large to detect 'high' heterogeneity, while the detection of 'moderate' heterogeneity requires a meta-analysis to have at least 20 trials in order to achieve a desired power of approximately 80%. For $\rho = 0.01$, a meta-analysis with 40 trials each with $(N_j, m) = (40, 100)$ is sufficiently large to detect 'high' heterogeneity. On the other hand, in Table 6.3, the higher observed power for $\rho = 0.05$ was only 28.4% for the largest sample in the simulation.

6.3 Heterogeneity variance estimators

Tables 6.7 through 6.14 show the empirical biases of the eight estimators listed in Table 3.1 at $\psi = 0.7$ based on 1000 simulations as a function of the parameters of interest described in Table 5.2, including the number of trials, number of clusters, cluster size, disease rates for the control group, intracluster correlation coefficient and degree of heterogeneity. Tables 6.15 through 6.30 present the empirical coverage, tail errors from the left and the right, and average interval width of the confidence intervals described in Section 3.3. The tables are tabulated by the degree of heterogeneity and the disease rates for the control group. The simulation results are not shown for $\psi = 1.0$, which were fairly similar to that of $\psi = 0.7$.

6.3.1 Convergence issues

Note that the two iterative estimation procedures that calculate ML and REML rarely required an excessive number of iterations in this simulation study. With large k, the two iterative procedures converged within 20 iterations for all parameter combinations, suggesting all 1000 replicates were used. However, for $k = 4$, the small numbers of replicates was mostly found at $\rho = 0.05$, roughly in the range 969 to 975. The empirical properties of the estimators were calculated on the basis of the actual number of replicates for each parameter combination, instead of the intended number of replicates (i.e. 1000).

6.3.2 Comparing bias and mean square error

The magnitude of the bias is relatively large for small control group disease rates (i.e. r_A) in Tables 6.7 to 6.10 as compared to large control group disease rates (i.e. r_B) in Tables 6.11 to 6.14, particularly when ρ is large. It is noted that at $\rho = 0.05$, the bias was approximately in the range of 0.05 to 0.07 for r_A in Table 6.10 as compared to 0.01 to 0.02 for r_B in Table 6.14.

An increase in the magnitude of the bias was observed with an increase in ρ, particularly for large k and large degrees of heterogeneity. For instance, for $(N_j, m) = (20, 100)$ at $\rho = 0.05$, the bias was roughly -0.014 at $k = 4$ as compared to -0.020 at $k = 40$ for r_B in Table 6.14. Similarly, for the same number of trials (i.e. $k = 40$), the bias was roughly 0.002 for 'no' heterogeneity in Table 6.11 as compared to -0.020 for 'high' heterogeneity in Table 6.14.

For a fixed number of subjects, the bias was considerately reduced for a large number of small clusters (i.e, $(N_j, m) = (40, 100)$) as compared to a small number of large clusters (i.e, $(N_j, m) = (20, 200)$). In particular, at $k = 4$ and $\rho = 0.05$, the bias of all the estimators was reduced to approximately half from $(N_j, m) = (20, 200)$ to $(N_j, m) = (40, 100)$.

Tables 6.7 to 6.14 show that the DLVC estimator had the largest average magnitude of bias for 'no' and 'low' heterogeneity but had the smallest average magnitude of bias for 'moderate' and 'high' heterogeneity. For the remaining estimators, the magnitudes of the bias were very similar. In particular, when there is 'no' heterogeneity, they tend to overestimate the true τ_c^2 with positive average biases. Otherwise, when heterogeneity is present, they underestimate the true τ_c^2 with negative average biases.

More specifically, it appears that the two-step estimator DL2 with the DL estimator as the initial weights gave relatively similar bias as compared to the DL estimator for all parameter combinations. On the other hand, another two-step estimator DLVC with the VC estimator as the initial weights had a large magnitude of bias for 'no' and 'low' heterogeneity but small bias for 'moderate' and 'high' heterogeneity as compared to the DL estimator, particularly for large ρ.

The VC and MVVC estimators were compared since they both use the VC estimator in their calculation. The MVVC estimator tends to have a slightly larger magnitude of bias as compared to the VC estimator for all parameter combinations. For instance, the largest difference between the average biases for VC and MVVC calculated from Table 6.9 was approximately 0.002 where $\overline{bias(VC)} = -0.0108$ and $\overline{bias(MVVC)} = -0.0128$. As expected, the MVVC estimator, an improved MV estimator, clearly outperforms the MV estimator. The average bias for MVVC calculated from Table 6.7 was 0.0034 as compared to 0.0056 for MV.

The magnitude of the bias for the intensive iterative estimators ML and REML decreases as k increases, with $k/(k-1)$ the only factor distinguishing these two estimators. For small k, it is noted that the ML estimator has a relatively small bias as compared to the REML estimator when there is 'no' heterogeneity, but has a relatively large negative bias as compared to the REML estimator when the degree of heterogeneity is 'low' to 'high'.

However, there was a small difference approximately of 0.001.

The mean square errors of the all estimators (not shown) were approximately zero to three decimal places for most parameter combinations, except for large ρ. In this case, the MV estimator resulted in a large MSE as compared to the other estimators. As for a fixed sample size, the mean square errors were smaller for a large number of small clusters (i.e. $(N_j, m) = (40, 100)$) than for a small number of large clusters (i.e. $(N_j, m) = (20, 200)$).

6.3.3 Confidence interval approaches

Empirical coverage ($\alpha = 0.05$), tail errors from the left and the right, and average interval widths for the Q profile (QP), Biggerstaff-Tweedie (BT), Sidik-Jonkman (SJ), nonparametric bootstraps (NB), the ML and REML profile likelihood (pML and pRE) and the ML and REML Wald-Type (wML and wRE) confidence intervals are presented in Tables 6.15 to 6.30 as a function of the parameters of interest listed in Table 5.2. The tables again differ by the degrees of heterogeneity with four levels: 'no', 'low', 'moderate' and 'high' and the disease rates for the control group (i.e. r_A and r_B).

Empirical coverage

Overall, the Q profile confidence interval approach yielded the empirical coverage most close to normal as compared to the other confidence interval approaches. Specifically, the empirical coverage of the Q profile confidence interval approach fell within the desired range (93.6% to 96.4%) for almost all parameter combinations for 'low' to 'high' heterogeneity with relatively small k. For 'no' heterogeneity, the empirical coverage was slightly above the nominal; otherwise, the empirical coverage was generally below the nominal.

The Biggerstaff-Tweedie confidence interval approach appears to have coverage similar to

the Q profile confidence interval approach when there is 'no' heterogeneity. The Biggerstaff-Tweedie confidence interval approach shows consistently high coverage throughout at almost all parameter combinations. However, the empirical coverage begins to drop below nominal when the heterogeneity is 'moderate' at $k = 40$.

The Sidik-Jonkman confidence interval approach yielded unacceptably low empirical coverage throughout all of the parameter combinations, showing slightly improvement as the degree of heterogeneity increases. The highest empirical coverage was 75%.

The empirical coverage of the bootstraps confidence interval approach was unacceptably high for 'no' heterogeneity and unacceptably low for 'low' to 'high' heterogeneity. For the latter, the coverage approaches nominal as the number of trials increases. However, the empirical coverage occasionally reached the desired range for large k at $\rho = 0$ for 'moderate' to 'high' heterogeneity.

The ML and REML profile likelihood confidence interval approaches begin to show reasonable coverage for large degrees of heterogeneity, relatively large k and small ρ. In addition, the large disease rates (i.e. r_B) tend to lead to relatively better performance as compared to the small disease rates (i.e. r_A). Otherwise, the empirical coverage is consistently higher than nominal.

The ML and REML Wald-type confidence interval approaches generally show poor empirical coverage, either too high or too low.

Tail errors

The imbalance is observed with the Q profile confidence interval approach, which misses the true parameter value more frequently on the left than on the right. On the other

hand, the ML and REML profile likelihood confidence interval approach generally misses the true parameter value more frequently on the right than on the left. Since the other approaches do not have valid empirical coverage results, there is no need to compare their tail errors.

Interval width

The Q profile and Biggerstaff-Tweedie approaches have relatively higher average interval widths, while the bootstraps approach has relatively small average interval width. Moreover, it is clearly seen that the approaches all demonstrate an increase in average interval width as ρ increases. Overall, an increase in the number of trials or the degree of heterogeneity reduces the average interval width.

6.4 Measures of heterogeneity

Tables 6.31 through Table 6.38, limited to $\psi = 0.7$ (with similar results for $\psi = 1.0$), show the empirical biases and mean square errors corresponding the true H_a, R_a and I_a^2 statistics based on 1000 simulations as a function of the parameters of interest described in Table 5.2. Tables 6.39 through Table 6.46 present the empirical coverage, tail errors from the left and the right, and average interval width of the confidence interval approaches described in Section 4.3. The tables are again tabulated by the degree of heterogeneity and the disease rates for the control group.

6.4.1 Bias and mean square error

In Tables 6.31 to 6.38, the bias and MSE for small disease rates (i.e. r_A) tend to be similar to those for large disease rates (i.e. r_B) for all parameter combinations. An increase in the

number of trials or the degree of heterogeneity results in a great reduction in bias and MSE of the three statistics. Moreover, it is noted that the bias and MSE tend to be relatively small when $\rho = 0$ as compared to $\rho \geq 0$. More specifically, among the three statistics, the I_a^2 statistic has the highest bias and MSE for almost all of parameter combinations, except for $k = 4$. A bias above 0.15 observed mainly at $k = 4$ may be a concern. For instance, 'no' heterogeneity ($H_a = 1.1$) may be interpreted as 'low' heterogeneity ($H_a = 1.25$) with a bias of 0.15.

6.4.2 Confidence interval approaches

The empirical coverage ($\alpha = 0.05$), tail errors from the left and the right, and average interval widths for the confidence intervals based on the MOVER, the Q distribution, the test-based method, τ_c^2, and the nonparametric bootstrap are presented in Tables 6.39 to Table 6.46 as a function of the parameters listed in Table 5.2. It is noted that the results for the confidence interval based on τ_c^2 were not shown for they were similar to ones for the MOVER.

Empirical coverage

The empirical coverage of the MOVER generally falls within the desired range (93.6%-96.4%) for 'low' to 'high' heterogeneity with small meta-analyses of large trials. When there is 'no' heterogeneity, the empirical coverage tends to be slightly above the nominal (approximately 97%-98%). Overall, the empirical coverage results are similar to those of the Q profile confidence interval approach, which is used for constructing the MOVER. Consequently, the confidence interval based on τ_c^2 tends to have identical coverage results as compared to the MOVER since it also uses the Q profile confidence intervals to construct the confidence limits for τ_c^2.

The confidence interval based on the Q distribution has coverage rates falling within the desired range for most of the parameter combinations with 'no' heterogeneity. For 'low' heterogeneity, the coverage was also reasonable with small k. Otherwise, the empirical coverage is usually below nominal.

The empirical coverage of the test-based confidence interval falls within the desired range 78% of the time for 'no' to 'low' heterogeneity at large control disease rates r_B with small k (Tables 6.43-6.44). Otherwise, the empirical coverage was either above the nominal for 'no' heterogeneity or below the nominal for 'low' to 'high' heterogeneity.

The bootstraps confidence interval for H_a generally performed poorly. The empirical coverage was close to 100% when there is 'no' heterogeneity but unacceptably small when the degree of heterogeneity is 'low' to 'high'. However, few reasonable empirical coverage levels were observed for $k = 40$, $\rho = 0.05$, and 'high' heterogeneity at small disease rates r_A (Table 6.42).

Tail errors

The MOVER resulted in unbalanced tail errors with the empirical coverage falling within the desired range. For $\rho = 0$, the confidence intervals miss from the left more often than from the right. On the contrary, the confidence intervals miss from the right more often than from the left for $\rho \neq 0$. Similar results apply to the confidence interval based on τ_c^2.

The skewness of the Q distribution, which follows the chi square distribution, depends on the degrees of freedom, with the distribution becoming less skewed as the degrees of freedom increase, where the distribution is more likely skewed to the right. Consequently, the confidence intervals miss less than 2.5% from the left and more than 2.5% from the right as the number of trials or the degree of heterogeneity increases. The degree of

imbalance in the tail errors tends to be reduced as the number of trials increases, as we have observed for $k = 40$. However, it is noted that when the number of trials is large, the confidence intervals tend to have more misses on the left then on the right at $\rho = 0$.

Interval width

The MOVER has the highest average interval width as compared to the confidence interval based on the Q distribution and the test-based confidence interval, particularly at small k. All the confidence interval approaches show a great improvement in the average interval width as the number of trials increase. In general, the average interval widths were fairly similar at the different values of ρ.

6.5 Discussion

In summary, the inflated Type I error rates of the unadjusted Q statistic may be explained by the inflated estimates of the within study variances that results from clustered data. However, after adjusting for the clustering, the adjusted Q statistic consistently maintained Type I error rates at nominal. It is noted that the truncation at zero in estimating ρ may reduce the estimated within study variances (Murray et al., 1998). As a result, the adjusted Q statistic with the truncated $\hat{\rho}$ tends to have Type I error rates below nominal at $\rho = 0$. The comparisons between the unadjusted and adjusted Q statistic in terms of Type I error clearly shows that the validity of the unadjusted Q statistic is in question for clustered data without properly adjusting for clustering (Darlington and Donner, 2007; Song, 2004).

The calculated power is fairly accurate as compared to the empirical power across all parameter combinations investigated. It depends on number of trials, number of clusters,

cluster size, intracluster correlation coefficient and degree of odds ratio heterogeneity across trials but is not affected by the disease rates for the control group. An increase in power is obtained by increasing the number of trials, number of clusters, cluster size and degree of odds ratio heterogeneity across trials. However, the power decreases dramatically for a small increase in intracluster correlation coefficient. Also, for a fixed sample size, the power is greater for a large number of small clusters than for a small number of large clusters.

In summary, when there is 'no' heterogeneity, the ML estimator tends to outperform the other estimators in terms of bias, especially for large ρ. On the other hand, when the heterogeneity is 'low' to 'high', the REML estimator appears to be the best estimator even for small k and large ρ. As for other simpler estimators including DL, VC, MV, MVVC, DLVC and DL2 estimators, the MVVC estimator may be recommended for small k but the DLVC may be recommended for large k.

The mean square errors are very close to zero, which may be due to the enforcement of the non-negativity that results in reducing the variances sufficiently to offset the squared bias. A similar explanation applies in the comparison of estimators for variance components (Swallow and Monahan, 1984). Therefore, it also implies that the bias may be more informative as compared to the mean square errors for assessing variance estimators, an argument given by Casella and Berger (2002, p.305).

The Q profile confidence interval may be recommended for 'low' to 'high' heterogeneity for small meta-analyses with large trials. However, for a large number of trials and a large degree of heterogeneity, the ML and REML profile likelihood confidence interval approaches slightly perform better.

Overall, the three statistics generally show similar simulation results. However, it is noted

that caution must be taken for a small number of trials, where a relatively large bias may lead to misleading interpretation.

As expected, the MOVER and the confidence interval based on τ_c^2 have the similar performance as compared to the Q profile confidence interval for τ_c^2, which is used to construct the confidence interval (Schuster and Metzger, 2010, CH 11). They should be again used for 'low' to 'high' heterogeneity with small meta-analyses with large trials. However, the confidence interval based on the Q distribution generally perform well for 'no' to 'low' heterogeneity. Although the test-based confidence interval is widely applied due to its simplicity, the coverage for the test-based confidence interval were generally inadequate for large heterogeneity according to the simulation results.

Table 6.1: Type I error (%) of Q statistic (U: unadjusted; A: adjusted): based on 1000 simulations of k trials each with N_j clusters per group of size m, control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$, $r_B = \{0.35, 0.40, 0.45, 0.50\}$, intracluster correlation ρ and odds ratio $\psi = 0.7$.

		r_A			r_B		
$k/N_j/m$	ρ	U	A[1]	A	U	A[1]	A
4 / 20 / 100	0	4.3	**3.3**	5.1	4.5	**3.3**	4.6
	0.01	**25.9**	4.6	4.8	**28.6**	5.6	5.6
	0.05	**71.3**	3.6	3.6	**71.6**	6.1	6.1
4 / 20 / 200	0	3.7	**2.9**	4.5	5.3	4.1	5.5
	0.01	**44.7**	4.2	4.3	**47.8**	5.0	5.0
	0.05	**82.7**	4.5	4.5	**87.6**	6.2	6.2
4 / 40 / 100	0	5.0	4.4	5.4	4.6	3.9	5.0
	0.01	**24.0**	5.3	5.3	**28.1**	5.5	5.5
	0.05	**71.8**	4.4	4.4	**71.8**	5.5	5.5
12 / 20 / 100	0	4.3	**3.2**	6.2	4.8	**2.6**	4.7
	0.01	**52.7**	4.5	4.8	**53.7**	5.8	5.8
	0.05	**98.8**	**2.7**	**2.9**	**98.5**	5.1	5.1
12 / 20 / 200	0	4.8	**3.0**	4.9	**3.2**	**1.8**	3.6
	0.01	**81.1**	4.4	4.4	**82.9**	6.2	6.2
	0.05	**99.9**	4.1	4.1	**99.9**	4.8	4.8
12 / 40 / 100	0	5.4	3.6	5.3	5.6	4.1	5.6
	0.01	**52.9**	4.3	4.4	**54.4**	5.6	5.6
	0.05	**98.4**	4.9	4.9	**98.2**	5.1	5.1
20 / 20 / 100	0	4.2	**2.2**	5.3	5.2	**2.8**	6.1
	0.01	**67.4**	3.9	4.1	**72.3**	**6.9**	**6.9**
	0.05	**100**	**3.0**	**3.0**	**99.9**	5.2	5.2
20 / 20 / 200	0	6.3	**2.8**	5.9	5.9	**3.4**	**6.5**
	0.01	**93.4**	4.5	4.6	**95.3**	5.0	5.0
	0.05	**100**	4.5	4.5	**100**	4.2	4.2
20 / 40 / 100	0	4.9	**2.6**	5.1	6.1	4.0	6.3
	0.01	**69.7**	4.9	4.9	**72.6**	4.6	4.6
	0.05	**99.8**	4.5	4.5	**100**	5.8	5.8
40 / 20 / 100	0	3.9	**1.8**	4.7	4.5	**2.2**	5.4
	0.01	**91.4**	3.7	4.1	**91.9**	5.4	5.4
	0.05	**100**	**3.1**	**3.1**	**100**	**6.6**	**6.6**
40 / 20 / 200	0	5.0	**1.8**	5.0	4.4	**2.2**	5.2
	0.01	**99.7**	4.4	4.7	**99.8**	5.3	5.3
	0.05	**100**	3.5	3.5	**100**	5.0	5.0
40 / 40 / 100	0	5.8	3.5	6.1	5.0	**2.6**	5.1
	0.01	**91.4**	4.5	4.5	**92.2**	4.7	4.7
	0.05	**100**	4.0	4.0	**100**	4.9	4.9

[1] Negative values of $\hat{\rho}_j$ set to zero.

Table 6.2: Type I error (%) of Q statistic (U: unadjusted; A: adjusted): based on 1000 simulations of k trials each with N_j clusters per group of size m, control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$, $r_B = \{0.35, 0.40, 0.45, 0.50\}$, intracluster correlation ρ and odds ratio $\psi = 1.0$.

		r_A			r_B		
$k/N_j/m$	ρ	U	A	A[1]	U	A	A[1]
4/20/100	0	5.6	4.4	**6.7**	4.2	**3.2**	4.8
	0.01	**25.3**	5.3	5.6	**25.4**	5.6	5.6
	0.05	**70.2**	4.1	4.1	**72.9**	5.4	5.4
4/20/200	0	4.6	**3.0**	5.1	5.3	4.2	5.8
	0.01	**48.0**	5.1	5.1	**46.5**	4.5	4.5
	0.05	**86.4**	4.8	4.8	**87.9**	4.3	4.3
4/40/100	0	5.1	4.1	5.4	5.4	4.4	5.7
	0.01	**27.2**	5.0	5.1	**26.4**	4.5	4.5
	0.05	**71.4**	4.2	4.2	**73.3**	5.5	5.5
12/20/100	0	4.0	**2.8**	4.5	6.4	**3.4**	**6.5**
	0.01	**51.0**	**3.1**	**3.1**	**54.6**	4.3	4.3
	0.05	**98.1**	4.4	4.4	**98.3**	4.4	4.4
12/20/200	0	4.6	3.6	5.5	4.8	**3.2**	5.5
	0.01	**81.8**	4.5	4.6	**82.2**	5.4	5.4
	0.05	**99.9**	**3.4**	**3.4**	**100**	6.1	6.1
12/40/100	0	4.6	3.7	4.7	4.5	**3.5**	5.0
	0.01	**52.2**	4.9	4.9	**53.3**	4.6	4.6
	0.05	**98.4**	4.4	4.4	**98.5**	6.0	6.0
20/20/100	0	**3.2**	**1.6**	3.9	5.3	**2.4**	5.1
	0.01	**67.5**	5.2	5.3	**71.6**	5.4	5.4
	0.05	**100**	4.5	4.6	**100**	5.9	5.9
20/20/200	0	5.3	**3.0**	5.7	5.9	**2.5**	5.9
	0.01	**95.6**	4.4	4.5	**95.4**	5.8	5.8
	0.05	**100**	4.2	4.2	**100**	4.9	4.9
20/40/100	0	4.7	**2.8**	4.9	5.0	**2.6**	5.9
	0.01	**71.6**	4.6	4.6	**70.1**	4.3	4.3
	0.05	**100**	4.1	4.1	**100**	4.5	4.5
40/20/100	0	5.4	**1.7**	6.0	**6.9**	**3.1**	**7.3**
	0.01	**90.7**	4.1	4.3	**92.0**	5.1	5.1
	0.05	**100**	4.4	4.4	**100**	4.5	4.5
40/20/200	0	5.8	**2.5**	5.6	4.5	**1.2**	4.6
	0.01	**99.9**	3.7	3.8	**99.7**	5.4	5.4
	0.05	**100**	4.0	4.1	**100**	6.2	6.2
40/40/100	0	5.1	**2.3**	5.2	5.9	**2.8**	5.7
	0.01	**91.6**	5.3	5.4	**93.8**	4.1	4.1
	0.05	**100**	**3.5**	**3.5**	**100**	5.3	5.3

[1] negative value of $\hat{\rho}_j$ was set equal to zero.

Table 6.3: Power (%) of adjusted Q statistic for odds ratio $\psi = 0.7$ with truncated intracluster correlation ρ based on 1000 simulations of k trials each with N_j clusters per group of size m, control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$, $r_B = \{0.35, 0.40, 0.45, 0.50\}$.

		Degree of heterogeneity								
		Low			Moderate			High		
$k/N_j/m$	ρ	$power^1$	r_A	r_B	$power^1$	r_A	r_B	$power^1$	r_A	r_B
4/20/100	0	11.2	7.1	8.4	18.9	15.1	17.8	27.1	21.9	24.6
	0.01	7.9	8.2	10.4	11.5	10.4	13.3	15.2	13.2	17.5
	0.05	5.9	5.8	5.9	7.0	5.1	7.2	8.1	6.0	8.5
4/20/200	0	18.3	15.2	14.6	35.3	29.2	32.7	51.1	43.6	47.0
	0.01	9.0	9.9	10.0	13.9	11.2	14.6	19.0	18.9	20.6
	0.05	6.0	4.8	6.7	7.2	6.3	8.0	8.4	5.8	7.2
4/40/100	0	18.3	16.1	16.4	35.3	30.8	31.2	51.1	48.2	48.8
	0.01	11.2	10.3	11.5	19.0	17.8	19.9	27.2	28.8	26.0
	0.05	6.9	6.4	7.1	9.2	8.9	10.3	11.5	11.0	10.8
12/20/100	0	15.1	10.2	8.9	30.3	22.3	22.7	46.3	35.3	40.5
	0.01	9.5	7.3	11.1	15.7	14.9	14.7	22.8	23.4	23.5
	0.05	6.4	4.7	7.0	8.0	6.3	6.6	9.7	7.1	11.9
12/20/200	0	29.0	24.7	21.9	61.0	52.2	52.0	82.6	74.1	77.2
	0.01	11.2	11.2	10.8	20.3	19.9	20.3	30.5	28.5	33.7
	0.05	6.5	4.6	7.6	8.3	5.8	7.7	10.2	8.7	9.3
12/40/100	0	29.0	21.2	24.5	61.0	53.5	55.3	82.6	78.3	76.9
	0.01	15.2	15.4	15.1	30.5	29.2	30.7	46.6	47.7	48.7
	0.05	7.9	6.5	7.0	11.6	10.5	14.4	15.8	14.7	17.0
20/20/100	0	18.7	10.9	10.3	40.4	31.3	29.0	61.4	48.9	49.9
	0.01	10.8	11.3	11.9	19.6	17.9	18.7	29.8	28.8	29.6
	0.05	6.7	4.0	7.8	8.8	6.8	8.2	11.2	8.0	10.6
20/20/200	0	38.6	25.2	28.3	77.6	71.1	68.9	94.5	90.5	91.2
	0.01	13.2	11.3	13.6	26.1	27.4	27.7	40.6	39.5	39.7
	0.05	6.8	5.5	7.3	9.2	7.6	9.8	11.8	8.9	13.9
20/40/100	0	38.6	31.0	33.1	77.6	72.0	70.7	94.5	90.8	91.7
	0.01	18.8	20.0	19.4	40.6	40.5	42.4	61.7	62.9	60.7
	0.05	8.6	7.1	8.9	13.7	11.4	14.6	19.7	18.2	21.8
40/20/100	0	26.7	13.6	15.9	60.5	40.4	44.4	84.2	68.3	72.3
	0.01	13.7	12.0	14.1	28.2	25.6	27.6	44.8	41.3	44.1
	0.05	7.3	5.5	7.4	10.5	7.0	12.3	14.2	9.3	15.8
40/20/200	0	58.0	41.6	40.3	95.1	88.0	90.4	99.8	99.0	99.1
	0.01	17.6	15.9	17.4	39.0	36.5	33.3	60.8	60.9	62.8
	0.05	7.6	7.3	7.9	11.1	8.7	11.3	15.3	10.7	16.8
40/40/100	0	58.0	43.4	44.3	95.1	92.1	92.8	99.8	99.4	99.9
	0.01	26.9	25.0	27.3	60.8	59.7	59.7	84.4	84.7	84.3
	0.05	10.3	6.2	11.3	18.4	15.4	19.1	28.3	25.5	28.4

[1] Power is computed using equation (2.11).

Table 6.4: Power (%) of adjusted Q statistic for odds ratio $\psi = 1.0$ with truncated intracluster correlation ρ based on 1000 simulations of k trials each with N_j clusters per group of size m, control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$, $r_B = \{0.35, 0.40, 0.45, 0.50\}$.

		Degree of heterogeneity								
		Low			Moderate			High		
$k/N_j/m$	ρ	$power^1$	r_A	r_B	$power^1$	r_A	r_B	$power^1$	r_A	r_B
4/20/100	0	11.2	9.5	7.9	18.9	12.8	14.1	27.1	26.7	25.7
	0.01	7.9	8.6	6.7	11.5	10.6	11.8	15.2	15.2	17.9
	0.05	5.9	5.1	6.5	7.0	6.0	7.3	8.1	8.5	8.1
4/20/200	0	18.3	15.6	16.1	35.2	29.3	27.7	51.1	44.9	49.4
	0.01	9.0	8.2	8.3	13.9	14.1	15.4	19.0	20.1	21.1
	0.05	6.0	4.9	5.6	7.2	6.7	8.5	8.4	8.6	9.8
4/40/100	0	18.3	17.6	17.7	35.2	31.3	31.9	51.1	47.9	48.9
	0.01	11.2	12.2	10.7	19.0	19.8	20.7	27.2	28.7	27.6
	0.05	6.9	6.6	7.3	9.2	8.9	8.6	11.5	11.9	12.1
12/20/100	0	15.1	10.2	10.1	30.3	21.8	23.1	46.4	37.0	37.6
	0.01	9.5	9.3	8.6	15.7	16.5	17.1	22.8	21.2	21.4
	0.05	6.4	4.7	5.6	8.0	7.1	8.3	9.7	8.4	9.6
12/20/200	0	29.1	22.1	20.9	60.9	51.9	52.6	82.6	76.7	74.7
	0.01	11.3	10.3	12.4	20.3	20.3	20.4	30.5	31.4	29.1
	0.05	6.5	4.2	6.9	8.3	5.8	8.8	10.2	8.4	11.2
12/40/100	0	29.1	24.5	25.5	60.9	52.2	52.0	82.6	78.7	78.7
	0.01	15.2	15.2	15.7	30.4	31.4	29.4	46.6	46.4	48.3
	0.05	7.9	7.1	7.4	11.6	8.4	12.6	15.8	14.7	14.5
20/20/100	0	18.7	9.3	11.8	40.3	29.4	29.6	61.4	52.2	49.3
	0.01	10.8	10.4	10.7	19.5	18.5	19.5	29.8	28.6	28.3
	0.05	6.7	5.7	6.5	8.8	6.1	8.0	11.2	9.6	9.0
20/20/200	0	38.7	29.1	26.3	77.5	66.0	69.2	94.5	90.2	90.9
	0.01	13.3	13.7	14.3	26.1	26.1	27.3	40.7	38.5	39.3
	0.05	6.8	6.2	6.8	9.2	6.4	9.4	11.8	9.0	13.7
20/40/100	0	38.7	28.8	31.6	77.5	68.3	71.5	94.5	91.5	91.2
	0.01	18.8	19.5	21.4	40.5	40.9	40.8	61.7	63.3	60.0
	0.05	8.6	7.1	8.2	13.7	10.8	13.9	19.7	18.7	18.8
40/20/100	0	26.8	14.5	14.1	60.3	44.5	45.3	84.2	73.7	74.3
	0.01	13.7	12.0	14.1	28.1	26.7	30.0	44.8	41.0	46.1
	0.05	7.4	5.6	9.4	10.5	7.5	10.8	14.2	12.3	16.3
40/20/200	0	58.1	41.3	43.0	95.0	89.6	91.0	99.8	98.9	99.7
	0.01	17.6	16.5	17.3	38.9	38.6	39.2	60.8	59.7	59.8
	0.05	7.6	6.1	7.4	11.1	8.2	10.9	15.3	14.6	15.6
40/40/100	0	58.1	49.5	44.8	95.0	91.4	91.8	99.8	99.2	99.3
	0.01	26.9	29.7	26.7	60.6	62.1	61.4	84.4	83.4	83.2
	0.05	10.3	9.3	9.7	18.4	17.1	15.6	28.3	27.0	27.1

[1] Power is computed using equation (2.11).

Table 6.5: Power (%) of adjusted Q statistic for odds ratio $\psi = 0.7$ omitting truncation of intracluster correlation ρ based on 1000 simulations of k trials each with N_j clusters per group of size m, control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$, $r_B = \{0.35, 0.40, 0.45, 0.50\}$.

		Degree of heterogeneity								
		Low			Moderate			High		
$k/N_j/m$	ρ	$power^1$	r_A	r_B	$power^1$	r_A	r_B	$power^1$	r_A	r_B
4/20/100	0	11.2	14.1	12.0	18.9	17.0	19.0	27.1	27.8	28.1
	0.01	7.9	7.2	8.9	11.5	9.9	11.9	15.2	14.6	14.9
	0.05	5.9	4.3	7.1	7.0	4.5	6.5	8.1	5.9	7.6
4/20/200	0	18.3	17.5	16.9	35.3	33.6	36.0	51.1	50.4	51.6
	0.01	9.0	7.7	9.7	13.9	15.0	15.3	19.0	20.6	18.7
	0.05	6.0	4.4	6.7	7.2	5.2	7.4	8.4	7.6	7.4
4/40/100	0	18.3	16.4	19.5	35.3	35.4	34.6	51.1	51.7	49.8
	0.01	11.2	11.0	11.0	19.0	16.6	18.3	27.2	26.1	27.8
	0.05	6.9	7.4	6.4	9.2	7.8	8.8	11.5	11.1	12.5
12/20/100	0	15.1	13.6	15.3	30.3	29.5	30.4	46.3	46.7	43.9
	0.01	9.5	10.4	11.6	15.7	13.4	16.2	22.8	21.9	26.8
	0.05	6.4	4.7	5.5	8.0	5.2	7.4	9.7	7.7	9.7
12/20/200	0	29.0	29.6	30.4	61.0	62.2	61.7	82.6	84.2	81.4
	0.01	11.2	10.3	9.8	20.3	19.2	19.6	30.5	29.8	29.4
	0.05	6.5	4.0	7.6	8.3	7.5	8.7	10.2	8.4	11.0
12/40/100	0	29.0	26.5	28.0	61.0	63.2	58.1	82.6	82.2	82.3
	0.01	15.2	17.1	16.3	30.5	31.2	30.2	46.6	47.0	46.9
	0.05	7.9	7.4	9.1	11.6	11.6	10.4	15.8	13.0	16.9
20/20/100	0	18.7	17.6	20.7	40.4	39.6	41.4	61.4	61.2	60.4
	0.01	10.8	10.1	12.3	19.6	18.3	21.0	29.8	29.4	27.0
	0.05	6.7	3.7	7.2	8.8	5.4	8.7	11.2	9.3	10.9
20/20/200	0	38.6	38.0	38.7	77.6	78.6	76.1	94.5	95.0	94.5
	0.01	13.2	13.0	13.3	26.1	24.9	27.9	40.6	40.8	41.9
	0.05	6.8	5.2	5.8	9.2	6.5	11.4	11.8	9.1	11.8
20/40/100	0	38.6	37.6	37.7	77.6	79.4	78.4	94.5	94.9	94.5
	0.01	18.8	16.3	17.5	40.6	37.4	41.1	61.7	63.2	62.3
	0.05	8.6	7.0	8.3	13.7	11.8	15.7	19.7	17.0	19.5
40/20/100	0	26.7	23.8	27.5	60.5	59.8	61.9	84.2	83.4	84.2
	0.01	13.7	10.9	12.3	28.2	26.0	29.1	44.8	42.3	45.2
	0.05	7.3	3.9	7.8	10.5	6.8	9.8	14.2	10.8	14.0
40/20/200	0	58.0	59.5	60.0	95.1	95.7	96.5	99.8	99.8	99.9
	0.01	17.6	16.0	17.6	39.0	37.0	39.8	60.8	60.2	59.9
	0.05	7.6	5.1	6.8	11.1	9.9	12.7	15.3	11.9	17.0
40/40/100	0	58.0	57.3	59.4	95.1	93.1	95.4	99.8	99.8	100.0
	0.01	26.9	26.0	26.3	60.8	57.5	64.5	84.4	84.7	84.0
	0.05	10.3	9.5	9.4	18.4	14.5	17.8	28.3	24.8	30.0

[1] Power is computed using equation (2.11).

Table 6.6: Power (%) of adjusted Q statistic for odds ratio $\psi = 1.0$ omitting truncation of intracluster correlation ρ based on 1000 simulations of k trials each with N_j clusters per group of size m, control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$, $r_B = \{0.35, 0.40, 0.45, 0.50\}$.

		Degree of heterogeneity								
		Low			Moderate			High		
$k/N_j/m$	ρ	$power^1$	r_A	r_B	$power^1$	r_A	r_B	$power^1$	r_A	r_B
4/20/100	0	11.2	10.3	13.0	18.9	17.4	17.7	27.1	26.1	26.4
	0.01	7.9	7.9	10.4	11.5	12.5	12.0	15.2	13.4	15.4
	0.05	5.9	4.8	6.2	7.0	5.5	7.5	8.1	8.1	8.3
4/20/200	0	18.3	19.0	20.0	35.2	36.5	34.8	51.1	49.0	51.3
	0.01	9.0	10.0	11.2	13.9	14.3	13.0	19.0	21.2	19.6
	0.05	6.0	5.3	7.5	7.2	7.5	6.9	8.4	8.3	9.4
4/40/100	0	18.3	15.9	19.6	35.2	35.5	34.1	51.1	54.6	52.4
	0.01	11.2	9.9	10.1	19.0	19.9	21.5	27.2	30.7	29.2
	0.05	6.9	6.6	7.0	9.2	9.0	9.7	11.5	13.6	12.2
12/20/100	0	15.1	14.4	16.7	30.3	29.4	30.9	46.4	48.2	44.6
	0.01	9.5	10.1	10.4	15.7	17.5	17.3	22.8	22.1	22.6
	0.05	6.4	5.0	6.1	8.0	6.6	8.4	9.7	8.0	10.3
12/20/200	0	29.1	30.5	28.4	60.9	64.1	62.5	82.6	83.2	83.8
	0.01	11.3	11.9	11.5	20.3	21.8	20.9	30.5	29.4	29.9
	0.05	6.5	4.9	6.3	8.3	8.2	9.6	10.2	8.1	11.4
12/40/100	0	29.1	29.1	28.9	60.9	60.1	62.9	82.6	82.6	82.2
	0.01	15.2	15.5	14.9	30.4	27.0	30.1	46.6	47.4	46.0
	0.05	7.9	6.2	6.9	11.6	10.1	13.2	15.8	15.8	13.7
20/20/100	0	18.7	19.6	18.0	40.3	43.3	38.8	61.4	58.2	62.8
	0.01	10.8	11.7	11.4	19.5	19.7	20.4	29.8	29.2	30.5
	0.05	6.7	6.5	7.0	8.8	6.2	10.8	11.2	9.3	11.7
20/20/200	0	38.7	37.6	43.6	77.5	78.5	77.4	94.5	93.7	94.1
	0.01	13.3	13.9	14.7	26.1	28.3	25.1	40.7	42.8	41.3
	0.05	6.8	6.2	7.3	9.2	7.9	10.0	11.8	11.3	14.2
20/40/100	0	38.7	38.4	39.3	77.5	73.8	76.6	94.5	94.8	93.7
	0.01	18.8	19.3	18.1	40.5	42.2	41.7	61.7	60.5	60.1
	0.05	8.6	7.3	8.6	13.7	12.5	14.8	19.7	19.0	17.9
40/20/100	0	26.8	26.5	28.4	60.3	58.4	62.3	84.2	83.8	84.2
	0.01	13.7	12.5	13.8	28.1	28.8	29.5	44.8	44.9	43.7
	0.05	7.4	5.0	6.4	10.5	8.9	10.7	14.2	11.2	13.8
40/20/200	0	58.1	57.8	59.9	95.0	94.9	95.4	99.8	99.7	99.7
	0.01	17.6	18.3	17.3	38.9	39.5	38.4	60.8	60.8	62.1
	0.05	7.6	6.0	8.9	11.1	9.3	11.7	15.3	14.3	15.2
40/40/100	0	58.1	56.8	58.5	95.0	95.2	94.0	99.8	99.6	99.8
	0.01	26.9	26.8	25.5	60.6	61.5	60.2	84.4	86.0	82.6
	0.05	10.3	9.3	10.4	18.4	17.1	19.6	28.3	26.5	30.0

[1] Power is computed using equation (2.11).

Table 6.7: Bias for τ_c^2 with 'no' heterogeneity and control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

$k/N_j/m$	ρ	VC	DL	DLVC	DL2	MV	MVVC	ML	REML
4/20/100	0	0.005	0.004	0.008	0.005	0.002	0.004	0.005	0.004
	0.01	0.010	0.009	0.017	0.010	0.004	0.009	0.010	0.009
	0.05	0.028	0.028	0.049	0.028	0.013	0.028	0.029	0.027
4/20/200	0	0.002	0.002	0.004	0.002	0.001	0.002	0.002	0.002
	0.01	0.008	0.007	0.013	0.008	0.003	0.007	0.008	0.007
	0.05	0.024	0.022	0.043	0.023	0.009	0.022	0.024	0.022
4/40/100	0	0.003	0.002	0.005	0.003	0.001	0.002	0.003	0.003
	0.01	0.005	0.004	0.008	0.005	0.002	0.004	0.005	0.004
	0.05	0.015	0.013	0.025	0.015	0.006	0.013	0.015	0.014
12/20/100	0	0.003	0.002	0.009	0.003	0.001	0.002	0.002	0.002
	0.01	0.006	0.005	0.017	0.005	0.003	0.005	0.005	0.005
	0.05	0.015	0.014	0.049	0.015	0.010	0.014	0.015	0.014
12/20/200	0	0.001	0.001	0.004	0.001	0.001	0.001	0.001	0.001
	0.01	0.005	0.004	0.013	0.004	0.003	0.004	0.004	0.004
	0.05	0.011	0.012	0.043	0.011	0.008	0.012	0.012	0.011
12/40/100	0	0.002	0.001	0.004	0.001	0.001	0.001	0.001	0.001
	0.01	0.003	0.003	0.009	0.003	0.002	0.003	0.003	0.003
	0.05	0.008	0.007	0.025	0.008	0.005	0.007	0.007	0.007
20/20/100	0	0.002	0.001	0.009	0.002	0.001	0.001	0.002	0.002
	0.01	0.005	0.004	0.017	0.004	0.003	0.004	0.004	0.004
	0.05	0.009	0.010	0.048	0.009	0.008	0.010	0.010	0.010
20/20/200	0	0.001	0.001	0.004	0.001	0.000	0.001	0.001	0.001
	0.01	0.003	0.003	0.013	0.003	0.002	0.003	0.003	0.003
	0.05	0.008	0.009	0.043	0.008	0.008	0.010	0.010	0.009
20/40/100	0	0.001	0.001	0.004	0.001	0.001	0.001	0.001	0.001
	0.01	0.002	0.002	0.009	0.002	0.001	0.002	0.002	0.002
	0.05	0.006	0.005	0.025	0.006	0.004	0.005	0.005	0.005
40/20/100	0	0.001	0.001	0.008	0.001	0.001	0.001	0.001	0.001
	0.01	0.003	0.003	0.017	0.003	0.002	0.003	0.003	0.003
	0.05	0.006	0.006	0.047	0.005	0.006	0.007	0.006	0.006
40/20/200	0	0.001	0.000	0.004	0.001	0.000	0.000	0.000	0.000
	0.01	0.002	0.002	0.012	0.002	0.002	0.002	0.002	0.002
	0.05	0.005	0.006	0.043	0.005	0.006	0.007	0.006	0.006
40/40/100	0	0.001	0.001	0.004	0.001	0.000	0.000	0.001	0.001
	0.01	0.002	0.001	0.009	0.002	0.001	0.001	0.001	0.001
	0.05	0.004	0.004	0.025	0.004	0.003	0.004	0.004	0.004

Table 6.8: Bias for τ_c^2 with 'low' heterogeneity and control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

$k/N_j/m$	ρ	VC	DL	DLVC	DL2	MV	MVVC	ML	REML
4/20/100	0	0.004	0.003	0.008	0.004	-0.001	0.003	0.004	0.003
	0.01	0.004	0.001	0.010	0.003	-0.005	0.001	0.003	0.002
	0.05	0.001	-0.004	0.021	-0.001	-0.019	-0.004	-0.001	-0.003
4/20/200	0	0.004	0.003	0.005	0.003	0.000	0.003	0.003	0.003
	0.01	0.005	0.003	0.010	0.004	-0.002	0.003	0.005	0.004
	0.05	-0.003	-0.005	0.017	-0.004	-0.018	-0.005	-0.003	-0.005
4/40/100	0	0.004	0.003	0.006	0.004	0.001	0.004	0.004	0.004
	0.01	0.004	0.002	0.007	0.004	-0.001	0.003	0.004	0.003
	0.05	0.004	0.001	0.014	0.003	-0.009	0.000	0.003	0.002
12/20/100	0	0.001	-0.001	0.007	0.000	-0.002	-0.001	-0.000	-0.000
	0.01	-0.002	-0.004	0.010	-0.002	-0.006	-0.004	-0.003	-0.003
	0.05	-0.015	-0.016	0.020	-0.015	-0.020	-0.016	-0.015	-0.016
12/20/200	0	0.002	0.001	0.005	0.002	0.000	0.001	0.002	0.002
	0.01	-0.001	-0.002	0.008	-0.001	-0.003	-0.002	-0.001	-0.002
	0.05	-0.014	-0.015	0.019	-0.014	-0.019	-0.014	-0.014	-0.015
12/40/100	0	0.002	0.001	0.005	0.002	0.000	0.001	0.002	0.002
	0.01	0.001	-0.000	0.007	0.001	-0.002	-0.000	0.000	0.000
	0.05	-0.005	-0.006	0.013	-0.005	-0.009	-0.007	-0.005	-0.006
20/20/100	0	-0.000	-0.002	0.007	-0.001	-0.002	-0.002	-0.001	-0.001
	0.01	-0.003	-0.004	0.010	-0.004	-0.006	-0.005	-0.004	-0.004
	0.05	-0.018	-0.019	0.021	-0.018	-0.021	-0.019	-0.018	-0.019
20/20/200	0	0.002	0.001	0.005	0.001	0.000	0.001	0.001	0.001
	0.01	-0.001	-0.002	0.008	-0.002	-0.004	-0.003	-0.002	-0.002
	0.05	-0.017	-0.017	0.019	-0.017	-0.018	-0.016	-0.016	-0.017
20/40/100	0	0.002	0.001	0.005	0.002	0.000	0.001	0.001	0.001
	0.01	0.001	-0.000	0.007	0.000	-0.001	-0.001	0.000	-0.000
	0.05	-0.006	-0.007	0.013	-0.007	-0.009	-0.008	-0.007	-0.007
40/20/100	0	-0.001	-0.002	0.007	-0.001	-0.003	-0.002	-0.002	-0.002
	0.01	-0.004	-0.005	0.011	-0.004	-0.006	-0.006	-0.005	-0.005
	0.05	-0.022	-0.022	0.021	-0.022	-0.023	-0.022	-0.022	-0.022
40/20/200	0	0.001	0.000	0.005	0.001	-0.000	0.000	0.001	0.001
	0.01	-0.003	-0.003	0.008	-0.003	-0.004	-0.004	-0.003	-0.003
	0.05	-0.021	-0.020	0.019	-0.021	-0.021	-0.020	-0.020	-0.020
40/40/100	0	0.001	0.001	0.005	0.001	0.000	0.000	0.001	0.001
	0.01	-0.000	-0.001	0.007	-0.000	-0.002	-0.001	-0.001	-0.001
	0.05	-0.008	-0.009	0.013	-0.009	-0.010	-0.009	-0.009	-0.009

Table 6.9: Bias for τ_c^2 with 'moderate' heterogeneity and control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

$k/N_j/m$	ρ	VC	DL	DLVC	DL2	MV	MVVC	ML	REML
4/20/100	0	0.003	0.000	0.006	0.002	-0.005	0.001	0.002	0.001
	0.01	-0.002	-0.005	0.005	-0.003	-0.013	-0.005	-0.003	-0.004
	0.05	-0.028	-0.030	-0.006	-0.029	-0.049	-0.030	-0.027	-0.030
4/20/200	0	0.007	0.005	0.008	0.007	0.001	0.006	0.007	0.006
	0.01	0.001	-0.002	0.006	0.000	-0.009	-0.002	0.000	-0.001
	0.05	-0.024	-0.027	-0.005	-0.025	-0.045	-0.028	-0.024	-0.027
4/40/100	0	0.007	0.005	0.008	0.007	0.001	0.006	0.007	0.006
	0.01	0.003	0.002	0.006	0.003	-0.004	0.002	0.003	0.002
	0.05	-0.009	-0.012	0.002	-0.010	-0.022	-0.012	-0.009	-0.011
12/20/100	0	0.000	-0.002	0.006	-0.000	-0.004	-0.002	-0.001	-0.001
	0.01	-0.008	-0.011	0.003	-0.009	-0.014	-0.011	-0.010	-0.010
	0.05	-0.041	-0.044	-0.006	-0.042	-0.050	-0.045	-0.042	-0.044
12/20/200	0	0.004	0.002	0.007	0.003	0.001	0.002	0.003	0.003
	0.01	-0.004	-0.006	0.005	-0.005	-0.009	-0.006	-0.005	-0.006
	0.05	-0.038	-0.039	-0.005	-0.039	-0.044	-0.039	-0.038	-0.039
12/40/100	0	0.004	0.002	0.007	0.004	0.001	0.003	0.003	0.003
	0.01	0.000	-0.002	0.006	-0.000	-0.004	-0.002	-0.000	-0.001
	0.05	-0.017	-0.019	0.001	-0.018	-0.023	-0.020	-0.018	-0.019
20/20/100	0	-0.001	-0.003	0.006	-0.002	-0.005	-0.004	-0.002	-0.002
	0.01	-0.010	-0.012	0.003	-0.010	-0.014	-0.013	-0.011	-0.011
	0.05	-0.046	-0.047	-0.006	-0.047	-0.050	-0.048	-0.046	-0.047
20/20/200	0	0.003	0.002	0.006	0.003	0.001	0.002	0.003	0.003
	0.01	-0.005	-0.007	0.005	-0.006	-0.009	-0.007	-0.006	-0.006
	0.05	-0.043	-0.043	-0.006	-0.043	-0.046	-0.043	-0.042	-0.043
20/40/100	0	0.004	0.002	0.006	0.003	0.001	0.002	0.003	0.003
	0.01	-0.001	-0.002	0.006	-0.001	-0.004	-0.003	-0.001	-0.002
	0.05	-0.018	-0.021	0.002	-0.019	-0.023	-0.021	-0.019	-0.020
40/20/100	0	-0.003	-0.004	0.005	-0.003	-0.006	-0.005	-0.004	-0.004
	0.01	-0.011	-0.012	0.004	-0.011	-0.014	-0.013	-0.012	-0.012
	0.05	-0.050	-0.050	-0.006	-0.051	-0.051	-0.050	-0.050	-0.050
40/20/200	0	0.003	0.002	0.006	0.003	0.001	0.002	0.003	0.002
	0.01	-0.006	-0.008	0.004	-0.007	-0.009	-0.008	-0.007	-0.007
	0.05	-0.046	-0.046	-0.006	-0.046	-0.047	-0.045	-0.045	-0.046
40/40/100	0	0.004	0.002	0.006	0.003	0.001	0.002	0.003	0.003
	0.01	-0.001	-0.002	0.006	-0.001	-0.003	-0.003	-0.002	-0.002
	0.05	-0.020	-0.022	0.002	-0.020	-0.024	-0.023	-0.021	-0.022

Table 6.10: Bias for τ_c^2 with 'high' heterogeneity and control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

$k/N_j/m$	ρ	VC	DL	DLVC	DL2	MV	MVVC	ML	REML
4/20/100	0	0.002	-0.000	0.005	0.001	-0.007	0.000	0.002	0.001
	0.01	-0.009	-0.012	-0.002	-0.009	-0.021	-0.011	-0.009	-0.010
	0.05	-0.059	-0.063	-0.037	-0.061	-0.080	-0.063	-0.060	-0.063
4/20/200	0	0.010	0.008	0.011	0.009	0.002	0.009	0.009	0.009
	0.01	-0.001	-0.005	0.003	-0.002	-0.013	-0.004	-0.002	-0.003
	0.05	-0.050	-0.056	-0.032	-0.052	-0.073	-0.057	-0.052	-0.055
4/40/100	0	0.010	0.008	0.011	0.009	0.002	0.009	0.010	0.009
	0.01	0.004	0.002	0.007	0.003	-0.006	0.002	0.004	0.003
	0.05	-0.019	-0.023	-0.009	-0.020	-0.036	-0.023	-0.020	-0.022
12/20/100	0	-0.000	-0.003	0.006	-0.001	-0.006	-0.003	-0.001	-0.002
	0.01	-0.014	-0.017	-0.002	-0.015	-0.021	-0.018	-0.015	-0.016
	0.05	-0.067	-0.070	-0.031	-0.069	-0.077	-0.071	-0.068	-0.070
12/20/200	0	0.006	0.004	0.009	0.006	0.002	0.004	0.006	0.005
	0.01	-0.007	-0.010	0.001	-0.008	-0.013	-0.011	-0.009	-0.009
	0.05	-0.062	-0.064	-0.029	-0.063	-0.070	-0.064	-0.062	-0.063
12/40/100	0	0.006	0.004	0.008	0.006	0.002	0.004	0.005	0.005
	0.01	0.000	-0.003	0.005	-0.001	-0.005	-0.003	-0.001	-0.001
	0.05	-0.027	-0.031	-0.009	-0.028	-0.035	-0.032	-0.029	-0.030
20/20/100	0	-0.002	-0.005	0.004	-0.003	-0.006	-0.005	-0.003	-0.004
	0.01	-0.014	-0.018	-0.002	-0.015	-0.020	-0.018	-0.016	-0.017
	0.05	-0.073	-0.074	-0.032	-0.074	-0.077	-0.074	-0.073	-0.074
20/20/200	0	0.006	0.003	0.008	0.005	0.003	0.004	0.005	0.005
	0.01	-0.009	-0.011	0.001	-0.009	-0.013	-0.012	-0.010	-0.010
	0.05	-0.066	-0.067	-0.029	-0.067	-0.070	-0.067	-0.066	-0.067
20/40/100	0	0.006	0.004	0.008	0.006	0.003	0.004	0.005	0.005
	0.01	-0.001	-0.003	0.005	-0.002	-0.005	-0.004	-0.002	-0.002
	0.05	-0.028	-0.031	-0.008	-0.029	-0.035	-0.032	-0.030	-0.031
40/20/100	0	-0.003	-0.005	0.004	-0.003	-0.006	-0.006	-0.004	-0.004
	0.01	-0.016	-0.019	-0.002	-0.017	-0.021	-0.020	-0.018	-0.018
	0.05	-0.076	-0.077	-0.031	-0.076	-0.078	-0.077	-0.076	-0.076
40/20/200	0	0.006	0.003	0.008	0.005	0.003	0.003	0.005	0.004
	0.01	-0.009	-0.011	0.001	-0.010	-0.013	-0.012	-0.010	-0.011
	0.05	-0.071	-0.071	-0.030	-0.072	-0.072	-0.071	-0.071	-0.071
40/40/100	0	0.005	0.003	0.008	0.005	0.003	0.003	0.005	0.004
	0.01	-0.001	-0.004	0.004	-0.002	-0.005	-0.004	-0.003	-0.003
	0.05	-0.031	-0.034	-0.010	-0.032	-0.036	-0.035	-0.033	-0.034

Table 6.11: Bias for τ_c^2 with 'no' heterogeneity and control group disease rates $r_B = \{0.35, 0.40, 0.45, 0.50\}$ based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

$k/N_j/m$	ρ	VC	DL	DLVC	DL2	MV	MVVC	ML	REML
4/20/100	0	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001
	0.01	0.003	0.003	0.005	0.003	0.001	0.003	0.003	0.003
	0.05	0.009	0.009	0.014	0.009	0.004	0.009	0.009	0.009
4/20/200	0	0.001	0.001	0.001	0.001	0.000	0.001	0.001	0.001
	0.01	0.002	0.002	0.003	0.002	0.001	0.002	0.002	0.002
	0.05	0.008	0.007	0.012	0.008	0.004	0.007	0.008	0.008
4/40/100	0	0.001	0.001	0.001	0.001	0.000	0.001	0.001	0.001
	0.01	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001
	0.05	0.005	0.005	0.007	0.005	0.002	0.005	0.005	0.005
12/20/100	0	0.001	0.001	0.002	0.001	0.000	0.001	0.001	0.001
	0.01	0.001	0.001	0.004	0.001	0.001	0.001	0.001	0.001
	0.05	0.005	0.004	0.013	0.004	0.003	0.004	0.004	0.004
12/20/200	0	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	0.01	0.001	0.001	0.003	0.001	0.001	0.001	0.001	0.001
	0.05	0.004	0.004	0.012	0.004	0.003	0.004	0.004	0.004
12/40/100	0	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	0.01	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001
	0.05	0.002	0.002	0.007	0.002	0.002	0.002	0.002	0.002
20/20/100	0	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.000
	0.01	0.001	0.001	0.004	0.001	0.001	0.001	0.001	0.001
	0.05	0.003	0.003	0.013	0.003	0.002	0.003	0.003	0.003
20/20/200	0	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	0.01	0.001	0.001	0.003	0.001	0.001	0.001	0.001	0.001
	0.05	0.003	0.003	0.012	0.003	0.002	0.003	0.003	0.003
20/40/100	0	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	0.01	0.001	0.001	0.002	0.001	0.000	0.001	0.001	0.001
	0.05	0.002	0.002	0.006	0.002	0.001	0.002	0.002	0.002
40/20/100	0	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000
	0.01	0.001	0.001	0.004	0.001	0.001	0.001	0.001	0.001
	0.05	0.002	0.002	0.013	0.002	0.002	0.002	0.002	0.002
40/20/200	0	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	0.01	0.001	0.001	0.003	0.001	0.000	0.001	0.001	0.001
	0.05	0.002	0.002	0.012	0.002	0.002	0.002	0.002	0.002
40/40/100	0	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	0.01	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000
	0.05	0.001	0.001	0.006	0.001	0.001	0.001	0.001	0.001

Table 6.12: Bias for τ_c^2 with 'low' heterogeneity and control group disease rates $r_B = \{0.35, 0.40, 0.45, 0.50\}$ based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

$k/N_j/m$	ρ	VC	DL	DLVC	DL2	MV	MVVC	ML	REML
4/20/100	0	0.001	0.001	0.002	0.001	-0.000	0.001	0.001	0.001
	0.01	0.001	0.001	0.003	0.001	-0.001	0.001	0.001	0.001
	0.05	0.001	0.001	0.006	0.001	-0.004	0.001	0.001	0.001
4/20/200	0	0.001	0.001	0.001	0.001	0.000	0.001	0.001	0.001
	0.01	0.001	0.001	0.002	0.001	-0.001	0.001	0.001	0.001
	0.05	0.000	-0.000	0.005	0.000	-0.004	-0.000	0.000	0.000
4/40/100	0	0.001	0.001	0.001	0.001	0.000	0.001	0.001	0.001
	0.01	0.001	0.001	0.002	0.001	-0.000	0.001	0.001	0.001
	0.05	0.000	0.000	0.003	0.000	-0.002	0.000	0.000	0.000
12/20/100	0	-0.000	-0.000	0.002	-0.000	-0.000	-0.000	-0.000	-0.000
	0.01	-0.000	-0.000	0.003	-0.000	-0.001	-0.000	-0.000	-0.000
	0.05	-0.003	-0.003	0.006	-0.003	-0.005	-0.003	-0.003	-0.003
12/20/200	0	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	0.01	-0.000	-0.000	0.002	-0.000	-0.001	-0.000	-0.000	-0.000
	0.05	-0.003	-0.003	0.005	-0.003	-0.004	-0.003	-0.003	-0.003
12/40/100	0	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	0.01	0.000	0.000	0.002	0.000	-0.000	0.000	0.000	0.000
	0.05	-0.001	-0.002	0.003	-0.001	-0.002	-0.002	-0.002	-0.002
20/20/100	0	-0.000	-0.000	0.001	-0.000	-0.001	-0.000	-0.000	-0.000
	0.01	-0.001	-0.001	0.002	-0.001	-0.001	-0.001	-0.001	-0.001
	0.05	-0.004	-0.004	0.005	-0.004	-0.005	-0.005	-0.004	-0.004
20/20/200	0	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	0.01	-0.001	-0.001	0.002	-0.001	-0.001	-0.001	-0.001	-0.001
	0.05	-0.004	-0.004	0.005	-0.004	-0.005	-0.004	-0.004	-0.004
20/40/100	0	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	0.01	-0.000	-0.000	0.002	-0.000	-0.000	-0.000	-0.000	-0.000
	0.05	-0.002	-0.002	0.003	-0.002	-0.002	-0.002	-0.002	-0.002
40/20/100	0	-0.001	-0.001	0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	0.01	-0.001	-0.001	0.002	-0.001	-0.002	-0.001	-0.001	-0.001
	0.05	-0.005	-0.005	0.005	-0.005	-0.006	-0.005	-0.005	-0.005
40/20/200	0	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	0.01	-0.001	-0.001	0.002	-0.001	-0.001	-0.001	-0.001	-0.001
	0.05	-0.005	-0.005	0.005	-0.005	-0.005	-0.005	-0.005	-0.005
40/40/100	0	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	0.01	-0.000	-0.000	0.002	-0.000	-0.000	-0.000	-0.000	-0.000
	0.05	-0.002	-0.002	0.003	-0.002	-0.003	-0.002	-0.002	-0.002

Table 6.13: Bias for τ_c^2 with 'moderate' heterogeneity and control group disease rates $r_B = \{0.35, 0.40, 0.45, 0.50\}$ based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

$k/N_j/m$	ρ	VC	DL	DLVC	DL2	MV	MVVC	ML	REML
4/20/100	0	0.000	0.000	0.001	0.000	-0.001	0.000	0.000	0.000
	0.01	-0.001	-0.001	0.001	-0.001	-0.003	-0.001	-0.001	-0.001
	0.05	-0.007	-0.007	-0.002	-0.007	-0.012	-0.007	-0.007	-0.007
4/20/200	0	0.002	0.001	0.002	0.001	0.000	0.001	0.001	0.001
	0.01	-0.000	-0.000	0.001	-0.000	-0.002	-0.000	-0.000	-0.000
	0.05	-0.006	-0.007	-0.002	-0.007	-0.011	-0.007	-0.007	-0.007
4/40/100	0	0.002	0.002	0.002	0.002	0.000	0.002	0.002	0.002
	0.01	0.001	0.001	0.002	0.001	-0.001	0.001	0.001	0.001
	0.05	-0.002	-0.002	0.000	-0.002	-0.005	-0.002	-0.002	-0.002
12/20/100	0	-0.001	-0.001	0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	0.01	-0.003	-0.003	0.000	-0.003	-0.004	-0.003	-0.003	-0.003
	0.05	-0.011	-0.011	-0.003	-0.011	-0.013	-0.011	-0.011	-0.011
12/20/200	0	0.001	0.001	0.001	0.001	0.000	0.001	0.001	0.001
	0.01	-0.002	-0.002	0.001	-0.002	-0.002	-0.002	-0.002	-0.002
	0.05	-0.010	-0.010	-0.002	-0.010	-0.012	-0.010	-0.010	-0.010
12/40/100	0	0.001	0.001	0.002	0.001	0.000	0.001	0.001	0.001
	0.01	-0.000	-0.000	0.001	-0.000	-0.001	-0.000	-0.000	-0.000
	0.05	-0.005	-0.005	-0.001	-0.005	-0.006	-0.005	-0.005	-0.005
20/20/100	0	-0.001	-0.001	0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	0.01	-0.003	-0.003	0.000	-0.003	-0.004	-0.003	-0.003	-0.003
	0.05	-0.012	-0.012	-0.002	-0.012	-0.013	-0.012	-0.012	-0.012
20/20/200	0	0.001	0.001	0.001	0.001	0.000	0.001	0.001	0.001
	0.01	-0.002	-0.002	0.001	-0.002	-0.002	-0.002	-0.002	-0.002
	0.05	-0.011	-0.012	-0.002	-0.011	-0.012	-0.012	-0.011	-0.011
20/40/100	0	0.001	0.001	0.001	0.001	0.000	0.001	0.001	0.001
	0.01	-0.000	-0.001	0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	0.05	-0.005	-0.005	-0.000	-0.005	-0.006	-0.005	-0.005	-0.005
40/20/100	0	-0.001	-0.001	0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	0.01	-0.003	-0.003	0.000	-0.003	-0.004	-0.003	-0.003	-0.003
	0.05	-0.013	-0.014	-0.002	-0.013	-0.014	-0.014	-0.014	-0.014
40/20/200	0	0.001	0.001	0.001	0.001	0.000	0.001	0.001	0.001
	0.01	-0.002	-0.002	0.001	-0.002	-0.002	-0.002	-0.002	-0.002
	0.05	-0.012	-0.012	-0.002	-0.012	-0.013	-0.012	-0.012	-0.012
40/40/100	0	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	0.01	-0.001	-0.001	0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	0.05	-0.006	-0.006	-0.001	-0.006	-0.006	-0.006	-0.006	-0.006

Table 6.14: Bias for τ_c^2 with 'high' heterogeneity and control group disease rates $r_B = \{0.35, 0.40, 0.45, 0.50\}$ based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

$k/N_j/m$	ρ	VC	DL	DLVC	DL2	MV	MVVC	ML	REML
4/20/100	0	0.000	0.000	0.001	0.000	-0.002	0.000	0.000	0.000
	0.01	-0.003	-0.003	-0.001	-0.003	-0.005	-0.003	-0.003	-0.003
	0.05	-0.014	-0.014	-0.009	-0.014	-0.020	-0.014	-0.014	-0.014
4/20/200	0	0.002	0.002	0.002	0.002	0.000	0.002	0.002	0.002
	0.01	-0.001	-0.001	0.000	-0.001	-0.004	-0.001	-0.001	-0.001
	0.05	-0.013	-0.013	-0.008	-0.013	-0.019	-0.013	-0.013	-0.013
4/40/100	0	0.002	0.002	0.002	0.002	0.001	0.002	0.002	0.002
	0.01	0.000	0.000	0.001	0.000	-0.002	0.000	0.000	0.000
	0.05	-0.005	-0.006	-0.003	-0.005	-0.009	-0.006	-0.005	-0.005
12/20/100	0	-0.001	-0.001	0.001	-0.001	-0.002	-0.001	-0.001	-0.001
	0.01	-0.004	-0.005	-0.002	-0.005	-0.005	-0.005	-0.005	-0.005
	0.05	-0.019	-0.019	-0.010	-0.019	-0.020	-0.019	-0.019	-0.019
12/20/200	0	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001
	0.01	-0.003	-0.003	-0.000	-0.003	-0.003	-0.003	-0.003	-0.003
	0.05	-0.017	-0.017	-0.009	-0.017	-0.019	-0.017	-0.017	-0.017
12/40/100	0	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001
	0.01	-0.001	-0.001	0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	0.05	-0.008	-0.008	-0.003	-0.008	-0.009	-0.008	-0.008	-0.008
20/20/100	0	-0.001	-0.001	0.000	-0.001	-0.002	-0.001	-0.001	-0.001
	0.01	-0.005	-0.005	-0.002	-0.005	-0.006	-0.005	-0.005	-0.005
	0.05	-0.020	-0.020	-0.010	-0.020	-0.021	-0.020	-0.020	-0.020
20/20/200	0	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001
	0.01	-0.003	-0.003	-0.001	-0.003	-0.003	-0.003	-0.003	-0.003
	0.05	-0.018	-0.019	-0.009	-0.018	-0.019	-0.019	-0.018	-0.018
20/40/100	0	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001
	0.01	-0.001	-0.001	0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	0.05	-0.009	-0.009	-0.004	-0.009	-0.009	-0.009	-0.009	-0.009
40/20/100	0	-0.001	-0.001	0.000	-0.001	-0.002	-0.001	-0.001	-0.001
	0.01	-0.005	-0.005	-0.002	-0.005	-0.006	-0.005	-0.005	-0.005
	0.05	-0.021	-0.021	-0.010	-0.021	-0.022	-0.021	-0.021	-0.021
40/20/200	0	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001
	0.01	-0.003	-0.003	-0.001	-0.003	-0.003	-0.003	-0.003	-0.003
	0.05	-0.020	-0.020	-0.010	-0.020	-0.020	-0.020	-0.020	-0.020
40/40/100	0	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001
	0.01	-0.001	-0.001	0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	0.05	-0.009	-0.009	-0.004	-0.009	-0.010	-0.009	-0.009	-0.009

Table 6.15: Confidence intervals for τ_c^2 with 'no' heterogeneity, control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ for Q profile(PQ), Biggerstaff-Tweedie(BT), Sidik-Jonkman(SJ) and nonparametric bootstraps(NB) based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

		QP	BT	SJ	NB
$k/N_j/m$	ρ	C(L,R)W	C(L,R)W	C(L,R)W	C(L,R)W
4/20/100	0	98.6(1.4, 0.0)0.22	99.3(0.7, 0.0)0.27	0.0(100, 0.0)0.06	100(0.0, 0.0)0.01
	0.01	98.0(2.0, 0.0)0.46	98.9(1.1, 0.0)0.52	0.0(100, 0.0)0.14	100(0.0, 0.0)0.02
	0.05	98.5(1.5, 0.0)1.29	99.4(0.6, 0.0)1.51	0.0(100, 0.0)0.37	100(0.0, 0.0)0.06
4/20/200	0	98.6(1.4, 0.0)0.11	99.2(0.8, 0.0)0.13	0.0(100, 0.0)0.03	100(0.0, 0.0)0.01
	0.01	97.5(2.5, 0.0)0.33	98.2(1.8, 0.0)0.38	0.0(100, 0.0)0.11	100(0.0, 0.0)0.02
	0.05	98.1(1.9, 0.0)1.15	99.3(0.7, 0.0)1.39	0.0(100, 0.0)0.33	100(0.0, 0.0)0.06
4/40/100	0	97.7(2.3, 0.0)0.12	98.7(1.3, 0.0)0.13	0.0(100, 0.0)0.04	100(0.0, 0.0)0.01
	0.01	97.8(2.2, 0.0)0.22	99.3(0.7, 0.0)0.25	0.0(100, 0.0)0.07	100(0.0, 0.0)0.01
	0.05	98.0(2.0, 0.0)0.68	99.0(1.0, 0.0)0.77	0.0(100, 0.0)0.21	100(0.0, 0.0)0.04
12/20/100	0	98.6(1.4, 0.0)0.03	99.1(0.9, 0.0)0.04	0.0(100, 0.0)0.01	99.9(0.1, 0.0)0.01
	0.01	98.3(1.7, 0.0)0.07	98.8(1.2, 0.0)0.07	0.0(100, 0.0)0.01	99.7(0.3, 0.0)0.02
	0.05	98.7(1.3, 0.0)0.17	99.2(0.8, 0.0)0.19	0.0(100, 0.0)0.03	99.8(0.2, 0.0)0.05
12/20/200	0	98.4(1.6, 0.0)0.02	99.1(0.9, 0.0)0.02	0.0(100, 0.0)0.00	99.9(0.1, 0.0)0.00
	0.01	97.9(2.1, 0.0)0.05	99.3(0.7, 0.0)0.05	0.0(100, 0.0)0.01	99.7(0.3, 0.0)0.01
	0.05	98.6(1.4, 0.0)0.16	99.6(0.4, 0.0)0.18	0.0(100, 0.0)0.03	100(0.0, 0.0)0.05
12/40/100	0	98.1(1.9, 0.0)0.02	99.0(1.0, 0.0)0.02	0.0(100, 0.0)0.00	99.9(0.1, 0.0)0.00
	0.01	97.6(2.4, 0.0)0.03	98.4(1.6, 0.0)0.03	0.0(100, 0.0)0.01	99.9(0.1, 0.0)0.01
	0.05	97.7(2.3, 0.0)0.10	98.9(1.1, 0.0)0.10	0.0(100, 0.0)0.02	99.8(0.2, 0.0)0.03
20/20/100	0	98.9(1.1, 0.0)0.02	99.6(0.4, 0.0)0.02	0.0(100, 0.0)0.00	99.8(0.2, 0.0)0.01
	0.01	97.6(2.4, 0.0)0.04	98.9(1.1, 0.0)0.04	0.0(100, 0.0)0.01	99.7(0.3, 0.0)0.02
	0.05	98.7(1.3, 0.0)0.11	99.1(0.9, 0.0)0.11	0.0(100, 0.0)0.01	99.6(0.4, 0.0)0.05
20/20/200	0	98.4(1.6, 0.0)0.01	98.8(1.2, 0.0)0.01	0.0(100, 0.0)0.00	99.6(0.4, 0.0)0.00
	0.01	98.2(1.8, 0.0)0.03	98.8(1.2, 0.0)0.03	0.0(100, 0.0)0.00	99.6(0.4, 0.0)0.01
	0.05	97.7(2.3, 0.0)0.10	98.5(1.5, 0.0)0.11	0.0(100, 0.0)0.01	99.7(0.3, 0.0)0.04
20/40/100	0	98.7(1.3, 0.0)0.01	99.0(1.0, 0.0)0.01	0.0(100, 0.0)0.00	99.7(0.3, 0.0)0.00
	0.01	97.2(2.8, 0.0)0.02	98.1(1.9, 0.0)0.02	0.0(100, 0.0)0.00	99.1(0.9, 0.0)0.01
	0.05	98.1(1.9, 0.0)0.06	98.8(1.2, 0.0)0.06	0.0(100, 0.0)0.01	99.4(0.6, 0.0)0.02
40/20/100	0	99.4(0.6, 0.0)0.01	99.7(0.3, 0.0)0.01	0.0(100, 0.0)0.00	99.8(0.2, 0.0)0.01
	0.01	98.3(1.7, 0.0)0.02	99.0(1.0, 0.0)0.02	0.0(100, 0.0)0.00	99.2(0.8, 0.0)0.01
	0.05	98.5(1.5, 0.0)0.06	99.2(0.8, 0.0)0.06	0.0(100, 0.0)0.01	99.6(0.4, 0.0)0.03
40/20/200	0	99.2(0.8, 0.0)0.01	99.7(0.3, 0.0)0.01	0.0(100, 0.0)0.00	99.8(0.2, 0.0)0.00
	0.01	97.8(2.2, 0.0)0.02	98.9(1.1, 0.0)0.02	0.0(100, 0.0)0.00	99.3(0.7, 0.0)0.01
	0.05	98.4(1.6, 0.0)0.05	99.0(1.0, 0.0)0.06	0.0(100, 0.0)0.01	99.5(0.5, 0.0)0.03
40/40/100	0	98.6(1.4, 0.0)0.01	99.0(1.0, 0.0)0.01	0.0(100, 0.0)0.00	99.7(0.3, 0.0)0.00
	0.01	97.8(2.2, 0.0)0.01	98.3(1.7, 0.0)0.01	0.0(100, 0.0)0.00	99.4(0.6, 0.0)0.01
	0.05	98.0(2.0, 0.0)0.03	99.1(0.9, 0.0)0.03	0.0(100, 0.0)0.00	99.5(0.5, 0.0)0.02

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.16: Confidence intervals for τ_c^2 with 'no' heterogeneity, control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ for profile likelihood, and Wald-Type based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

$k/N_j/m$	ρ	pML C(L,R)W	pRE C(L,R)W	wML C(L,R)W	wRE C(L,R)W
4/20/100	0	99.9(0.1, 0.0)0.07	99.3(0.7, 0.0)0.14	100(0.0, 0.0)0.05	100(0.0, 0.0)0.07
	0.01	99.5(0.5, 0.0)0.14	98.8(1.2, 0.0)0.28	100(0.0, 0.0)0.09	100(0.0, 0.0)0.13
	0.05	99.7(0.3, 0.0)0.38	99.2(0.8, 0.0)0.80	100(0.0, 0.0)0.26	100(0.0, 0.0)0.37
4/20/200	0	100(0.0, 0.0)0.03	99.6(0.4, 0.0)0.07	100(0.0, 0.0)0.02	100(0.0, 0.0)0.03
	0.01	99.2(0.8, 0.0)0.10	98.1(1.9, 0.0)0.21	100(0.0, 0.0)0.07	100(0.0, 0.0)0.10
	0.05	99.4(0.6, 0.0)0.36	98.7(1.3, 0.0)0.73	100(0.0, 0.0)0.24	100(0.0, 0.0)0.34
4/40/100	0	99.5(0.5, 0.0)0.04	98.9(1.1, 0.0)0.07	100(0.0, 0.0)0.02	100(0.0, 0.0)0.03
	0.01	99.9(0.1, 0.0)0.07	99.3(0.7, 0.0)0.14	100(0.0, 0.0)0.05	100(0.0, 0.0)0.06
	0.05	99.4(0.6, 0.0)0.20	98.8(1.2, 0.0)0.42	100(0.0, 0.0)0.13	100(0.0, 0.0)0.19
12/20/100	0	99.4(0.6, 0.0)0.02	99.1(0.9, 0.0)0.03	100(0.0, 0.0)0.03	100(0.0, 0.0)0.03
	0.01	99.0(1.0, 0.0)0.04	98.5(1.5, 0.0)0.06	100(0.0, 0.0)0.05	100(0.0, 0.0)0.06
	0.05	99.2(0.8, 0.0)0.13	99.1(0.9, 0.0)0.15	100(0.0, 0.0)0.15	100(0.0, 0.0)0.16
12/20/200	0	99.4(0.6, 0.0)0.01	98.8(1.2, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.02
	0.01	99.2(0.8, 0.0)0.03	98.9(1.1, 0.0)0.04	100(0.0, 0.0)0.04	100(0.0, 0.0)0.04
	0.05	99.4(0.6, 0.0)0.12	98.5(1.5, 0.0)0.14	100(0.0, 0.0)0.14	100(0.0, 0.0)0.15
12/40/100	0	99.4(0.6, 0.0)0.01	98.9(1.1, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	98.4(1.6, 0.0)0.02	97.9(2.1, 0.0)0.03	100(0.0, 0.0)0.03	100(0.0, 0.0)0.03
	0.05	98.6(1.4, 0.0)0.07	98.0(2.0, 0.0)0.08	100(0.0, 0.0)0.08	100(0.0, 0.0)0.09
20/20/100	0	99.4(0.6, 0.0)0.01	99.1(0.9, 0.0)0.02	99.9(0.1, 0.0)0.02	99.9(0.1, 0.0)0.02
	0.01	99.0(1.0, 0.0)0.03	98.3(1.7, 0.0)0.03	100(0.0, 0.0)0.04	100(0.0, 0.0)0.04
	0.05	99.1(0.9, 0.0)0.08	98.4(1.6, 0.0)0.10	100(0.0, 0.0)0.11	100(0.0, 0.0)0.12
20/20/200	0	98.9(1.1, 0.0)0.01	98.8(1.2, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	99.0(1.0, 0.0)0.02	98.5(1.5, 0.0)0.02	100(0.0, 0.0)0.03	100(0.0, 0.0)0.03
	0.05	98.6(1.4, 0.0)0.08	97.5(2.5, 0.0)0.09	100(0.0, 0.0)0.10	100(0.0, 0.0)0.11
20/40/100	0	99.0(1.0, 0.0)0.01	98.8(1.2, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	98.0(2.0, 0.0)0.01	97.6(2.4, 0.0)0.02	100(0.0, 0.0)0.02	100(0.0, 0.0)0.02
	0.05	98.7(1.3, 0.0)0.04	98.0(2.0, 0.0)0.05	100(0.0, 0.0)0.06	100(0.0, 0.0)0.06
40/20/100	0	99.6(0.4, 0.0)0.01	99.5(0.5, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	99.2(0.8, 0.0)0.02	98.7(1.3, 0.0)0.02	99.9(0.1, 0.0)0.03	99.9(0.1, 0.0)0.03
	0.05	98.4(1.6, 0.0)0.05	97.9(2.1, 0.0)0.06	99.9(0.1, 0.0)0.08	99.9(0.1, 0.0)0.08
40/20/200	0	99.7(0.3, 0.0)0.00	99.5(0.5, 0.0)0.00	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	98.5(1.5, 0.0)0.01	97.9(2.1, 0.0)0.01	99.9(0.1, 0.0)0.02	99.9(0.1, 0.0)0.02
	0.05	97.5(2.5, 0.0)0.05	97.0(3.0, 0.0)0.05	100(0.0, 0.0)0.07	100(0.0, 0.0)0.07
40/40/100	0	98.7(1.3, 0.0)0.00	98.5(1.5, 0.0)0.00	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	98.3(1.7, 0.0)0.01	97.8(2.2, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.05	98.4(1.6, 0.0)0.03	97.9(2.1, 0.0)0.03	100(0.0, 0.0)0.04	99.9(0.1, 0.0)0.04

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.17: Confidence intervals for τ_c^2 with 'low' heterogeneity, control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ for Q profile(PQ), Biggerstaff-Tweedie(BT), Sidik-Jonkman(SJ) and nonparametric bootstraps(NB) based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

		QP	BT	SJ	NB
$k/N_j/m$	ρ	C(L,R)W	C(L,R)W	C(L,R)W	C(L,R)W
4/20/100	0	94.1(1.7, 4.2)0.29	99.3(0.7, 0.0)0.32	23.8(17.0,59.2)0.11	50.9(0.0,49.1)0.02
	0.01	95.6(1.0, 3.3)0.55	99.4(0.6, 0.0)0.59	27.0(16.6,56.4)0.21	50.2(0.0,49.8)0.03
	0.05	95.9(0.3, 3.8)1.34	99.8(0.2, 0.0)1.58	23.4(11.3,65.3)0.40	44.2(0.0,55.8)0.07
4/20/200	0	93.9(4.5, 1.6)0.20	98.2(1.8, 0.0)0.20	26.6(30.4,43.0)0.09	63.3(0.0,36.7)0.01
	0.01	97.3(1.3, 1.4)0.44	99.4(0.6, 0.0)0.46	30.7(17.4,52.0)0.17	55.7(0.0,44.3)0.03
	0.05	96.9(0.4, 2.7)1.22	99.9(0.1, 0.0)1.43	28.6(10.6,60.9)0.34	46.1(0.0,53.9)0.06
4/40/100	0	92.7(6.0, 1.3)0.20	97.6(2.4, 0.0)0.19	26.9(30.9,42.2)0.10	66.3(0.0,33.7)0.01
	0.01	95.7(1.8, 2.5)0.32	98.7(1.3, 0.0)0.32	31.1(21.2,47.6)0.13	59.5(0.0,40.5)0.02
	0.05	95.3(0.6, 4.1)0.72	99.7(0.3, 0.0)0.83	25.5(14.3,60.2)0.24	46.0(0.0,54.0)0.04
12/20/100	0	95.7(0.9, 3.4)0.05	99.8(0.2, 0.0)0.04	25.9(21.9,52.2)0.01	73.4(0.1,26.5)0.01
	0.01	96.5(0.6, 2.9)0.08	99.7(0.3, 0.0)0.07	29.2(14.5,56.3)0.02	69.1(0.1,30.8)0.02
	0.05	95.0(0.0, 5.0)0.20	100(0.0, 0.0)0.20	26.9(8.5,64.6)0.04	63.6(0.0,36.4)0.06
12/20/200	0	94.0(4.7, 1.3)0.03	97.9(2.1, 0.0)0.03	29.9(38.6,31.5)0.01	87.2(0.4,12.4)0.01
	0.01	96.1(1.0, 2.9)0.06	99.7(0.3, 0.0)0.06	30.8(19.4,49.8)0.02	72.8(0.0,27.2)0.02
	0.05	94.9(0.1, 5.0)0.18	100(0.0, 0.0)0.18	25.8(7.9,66.3)0.03	64.1(0.0,35.9)0.06
12/40/100	0	94.5(3.9, 1.6)0.03	98.4(1.6, 0.0)0.02	29.6(34.8,35.6)0.01	83.9(0.6,15.5)0.01
	0.01	95.1(1.6, 3.3)0.05	98.9(1.1, 0.0)0.04	31.8(25.0,43.2)0.01	78.4(0.4,21.2)0.02
	0.05	94.8(0.3, 4.9)0.11	99.9(0.1, 0.0)0.11	29.0(13.0,58.0)0.03	67.8(0.1,32.1)0.04
20/20/100	0	94.7(0.6, 4.7)0.03	99.8(0.2, 0.0)0.03	27.7(20.9,51.4)0.01	77.3(0.1,22.6)0.01
	0.01	93.7(0.2, 6.1)0.05	99.7(0.3, 0.0)0.04	27.7(16.0,56.3)0.01	70.2(0.1,29.7)0.02
	0.05	92.0(0.0, 8.0)0.11	100(0.0, 0.0)0.12	23.1(5.2,71.7)0.02	60.6(0.0,39.4)0.05
20/20/200	0	94.0(4.5, 1.5)0.02	98.5(1.5, 0.0)0.02	30.3(37.6,32.1)0.01	88.1(0.6,11.3)0.01
	0.01	94.8(0.7, 4.5)0.04	99.8(0.2, 0.0)0.03	31.5(14.2,54.3)0.01	75.5(0.1,24.4)0.02
	0.05	91.3(0.1, 8.6)0.10	99.9(0.1, 0.0)0.11	23.8(5.0,71.2)0.02	63.2(0.0,36.8)0.05
20/40/100	0	92.8(6.0, 1.2)0.02	97.4(2.6, 0.0)0.02	27.2(41.1,31.7)0.01	88.5(0.8,10.7)0.01
	0.01	95.3(1.6, 3.1)0.03	99.2(0.8, 0.0)0.03	29.5(26.4,44.1)0.01	81.9(0.5,17.6)0.01
	0.05	93.4(0.1, 6.5)0.06	99.9(0.1, 0.0)0.06	24.8(10.6,64.6)0.01	67.7(0.1,32.2)0.03
40/20/100	0	93.2(0.7, 6.1)0.02	99.6(0.4, 0.0)0.01	26.7(16.8,56.5)0.00	77.0(0.2,22.8)0.01
	0.01	93.2(0.3, 6.5)0.03	99.9(0.1, 0.0)0.03	25.6(10.8,63.6)0.01	72.8(0.1,27.1)0.02
	0.05	86.1(0.0,13.9)0.07	100(0.0, 0.0)0.07	17.9(2.4,79.7)0.01	60.9(0.0,39.1)0.04
40/20/200	0	94.5(5.1, 0.4)0.01	98.2(1.8, 0.0)0.01	31.6(41.9,26.5)0.00	93.5(0.9, 5.6)0.01
	0.01	92.2(0.5, 7.3)0.02	99.9(0.1, 0.0)0.02	30.3(12.4,57.3)0.00	77.1(0.0,22.9)0.01
	0.05	86.6(0.0,13.4)0.06	100(0.0, 0.0)0.06	17.5(2.3,80.2)0.01	60.8(0.0,39.2)0.04
40/40/100	0	91.9(7.4, 0.7)0.01	96.7(3.3, 0.0)0.01	30.0(44.5,25.5)0.00	91.9(2.2, 5.9)0.01
	0.01	96.2(1.5, 2.3)0.02	99.9(0.1, 0.0)0.02	33.2(22.3,44.5)0.00	85.0(0.1,14.9)0.01
	0.05	88.3(0.0,11.7)0.04	100(0.0, 0.0)0.04	20.3(6.6,73.1)0.01	65.3(0.0,34.7)0.02

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.18: Confidence intervals for τ_c^2 with 'low' heterogeneity, control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ for profile likelihood, and Wald-Type based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

$k/N_j/m$	ρ	pML C(L,R)W	pRE C(L,R)W	wML C(L,R)W	wRE C(L,R)W
4/20/100	0	99.5(0.5, 0.0)0.09	99.3(0.7, 0.0)0.18	100(0.0, 0.0)0.05	100(0.0, 0.0)0.08
	0.01	99.7(0.3, 0.0)0.16	99.4(0.6, 0.0)0.33	100(0.0, 0.0)0.10	100(0.0, 0.0)0.14
	0.05	100(0.0, 0.0)0.40	99.9(0.1, 0.0)0.83	100(0.0, 0.0)0.26	100(0.0, 0.0)0.37
4/20/200	0	98.9(1.1, 0.0)0.06	98.3(1.7, 0.0)0.12	100(0.0, 0.0)0.03	100(0.0, 0.0)0.05
	0.01	99.6(0.4, 0.0)0.13	99.4(0.6, 0.0)0.26	100(0.0, 0.0)0.08	100(0.0, 0.0)0.11
	0.05	99.9(0.1, 0.0)0.36	99.9(0.1, 0.0)0.75	100(0.0, 0.0)0.24	100(0.0, 0.0)0.34
4/40/100	0	99.6(0.4, 0.0)0.06	97.9(2.1, 0.0)0.12	100(0.0, 0.0)0.03	100(0.0, 0.0)0.05
	0.01	99.7(0.3, 0.0)0.09	99.0(1.0, 0.0)0.19	100(0.0, 0.0)0.05	100(0.0, 0.0)0.08
	0.05	100(0.0, 0.0)0.21	99.7(0.3, 0.0)0.44	100(0.0, 0.0)0.14	100(0.0, 0.0)0.20
12/20/100	0	99.9(0.1, 0.0)0.03	99.7(0.3, 0.0)0.04	100(0.0, 0.0)0.03	100(0.0, 0.0)0.03
	0.01	100(0.0, 0.0)0.05	99.7(0.3, 0.0)0.06	100(0.0, 0.0)0.05	100(0.0, 0.0)0.06
	0.05	100(0.0, 0.0)0.13	100(0.0, 0.0)0.16	100(0.0, 0.0)0.15	100(0.0, 0.0)0.17
12/20/200	0	98.5(1.5, 0.0)0.02	97.7(2.3, 0.0)0.02	100(0.0, 0.0)0.02	100(0.0, 0.0)0.02
	0.01	99.9(0.1, 0.0)0.04	99.4(0.6, 0.0)0.05	100(0.0, 0.0)0.04	100(0.0, 0.0)0.05
	0.05	100(0.0, 0.0)0.12	100(0.0, 0.0)0.15	100(0.0, 0.0)0.14	100(0.0, 0.0)0.15
12/40/100	0	99.1(0.9, 0.0)0.02	98.5(1.5, 0.0)0.02	100(0.0, 0.0)0.02	100(0.0, 0.0)0.02
	0.01	99.2(0.8, 0.0)0.03	98.7(1.3, 0.0)0.04	100(0.0, 0.0)0.03	100(0.0, 0.0)0.03
	0.05	99.9(0.1, 0.0)0.07	99.8(0.2, 0.0)0.09	100(0.0, 0.0)0.08	100(0.0, 0.0)0.09
20/20/100	0	99.8(0.0, 0.2)0.02	99.8(0.2, 0.0)0.02	100(0.0, 0.0)0.02	100(0.0, 0.0)0.02
	0.01	98.9(0.3, 0.8)0.04	99.4(0.3, 0.3)0.04	100(0.0, 0.0)0.04	100(0.0, 0.0)0.04
	0.05	99.0(0.0, 1.0)0.09	99.5(0.0, 0.5)0.10	100(0.0, 0.0)0.11	100(0.0, 0.0)0.12
20/20/200	0	99.1(0.9, 0.0)0.01	98.0(2.0, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	99.1(0.2, 0.7)0.03	99.4(0.3, 0.3)0.03	100(0.0, 0.0)0.03	100(0.0, 0.0)0.03
	0.05	98.5(0.1, 1.4)0.08	99.3(0.1, 0.6)0.09	100(0.0, 0.0)0.10	100(0.0, 0.0)0.11
20/40/100	0	98.1(1.8, 0.1)0.01	97.4(2.6, 0.0)0.02	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	98.9(0.7, 0.4)0.02	99.1(0.9, 0.0)0.02	100(0.0, 0.0)0.02	100(0.0, 0.0)0.02
	0.05	98.7(0.1, 1.2)0.05	99.5(0.1, 0.4)0.06	100(0.0, 0.0)0.06	100(0.0, 0.0)0.06
40/20/100	0	94.3(0.2, 5.5)0.01	96.1(0.3, 3.6)0.01	100(0.0, 0.0)0.02	100(0.0, 0.0)0.02
	0.01	91.8(0.1, 8.1)0.02	94.4(0.1, 5.5)0.02	100(0.0, 0.0)0.03	100(0.0, 0.0)0.03
	0.05	85.2(0.0,14.8)0.06	89.2(0.0,10.8)0.06	99.9(0.0, 0.1)0.08	100(0.0, 0.0)0.08
40/20/200	0	97.7(1.6, 0.7)0.01	97.6(2.0, 0.4)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	92.1(0.1, 7.8)0.02	93.2(0.2, 6.6)0.02	99.9(0.0, 0.1)0.02	99.9(0.0, 0.1)0.02
	0.05	85.9(0.0,14.1)0.05	89.7(0.0,10.3)0.06	100(0.0, 0.0)0.07	100(0.0, 0.0)0.08
40/40/100	0	96.5(2.6, 0.9)0.01	95.9(3.5, 0.6)0.01	99.5(0.5, 0.0)0.01	99.5(0.5, 0.0)0.01
	0.01	95.6(0.3, 4.1)0.01	96.8(0.5, 2.7)0.01	100(0.0, 0.0)0.02	100(0.0, 0.0)0.02
	0.05	87.5(0.0,12.5)0.03	90.0(0.0,10.0)0.03	100(0.0, 0.0)0.04	100(0.0, 0.0)0.04

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.19: Confidence intervals for τ_c^2 with 'moderate' heterogeneity, control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ for Q profile(PQ), Biggerstaff-Tweedie(BT), Sidik-Jonkman(SJ) and nonparametric bootstraps(NB) based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

		QP	BT	SJ	NB
$k/N_j/m$	ρ	C(L,R)W	C(L,R)W	C(L,R)W	C(L,R)W
4/20/100	0	96.6(1.5, 1.9)0.42	99.4(0.6, 0.0)0.40	45.5(13.3,41.2)0.19	59.4(0.0,40.6)0.03
	0.01	96.5(0.8, 2.7)0.61	99.8(0.2, 0.0)0.64	41.1(7.6,51.4)0.24	46.8(0.0,53.2)0.04
	0.05	95.7(0.1, 4.2)1.42	99.9(0.1, 0.0)1.62	34.4(3.2,62.4)0.43	36.3(0.0,63.7)0.07
4/20/200	0	94.6(4.6, 0.8)0.28	98.1(1.9, 0.0)0.26	44.7(27.8,27.6)0.16	72.6(0.0,27.4)0.02
	0.01	96.6(0.5, 2.9)0.49	99.7(0.3, 0.0)0.49	45.9(8.5,45.7)0.20	50.9(0.0,49.1)0.03
	0.05	95.8(0.1, 4.1)1.35	100(0.0, 0.0)1.54	37.3(3.2,59.5)0.44	36.7(0.0,63.3)0.07
4/40/100	0	93.6(5.5, 0.9)0.29	97.5(2.5, 0.0)0.26	49.5(26.7,23.8)0.16	75.3(0.0,24.7)0.02
	0.01	96.7(0.9, 2.4)0.38	99.6(0.4, 0.0)0.37	45.7(13.8,40.5)0.18	58.8(0.0,41.2)0.02
	0.05	96.2(0.2, 3.6)0.84	99.9(0.1, 0.0)0.90	40.8(5.2,54.1)0.30	45.0(0.0,55.0)0.05
12/20/100	0	97.0(1.0, 2.0)0.06	99.7(0.3, 0.0)0.05	48.9(14.0,37.1)0.02	75.5(0.0,24.5)0.02
	0.01	94.5(0.1, 5.4)0.09	100(0.0, 0.0)0.08	39.8(7.5,52.7)0.03	60.9(0.0,39.1)0.03
	0.05	90.8(0.0, 9.2)0.21	100(0.0, 0.0)0.21	29.0(1.6,69.4)0.05	46.6(0.0,53.4)0.07
12/20/200	0	94.9(4.8, 0.3)0.04	97.6(2.4, 0.0)0.04	49.6(34.4,16.0)0.02	90.7(0.6, 8.7)0.02
	0.01	95.6(0.3, 4.1)0.07	99.9(0.1, 0.0)0.07	48.7(6.8,44.5)0.03	66.6(0.0,33.4)0.03
	0.05	89.6(0.0,10.4)0.19	100(0.0, 0.0)0.19	26.8(1.2,72.0)0.04	44.5(0.0,55.5)0.06
12/40/100	0	94.4(5.5, 0.1)0.04	97.6(2.4, 0.0)0.04	48.4(36.2,15.4)0.02	91.4(0.4, 8.2)0.02
	0.01	96.2(1.3, 2.5)0.06	99.7(0.3, 0.0)0.05	50.8(14.7,34.5)0.02	76.1(0.0,23.9)0.02
	0.05	92.5(0.0, 7.5)0.13	100(0.0, 0.0)0.12	39.0(3.2,57.8)0.04	55.6(0.0,44.4)0.04
20/20/100	0	94.8(0.7, 4.5)0.04	99.7(0.3, 0.0)0.03	48.3(13.5,38.2)0.01	76.8(0.3,22.9)0.02
	0.01	89.0(0.1,10.9)0.06	99.9(0.1, 0.0)0.05	37.5(4.2,58.3)0.02	58.7(0.1,41.2)0.03
	0.05	82.1(0.0,17.9)0.12	100(0.0, 0.0)0.12	20.6(0.3,79.1)0.02	40.2(0.0,59.8)0.06
20/20/200	0	93.4(6.5, 0.1)0.03	97.3(2.7, 0.0)0.02	45.6(45.1, 9.3)0.01	95.3(1.1, 3.6)0.01
	0.01	93.3(0.2, 6.5)0.05	99.9(0.1, 0.0)0.04	44.7(6.0,49.3)0.02	67.4(0.1,32.5)0.02
	0.05	82.2(0.0,17.8)0.11	100(0.0, 0.0)0.11	20.7(0.6,78.7)0.02	41.6(0.0,58.4)0.05
20/40/100	0	90.7(9.3, 0.0)0.03	95.5(4.5, 0.0)0.02	43.8(45.8,10.4)0.01	92.4(2.4, 5.2)0.01
	0.01	97.0(0.8, 2.2)0.04	99.7(0.3, 0.0)0.03	49.7(15.7,34.6)0.01	80.4(0.0,19.6)0.02
	0.05	87.1(0.0,12.9)0.07	100(0.0, 0.0)0.07	31.1(2.1,66.8)0.02	51.8(0.0,48.2)0.03
40/20/100	0	93.5(0.0, 6.5)0.02	90.5(0.0, 9.5)0.02	43.0(10.5,46.5)0.01	72.8(0.0,27.2)0.01
	0.01	83.0(0.0,17.0)0.03	76.3(0.0,23.7)0.03	30.4(2.4,67.2)0.01	53.9(0.0,46.1)0.02
	0.05	56.8(0.0,43.2)0.07	45.6(0.0,54.4)0.07	7.1(0.2,92.7)0.01	21.3(0.0,78.7)0.04
40/20/200	0	92.3(7.4, 0.3)0.02	96.7(2.8, 0.5)0.01	39.6(52.0, 8.4)0.01	94.8(2.3, 2.9)0.01
	0.01	89.6(0.2,10.2)0.03	83.8(0.1,16.1)0.02	38.0(3.5,58.5)0.01	63.2(0.0,36.8)0.02
	0.05	61.8(0.0,38.2)0.07	49.2(0.0,50.8)0.06	9.0(0.0,91.0)0.01	24.2(0.0,75.8)0.04
40/40/100	0	90.1(9.8, 0.1)0.02	94.6(5.0, 0.4)0.01	41.0(54.2, 4.8)0.01	94.0(4.4, 1.6)0.01
	0.01	96.2(0.6, 3.2)0.02	94.3(0.2, 5.5)0.02	51.1(13.5,35.4)0.01	83.3(0.2,16.5)0.02
	0.05	73.9(0.0,26.1)0.04	65.7(0.0,34.3)0.04	21.4(0.6,78.0)0.01	42.1(0.0,57.9)0.03

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.20: Confidence intervals for τ_c^2 with 'moderate' heterogeneity, control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ for profile likelihood, and Wald-Type based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

$k/N_j/m$	ρ	pML C(L,R)W	pRE C(L,R)W	wML C(L,R)W	wRE C(L,R)W
4/20/100	0	99.8(0.2, 0.0)0.12	99.5(0.5, 0.0)0.25	100(0.0, 0.0)0.06	100(0.0, 0.0)0.10
	0.01	100(0.0, 0.0)0.18	99.8(0.2, 0.0)0.37	100(0.0, 0.0)0.10	100(0.0, 0.0)0.15
	0.05	100(0.0, 0.0)0.40	99.9(0.1, 0.0)0.85	98.9(0.0, 1.1)0.26	99.7(0.0, 0.3)0.37
4/20/200	0	99.4(0.6, 0.0)0.08	97.6(2.4, 0.0)0.17	100(0.0, 0.0)0.04	100(0.0, 0.0)0.06
	0.01	99.8(0.2, 0.0)0.14	99.7(0.3, 0.0)0.29	99.6(0.0, 0.4)0.08	100(0.0, 0.0)0.12
	0.05	100(0.0, 0.0)0.39	100(0.0, 0.0)0.82	98.7(0.0, 1.3)0.24	99.7(0.0, 0.3)0.36
4/40/100	0	98.9(1.1, 0.0)0.08	97.5(2.5, 0.0)0.17	100(0.0, 0.0)0.04	100(0.0, 0.0)0.06
	0.01	99.7(0.3, 0.0)0.11	99.6(0.4, 0.0)0.23	100(0.0, 0.0)0.06	100(0.0, 0.0)0.09
	0.05	100(0.0, 0.0)0.24	99.9(0.1, 0.0)0.50	100(0.0, 0.0)0.14	100(0.0, 0.0)0.21
12/20/100	0	98.5(0.2, 1.3)0.04	99.5(0.4, 0.1)0.05	100(0.0, 0.0)0.04	100(0.0, 0.0)0.04
	0.01	95.4(0.0, 4.6)0.06	98.4(0.1, 1.5)0.07	90.1(0.0, 9.9)0.06	97.6(0.0, 2.4)0.07
	0.05	90.3(0.0, 9.7)0.14	96.9(0.0, 3.1)0.18	80.4(0.0,19.6)0.15	91.3(0.0, 8.7)0.17
12/20/200	0	98.3(1.5, 0.2)0.03	96.6(3.4, 0.0)0.04	100(0.0, 0.0)0.02	100(0.0, 0.0)0.03
	0.01	95.9(0.1, 4.0)0.05	97.7(0.2, 2.1)0.06	90.2(0.0, 9.8)0.05	97.0(0.0, 3.0)0.05
	0.05	89.3(0.0,10.7)0.13	96.8(0.0, 3.2)0.16	78.7(0.0,21.3)0.14	90.7(0.0, 9.3)0.15
12/40/100	0	98.7(1.3, 0.0)0.03	96.8(3.2, 0.0)0.04	100(0.0, 0.0)0.02	100(0.0, 0.0)0.03
	0.01	97.6(0.2, 2.2)0.04	99.1(0.5, 0.4)0.05	96.6(0.0, 3.4)0.03	99.5(0.0, 0.5)0.04
	0.05	92.5(0.0, 7.5)0.08	97.0(0.0, 3.0)0.10	88.6(0.0,11.4)0.08	97.4(0.0, 2.6)0.09
20/20/100	0	92.5(0.2, 7.3)0.03	95.3(0.3, 4.4)0.03	71.8(0.0,28.2)0.03	80.9(0.0,19.1)0.03
	0.01	82.9(0.1,17.0)0.04	88.6(0.1,11.3)0.05	54.1(0.0,45.9)0.04	65.2(0.0,34.8)0.05
	0.05	71.6(0.0,28.4)0.10	78.4(0.0,21.6)0.11	38.4(0.0,61.6)0.12	48.5(0.0,51.5)0.12
20/20/200	0	97.2(2.3, 0.5)0.02	95.8(4.0, 0.2)0.02	94.8(0.1, 5.1)0.02	96.7(0.1, 3.2)0.02
	0.01	87.5(0.1,12.4)0.04	91.9(0.1, 8.0)0.04	64.3(0.0,35.7)0.03	71.4(0.0,28.6)0.04
	0.05	71.3(0.0,28.7)0.09	79.8(0.0,20.2)0.10	40.5(0.0,59.5)0.11	49.1(0.0,50.9)0.11
20/40/100	0	94.9(4.1, 1.0)0.02	93.2(6.6, 0.2)0.02	94.0(0.0, 6.0)0.02	96.0(0.0, 4.0)0.02
	0.01	94.7(0.3, 5.0)0.03	96.6(0.4, 3.0)0.03	74.5(0.0,25.5)0.03	81.2(0.0,18.8)0.03
	0.05	77.5(0.0,22.5)0.06	85.1(0.0,14.9)0.06	47.0(0.0,53.0)0.06	55.6(0.0,44.4)0.07
40/20/100	0	84.1(0.0,15.9)0.02	87.0(0.0,13.0)0.02	68.0(0.0,32.0)0.02	72.0(0.0,28.0)0.02
	0.01	69.6(0.0,30.4)0.03	74.5(0.0,25.5)0.03	48.0(0.0,52.0)0.03	53.1(0.0,46.9)0.03
	0.05	41.8(0.0,58.2)0.06	47.9(0.0,52.1)0.06	20.7(0.0,79.3)0.08	25.4(0.0,74.6)0.08
40/20/200	0	96.5(2.6, 0.9)0.01	95.1(4.0, 0.9)0.01	95.6(0.1, 4.3)0.01	96.9(0.1, 3.0)0.01
	0.01	78.4(0.1,21.5)0.02	81.9(0.1,18.0)0.02	56.2(0.0,43.8)0.02	62.2(0.0,37.8)0.03
	0.05	44.9(0.0,55.1)0.06	51.0(0.0,49.0)0.06	25.6(0.0,74.4)0.07	29.8(0.0,70.2)0.08
40/40/100	0	94.3(5.1, 0.6)0.01	93.4(6.2, 0.4)0.02	97.0(0.5, 2.5)0.01	97.4(0.9, 1.7)0.01
	0.01	91.3(0.3, 8.4)0.02	93.5(0.3, 6.2)0.02	79.1(0.0,20.9)0.02	82.3(0.0,17.7)0.02
	0.05	59.4(0.0,40.6)0.04	65.0(0.0,35.0)0.04	37.5(0.0,62.5)0.04	43.8(0.0,56.2)0.04

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.21: Confidence intervals for τ_c^2 with 'high' heterogeneity, control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ for Q profile(PQ), Biggerstaff-Tweedie(BT), Sidik-Jonkman(SJ) and nonparametric bootstraps(NB) based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

		QP	BT	SJ	NB
$k/N_j/m$	ρ	C(L,R)W	C(L,R)W	C(L,R)W	C(L,R)W
4/20/100	0	98.1(0.8, 1.1)0.48	99.9(0.1, 0.0)0.45	55.1(9.5,35.4)0.24	57.6(0.0,42.4)0.03
	0.01	97.1(0.2, 2.7)0.69	100(0.0, 0.0)0.69	49.4(4.2,46.4)0.29	44.5(0.0,55.5)0.04
	0.05	96.9(0.0, 3.1)1.55	100(0.0, 0.0)1.71	40.7(0.5,58.8)0.49	32.2(0.0,67.8)0.08
4/20/200	0	93.0(6.1, 0.9)0.36	98.5(1.5, 0.0)0.33	56.0(25.9,18.1)0.23	78.7(0.0,21.3)0.03
	0.01	97.5(0.3, 2.2)0.58	100(0.0, 0.0)0.56	54.0(4.8,41.1)0.27	52.2(0.0,47.8)0.04
	0.05	96.3(0.0, 3.7)1.41	100(0.0, 0.0)1.57	42.1(1.1,56.8)0.46	31.6(0.0,68.4)0.07
4/40/100	0	92.8(6.7, 0.5)0.37	97.8(2.2, 0.0)0.33	56.0(27.8,16.2)0.24	80.3(0.0,19.7)0.03
	0.01	97.3(1.5, 1.2)0.48	99.6(0.4, 0.0)0.45	56.7(11.8,31.5)0.27	60.7(0.0,39.3)0.03
	0.05	95.2(0.3, 4.4)0.93	99.9(0.1, 0.0)0.96	45.7(2.8,51.5)0.37	39.4(0.0,60.6)0.05
12/20/100	0	97.2(0.4, 2.4)0.07	99.9(0.1, 0.0)0.06	58.8(7.9,33.3)0.03	71.7(0.1,28.2)0.03
	0.01	93.2(0.1, 6.7)0.11	100(0.0, 0.0)0.09	48.3(3.1,48.5)0.04	54.7(0.0,45.3)0.04
	0.05	85.0(0.0,15.0)0.22	100(0.0, 0.0)0.21	25.5(0.0,74.5)0.05	28.8(0.0,71.2)0.07
12/20/200	0	91.4(8.4, 0.2)0.05	97.8(2.2, 0.0)0.05	57.8(33.7, 8.5)0.03	92.2(0.6, 7.2)0.02
	0.01	95.5(0.2, 4.3)0.09	100(0.0, 0.0)0.08	55.0(3.4,41.6)0.03	60.5(0.0,39.5)0.03
	0.05	83.1(0.0,16.9)0.20	100(0.0, 0.0)0.20	26.8(0.1,73.1)0.05	30.0(0.0,70.0)0.06
12/40/100	0	91.4(8.4, 0.2)0.06	97.6(2.4, 0.0)0.05	58.1(36.3, 5.6)0.03	95.0(0.5, 4.5)0.02
	0.01	97.5(1.2, 1.3)0.07	99.7(0.3, 0.0)0.06	64.9(10.5,24.6)0.03	77.3(0.0,22.7)0.03
	0.05	90.0(0.0,10.0)0.14	100(0.0, 0.0)0.12	38.0(0.6,61.4)0.04	44.9(0.0,55.1)0.05
20/20/100	0	95.3(0.2, 4.5)0.05	100(0.0, 0.0)0.04	57.4(9.2,33.4)0.02	74.6(0.1,25.3)0.02
	0.01	88.2(0.0,11.8)0.07	94.6(0.0, 5.4)0.06	43.0(1.4,55.6)0.02	53.1(0.0,46.9)0.03
	0.05	68.3(0.0,31.7)0.13	82.1(0.0,17.9)0.13	16.1(0.1,83.8)0.03	20.7(0.0,79.3)0.06
20/20/200	0	87.8(12.2, 0.0)0.04	97.0(3.0, 0.0)0.03	50.8(44.5, 4.7)0.02	95.1(1.8, 3.1)0.02
	0.01	94.2(0.1, 5.7)0.06	97.5(0.0, 2.5)0.05	50.7(3.0,46.3)0.02	59.8(0.0,40.2)0.03
	0.05	71.4(0.0,28.6)0.12	80.2(0.0,19.8)0.12	14.7(0.0,85.3)0.03	23.5(0.0,76.5)0.06
20/40/100	0	87.3(12.6, 0.1)0.04	96.6(3.4, 0.0)0.03	47.9(47.2, 4.9)0.02	93.8(1.9, 4.3)0.02
	0.01	97.3(0.5, 2.2)0.05	99.3(0.1, 0.6)0.04	63.4(10.3,26.3)0.02	79.8(0.1,20.1)0.02
	0.05	82.1(0.0,17.9)0.08	91.9(0.0, 8.1)0.07	29.7(0.4,69.9)0.02	36.7(0.0,63.3)0.04
40/20/100	0	93.8(0.0, 6.2)0.03	89.2(0.0,10.8)0.02	54.3(5.5,40.2)0.01	69.8(0.0,30.2)0.02
	0.01	74.3(0.0,25.7)0.04	60.8(0.0,39.2)0.03	25.8(0.9,73.3)0.01	37.3(0.0,62.7)0.03
	0.05	33.2(0.0,66.8)0.07	21.5(0.0,78.5)0.07	2.2(0.0,97.8)0.01	7.1(0.0,92.9)0.05
40/20/200	0	80.2(19.8, 0.0)0.02	95.0(5.0, 0.0)0.02	39.4(58.3, 2.3)0.01	92.6(6.4, 1.0)0.01
	0.01	85.9(0.0,14.1)0.04	77.3(0.0,22.7)0.03	43.1(1.0,55.9)0.01	53.2(0.0,46.8)0.02
	0.05	38.4(0.0,61.6)0.07	25.9(0.0,74.1)0.07	2.9(0.0,97.1)0.01	7.7(0.0,92.3)0.04
40/40/100	0	77.9(22.1, 0.0)0.02	94.0(5.8, 0.2)0.02	39.4(58.9, 1.7)0.01	92.8(6.6, 0.6)0.01
	0.01	95.8(0.2, 4.0)0.03	94.0(0.0, 6.0)0.02	66.5(7.1,26.4)0.01	80.5(0.0,19.5)0.02
	0.05	59.1(0.0,40.9)0.05	44.8(0.0,55.2)0.04	15.0(0.0,85.0)0.01	20.7(0.0,79.3)0.03

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.22: Confidence intervals for τ_c^2 with 'high' heterogeneity, control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ for profile likelihood, and Wald-Type based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

$k/N_j/m$	ρ	pML C(L,R)W	pRE C(L,R)W	wML C(L,R)W	wRE C(L,R)W
4/20/100	0	100(0.0, 0.0)0.14	99.8(0.2, 0.0)0.28	100(0.0, 0.0)0.07	100(0.0, 0.0)0.11
	0.01	100(0.0, 0.0)0.19	100(0.0, 0.0)0.41	90.9(0.0, 9.1)0.11	99.9(0.0, 0.1)0.16
	0.05	99.9(0.0, 0.1)0.44	100(0.0, 0.0)0.92	82.0(0.0,18.0)0.27	97.1(0.0, 2.9)0.39
4/20/200	0	99.5(0.5, 0.0)0.11	98.2(1.8, 0.0)0.22	100(0.0, 0.0)0.05	100(0.0, 0.0)0.08
	0.01	100(0.0, 0.0)0.16	100(0.0, 0.0)0.34	89.6(0.0,10.4)0.09	99.1(0.0, 0.9)0.13
	0.05	99.9(0.0, 0.1)0.40	100(0.0, 0.0)0.84	80.6(0.0,19.4)0.24	97.0(0.0, 3.0)0.36
4/40/100	0	99.2(0.8, 0.0)0.11	97.4(2.6, 0.0)0.22	100(0.0, 0.0)0.05	100(0.0, 0.0)0.08
	0.01	99.8(0.2, 0.0)0.14	99.6(0.4, 0.0)0.29	97.7(0.0, 2.3)0.07	100(0.0, 0.0)0.11
	0.05	100(0.0, 0.0)0.26	99.9(0.1, 0.0)0.55	91.1(0.0, 8.9)0.15	99.6(0.0, 0.4)0.22
12/20/100	0	95.2(0.1, 4.7)0.05	97.8(0.1, 2.1)0.06	70.1(0.0,29.9)0.04	78.9(0.0,21.1)0.05
	0.01	86.7(0.0,13.3)0.07	93.5(0.0, 6.5)0.09	53.1(0.0,46.9)0.06	65.4(0.0,34.6)0.07
	0.05	69.1(0.0,30.9)0.14	82.4(0.0,17.6)0.18	27.7(0.0,72.3)0.15	39.6(0.0,60.4)0.17
12/20/200	0	97.5(2.1, 0.4)0.04	96.0(3.8, 0.2)0.05	92.4(0.0, 7.6)0.03	95.9(0.0, 4.1)0.03
	0.01	89.4(0.1,10.5)0.06	94.5(0.1, 5.4)0.07	58.2(0.0,41.8)0.05	70.2(0.0,29.8)0.06
	0.05	69.5(0.0,30.5)0.14	80.0(0.0,20.0)0.17	30.1(0.0,69.9)0.14	41.8(0.0,58.2)0.16
12/40/100	0	96.9(2.3, 0.8)0.04	96.2(3.7, 0.1)0.05	95.0(0.0, 5.0)0.03	97.2(0.0, 2.8)0.03
	0.01	96.2(0.3, 3.5)0.05	97.9(0.5, 1.6)0.06	76.2(0.0,23.8)0.04	83.8(0.0,16.2)0.05
	0.05	77.9(0.0,22.1)0.09	88.7(0.0,11.3)0.11	41.5(0.0,58.5)0.08	52.8(0.0,47.2)0.10
20/20/100	0	89.2(0.1,10.7)0.03	92.7(0.1, 7.2)0.04	70.6(0.0,29.4)0.03	77.1(0.0,22.9)0.03
	0.01	75.5(0.0,24.5)0.05	82.2(0.0,17.8)0.06	49.0(0.0,51.0)0.05	57.4(0.0,42.6)0.05
	0.05	47.4(0.0,52.6)0.10	56.8(0.0,43.2)0.11	20.7(0.0,79.3)0.12	27.1(0.0,72.9)0.12
20/20/200	0	96.0(3.3, 0.7)0.03	94.3(5.2, 0.5)0.03	96.4(0.1, 3.5)0.02	97.5(0.1, 2.4)0.03
	0.01	82.9(0.0,17.1)0.04	89.4(0.1,10.5)0.05	56.5(0.0,43.5)0.04	65.3(0.0,34.7)0.04
	0.05	51.7(0.0,48.3)0.09	59.5(0.0,40.5)0.11	24.0(0.0,76.0)0.11	30.5(0.0,69.5)0.12
20/40/100	0	95.6(3.8, 0.6)0.03	93.9(5.6, 0.5)0.03	95.8(0.0, 4.2)0.02	97.5(0.0, 2.5)0.03
	0.01	92.5(0.1, 7.4)0.04	95.4(0.1, 4.5)0.04	76.3(0.0,23.7)0.03	81.9(0.0,18.1)0.03
	0.05	62.2(0.0,37.8)0.06	71.7(0.0,28.3)0.07	33.7(0.0,66.3)0.06	42.5(0.0,57.5)0.07
40/20/100	0	84.9(0.0,15.1)0.02	88.4(0.0,11.6)0.02	65.1(0.0,34.9)0.02	71.7(0.0,28.3)0.02
	0.01	54.0(0.0,46.0)0.03	60.3(0.0,39.7)0.03	32.4(0.0,67.6)0.03	37.8(0.0,62.2)0.03
	0.05	18.7(0.0,81.3)0.06	22.7(0.0,77.3)0.07	5.8(0.0,94.2)0.08	8.4(0.0,91.6)0.08
40/20/200	0	93.0(6.9, 0.1)0.02	90.8(9.2, 0.0)0.02	97.8(1.0, 1.2)0.02	97.9(1.4, 0.7)0.02
	0.01	71.6(0.0,28.4)0.03	76.0(0.0,24.0)0.03	48.9(0.0,51.1)0.03	55.2(0.0,44.8)0.03
	0.05	22.3(0.0,77.7)0.06	28.4(0.0,71.6)0.07	7.9(0.0,92.1)0.08	10.2(0.0,89.8)0.08
40/40/100	0	92.6(7.2, 0.2)0.02	90.4(9.5, 0.1)0.02	98.3(0.9, 0.8)0.02	98.1(1.4, 0.5)0.02
	0.01	91.3(0.0, 8.7)0.02	93.2(0.0, 6.8)0.03	77.5(0.0,22.5)0.02	81.6(0.0,18.4)0.02
	0.05	38.4(0.0,61.6)0.04	43.4(0.0,56.6)0.04	17.8(0.0,82.2)0.04	22.2(0.0,77.8)0.05

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.23: Confidence intervals for τ_c^2 with 'no' heterogeneity, control group disease rates $r_B = \{0.35, 0.40, 0.45, 0.50\}$ for Q profile(PQ), Biggerstaff-Tweedie(BT), Sidik-Jonkman(SJ) and nonparametric bootstraps(NB) based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

		QP	BT	SJ	NB
$k/N_j/m$	ρ	C(L,R)W	C(L,R)W	C(L,R)W	C(L,R)W
4/20/100	0	98.4(1.6, 0.0)0.06	99.3(0.7, 0.0)0.06	0.0(100, 0.0)0.02	100(0.0, 0.0)0.00
	0.01	97.0(3.0, 0.0)0.12	98.0(2.0, 0.0)0.12	0.0(100, 0.0)0.04	100(0.0, 0.0)0.01
	0.05	97.4(2.6, 0.0)0.33	98.6(1.4, 0.0)0.35	0.0(100, 0.0)0.11	100(0.0, 0.0)0.02
4/20/200	0	98.2(1.8, 0.0)0.03	98.9(1.1, 0.0)0.03	0.0(100, 0.0)0.01	100(0.0, 0.0)0.00
	0.01	97.0(3.0, 0.0)0.09	98.2(1.8, 0.0)0.09	0.0(100, 0.0)0.03	100(0.0, 0.0)0.00
	0.05	97.2(2.8, 0.0)0.32	97.9(2.1, 0.0)0.33	0.0(100, 0.0)0.11	100(0.0, 0.0)0.02
4/40/100	0	98.1(1.9, 0.0)0.03	98.7(1.3, 0.0)0.03	0.0(100, 0.0)0.01	100(0.0, 0.0)0.00
	0.01	97.3(2.7, 0.0)0.06	98.3(1.7, 0.0)0.06	0.0(100, 0.0)0.02	100(0.0, 0.0)0.00
	0.05	97.0(3.0, 0.0)0.17	98.1(1.9, 0.0)0.17	0.0(100, 0.0)0.06	100(0.0, 0.0)0.01
12/20/100	0	99.1(0.9, 0.0)0.01	99.2(0.8, 0.0)0.01	0.0(100, 0.0)0.00	100(0.0, 0.0)0.00
	0.01	96.4(3.6, 0.0)0.02	98.2(1.8, 0.0)0.02	0.0(100, 0.0)0.00	99.7(0.3, 0.0)0.01
	0.05	96.4(3.6, 0.0)0.05	97.9(2.1, 0.0)0.05	0.0(100, 0.0)0.01	99.7(0.3, 0.0)0.02
12/20/200	0	99.3(0.7, 0.0)0.00	99.5(0.5, 0.0)0.00	0.0(100, 0.0)0.00	100(0.0, 0.0)0.00
	0.01	97.1(2.9, 0.0)0.01	98.0(2.0, 0.0)0.01	0.0(100, 0.0)0.00	99.6(0.4, 0.0)0.00
	0.05	97.3(2.7, 0.0)0.05	98.1(1.9, 0.0)0.05	0.0(100, 0.0)0.01	99.8(0.2, 0.0)0.02
12/40/100	0	98.0(2.0, 0.0)0.00	98.7(1.3, 0.0)0.00	0.0(100, 0.0)0.00	99.7(0.3, 0.0)0.00
	0.01	97.5(2.5, 0.0)0.01	98.3(1.7, 0.0)0.01	0.0(100, 0.0)0.00	99.4(0.6, 0.0)0.00
	0.05	97.2(2.8, 0.0)0.03	97.7(2.3, 0.0)0.03	0.0(100, 0.0)0.01	99.3(0.7, 0.0)0.01
20/20/100	0	98.1(1.9, 0.0)0.00	98.8(1.2, 0.0)0.01	0.0(100, 0.0)0.00	99.3(0.7, 0.0)0.00
	0.01	96.2(3.8, 0.0)0.01	96.9(3.1, 0.0)0.01	0.0(100, 0.0)0.00	99.1(0.9, 0.0)0.01
	0.05	96.8(3.2, 0.0)0.03	98.3(1.7, 0.0)0.03	0.0(100, 0.0)0.01	98.9(1.1, 0.0)0.01
20/20/200	0	98.8(1.2, 0.0)0.00	99.1(0.9, 0.0)0.00	0.0(100, 0.0)0.00	99.7(0.3, 0.0)0.00
	0.01	97.4(2.6, 0.0)0.01	98.5(1.5, 0.0)0.01	0.0(100, 0.0)0.00	99.7(0.3, 0.0)0.00
	0.05	97.7(2.3, 0.0)0.03	98.8(1.2, 0.0)0.03	0.0(100, 0.0)0.00	99.5(0.5, 0.0)0.01
20/40/100	0	97.8(2.2, 0.0)0.00	98.3(1.7, 0.0)0.00	0.0(100, 0.0)0.00	99.7(0.3, 0.0)0.00
	0.01	97.5(2.5, 0.0)0.01	98.6(1.4, 0.0)0.01	0.0(100, 0.0)0.00	99.4(0.6, 0.0)0.00
	0.05	96.4(3.6, 0.0)0.02	97.7(2.3, 0.0)0.02	0.0(100, 0.0)0.00	98.9(1.1, 0.0)0.01
40/20/100	0	98.7(1.3, 0.0)0.00	99.0(1.0, 0.0)0.00	0.0(100, 0.0)0.00	99.4(0.6, 0.0)0.00
	0.01	97.3(2.7, 0.0)0.01	98.2(1.8, 0.0)0.01	0.0(100, 0.0)0.00	98.9(1.1, 0.0)0.00
	0.05	96.8(3.2, 0.0)0.02	97.9(2.1, 0.0)0.02	0.0(100, 0.0)0.00	99.2(0.8, 0.0)0.01
40/20/200	0	98.7(1.3, 0.0)0.00	99.2(0.8, 0.0)0.00	0.0(100, 0.0)0.00	99.7(0.3, 0.0)0.00
	0.01	97.1(2.9, 0.0)0.00	97.8(2.2, 0.0)0.00	0.0(100, 0.0)0.00	99.0(1.0, 0.0)0.00
	0.05	97.5(2.5, 0.0)0.02	98.4(1.6, 0.0)0.02	0.0(100, 0.0)0.00	99.3(0.7, 0.0)0.01
40/40/100	0	98.6(1.4, 0.0)0.00	99.0(1.0, 0.0)0.00	0.0(100, 0.0)0.00	99.5(0.5, 0.0)0.00
	0.01	97.7(2.3, 0.0)0.00	98.4(1.6, 0.0)0.00	0.0(100, 0.0)0.00	99.3(0.7, 0.0)0.00
	0.05	97.2(2.8, 0.0)0.01	98.1(1.9, 0.0)0.01	0.0(100, 0.0)0.00	98.6(1.4, 0.0)0.01

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.24: Confidence intervals for τ_c^2 with 'no' heterogeneity, control group disease rates $r_B = \{0.35, 0.40, 0.45, 0.50\}$ for profile likelihood, and Wald-Type based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

$k/N_j/m$	ρ	pML C(L,R)W	pRE C(L,R)W	wML C(L,R)W	wRE C(L,R)W
4/20/100	0	99.5(0.5, 0.0)0.02	99.5(0.5, 0.0)0.04	100(0.0, 0.0)0.01	100(0.0, 0.0)0.02
	0.01	98.9(1.1, 0.0)0.04	98.2(1.8, 0.0)0.08	100(0.0, 0.0)0.03	100(0.0, 0.0)0.04
	0.05	99.0(1.0, 0.0)0.11	98.7(1.3, 0.0)0.22	100(0.0, 0.0)0.08	100(0.0, 0.0)0.11
4/20/200	0	99.6(0.4, 0.0)0.01	99.1(0.9, 0.0)0.02	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	99.1(0.9, 0.0)0.03	98.5(1.5, 0.0)0.06	100(0.0, 0.0)0.02	100(0.0, 0.0)0.03
	0.05	98.9(1.1, 0.0)0.11	98.1(1.9, 0.0)0.21	100(0.0, 0.0)0.08	100(0.0, 0.0)0.10
4/40/100	0	99.4(0.6, 0.0)0.01	98.9(1.1, 0.0)0.02	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	99.2(0.8, 0.0)0.02	98.5(1.5, 0.0)0.04	100(0.0, 0.0)0.01	100(0.0, 0.0)0.02
	0.05	99.2(0.8, 0.0)0.06	98.2(1.8, 0.0)0.11	100(0.0, 0.0)0.04	100(0.0, 0.0)0.05
12/20/100	0	99.4(0.6, 0.0)0.01	99.2(0.8, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	98.4(1.6, 0.0)0.01	97.7(2.3, 0.0)0.02	100(0.0, 0.0)0.02	100(0.0, 0.0)0.02
	0.05	98.5(1.5, 0.0)0.04	97.7(2.3, 0.0)0.04	100(0.0, 0.0)0.05	100(0.0, 0.0)0.05
12/20/200	0	99.8(0.2, 0.0)0.00	99.5(0.5, 0.0)0.00	100(0.0, 0.0)0.00	100(0.0, 0.0)0.00
	0.01	98.6(1.4, 0.0)0.01	97.7(2.3, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.05	98.6(1.4, 0.0)0.03	97.7(2.3, 0.0)0.04	100(0.0, 0.0)0.04	100(0.0, 0.0)0.05
12/40/100	0	98.9(1.1, 0.0)0.00	98.5(1.5, 0.0)0.00	100(0.0, 0.0)0.00	100(0.0, 0.0)0.00
	0.01	98.8(1.2, 0.0)0.01	98.2(1.8, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.05	98.4(1.6, 0.0)0.02	97.6(2.4, 0.0)0.02	100(0.0, 0.0)0.02	100(0.0, 0.0)0.03
20/20/100	0	99.2(0.8, 0.0)0.00	98.4(1.6, 0.0)0.00	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	97.4(2.6, 0.0)0.01	96.7(3.3, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.05	98.4(1.6, 0.0)0.03	97.7(2.3, 0.0)0.03	100(0.0, 0.0)0.04	100(0.0, 0.0)0.04
20/20/200	0	99.3(0.7, 0.0)0.00	99.1(0.9, 0.0)0.00	100(0.0, 0.0)0.00	100(0.0, 0.0)0.00
	0.01	98.9(1.1, 0.0)0.01	98.2(1.8, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.05	99.1(0.9, 0.0)0.02	98.1(1.9, 0.0)0.03	100(0.0, 0.0)0.03	100(0.0, 0.0)0.03
20/40/100	0	99.3(0.7, 0.0)0.00	98.2(1.8, 0.0)0.00	100(0.0, 0.0)0.00	100(0.0, 0.0)0.00
	0.01	98.7(1.3, 0.0)0.00	98.2(1.8, 0.0)0.00	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.05	97.9(2.1, 0.0)0.01	96.7(3.3, 0.0)0.01	100(0.0, 0.0)0.02	100(0.0, 0.0)0.02
40/20/100	0	99.1(0.9, 0.0)0.00	98.9(1.1, 0.0)0.00	99.9(0.1, 0.0)0.00	99.9(0.1, 0.0)0.00
	0.01	98.3(1.7, 0.0)0.01	97.9(2.1, 0.0)0.01	99.7(0.3, 0.0)0.01	99.6(0.4, 0.0)0.01
	0.05	98.1(1.9, 0.0)0.02	97.3(2.7, 0.0)0.02	99.8(0.2, 0.0)0.02	99.8(0.2, 0.0)0.03
40/20/200	0	99.4(0.6, 0.0)0.00	99.0(1.0, 0.0)0.00	100(0.0, 0.0)0.00	100(0.0, 0.0)0.00
	0.01	97.8(2.2, 0.0)0.00	97.5(2.5, 0.0)0.00	99.9(0.1, 0.0)0.01	99.9(0.1, 0.0)0.01
	0.05	98.3(1.7, 0.0)0.01	97.9(2.1, 0.0)0.02	99.9(0.1, 0.0)0.02	99.8(0.2, 0.0)0.02
40/40/100	0	99.4(0.6, 0.0)0.00	99.0(1.0, 0.0)0.00	100(0.0, 0.0)0.00	100(0.0, 0.0)0.00
	0.01	98.7(1.3, 0.0)0.00	98.2(1.8, 0.0)0.00	99.9(0.1, 0.0)0.00	99.9(0.1, 0.0)0.00
	0.05	98.1(1.9, 0.0)0.01	97.4(2.6, 0.0)0.01	99.8(0.2, 0.0)0.01	99.8(0.2, 0.0)0.01

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.25: Confidence intervals for τ_c^2 with 'low' heterogeneity, control group disease rates $r_B = \{0.35, 0.40, 0.45, 0.50\}$ for Q profile(PQ), Biggerstaff-Tweedie(BT), Sidik-Jonkman(SJ) and nonparametric bootstraps(NB) based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

		QP	BT	SJ	NB
$k/N_j/m$	ρ	C(L,R)W	C(L,R)W	C(L,R)W	C(L,R)W
4/20/100	0	96.1(2.4, 1.5)0.08	98.7(1.3, 0.0)0.07	31.6(19.0,49.4)0.03	56.0(0.0,44.0)0.00
	0.01	95.2(1.3, 3.5)0.14	99.6(0.4, 0.0)0.13	30.0(17.6,52.4)0.05	51.6(0.0,48.4)0.01
	0.05	95.1(1.0, 3.9)0.35	99.5(0.5, 0.0)0.36	26.4(12.2,61.4)0.12	44.5(0.0,55.5)0.02
4/20/200	0	89.9(9.0, 1.1)0.05	98.4(1.6, 0.0)0.04	51.9(28.8,19.3)0.02	64.4(0.0,35.6)0.00
	0.01	95.4(1.7, 2.9)0.11	98.8(1.2, 0.0)0.10	30.0(19.2,50.8)0.04	53.8(0.0,46.2)0.01
	0.05	96.6(1.2, 2.2)0.33	99.4(0.6, 0.0)0.33	29.0(12.6,58.4)0.11	46.0(0.0,54.0)0.02
4/40/100	0	88.5(10.1, 1.4)0.05	97.4(2.6, 0.0)0.04	53.7(28.5,17.8)0.02	66.0(0.0,34.0)0.00
	0.01	95.3(2.7, 2.0)0.08	98.8(1.2, 0.0)0.07	31.5(20.6,47.9)0.03	56.0(0.0,44.0)0.00
	0.05	95.6(1.0, 3.4)0.19	99.6(0.4, 0.0)0.18	27.8(14.7,57.5)0.07	48.0(0.0,52.0)0.01
12/20/100	0	95.1(1.6, 3.3)0.01	99.8(0.2, 0.0)0.01	33.2(15.8,51.0)0.00	72.3(0.0,27.7)0.00
	0.01	94.6(1.8, 3.6)0.02	99.7(0.3, 0.0)0.02	31.0(15.1,53.9)0.01	70.3(0.0,29.7)0.01
	0.05	93.4(0.7, 5.9)0.05	99.7(0.3, 0.0)0.06	31.5(10.7,57.8)0.01	63.5(0.1,36.4)0.02
12/20/200	0	84.8(14.5, 0.7)0.01	97.7(2.3, 0.0)0.01	32.5(31.5,36.0)0.00	80.5(1.0,18.5)0.00
	0.01	96.5(0.9, 2.6)0.02	99.6(0.4, 0.0)0.02	33.8(15.3,50.9)0.00	72.1(0.1,27.8)0.01
	0.05	92.9(0.5, 6.6)0.05	99.6(0.4, 0.0)0.05	29.3(10.3,60.4)0.01	63.1(0.0,36.9)0.02
12/40/100	0	83.4(16.0, 0.6)0.01	98.0(2.0, 0.0)0.01	33.6(35.7,30.7)0.00	85.3(0.4,14.3)0.00
	0.01	95.3(3.2, 1.5)0.01	99.1(0.9, 0.0)0.01	34.1(20.4,45.5)0.00	76.6(0.2,23.2)0.00
	0.05	95.7(0.4, 3.9)0.03	99.7(0.3, 0.0)0.03	32.5(10.9,56.6)0.01	66.9(0.0,33.1)0.01
20/20/100	0	95.3(1.4, 3.3)0.01	99.5(0.5, 0.0)0.01	33.2(13.3,53.5)0.00	75.8(0.3,23.9)0.00
	0.01	95.7(0.9, 3.4)0.01	99.9(0.1, 0.0)0.01	32.9(12.7,54.4)0.00	75.5(0.0,24.5)0.01
	0.05	92.7(0.1, 7.2)0.03	100(0.0, 0.0)0.03	27.6(8.7,63.7)0.01	67.4(0.0,32.6)0.02
20/20/200	0	81.1(18.5, 0.4)0.00	97.7(2.3, 0.0)0.00	34.7(33.7,31.6)0.00	87.8(0.8,11.4)0.00
	0.01	93.8(1.0, 5.2)0.01	99.4(0.6, 0.0)0.01	34.3(14.5,51.2)0.00	78.4(0.1,21.5)0.01
	0.05	92.6(0.4, 7.0)0.03	99.9(0.1, 0.0)0.03	26.8(7.9,65.3)0.01	67.7(0.0,32.3)0.01
20/40/100	0	78.3(21.7, 0.0)0.00	97.8(2.2, 0.0)0.00	35.2(38.1,26.7)0.00	89.7(1.0, 9.3)0.00
	0.01	94.4(3.6, 2.0)0.01	99.0(1.0, 0.0)0.01	36.2(20.4,43.4)0.00	83.4(0.3,16.3)0.00
	0.05	91.9(0.6, 7.5)0.02	99.5(0.5, 0.0)0.02	28.3(9.2,62.5)0.00	68.2(0.0,31.8)0.01
40/20/100	0	94.8(1.0, 4.2)0.00	99.6(0.4, 0.0)0.00	28.8(12.7,58.5)0.00	78.3(0.2,21.5)0.00
	0.01	93.8(0.3, 5.9)0.01	100(0.0, 0.0)0.01	25.8(8.8,65.4)0.00	74.1(0.0,25.9)0.00
	0.05	86.5(0.0,13.5)0.02	100(0.0, 0.0)0.02	20.4(4.5,75.1)0.00	65.2(0.0,34.8)0.01
40/20/200	0	70.2(29.8, 0.0)0.00	97.8(2.2, 0.0)0.00	36.7(35.1,28.2)0.00	92.4(1.5, 6.1)0.00
	0.01	93.3(0.6, 6.1)0.01	99.8(0.2, 0.0)0.01	29.5(11.3,59.2)0.00	77.9(0.1,22.0)0.00
	0.05	87.1(0.1,12.8)0.02	100(0.0, 0.0)0.02	21.7(5.1,73.2)0.00	65.2(0.0,34.8)0.01
40/40/100	0	67.6(32.4, 0.0)0.00	97.3(2.7, 0.0)0.00	34.8(37.4,27.8)0.00	92.4(1.1, 6.5)0.00
	0.01	95.8(2.6, 1.6)0.00	99.4(0.6, 0.0)0.00	36.1(19.2,44.7)0.00	85.5(0.3,14.2)0.00
	0.05	87.7(0.4,11.9)0.01	99.7(0.3, 0.0)0.01	25.3(6.6,68.1)0.00	70.5(0.1,29.4)0.01

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.26: Confidence intervals for τ_c^2 with 'low' heterogeneity, control group disease rates $r_B = \{0.35, 0.40, 0.45, 0.50\}$ for profile likelihood, and Wald-Type based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

$k/N_j/m$	ρ	pML C(L,R)W	pRE C(L,R)W	wML C(L,R)W	wRE C(L,R)W
4/20/100	0	99.7(0.3, 0.0)0.02	98.8(1.2, 0.0)0.05	100(0.0, 0.0)0.02	100(0.0, 0.0)0.02
	0.01	99.9(0.1, 0.0)0.04	99.6(0.4, 0.0)0.09	100(0.0, 0.0)0.03	100(0.0, 0.0)0.04
	0.05	100(0.0, 0.0)0.11	99.6(0.4, 0.0)0.23	100(0.0, 0.0)0.08	100(0.0, 0.0)0.11
4/20/200	0	99.5(0.5, 0.0)0.01	98.7(1.3, 0.0)0.03	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	99.7(0.3, 0.0)0.03	99.0(1.0, 0.0)0.07	100(0.0, 0.0)0.02	100(0.0, 0.0)0.03
	0.05	99.9(0.1, 0.0)0.11	99.5(0.5, 0.0)0.21	100(0.0, 0.0)0.08	100(0.0, 0.0)0.10
4/40/100	0	99.2(0.8, 0.0)0.01	97.9(2.1, 0.0)0.03	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	99.6(0.4, 0.0)0.02	98.8(1.2, 0.0)0.05	100(0.0, 0.0)0.02	100(0.0, 0.0)0.02
	0.05	100(0.0, 0.0)0.06	99.6(0.4, 0.0)0.12	100(0.0, 0.0)0.04	100(0.0, 0.0)0.06
12/20/100	0	99.8(0.2, 0.0)0.01	99.8(0.2, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	99.8(0.2, 0.0)0.02	99.3(0.7, 0.0)0.02	100(0.0, 0.0)0.02	100(0.0, 0.0)0.02
	0.05	99.9(0.1, 0.0)0.04	99.6(0.4, 0.0)0.05	100(0.0, 0.0)0.05	100(0.0, 0.0)0.05
12/20/200	0	98.2(1.8, 0.0)0.01	97.4(2.6, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	99.6(0.4, 0.0)0.01	99.4(0.6, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.05	99.6(0.4, 0.0)0.04	99.6(0.4, 0.0)0.04	100(0.0, 0.0)0.04	100(0.0, 0.0)0.05
12/40/100	0	98.6(1.4, 0.0)0.01	97.5(2.5, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	99.3(0.7, 0.0)0.01	99.1(0.9, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.05	99.8(0.2, 0.0)0.02	99.7(0.3, 0.0)0.03	100(0.0, 0.0)0.02	100(0.0, 0.0)0.03
20/20/100	0	99.6(0.4, 0.0)0.01	99.5(0.5, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	99.9(0.1, 0.0)0.01	99.8(0.2, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.05	99.8(0.0, 0.2)0.03	99.9(0.1, 0.0)0.03	100(0.0, 0.0)0.04	100(0.0, 0.0)0.04
20/20/200	0	97.9(2.1, 0.0)0.00	97.5(2.5, 0.0)0.00	100(0.0, 0.0)0.00	100(0.0, 0.0)0.00
	0.01	99.3(0.5, 0.2)0.01	99.3(0.7, 0.0)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.05	99.7(0.1, 0.2)0.03	99.8(0.1, 0.1)0.03	100(0.0, 0.0)0.03	100(0.0, 0.0)0.03
20/40/100	0	98.4(1.6, 0.0)0.00	97.1(2.9, 0.0)0.00	100(0.0, 0.0)0.00	100(0.0, 0.0)0.00
	0.01	98.6(1.0, 0.4)0.01	98.5(1.3, 0.2)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.05	99.4(0.5, 0.1)0.01	99.5(0.5, 0.0)0.02	100(0.0, 0.0)0.02	100(0.0, 0.0)0.02
40/20/100	0	95.4(0.4, 4.2)0.00	96.6(0.4, 3.0)0.00	100(0.0, 0.0)0.00	100(0.0, 0.0)0.00
	0.01	92.7(0.0, 7.3)0.01	94.4(0.0, 5.6)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.05	88.1(0.0,11.9)0.02	89.6(0.0,10.4)0.02	100(0.0, 0.0)0.02	100(0.0, 0.0)0.03
40/20/200	0	97.7(2.1, 0.2)0.00	97.0(2.8, 0.2)0.00	99.6(0.4, 0.0)0.00	99.6(0.4, 0.0)0.00
	0.01	94.5(0.2, 5.3)0.01	96.2(0.5, 3.3)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.05	87.5(0.0,12.5)0.02	90.4(0.0, 9.6)0.02	100(0.0, 0.0)0.02	100(0.0, 0.0)0.02
40/40/100	0	96.9(2.5, 0.6)0.00	96.7(2.9, 0.4)0.00	99.6(0.4, 0.0)0.00	99.6(0.4, 0.0)0.00
	0.01	96.9(0.6, 2.5)0.00	97.6(0.7, 1.7)0.00	100(0.0, 0.0)0.00	100(0.0, 0.0)0.00
	0.05	89.1(0.3,10.6)0.01	92.4(0.4, 7.2)0.01	100(0.0, 0.0)0.01	99.9(0.1, 0.0)0.01

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.27: Confidence intervals for τ_c^2 with 'moderate' heterogeneity, control group disease rates $r_B = \{0.35, 0.40, 0.45, 0.50\}$ for Q profile(PQ), Biggerstaff-Tweedie(BT), Sidik-Jonkman(SJ) and nonparametric bootstraps(NB) based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

		QP	BT	SJ	NB
$k/N_j/m$	ρ	C(L,R)W	C(L,R)W	C(L,R)W	C(L,R)W
4/20/100	0	94.6(3.0, 2.4)0.10	99.0(1.0, 0.0)0.09	46.9(13.4,39.7)0.05	56.2(0.0,43.8)0.01
	0.01	95.8(0.9, 3.3)0.15	99.7(0.3, 0.0)0.14	42.5(7.8,49.7)0.07	45.5(0.0,54.5)0.01
	0.05	94.8(0.3, 4.9)0.38	99.9(0.1, 0.0)0.37	38.1(3.8,58.1)0.14	38.6(0.0,61.4)0.02
4/20/200	0	90.7(8.6, 0.7)0.07	96.6(3.4, 0.0)0.06	48.6(27.8,23.6)0.04	73.7(0.0,26.3)0.01
	0.01	95.9(0.9, 3.2)0.12	99.7(0.3, 0.0)0.11	43.9(9.8,46.3)0.06	50.3(0.0,49.7)0.01
	0.05	95.9(0.5, 3.6)0.35	99.6(0.4, 0.0)0.35	38.4(4.8,56.8)0.13	38.0(0.0,62.0)0.02
4/40/100	0	90.7(8.5, 0.8)0.07	96.9(3.1, 0.0)0.06	51.4(24.8,23.8)0.04	73.4(0.0,26.6)0.00
	0.01	95.2(3.5, 1.3)0.10	98.7(1.3, 0.0)0.08	49.2(13.3,37.5)0.05	57.1(0.0,42.9)0.01
	0.05	94.9(0.6, 4.5)0.21	99.9(0.1, 0.0)0.20	40.7(5.4,53.9)0.09	41.7(0.0,58.3)0.01
12/20/100	0	95.4(1.8, 2.8)0.01	100(0.0, 0.0)0.01	49.6(8.9,41.5)0.01	69.8(0.1,30.1)0.01
	0.01	94.5(0.4, 5.1)0.02	100(0.0, 0.0)0.02	42.9(4.1,53.0)0.01	59.1(0.0,40.9)0.01
	0.05	87.1(0.2,12.7)0.06	99.9(0.1, 0.0)0.06	33.9(1.6,64.5)0.01	47.6(0.0,52.4)0.02
12/20/200	0	88.2(11.4, 0.4)0.01	96.7(3.3, 0.0)0.01	52.9(31.9,15.2)0.01	89.9(0.7, 9.4)0.00
	0.01	95.0(0.9, 4.1)0.02	99.7(0.3, 0.0)0.02	49.8(5.5,44.7)0.01	68.0(0.0,32.0)0.01
	0.05	90.1(0.0, 9.9)0.05	100(0.0, 0.0)0.05	32.0(1.5,66.5)0.01	49.1(0.0,50.9)0.02
12/40/100	0	89.4(10.2, 0.4)0.01	97.3(2.7, 0.0)0.01	56.4(29.1,14.5)0.01	90.1(0.3, 9.6)0.00
	0.01	96.0(3.0, 1.0)0.02	99.5(0.5, 0.0)0.01	55.3(11.1,33.6)0.01	75.4(0.1,24.5)0.01
	0.05	91.8(0.1, 8.1)0.03	99.9(0.1, 0.0)0.03	37.2(3.6,59.2)0.01	54.3(0.0,45.7)0.01
20/20/100	0	96.4(0.9, 2.7)0.01	100(0.0, 0.0)0.01	46.7(6.8,46.5)0.00	75.1(0.0,24.9)0.00
	0.01	91.6(0.2, 8.2)0.01	100(0.0, 0.0)0.01	38.0(2.0,60.0)0.00	61.1(0.0,38.9)0.01
	0.05	82.2(0.0,17.8)0.04	100(0.0, 0.0)0.04	25.7(0.3,74.0)0.01	47.1(0.0,52.9)0.02
20/20/200	0	84.7(15.0, 0.3)0.01	96.6(3.4, 0.0)0.01	52.8(34.3,12.9)0.00	93.3(1.8, 4.9)0.00
	0.01	94.0(0.5, 5.5)0.01	100(0.0, 0.0)0.01	43.4(5.7,50.9)0.00	67.6(0.0,32.4)0.01
	0.05	82.7(0.0,17.3)0.03	100(0.0, 0.0)0.03	25.1(0.8,74.1)0.01	45.4(0.0,54.6)0.02
20/40/100	0	86.7(13.0, 0.3)0.01	96.7(3.3, 0.0)0.01	56.7(34.1, 9.2)0.00	94.1(1.4, 4.5)0.00
	0.01	94.9(4.1, 1.0)0.01	99.3(0.7, 0.0)0.01	53.6(11.6,34.8)0.00	78.9(0.4,20.7)0.01
	0.05	86.1(0.1,13.8)0.02	100(0.0, 0.0)0.02	32.5(1.7,65.8)0.01	53.7(0.0,46.3)0.01
40/20/100	0	95.5(1.3, 3.2)0.01	89.2(0.0,10.8)0.01	43.9(4.9,51.2)0.00	75.8(0.0,24.2)0.00
	0.01	85.3(0.0,14.7)0.01	75.9(0.0,24.1)0.01	27.1(0.9,72.0)0.00	57.8(0.0,42.2)0.01
	0.05	59.6(0.0,40.4)0.02	52.1(0.0,47.9)0.02	11.5(0.0,88.5)0.00	31.4(0.0,68.6)0.01
40/20/200	0	78.2(21.8, 0.0)0.00	95.1(4.9, 0.0)0.00	51.2(40.9, 7.9)0.00	95.2(2.6, 2.2)0.00
	0.01	90.9(0.0, 9.1)0.01	85.1(0.0,14.9)0.01	30.6(2.8,66.6)0.00	64.3(0.0,35.7)0.01
	0.05	64.5(0.1,35.4)0.02	54.2(0.0,45.8)0.02	11.4(0.2,88.4)0.00	34.3(0.0,65.7)0.01
40/40/100	0	72.6(27.4, 0.0)0.00	93.8(5.9, 0.3)0.00	49.9(44.0, 6.1)0.00	94.4(3.9, 1.7)0.00
	0.01	95.1(3.2, 1.7)0.01	92.9(0.1, 7.0)0.01	52.1(7.5,40.4)0.00	82.0(0.0,18.0)0.00
	0.05	76.6(0.0,23.4)0.01	65.5(0.0,34.5)0.01	18.5(0.1,81.4)0.00	44.2(0.0,55.8)0.01

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.28: Confidence intervals for τ_c^2 with 'moderate' heterogeneity, control group disease rates $r_B = \{0.35, 0.40, 0.45, 0.50\}$ for profile likelihood, and Wald-Type based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

$k/N_j/m$	ρ	pML C(L,R)W	pRE C(L,R)W	wML C(L,R)W	wRE C(L,R)W
4/20/100	0	99.6(0.4, 0.0)0.03	99.0(1.0, 0.0)0.06	100(0.0, 0.0)0.02	100(0.0, 0.0)0.03
	0.01	100(0.0, 0.0)0.05	99.7(0.3, 0.0)0.10	100(0.0, 0.0)0.03	100(0.0, 0.0)0.04
	0.05	100(0.0, 0.0)0.12	99.9(0.1, 0.0)0.24	100(0.0, 0.0)0.09	100(0.0, 0.0)0.12
4/20/200	0	98.7(1.3, 0.0)0.02	96.7(3.3, 0.0)0.04	100(0.0, 0.0)0.01	100(0.0, 0.0)0.02
	0.01	100(0.0, 0.0)0.04	99.7(0.3, 0.0)0.08	100(0.0, 0.0)0.02	100(0.0, 0.0)0.03
	0.05	99.8(0.2, 0.0)0.11	99.7(0.3, 0.0)0.23	100(0.0, 0.0)0.08	100(0.0, 0.0)0.11
4/40/100	0	98.9(1.1, 0.0)0.02	97.0(3.0, 0.0)0.04	100(0.0, 0.0)0.01	100(0.0, 0.0)0.02
	0.01	99.4(0.6, 0.0)0.03	98.7(1.3, 0.0)0.06	100(0.0, 0.0)0.02	100(0.0, 0.0)0.03
	0.05	100(0.0, 0.0)0.07	100(0.0, 0.0)0.13	100(0.0, 0.0)0.05	100(0.0, 0.0)0.06
12/20/100	0	98.1(0.0, 1.9)0.01	99.5(0.1, 0.4)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	96.0(0.0, 4.0)0.02	98.8(0.0, 1.2)0.02	99.9(0.0, 0.1)0.02	100(0.0, 0.0)0.02
	0.05	90.8(0.0, 9.2)0.04	96.1(0.2, 3.7)0.05	100(0.0, 0.0)0.05	100(0.0, 0.0)0.05
12/20/200	0	97.0(2.9, 0.1)0.01	95.4(4.5, 0.1)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	96.4(0.2, 3.4)0.01	97.8(0.6, 1.6)0.02	100(0.0, 0.0)0.01	100(0.0, 0.0)0.02
	0.05	92.0(0.0, 8.0)0.04	96.5(0.0, 3.5)0.05	100(0.0, 0.0)0.04	100(0.0, 0.0)0.05
12/40/100	0	97.8(1.9, 0.3)0.01	96.6(3.3, 0.1)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.01	98.8(0.3, 0.9)0.01	99.2(0.7, 0.1)0.01	100(0.0, 0.0)0.01	100(0.0, 0.0)0.01
	0.05	93.9(0.1, 6.0)0.02	97.2(0.1, 2.7)0.03	100(0.0, 0.0)0.03	100(0.0, 0.0)0.03
20/20/100	0	93.1(0.0, 6.9)0.01	95.9(0.0, 4.1)0.01	89.3(0.0,10.7)0.01	96.9(0.0, 3.1)0.01
	0.01	85.1(0.0,14.9)0.01	90.4(0.1, 9.5)0.01	64.3(0.0,35.7)0.01	74.7(0.0,25.3)0.01
	0.05	76.0(0.0,24.0)0.03	82.7(0.0,17.3)0.03	49.1(0.0,50.9)0.04	59.2(0.0,40.8)0.04
20/20/200	0	96.3(3.2, 0.5)0.01	94.6(5.0, 0.4)0.01	99.0(0.0, 1.0)0.01	99.8(0.0, 0.2)0.01
	0.01	90.3(0.0, 9.7)0.01	93.9(0.2, 5.9)0.01	70.2(0.0,29.8)0.01	78.3(0.0,21.7)0.01
	0.05	74.5(0.0,25.5)0.03	82.0(0.0,18.0)0.03	44.3(0.0,55.7)0.03	53.2(0.0,46.8)0.04
20/40/100	0	96.3(3.1, 0.6)0.01	95.3(4.2, 0.5)0.01	97.3(0.1, 2.6)0.01	99.2(0.1, 0.7)0.01
	0.01	95.0(0.6, 4.4)0.01	96.6(0.9, 2.5)0.01	81.6(0.0,18.4)0.01	85.6(0.0,14.4)0.01
	0.05	79.3(0.0,20.7)0.02	85.3(0.0,14.7)0.02	55.4(0.0,44.6)0.02	63.5(0.0,36.5)0.02
40/20/100	0	87.5(0.0,12.5)0.01	89.8(0.0,10.2)0.01	76.3(0.0,23.7)0.01	79.0(0.0,21.0)0.01
	0.01	74.6(0.0,25.4)0.01	77.4(0.0,22.6)0.01	57.2(0.0,42.8)0.01	61.6(0.0,38.4)0.01
	0.05	48.6(0.0,51.4)0.02	56.0(0.0,44.0)0.02	29.1(0.0,70.9)0.03	34.1(0.0,65.9)0.03
40/20/200	0	94.4(5.1, 0.5)0.00	93.4(6.4, 0.2)0.00	98.1(0.1, 1.8)0.00	98.3(0.5, 1.2)0.00
	0.01	82.9(0.0,17.1)0.01	86.4(0.0,13.6)0.01	64.3(0.0,35.7)0.01	70.2(0.0,29.8)0.01
	0.05	51.4(0.0,48.6)0.02	57.0(0.0,43.0)0.02	31.2(0.0,68.8)0.02	37.3(0.0,62.7)0.02
40/40/100	0	93.5(6.2, 0.3)0.00	92.4(7.3, 0.3)0.00	98.5(0.4, 1.1)0.00	99.1(0.4, 0.5)0.00
	0.01	91.6(0.1, 8.3)0.01	94.2(0.2, 5.6)0.01	83.2(0.0,16.8)0.01	86.5(0.0,13.5)0.01
	0.05	62.2(0.0,37.8)0.01	67.7(0.0,32.3)0.01	41.9(0.0,58.1)0.01	48.1(0.0,51.9)0.01

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.29: Confidence intervals for τ_c^2 with 'high' heterogeneity, control group disease rates $r_B = \{0.35, 0.40, 0.45, 0.50\}$ for Q profile(PQ), Biggerstaff-Tweedie(BT), Sidik-Jonkman(SJ) and nonparametric bootstraps(NB) based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

		QP	BT	SJ	NB
$k/N_j/m$	ρ	C(L,R)W	C(L,R)W	C(L,R)W	C(L,R)W
4/20/100	0	96.6(1.6, 1.8)0.12	99.4(0.6, 0.0)0.10	56.3(8.5,35.2)0.06	53.5(0.0,46.5)0.01
	0.01	96.0(0.5, 3.5)0.18	99.8(0.2, 0.0)0.16	52.8(3.2,44.0)0.09	44.5(0.0,55.5)0.01
	0.05	95.0(0.1, 4.9)0.40	99.9(0.1, 0.0)0.38	41.4(2.2,56.4)0.16	30.7(0.0,69.3)0.02
4/20/200	0	94.3(4.8, 0.9)0.09	97.8(2.2, 0.0)0.07	63.1(23.6,13.3)0.06	78.3(0.0,21.7)0.01
	0.01	97.2(0.7, 2.1)0.14	99.7(0.3, 0.0)0.12	56.8(5.1,38.1)0.07	47.7(0.0,52.3)0.01
	0.05	94.2(0.0, 5.8)0.34	100(0.0, 0.0)0.34	39.8(0.5,59.7)0.12	28.8(0.0,71.2)0.02
4/40/100	0	94.4(5.2, 0.4)0.09	97.9(2.1, 0.0)0.07	64.4(21.3,14.3)0.06	76.6(0.0,23.4)0.01
	0.01	96.3(2.0, 1.7)0.12	99.2(0.8, 0.0)0.10	61.6(8.4,30.0)0.07	58.9(0.0,41.1)0.01
	0.05	95.7(0.4, 3.9)0.22	99.8(0.2, 0.0)0.21	46.4(2.6,51.0)0.10	36.3(0.0,63.7)0.01
12/20/100	0	97.7(0.4, 1.9)0.02	99.9(0.1, 0.0)0.02	64.5(5.5,30.0)0.01	72.6(0.0,27.4)0.01
	0.01	93.6(0.2, 6.2)0.03	100(0.0, 0.0)0.02	47.0(1.1,51.9)0.01	52.7(0.0,47.3)0.01
	0.05	85.5(0.0,14.5)0.06	100(0.0, 0.0)0.06	31.8(0.2,68.0)0.02	38.4(0.0,61.6)0.02
12/20/200	0	93.4(6.3, 0.3)0.01	96.7(3.3, 0.0)0.01	63.0(28.4, 8.6)0.01	92.2(0.4, 7.4)0.01
	0.01	94.7(0.4, 4.9)0.02	99.9(0.1, 0.0)0.02	58.3(2.9,38.8)0.01	65.0(0.0,35.0)0.01
	0.05	84.9(0.0,15.1)0.06	100(0.0, 0.0)0.06	32.8(0.1,67.1)0.02	37.7(0.0,62.3)0.02
12/40/100	0	92.2(7.7, 0.1)0.01	96.9(3.1, 0.0)0.01	64.3(29.2, 6.5)0.01	92.5(0.5, 7.0)0.01
	0.01	96.3(0.7, 3.0)0.02	99.7(0.3, 0.0)0.02	66.7(7.4,25.9)0.01	75.3(0.0,24.7)0.01
	0.05	90.3(0.0, 9.7)0.04	100(0.0, 0.0)0.03	40.1(0.6,59.3)0.01	44.9(0.0,55.1)0.01
20/20/100	0	97.6(0.1, 2.3)0.01	100(0.0, 0.0)0.01	59.3(4.1,36.6)0.00	72.7(0.0,27.3)0.01
	0.01	88.2(0.1,11.7)0.02	92.3(0.0, 7.7)0.02	39.5(0.5,60.0)0.01	52.1(0.0,47.9)0.01
	0.05	70.5(0.0,29.5)0.04	88.7(0.0,11.3)0.04	17.2(0.0,82.8)0.01	28.1(0.0,71.9)0.02
20/20/200	0	91.3(8.7, 0.0)0.01	96.0(4.0, 0.0)0.01	63.1(32.5, 4.4)0.00	94.9(1.7, 3.4)0.01
	0.01	93.0(0.1, 6.9)0.01	97.2(0.0, 2.8)0.01	52.4(0.8,46.8)0.01	63.8(0.0,36.2)0.01
	0.05	71.8(0.0,28.2)0.04	92.9(0.0, 7.1)0.04	20.9(0.1,79.0)0.01	30.5(0.0,69.5)0.02
20/40/100	0	88.8(11.1, 0.1)0.01	94.6(5.4, 0.0)0.01	59.3(35.9, 4.8)0.01	94.1(2.5, 3.4)0.01
	0.01	97.1(0.6, 2.3)0.01	98.6(0.1, 1.3)0.01	65.5(5.3,29.2)0.01	77.6(0.0,22.4)0.01
	0.05	79.9(0.0,20.1)0.02	91.4(0.0, 8.6)0.02	30.2(0.3,69.5)0.01	41.3(0.0,58.7)0.01
40/20/100	0	92.5(0.0, 7.5)0.01	87.6(0.0,12.4)0.01	54.8(1.1,44.1)0.00	72.4(0.0,27.6)0.01
	0.01	74.0(0.0,26.0)0.01	62.5(0.0,37.5)0.01	22.1(0.1,77.8)0.00	40.1(0.0,59.9)0.01
	0.05	39.5(0.0,60.5)0.02	30.5(0.0,69.5)0.02	4.8(0.0,95.2)0.00	13.5(0.0,86.5)0.02
40/20/200	0	84.9(15.0, 0.1)0.01	93.1(6.6, 0.3)0.01	53.8(44.2, 2.0)0.00	93.9(5.2, 0.9)0.00
	0.01	86.3(0.0,13.7)0.01	78.4(0.0,21.6)0.01	39.3(0.1,60.6)0.00	57.9(0.0,42.1)0.01
	0.05	39.3(0.0,60.7)0.02	29.7(0.0,70.3)0.02	5.2(0.0,94.8)0.00	14.5(0.0,85.5)0.01
40/40/100	0	83.8(16.2, 0.0)0.01	92.8(7.2, 0.0)0.01	49.8(48.6, 1.6)0.00	93.6(6.0, 0.4)0.00
	0.01	95.5(0.2, 4.3)0.01	93.4(0.0, 6.6)0.01	62.0(2.4,35.6)0.00	78.8(0.0,21.2)0.01
	0.05	61.0(0.0,39.0)0.01	46.8(0.0,53.2)0.01	10.8(0.0,89.2)0.00	25.7(0.0,74.3)0.01

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.30: Confidence intervals for τ_c^2 with 'high' heterogeneity, control group disease rates $r_B = \{0.35, 0.40, 0.45, 0.50\}$ for profile likelihood, and Wald-Type based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ (truncated), and odds ratio $\psi = 0.7$.

		pML	pRE	wML	wRE
$k/N_j/m$	ρ	C(L,R)W	C(L,R)W	C(L,R)W	C(L,R)W
4/20/100	0	99.9(0.1, 0.0)0.04	99.5(0.5, 0.0)0.07	100(0.0, 0.0)0.02	100(0.0, 0.0)0.03
	0.01	99.9(0.1, 0.0)0.06	99.8(0.2, 0.0)0.11	100(0.0, 0.0)0.03	100(0.0, 0.0)0.05
	0.05	100(0.0, 0.0)0.13	99.9(0.1, 0.0)0.26	100(0.0, 0.0)0.09	100(0.0, 0.0)0.12
4/20/200	0	99.0(1.0, 0.0)0.03	97.8(2.2, 0.0)0.05	100(0.0, 0.0)0.01	100(0.0, 0.0)0.02
	0.01	99.9(0.1, 0.0)0.05	99.7(0.3, 0.0)0.09	100(0.0, 0.0)0.03	100(0.0, 0.0)0.04
	0.05	100(0.0, 0.0)0.11	100(0.0, 0.0)0.22	100(0.0, 0.0)0.08	100(0.0, 0.0)0.10
4/40/100	0	99.0(1.0, 0.0)0.03	98.1(1.9, 0.0)0.05	100(0.0, 0.0)0.01	100(0.0, 0.0)0.02
	0.01	99.9(0.1, 0.0)0.04	99.2(0.8, 0.0)0.07	100(0.0, 0.0)0.02	100(0.0, 0.0)0.03
	0.05	100(0.0, 0.0)0.07	99.9(0.1, 0.0)0.14	100(0.0, 0.0)0.05	100(0.0, 0.0)0.06
12/20/100	0	95.8(0.1, 4.1)0.01	98.3(0.1, 1.6)0.02	77.6(0.0,22.4)0.01	84.9(0.0,15.1)0.01
	0.01	87.2(0.0,12.8)0.02	93.4(0.0, 6.6)0.02	56.8(0.0,43.2)0.02	66.7(0.0,33.3)0.02
	0.05	74.3(0.0,25.7)0.05	85.2(0.0,14.8)0.06	40.7(0.0,59.3)0.05	50.4(0.0,49.6)0.06
12/20/200	0	97.0(2.5, 0.5)0.01	95.8(3.8, 0.4)0.01	95.1(0.0, 4.9)0.01	97.1(0.0, 2.9)0.01
	0.01	91.5(0.1, 8.4)0.02	94.6(0.1, 5.3)0.02	67.7(0.0,32.3)0.02	76.3(0.0,23.7)0.02
	0.05	74.9(0.0,25.1)0.04	84.4(0.0,15.6)0.05	41.4(0.0,58.6)0.05	51.4(0.0,48.6)0.05
12/40/100	0	96.8(2.8, 0.4)0.01	95.7(4.2, 0.1)0.01	96.3(0.0, 3.7)0.01	97.1(0.0, 2.9)0.01
	0.01	94.4(0.2, 5.4)0.01	96.4(0.3, 3.3)0.02	80.9(0.0,19.1)0.01	86.2(0.0,13.8)0.01
	0.05	81.2(0.0,18.8)0.03	89.4(0.0,10.6)0.03	50.2(0.0,49.8)0.03	59.8(0.0,40.2)0.03
20/20/100	0	93.3(0.0, 6.7)0.01	95.6(0.0, 4.4)0.01	77.4(0.0,22.6)0.01	83.4(0.0,16.6)0.01
	0.01	75.7(0.0,24.3)0.01	83.0(0.0,17.0)0.02	53.7(0.0,46.3)0.01	60.8(0.0,39.2)0.02
	0.05	55.0(0.0,45.0)0.03	63.9(0.0,36.1)0.03	28.7(0.0,71.3)0.04	35.8(0.0,64.2)0.04
20/20/200	0	95.9(3.9, 0.2)0.01	94.9(5.0, 0.1)0.01	98.4(0.0, 1.6)0.01	99.0(0.0, 1.0)0.01
	0.01	85.9(0.0,14.1)0.01	90.1(0.0, 9.9)0.01	66.6(0.0,33.4)0.01	72.5(0.0,27.5)0.01
	0.05	55.9(0.0,44.1)0.03	64.7(0.0,35.3)0.03	32.3(0.0,67.7)0.04	39.7(0.0,60.3)0.04
20/40/100	0	94.3(5.1, 0.6)0.01	92.2(7.5, 0.3)0.01	97.7(0.0, 2.3)0.01	98.2(0.0, 1.8)0.01
	0.01	93.5(0.1, 6.4)0.01	95.8(0.3, 3.9)0.01	81.9(0.0,18.1)0.01	86.2(0.0,13.8)0.01
	0.05	66.1(0.0,33.9)0.02	72.5(0.0,27.5)0.02	41.4(0.0,58.6)0.02	50.4(0.0,49.6)0.02
40/20/100	0	87.1(0.0,12.9)0.01	88.6(0.0,11.4)0.01	75.0(0.0,25.0)0.01	78.1(0.0,21.9)0.01
	0.01	61.3(0.0,38.7)0.01	65.4(0.0,34.6)0.01	40.6(0.0,59.4)0.01	46.3(0.0,53.7)0.01
	0.05	28.0(0.0,72.0)0.02	32.5(0.0,67.5)0.02	13.2(0.0,86.8)0.03	16.2(0.0,83.8)0.03
40/20/200	0	92.2(7.5, 0.3)0.01	90.6(9.2, 0.2)0.01	98.4(0.9, 0.7)0.00	98.2(1.1, 0.7)0.00
	0.01	76.6(0.0,23.4)0.01	80.3(0.0,19.7)0.01	58.8(0.0,41.2)0.01	63.7(0.0,36.3)0.01
	0.05	28.1(0.0,71.9)0.02	31.9(0.0,68.1)0.02	13.5(0.0,86.5)0.02	16.8(0.0,83.2)0.02
40/40/100	0	91.9(8.1, 0.0)0.01	90.1(9.9, 0.0)0.01	99.5(0.4, 0.1)0.00	99.4(0.5, 0.1)0.00
	0.01	92.5(0.0, 7.5)0.01	93.8(0.1, 6.1)0.01	80.7(0.0,19.3)0.01	84.9(0.0,15.1)0.01
	0.05	43.9(0.0,56.1)0.01	50.1(0.0,49.9)0.01	24.1(0.0,75.9)0.01	29.1(0.0,70.9)0.01

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.31: Bias and MSE for the measures of heterogeneity with 'no' heterogeneity and control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ(truncated), and odds ratio $\psi = 0.7$.

		Bias			MSE		
$k/N_j/m$	ρ	H_a	R_a	I_a^2	H_a	R_a	I_a^2
4/20/100	0	0.117	0.099	0.128	0.070	0.050	0.065
	0.01	0.156	0.132	0.165	0.100	0.071	0.088
	0.05	0.148	0.127	0.160	0.089	0.065	0.084
4/20/200	0	0.131	0.110	0.145	0.076	0.053	0.074
	0.01	0.163	0.138	0.171	0.109	0.077	0.091
	0.05	0.150	0.129	0.160	0.094	0.069	0.085
4/40/100	0	0.143	0.121	0.152	0.092	0.065	0.080
	0.01	0.145	0.122	0.150	0.094	0.066	0.082
	0.05	0.158	0.134	0.167	0.102	0.073	0.088
12/20/100	0	0.059	0.059	0.084	0.016	0.016	0.028
	0.01	0.080	0.081	0.114	0.023	0.023	0.039
	0.05	0.067	0.068	0.098	0.017	0.018	0.031
12/20/200	0	0.059	0.059	0.085	0.016	0.016	0.028
	0.01	0.072	0.073	0.103	0.020	0.020	0.036
	0.05	0.075	0.077	0.108	0.020	0.021	0.037
12/40/100	0	0.064	0.064	0.091	0.018	0.018	0.031
	0.01	0.081	0.082	0.113	0.024	0.024	0.040
	0.05	0.083	0.084	0.117	0.024	0.024	0.041
20/20/100	0	0.041	0.043	0.065	0.008	0.009	0.017
	0.01	0.053	0.056	0.082	0.011	0.012	0.023
	0.05	0.051	0.055	0.081	0.010	0.011	0.021
20/20/200	0	0.044	0.046	0.068	0.009	0.010	0.019
	0.01	0.058	0.061	0.090	0.012	0.013	0.025
	0.05	0.059	0.063	0.090	0.012	0.014	0.026
20/40/100	0	0.045	0.047	0.070	0.009	0.009	0.019
	0.01	0.061	0.064	0.093	0.013	0.015	0.027
	0.05	0.060	0.063	0.092	0.013	0.014	0.026
40/20/100	0	0.022	0.024	0.037	0.003	0.003	0.007
	0.01	0.040	0.043	0.066	0.005	0.006	0.014
	0.05	0.035	0.038	0.058	0.005	0.006	0.012
40/20/200	0	0.023	0.025	0.039	0.003	0.003	0.007
	0.01	0.044	0.047	0.073	0.006	0.007	0.015
	0.05	0.039	0.043	0.066	0.005	0.006	0.013
40/40/100	0	0.030	0.032	0.049	0.004	0.005	0.010
	0.01	0.042	0.045	0.069	0.006	0.007	0.015
	0.05	0.039	0.043	0.065	0.006	0.006	0.014

Table 6.32: Bias and MSE for the measures of heterogeneity with 'low' heterogeneity and control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ(truncated), and odds ratio $\psi = 0.7$.

		Bias			MSE		
$k/N_j/m$	ρ	H_a	R_a	I_a^2	H_a	R_a	I_a^2
4/20/100	0	0.004	0.005	0.117	0.116	0.083	0.088
	0.01	0.012	0.008	0.100	0.119	0.085	0.087
	0.05	0.050	0.045	0.151	0.081	0.060	0.084
4/20/200	0	0.133	0.110	0.007	0.200	0.142	0.093
	0.01	0.040	0.032	0.079	0.140	0.100	0.088
	0.05	0.048	0.042	0.147	0.083	0.062	0.082
4/40/100	0	0.148	0.122	0.002	0.210	0.148	0.095
	0.01	0.062	0.051	0.055	0.141	0.100	0.086
	0.05	0.021	0.020	0.126	0.096	0.070	0.085
12/20/100	0	0.047	0.050	0.104	0.028	0.029	0.048
	0.01	0.061	0.063	0.119	0.025	0.026	0.047
	0.05	0.086	0.089	0.150	0.023	0.024	0.048
12/20/200	0	0.053	0.050	0.007	0.048	0.046	0.048
	0.01	0.035	0.037	0.087	0.028	0.029	0.045
	0.05	0.084	0.087	0.148	0.023	0.025	0.048
12/40/100	0	0.037	0.033	0.009	0.042	0.041	0.046
	0.01	0.001	0.004	0.051	0.036	0.036	0.046
	0.05	0.067	0.070	0.127	0.025	0.026	0.047
20/20/100	0	0.058	0.063	0.108	0.019	0.021	0.038
	0.01	0.065	0.070	0.119	0.020	0.022	0.041
	0.05	0.103	0.109	0.168	0.019	0.021	0.045
20/20/200	0	0.014	0.012	0.017	0.024	0.025	0.033
	0.01	0.059	0.064	0.109	0.019	0.021	0.038
	0.05	0.094	0.100	0.156	0.019	0.021	0.043
20/40/100	0	0.035	0.033	0.004	0.029	0.030	0.035
	0.01	0.019	0.022	0.060	0.023	0.025	0.036
	0.05	0.082	0.087	0.139	0.018	0.021	0.041
40/20/100	0	0.075	0.082	0.124	0.014	0.016	0.032
	0.01	0.083	0.090	0.134	0.014	0.017	0.033
	0.05	0.109	0.118	0.173	0.016	0.019	0.041
40/20/200	0	0.010	0.008	0.009	0.013	0.015	0.021
	0.01	0.069	0.075	0.114	0.013	0.015	0.030
	0.05	0.103	0.112	0.165	0.016	0.019	0.039
40/40/100	0	0.020	0.019	0.001	0.017	0.018	0.023
	0.01	0.039	0.043	0.072	0.012	0.014	0.025
	0.05	0.100	0.108	0.159	0.015	0.018	0.038

Table 6.33: Bias and MSE for the measures of heterogeneity with 'moderate' heterogeneity and control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ(truncated), and odds ratio $\psi = 0.7$.

		Bias			MSE		
$k/N_j/m$	ρ	H_a	R_a	I_a^2	H_a	R_a	I_a^2
4/20/100	0	0.046	0.040	0.168	0.178	0.128	0.120
	0.01	0.129	0.111	0.233	0.158	0.115	0.137
	0.05	0.220	0.191	0.306	0.133	0.100	0.157
4/20/200	0	0.153	0.126	0.046	0.296	0.209	0.101
	0.01	0.098	0.085	0.208	0.165	0.119	0.131
	0.05	0.198	0.171	0.286	0.134	0.100	0.150
4/40/100	0	0.187	0.155	0.021	0.306	0.216	0.095
	0.01	0.017	0.016	0.151	0.194	0.139	0.117
	0.05	0.164	0.141	0.258	0.142	0.104	0.143
12/20/100	0	0.101	0.103	0.151	0.053	0.053	0.068
	0.01	0.161	0.164	0.221	0.061	0.062	0.092
	0.05	0.220	0.223	0.286	0.067	0.069	0.111
12/20/200	0	0.068	0.064	0.004	0.067	0.063	0.042
	0.01	0.123	0.125	0.176	0.055	0.056	0.077
	0.05	0.224	0.227	0.293	0.070	0.072	0.115
12/40/100	0	0.085	0.080	0.014	0.073	0.069	0.044
	0.01	0.061	0.063	0.116	0.057	0.056	0.063
	0.05	0.184	0.186	0.243	0.061	0.062	0.097
20/20/100	0	0.110	0.115	0.151	0.040	0.043	0.058
	0.01	0.171	0.178	0.224	0.051	0.055	0.083
	0.05	0.229	0.239	0.297	0.063	0.070	0.109
20/20/200	0	0.074	0.073	0.032	0.040	0.040	0.025
	0.01	0.134	0.139	0.177	0.043	0.046	0.066
	0.05	0.226	0.235	0.291	0.062	0.068	0.106
20/40/100	0	0.095	0.093	0.043	0.050	0.050	0.029
	0.01	0.068	0.072	0.107	0.038	0.040	0.048
	0.05	0.201	0.209	0.260	0.056	0.061	0.094
40/20/100	0	0.133	0.140	0.164	0.032	0.036	0.049
	0.01	0.178	0.188	0.222	0.042	0.048	0.069
	0.05	0.245	0.261	0.319	0.065	0.075	0.114
40/20/200	0	0.050	0.050	0.026	0.020	0.021	0.014
	0.01	0.147	0.155	0.182	0.035	0.039	0.055
	0.05	0.238	0.253	0.308	0.063	0.071	0.108
40/40/100	0	0.074	0.075	0.046	0.024	0.025	0.014
	0.01	0.075	0.079	0.095	0.021	0.023	0.029
	0.05	0.206	0.218	0.260	0.051	0.058	0.085

Table 6.34: Bias and MSE for the measures of heterogeneity with 'high' heterogeneity and control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ(truncated), and odds ratio $\psi = 0.7$.

		Bias			MSE		
$k/N_j/m$	ρ	H_a	R_a	I_a^2	H_a	R_a	I_a^2
4/20/100	0	0.116	0.100	0.210	0.226	0.163	0.143
	0.01	0.242	0.207	0.297	0.218	0.158	0.177
	0.05	0.350	0.302	0.380	0.217	0.162	0.213
4/20/200	0	0.207	0.173	0.034	0.377	0.267	0.089
	0.01	0.167	0.142	0.244	0.217	0.157	0.158
	0.05	0.344	0.295	0.376	0.220	0.164	0.212
4/40/100	0	0.252	0.210	0.008	0.397	0.281	0.080
	0.01	0.038	0.033	0.169	0.267	0.191	0.132
	0.05	0.288	0.246	0.332	0.217	0.159	0.192
12/20/100	0	0.161	0.161	0.180	0.076	0.076	0.079
	0.01	0.235	0.234	0.256	0.098	0.098	0.112
	0.05	0.349	0.350	0.383	0.142	0.143	0.178
12/20/200	0	0.103	0.097	0.017	0.087	0.081	0.032
	0.01	0.197	0.196	0.218	0.089	0.088	0.096
	0.05	0.337	0.338	0.369	0.136	0.138	0.170
12/40/100	0	0.136	0.129	0.040	0.092	0.085	0.028
	0.01	0.087	0.087	0.117	0.070	0.068	0.058
	0.05	0.293	0.292	0.317	0.118	0.118	0.141
20/20/100	0	0.161	0.164	0.169	0.060	0.063	0.063
	0.01	0.244	0.249	0.256	0.086	0.090	0.100
	0.05	0.350	0.360	0.387	0.135	0.144	0.173
20/20/200	0	0.102	0.100	0.038	0.052	0.051	0.018
	0.01	0.197	0.201	0.207	0.071	0.075	0.078
	0.05	0.338	0.348	0.371	0.128	0.137	0.162
20/40/100	0	0.125	0.122	0.050	0.058	0.057	0.020
	0.01	0.101	0.103	0.112	0.047	0.049	0.043
	0.05	0.299	0.306	0.321	0.109	0.116	0.134
40/20/100	0	0.169	0.175	0.162	0.045	0.049	0.045
	0.01	0.259	0.269	0.265	0.081	0.088	0.092
	0.05	0.357	0.375	0.398	0.134	0.148	0.172
40/20/200	0	0.096	0.097	0.049	0.029	0.029	0.010
	0.01	0.198	0.205	0.194	0.056	0.061	0.059
	0.05	0.346	0.363	0.381	0.127	0.140	0.160
40/40/100	0	0.111	0.112	0.057	0.033	0.033	0.010
	0.01	0.103	0.106	0.097	0.028	0.030	0.025
	0.05	0.304	0.317	0.323	0.103	0.113	0.124

Table 6.35: Bias and MSE for the measures of heterogeneity with 'no' heterogeneity and control group disease rates $r_B = \{0.35, 0.45, 0.50, 0.55\}$ based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ(truncated), and odds ratio $\psi = 0.7$.

		Bias			MSE		
$k/N_j/m$	ρ	H_a	R_a	I_a^2	H_a	R_a	I_a^2
4/20/100	0	0.123	0.097	0.135	0.073	0.046	0.068
	0.01	0.168	0.133	0.174	0.118	0.075	0.094
	0.05	0.156	0.124	0.160	0.108	0.070	0.087
4/20/200	0	0.141	0.111	0.153	0.086	0.054	0.080
	0.01	0.162	0.128	0.170	0.106	0.068	0.091
	0.05	0.178	0.141	0.180	0.125	0.081	0.100
4/40/100	0	0.137	0.108	0.148	0.086	0.054	0.077
	0.01	0.167	0.132	0.174	0.109	0.069	0.093
	0.05	0.161	0.127	0.167	0.107	0.069	0.090
12/20/100	0	0.060	0.055	0.087	0.015	0.013	0.028
	0.01	0.082	0.076	0.115	0.025	0.022	0.041
	0.05	0.080	0.074	0.111	0.024	0.021	0.040
12/20/200	0	0.050	0.046	0.075	0.012	0.010	0.023
	0.01	0.092	0.085	0.129	0.027	0.024	0.046
	0.05	0.081	0.075	0.114	0.024	0.021	0.040
12/40/100	0	0.065	0.060	0.092	0.019	0.016	0.032
	0.01	0.081	0.075	0.114	0.023	0.020	0.040
	0.05	0.084	0.078	0.117	0.026	0.022	0.042
20/20/100	0	0.042	0.040	0.065	0.009	0.008	0.018
	0.01	0.070	0.067	0.104	0.017	0.016	0.032
	0.05	0.059	0.057	0.090	0.013	0.012	0.026
20/20/200	0	0.043	0.041	0.066	0.009	0.008	0.018
	0.01	0.062	0.060	0.095	0.013	0.012	0.028
	0.05	0.057	0.054	0.086	0.012	0.011	0.025
20/40/100	0	0.047	0.045	0.072	0.010	0.009	0.020
	0.01	0.061	0.058	0.093	0.013	0.012	0.027
	0.05	0.062	0.060	0.094	0.014	0.013	0.028
40/20/100	0	0.025	0.024	0.041	0.003	0.003	0.008
	0.01	0.045	0.044	0.074	0.007	0.006	0.016
	0.05	0.047	0.046	0.076	0.007	0.007	0.017
40/20/200	0	0.025	0.024	0.041	0.003	0.003	0.008
	0.01	0.045	0.044	0.074	0.007	0.006	0.016
	0.05	0.042	0.041	0.069	0.006	0.006	0.015
40/40/100	0	0.029	0.029	0.049	0.004	0.004	0.010
	0.01	0.043	0.043	0.072	0.006	0.006	0.015
	0.05	0.039	0.038	0.064	0.006	0.006	0.014

Table 6.36: Bias and MSE for the measures of heterogeneity with 'low' heterogeneity and control group disease rates $r_B = \{0.35, 0.45, 0.50, 0.55\}$ based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ(truncated), and odds ratio $\psi = 0.7$.

		Bias			MSE		
$k/N_j/m$	ρ	H_a	R_a	I_a^2	H_a	R_a	I_a^2
4/20/100	0	0.033	0.031	0.073	0.119	0.077	0.080
	0.01	0.029	0.028	0.079	0.113	0.073	0.085
	0.05	0.030	0.019	0.133	0.092	0.059	0.083
4/20/200	0	0.126	0.105	0.001	0.170	0.111	0.089
	0.01	0.044	0.040	0.069	0.131	0.085	0.085
	0.05	0.022	0.013	0.124	0.095	0.061	0.081
4/40/100	0	0.142	0.119	0.009	0.191	0.126	0.090
	0.01	0.072	0.063	0.047	0.145	0.095	0.088
	0.05	0.007	0.001	0.109	0.097	0.062	0.081
12/20/100	0	0.044	0.040	0.097	0.027	0.023	0.045
	0.01	0.043	0.039	0.100	0.029	0.025	0.047
	0.05	0.065	0.060	0.126	0.026	0.023	0.048
12/20/200	0	0.036	0.034	0.012	0.044	0.039	0.048
	0.01	0.038	0.035	0.093	0.030	0.026	0.046
	0.05	0.070	0.064	0.132	0.026	0.023	0.049
12/40/100	0	0.056	0.053	0.012	0.046	0.040	0.047
	0.01	0.008	0.007	0.059	0.034	0.030	0.046
	0.05	0.061	0.056	0.119	0.025	0.022	0.046
20/20/100	0	0.060	0.057	0.109	0.018	0.017	0.037
	0.01	0.057	0.055	0.107	0.019	0.017	0.038
	0.05	0.082	0.079	0.141	0.019	0.017	0.042
20/20/200	0	0.021	0.021	0.011	0.027	0.025	0.035
	0.01	0.046	0.044	0.092	0.019	0.018	0.037
	0.05	0.084	0.081	0.144	0.019	0.017	0.042
20/40/100	0	0.041	0.040	0.014	0.028	0.026	0.034
	0.01	0.018	0.017	0.059	0.023	0.021	0.036
	0.05	0.076	0.072	0.132	0.020	0.018	0.041
40/20/100	0	0.071	0.069	0.117	0.014	0.013	0.031
	0.01	0.080	0.078	0.132	0.014	0.014	0.034
	0.05	0.100	0.098	0.160	0.016	0.015	0.039
40/20/200	0	0.011	0.011	0.007	0.014	0.013	0.021
	0.01	0.065	0.063	0.109	0.013	0.013	0.030
	0.05	0.098	0.096	0.158	0.016	0.015	0.039
40/40/100	0	0.019	0.019	0.001	0.016	0.015	0.023
	0.01	0.031	0.030	0.062	0.013	0.012	0.025
	0.05	0.086	0.085	0.141	0.015	0.015	0.036

Table 6.37: Bias and MSE for the measures of heterogeneity with 'moderate' heterogeneity and control group disease rates $r_B = \{0.35, 0.45, 0.50, 0.55\}$ based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ(truncated), and odds ratio $\psi = 0.7$.

		Bias			MSE		
$k/N_j/m$	ρ	H_a	R_a	I_a^2	H_a	R_a	I_a^2
4/20/100	0	0.020	0.009	0.153	0.185	0.121	0.118
	0.01	0.097	0.072	0.211	0.157	0.101	0.132
	0.05	0.175	0.135	0.269	0.130	0.083	0.144
4/20/200	0	0.194	0.164	0.016	0.312	0.210	0.097
	0.01	0.066	0.046	0.188	0.168	0.109	0.126
	0.05	0.176	0.135	0.273	0.141	0.090	0.145
4/40/100	0	0.183	0.156	0.020	0.304	0.204	0.096
	0.01	0.000	0.007	0.139	0.200	0.132	0.115
	0.05	0.142	0.108	0.246	0.144	0.092	0.140
12/20/100	0	0.107	0.099	0.161	0.054	0.047	0.073
	0.01	0.153	0.142	0.213	0.059	0.051	0.089
	0.05	0.210	0.195	0.277	0.067	0.058	0.109
12/20/200	0	0.085	0.081	0.012	0.076	0.067	0.045
	0.01	0.116	0.107	0.172	0.056	0.049	0.076
	0.05	0.208	0.193	0.275	0.067	0.058	0.109
12/40/100	0	0.088	0.083	0.019	0.071	0.062	0.042
	0.01	0.055	0.051	0.111	0.056	0.048	0.062
	0.05	0.172	0.160	0.236	0.063	0.055	0.098
20/20/100	0	0.123	0.118	0.164	0.040	0.037	0.060
	0.01	0.167	0.160	0.217	0.048	0.044	0.078
	0.05	0.217	0.208	0.280	0.059	0.054	0.101
20/20/200	0	0.068	0.066	0.022	0.044	0.041	0.029
	0.01	0.126	0.120	0.172	0.044	0.041	0.066
	0.05	0.216	0.207	0.282	0.061	0.056	0.105
20/40/100	0	0.079	0.076	0.034	0.042	0.039	0.026
	0.01	0.059	0.056	0.099	0.039	0.036	0.048
	0.05	0.189	0.181	0.246	0.054	0.049	0.090
40/20/100	0	0.124	0.121	0.153	0.030	0.028	0.045
	0.01	0.171	0.167	0.214	0.041	0.039	0.066
	0.05	0.231	0.226	0.299	0.061	0.058	0.105
40/20/200	0	0.060	0.059	0.035	0.022	0.021	0.014
	0.01	0.148	0.145	0.184	0.035	0.034	0.055
	0.05	0.228	0.223	0.294	0.059	0.057	0.102
40/40/100	0	0.081	0.079	0.052	0.025	0.024	0.015
	0.01	0.075	0.073	0.097	0.022	0.021	0.030
	0.05	0.203	0.199	0.258	0.051	0.049	0.085

Table 6.38: Bias and MSE for the measures of heterogeneity with 'high' heterogeneity and control group disease rates $r_B = \{0.35, 0.45, 0.50, 0.55\}$ based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ(truncated), and odds ratio $\psi = 0.7$.

		Bias			MSE		
$k/N_j/m$	ρ	H_a	R_a	I_a^2	H_a	R_a	I_a^2
4/20/100	0	0.091	0.066	0.201	0.236	0.154	0.141
	0.01	0.186	0.143	0.264	0.213	0.137	0.164
	0.05	0.312	0.245	0.360	0.214	0.136	0.204
4/20/200	0	0.233	0.198	0.010	0.361	0.246	0.081
	0.01	0.143	0.108	0.235	0.226	0.147	0.152
	0.05	0.340	0.268	0.380	0.207	0.131	0.213
4/40/100	0	0.229	0.194	0.013	0.358	0.243	0.083
	0.01	0.033	0.019	0.161	0.252	0.167	0.125
	0.05	0.267	0.209	0.324	0.211	0.135	0.187
12/20/100	0	0.130	0.121	0.156	0.074	0.064	0.072
	0.01	0.235	0.219	0.260	0.101	0.088	0.116
	0.05	0.312	0.291	0.344	0.126	0.110	0.158
12/20/200	0	0.117	0.111	0.026	0.089	0.079	0.032
	0.01	0.171	0.159	0.197	0.085	0.074	0.090
	0.05	0.317	0.296	0.347	0.126	0.110	0.157
12/40/100	0	0.138	0.131	0.038	0.096	0.085	0.031
	0.01	0.075	0.070	0.113	0.073	0.064	0.060
	0.05	0.276	0.257	0.302	0.112	0.097	0.135
20/20/100	0	0.148	0.142	0.155	0.055	0.050	0.056
	0.01	0.241	0.231	0.257	0.087	0.080	0.103
	0.05	0.332	0.319	0.366	0.126	0.116	0.160
20/20/200	0	0.107	0.103	0.043	0.051	0.047	0.017
	0.01	0.196	0.188	0.203	0.067	0.062	0.074
	0.05	0.321	0.308	0.353	0.121	0.112	0.154
20/40/100	0	0.131	0.126	0.053	0.064	0.060	0.020
	0.01	0.095	0.092	0.110	0.048	0.045	0.044
	0.05	0.286	0.275	0.308	0.104	0.096	0.128
40/20/100	0	0.160	0.156	0.153	0.042	0.041	0.043
	0.01	0.246	0.241	0.251	0.076	0.073	0.085
	0.05	0.337	0.331	0.372	0.123	0.118	0.156
40/20/200	0	0.099	0.097	0.050	0.030	0.029	0.010
	0.01	0.189	0.185	0.185	0.052	0.050	0.055
	0.05	0.336	0.330	0.369	0.121	0.117	0.154
40/40/100	0	0.124	0.122	0.066	0.033	0.032	0.010
	0.01	0.101	0.099	0.097	0.028	0.027	0.025
	0.05	0.293	0.287	0.310	0.098	0.094	0.118

Table 6.39: Confidence interval for H_a with 'no' heterogeneity, control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ for MOVER, Q distribution, test-based, based on τ_c^2, nonparametric bootstrap based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ(truncated), and odds ratio $\psi = 0.7$

$k/N_j/m$	ρ	MOVER/τ_c^2 C(L,R)W	Q C(L,R)W	Test-Based C(L,R)W	NB C(L,R)W
4/20/100	0	98.6(1.4, 0.0)2.5	96.7(3.3, 0.0)1.5	95.9(4.1, 0.0)1.9	100(0.0, 0.0)0.28
	0.01	98.0(2.0, 0.0)2.7	95.3(4.7, 0.0)1.5	94.5(5.5, 0.0)1.9	100(0.0, 0.0)0.35
	0.05	98.5(1.5, 0.0)2.6	96.3(3.7, 0.0)1.5	95.2(4.8, 0.0)1.9	100(0.0, 0.0)0.33
4/20/200	0	98.6(1.4, 0.0)2.5	97.1(2.9, 0.0)1.5	96.4(3.6, 0.0)1.9	100(0.0, 0.0)0.31
	0.01	97.5(2.5, 0.0)2.7	95.7(4.3, 0.0)1.5	94.5(5.5, 0.0)1.9	100(0.0, 0.0)0.35
	0.05	98.1(1.9, 0.0)2.6	95.5(4.5, 0.0)1.5	95.2(4.8, 0.0)1.9	100(0.0, 0.0)0.33
4/40/100	0	97.7(2.3, 0.0)2.6	95.6(4.4, 0.0)1.5	94.9(5.1, 0.0)1.9	100(0.0, 0.0)0.32
	0.01	97.8(2.2, 0.0)2.6	94.7(5.3, 0.0)1.4	93.6(6.4, 0.0)1.9	100(0.0, 0.0)0.32
	0.05	98.0(2.0, 0.0)2.7	95.5(4.5, 0.0)1.5	94.2(5.8, 0.0)1.8	100(0.0, 0.0)0.35
12/20/100	0	98.6(1.4, 0.0)0.67	97.2(2.7, 0.1)1.0	97.9(2.1, 0.0)0.83	100(0.0, 0.0)0.25
	0.01	98.3(1.7, 0.0)0.74	96.4(3.6, 0.0)0.99	96.5(3.5, 0.0)0.83	99.8(0.2, 0.0)0.30
	0.05	98.7(1.3, 0.0)0.69	98.0(2.0, 0.0)1.0	98.1(1.9, 0.0)0.83	99.6(0.4, 0.0)0.27
12/20/200	0	98.4(1.6, 0.0)0.67	97.3(2.5, 0.2)1.0	97.9(2.1, 0.0)0.83	99.9(0.1, 0.0)0.25
	0.01	97.9(2.1, 0.0)0.71	96.2(3.8, 0.0)1.0	96.2(3.8, 0.0)0.84	99.8(0.2, 0.0)0.28
	0.05	98.6(1.4, 0.0)0.69	96.5(3.4, 0.1)0.99	97.2(2.8, 0.0)0.83	100(0.0, 0.0)0.28
12/40/100	0	98.1(1.9, 0.0)0.68	96.5(3.3, 0.2)1.0	97.2(2.8, 0.0)0.84	100(0.0, 0.0)0.25
	0.01	97.6(2.4, 0.0)0.73	96.1(3.8, 0.1)0.99	96.7(3.3, 0.0)0.83	99.8(0.2, 0.0)0.29
	0.05	97.7(2.3, 0.0)0.74	95.7(4.1, 0.2)0.98	96.2(3.8, 0.0)0.83	99.6(0.4, 0.0)0.30
20/20/100	0	98.9(1.1, 0.0)0.42	97.3(1.7, 1.0)0.78	98.3(1.7, 0.0)0.63	99.8(0.2, 0.0)0.20
	0.01	97.6(2.4, 0.0)0.46	96.2(3.2, 0.6)0.75	97.1(2.9, 0.0)0.62	99.7(0.3, 0.0)0.23
	0.05	98.7(1.3, 0.0)0.46	97.7(2.3, 0.0)0.74	98.0(2.0, 0.0)0.61	99.5(0.5, 0.0)0.23
20/20/200	0	98.4(1.6, 0.0)0.44	96.4(2.3, 1.3)0.77	98.3(1.7, 0.0)0.62	99.6(0.4, 0.0)0.21
	0.01	98.2(1.8, 0.0)0.47	96.1(3.2, 0.7)0.74	97.2(2.8, 0.0)0.62	99.9(0.1, 0.0)0.24
	0.05	97.7(2.3, 0.0)0.46	96.2(3.4, 0.4)0.74	96.7(3.3, 0.0)0.62	99.6(0.4, 0.0)0.24
20/40/100	0	98.7(1.3, 0.0)0.45	97.0(2.1, 0.9)0.76	98.4(1.6, 0.0)0.62	99.6(0.4, 0.0)0.21
	0.01	97.2(2.8, 0.0)0.49	95.2(4.1, 0.7)0.74	96.1(3.9, 0.0)0.62	99.5(0.5, 0.0)0.25
	0.05	98.1(1.9, 0.0)0.47	95.5(3.8, 0.7)0.74	96.6(3.4, 0.0)0.62	99.3(0.7, 0.0)0.24
40/20/100	0	99.4(0.6, 0.0)0.25	96.6(1.0, 2.4)0.50	99.3(0.7, 0.0)0.43	99.9(0.1, 0.0)0.14
	0.01	98.3(1.7, 0.0)0.29	97.0(2.5, 0.5)0.47	97.6(2.4, 0.0)0.43	99.3(0.7, 0.0)0.19
	0.05	98.5(1.5, 0.0)0.28	96.8(2.4, 0.8)0.47	98.0(2.0, 0.0)0.43	99.8(0.2, 0.0)0.18
40/20/200	0	99.2(0.8, 0.0)0.26	96.1(1.2, 2.7)0.50	99.1(0.9, 0.0)0.43	99.9(0.1, 0.0)0.14
	0.01	97.8(2.2, 0.0)0.30	96.1(3.1, 0.8)0.47	97.4(2.6, 0.0)0.43	99.1(0.9, 0.0)0.19
	0.05	98.4(1.6, 0.0)0.28	95.5(3.1, 1.4)0.47	97.6(2.4, 0.0)0.43	99.6(0.4, 0.0)0.18
40/40/100	0	98.6(1.4, 0.0)0.27	94.8(2.4, 2.8)0.49	97.9(2.1, 0.0)0.43	99.8(0.2, 0.0)0.16
	0.01	97.8(2.2, 0.0)0.30	95.4(3.4, 1.2)0.47	97.1(2.9, 0.0)0.43	99.5(0.5, 0.0)0.19
	0.05	98.0(2.0, 0.0)0.29	95.8(3.0, 1.2)0.47	97.3(2.7, 0.0)0.43	99.6(0.4, 0.0)0.18

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.40: Confidence interval for H_a with 'low' heterogeneity, control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ for MOVER, Q distribution, test-based, based on τ_c^2, nonparametric bootstrap based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ(truncated), and odds ratio $\psi = 0.7$

$k/N_j/m$	ρ	MOVER/τ_c^2 C(L,R)W	Q C(L,R)W	Test-Based C(L,R)W	NB C(L,R)W
4/20/100	0	94.5(0.9, 4.6)2.9	97.4(2.6, 0.0)1.6	94.2(3.5, 2.3)1.9	51.5(0.0,48.5)0.43
	0.01	95.7(0.7, 3.5)3.0	96.6(3.4, 0.0)1.5	95.4(3.4, 1.1)1.9	51.6(0.0,48.4)0.45
	0.05	96.2(0.1, 3.7)2.7	98.1(1.9, 0.0)1.5	96.6(1.3, 2.1)1.9	46.4(0.0,53.6)0.34
4/20/200	0	95.5(2.6, 1.9)3.6	95.5(4.5, 0.0)1.7	91.1(7.2, 1.7)1.9	64.3(0.0,35.7)0.65
	0.01	97.4(1.0, 1.6)3.2	95.6(4.4, 0.0)1.5	93.3(4.9, 1.8)1.8	57.4(0.0,42.6)0.50
	0.05	97.0(0.2, 2.8)2.7	97.7(2.3, 0.0)1.5	95.5(2.3, 2.2)1.9	48.2(0.0,51.8)0.36
4/40/100	0	94.9(3.7, 1.4)3.7	94.2(5.8, 0.0)1.7	89.4(9.0, 1.5)1.9	68.0(0.0,32.0)0.68
	0.01	95.8(1.4, 2.8)3.4	96.5(3.5, 0.0)1.6	93.2(4.9, 1.9)1.8	62.2(0.0,37.8)0.54
	0.05	94.9(0.4, 4.7)2.8	97.3(2.7, 0.0)1.5	95.5(2.5, 2.0)1.9	48.4(0.0,51.6)0.39
12/20/100	0	95.2(0.7, 4.1)0.88	92.1(1.1, 6.8)1.1	96.4(1.0, 2.6)0.83	73.2(0.1,26.7)0.37
	0.01	96.6(0.5, 2.9)0.85	92.4(1.1, 6.5)1.0	96.8(0.9, 2.3)0.83	69.7(0.1,30.2)0.35
	0.05	94.8(0.0, 5.2)0.76	91.6(0.3, 8.1)1.0	97.1(0.0, 2.9)0.81	65.9(0.0,34.1)0.30
12/20/200	0	94.8(3.8, 1.4)1.1	94.1(3.7, 2.2)1.1	93.8(5.4, 0.8)0.85	86.2(0.5,13.3)0.53
	0.01	96.3(0.8, 2.9)0.90	92.5(1.9, 5.6)1.0	97.2(1.2, 1.6)0.83	74.2(0.0,25.8)0.39
	0.05	94.8(0.0, 5.2)0.76	91.1(0.8, 8.1)1.0	96.6(0.4, 3.0)0.82	65.9(0.0,34.1)0.30
12/40/100	0	94.6(3.8, 1.6)1.1	93.7(3.3, 3.0)1.1	94.8(3.9, 1.3)0.86	83.9(0.6,15.5)0.50
	0.01	95.0(1.7, 3.3)0.98	91.3(3.3, 5.4)1.0	95.3(3.0, 1.7)0.84	79.7(0.3,20.0)0.45
	0.05	94.4(0.3, 5.3)0.82	90.8(1.6, 7.6)0.99	96.8(0.7, 2.5)0.82	70.3(0.1,29.6)0.34
20/20/100	0	94.6(0.6, 4.8)0.62	89.9(1.2, 8.9)0.76	96.6(1.0, 2.4)0.62	76.6(0.2,23.2)0.33
	0.01	93.2(0.2, 6.6)0.58	87.3(1.3,11.4)0.72	98.1(0.4, 1.5)0.63	71.9(0.2,27.9)0.31
	0.05	91.6(0.0, 8.4)0.48	83.3(0.4,16.3)0.74	96.6(0.3, 3.1)0.61	63.3(0.0,36.7)0.24
20/20/200	0	94.0(4.3, 1.7)0.76	92.7(2.9, 4.4)0.81	94.5(3.7, 1.8)0.64	88.6(0.5,10.9)0.44
	0.01	94.8(0.6, 4.6)0.60	88.8(1.2,10.0)0.73	96.7(0.8, 2.5)0.62	76.1(0.1,23.8)0.32
	0.05	91.7(0.1, 8.2)0.50	83.1(0.5,16.4)0.73	97.1(0.2, 2.7)0.62	64.9(0.0,35.1)0.26
20/40/100	0	92.8(6.0, 1.2)0.78	92.4(4.4, 3.2)0.79	92.5(5.7, 1.8)0.65	89.1(1.3, 9.6)0.46
	0.01	95.2(1.8, 3.0)0.68	89.2(3.9, 6.9)0.73	95.0(3.0, 2.0)0.63	83.1(0.6,16.3)0.39
	0.05	93.3(0.1, 6.6)0.54	87.5(0.7,11.8)0.72	96.6(0.3, 3.1)0.62	68.5(0.1,31.4)0.28
40/20/100	0	93.2(0.7, 6.1)0.40	81.2(0.7,18.1)0.50	96.2(0.7, 3.1)0.43	74.7(0.2,25.1)0.26
	0.01	91.6(0.2, 8.2)0.38	80.9(0.5,18.6)0.47	96.2(0.3, 3.5)0.43	74.0(0.1,25.9)0.25
	0.05	85.9(0.0,14.1)0.31	72.6(0.0,27.4)0.47	94.7(0.0, 5.3)0.43	63.3(0.0,36.7)0.20
40/20/200	0	94.5(5.1, 0.4)0.53	93.8(2.2, 4.0)0.54	94.7(3.1, 2.2)0.46	92.9(1.1, 6.0)0.37
	0.01	92.0(0.4, 7.6)0.40	81.6(0.8,17.6)0.48	96.7(0.5, 2.8)0.43	77.1(0.1,22.8)0.27
	0.05	87.0(0.0,13.0)0.32	73.3(0.4,26.3)0.47	97.1(0.1, 2.8)0.43	62.5(0.0,37.5)0.21
40/40/100	0	91.9(7.4, 0.7)0.53	92.2(4.1, 3.7)0.54	93.2(5.3, 1.5)0.46	92.1(2.1, 5.8)0.37
	0.01	95.3(1.0, 3.7)0.45	88.7(1.3,10.0)0.49	96.3(1.2, 2.5)0.44	85.7(0.2,14.1)0.31
	0.05	87.4(0.0,12.6)0.34	74.4(0.2,25.4)0.47	96.7(0.1, 3.2)0.43	65.9(0.0,34.1)0.22

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.41: Confidence interval for H_a with 'moderate' heterogeneity, control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ for MOVER, Q distribution, test-based, based on τ_c^2, nonparametric bootstrap based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ(truncated), and odds ratio $\psi = 0.7$

$k/N_j/m$	ρ	MOVER/τ_c^2 C(L,R)W	Q C(L,R)W	Test-Based C(L,R)W	NB C(L,R)W
4/20/100	0	97.1(0.7, 2.2)3.7	88.8(2.1, 9.0)1.7	92.5(3.3, 4.2)1.8	59.2(0.0,40.8)0.65
	0.01	96.5(0.2, 3.3)3.3	83.1(2.5,14.4)1.6	93.4(2.0, 4.6)1.8	49.9(0.0,50.1)0.53
	0.05	95.4(0.1, 4.5)2.9	83.0(1.0,16.0)1.5	94.8(0.7, 4.5)1.8	38.9(0.0,61.1)0.37
4/20/200	0	96.6(2.5, 0.9)4.5	91.5(4.3, 4.2)1.9	86.8(10.1, 3.1)1.9	72.9(0.0,27.1)0.94
	0.01	96.1(0.5, 3.4)3.4	86.5(2.4,11.1)1.6	92.9(2.5, 4.6)1.8	53.3(0.0,46.7)0.57
	0.05	95.5(0.1, 4.4)2.9	81.7(1.4,16.9)1.5	95.5(0.9, 3.6)1.9	40.2(0.0,59.8)0.40
4/40/100	0	96.2(2.8, 1.0)4.6	91.5(4.3, 4.1)1.9	87.6(10.0, 2.4)1.9	76.4(0.0,23.6)0.99
	0.01	96.8(0.4, 2.8)3.8	88.0(3.4, 8.6)1.7	92.1(4.6, 3.3)1.8	59.6(0.0,40.4)0.69
	0.05	95.8(0.1, 4.1)3.2	85.4(1.4,13.2)1.5	95.0(1.1, 3.9)1.8	46.0(0.0,54.0)0.47
12/20/100	0	96.2(0.9, 2.9)1.1	82.7(1.4,15.9)1.1	93.7(1.4, 4.9)0.85	73.5(0.0,26.5)0.53
	0.01	93.9(0.1, 6.0)0.97	72.2(1.3,26.5)1.0	94.4(0.4, 5.2)0.84	62.0(0.1,37.9)0.43
	0.05	90.5(0.0, 9.5)0.82	63.4(0.2,36.4)1.0	92.7(0.1, 7.2)0.81	49.4(0.0,50.6)0.33
12/20/200	0	95.6(4.1, 0.3)1.4	91.5(3.9, 4.6)1.3	91.6(6.4, 2.0)0.93	90.9(0.6, 8.5)0.73
	0.01	95.0(0.3, 4.7)1.0	78.0(1.4,20.6)1.1	94.4(0.6, 5.0)0.85	69.1(0.0,30.9)0.49
	0.05	89.8(0.0,10.2)0.79	61.0(0.4,38.6)1.0	94.1(0.1, 5.8)0.82	46.8(0.0,53.2)0.32
12/40/100	0	95.4(4.4, 0.2)1.4	91.0(4.3, 4.7)1.2	90.2(7.6, 2.2)0.94	90.7(0.4, 8.9)0.75
	0.01	96.2(1.1, 2.7)1.2	83.4(2.5,14.1)1.1	93.9(2.5, 3.6)0.87	77.0(0.2,22.8)0.58
	0.05	91.7(0.0, 8.3)0.91	67.6(0.2,32.2)1.0	94.8(0.2, 5.0)0.83	56.7(0.0,43.3)0.39
20/20/100	0	93.6(0.6, 5.8)0.79	82.0(0.9,17.1)0.80	93.0(0.6, 6.4)0.65	74.2(0.3,25.5)0.46
	0.01	88.8(0.1,11.1)0.67	69.9(0.6,29.5)0.74	91.3(0.1, 8.6)0.63	59.2(0.1,40.7)0.37
	0.05	81.9(0.0,18.1)0.53	52.7(0.1,47.2)0.73	92.0(0.0, 8.0)0.62	42.4(0.0,57.6)0.28
20/20/200	0	93.9(6.0, 0.1)1.0	93.7(4.4, 1.9)0.89	92.0(7.1, 0.9)0.72	94.8(1.0, 4.2)0.66
	0.01	92.7(0.2, 7.1)0.74	76.5(0.5,23.0)0.75	93.3(0.3, 6.4)0.65	67.7(0.1,32.2)0.43
	0.05	81.6(0.0,18.4)0.54	55.8(0.1,44.1)0.73	91.1(0.0, 8.9)0.62	44.3(0.0,55.7)0.29
20/40/100	0	90.7(9.3, 0.0)1.0	90.0(7.7, 2.3)0.88	87.3(10.7, 2.0)0.73	92.5(2.1, 5.4)0.67
	0.01	97.1(0.7, 2.2)0.85	86.4(2.0,11.6)0.78	92.0(1.5, 6.5)0.66	80.7(0.0,19.3)0.51
	0.05	86.6(0.0,13.4)0.61	63.6(0.3,36.1)0.73	92.3(0.0, 7.7)0.62	52.7(0.0,47.3)0.33
40/20/100	0	91.6(0.0, 8.4)0.53	84.3(0.0,15.7)0.55	75.7(0.0,24.3)0.46	67.9(0.0,32.1)0.37
	0.01	83.2(0.0,16.8)0.46	61.9(0.0,38.1)0.49	62.7(0.0,37.3)0.44	54.3(0.0,45.7)0.31
	0.05	58.2(0.0,41.8)0.33	21.7(0.0,78.3)0.48	32.2(0.0,67.8)0.42	24.1(0.0,75.9)0.21
40/20/200	0	92.3(7.4, 0.3)0.68	95.7(3.8, 0.5)0.60	92.6(5.7, 1.7)0.51	94.6(2.0, 3.4)0.50
	0.01	89.5(0.2,10.3)0.50	74.2(0.4,25.4)0.51	71.7(0.3,28.0)0.45	62.2(0.2,37.6)0.35
	0.05	62.2(0.0,37.8)0.34	26.2(0.0,73.8)0.47	36.8(0.0,63.2)0.43	27.8(0.0,72.2)0.23
40/40/100	0	90.1(9.8, 0.1)0.70	93.7(6.0, 0.3)0.58	89.7(9.3, 1.0)0.51	93.9(4.1, 2.0)0.51
	0.01	96.5(0.7, 2.8)0.59	93.7(0.8, 5.5)0.53	87.9(0.7,11.4)0.47	83.3(0.3,16.4)0.42
	0.05	74.4(0.0,25.6)0.41	48.1(0.0,51.9)0.48	53.8(0.0,46.2)0.44	44.1(0.0,55.9)0.27

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.42: Confidence interval for H_a with 'high' heterogeneity, control group disease rates $r_A = \{0.04, 0.07, 0.10, 0.13\}$ for MOVER, Q distribution, test-based, based on τ_c^2, nonparametric bootstrap based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ(truncated), and odds ratio $\psi = 0.7$

$k/N_j/m$	ρ	MOVER/τ_c^2 C(L,R)W	Q C(L,R)W	Test-Based C(L,R)W	NB C(L,R)W
4/20/100	0	97.8(0.3, 1.9)4.1	73.8(1.4,24.8)1.8	92.1(2.1, 5.8)1.9	57.0(0.0,43.0)0.78
	0.01	96.6(0.0, 3.4)3.5	64.7(1.6,33.6)1.7	92.0(1.4, 6.5)1.8	47.1(0.0,52.9)0.60
	0.05	96.6(0.0, 3.4)3.1	58.6(0.7,40.7)1.5	91.6(0.5, 7.9)1.8	32.7(0.0,67.3)0.42
4/20/200	0	96.4(2.3, 1.3)5.3	85.8(4.1,10.0)2.1	85.5(10.2, 4.2)2.0	77.8(0.0,22.2)1.2
	0.01	97.3(0.0, 2.7)3.9	70.1(1.2,28.7)1.7	92.5(1.3, 6.2)1.9	53.3(0.0,46.7)0.70
	0.05	95.7(0.0, 4.3)3.1	57.7(1.0,41.3)1.5	92.6(0.5, 6.9)1.8	34.5(0.0,65.5)0.43
4/40/100	0	96.6(2.7, 0.7)5.4	87.7(4.2, 8.1)2.1	85.2(11.0, 3.7)2.0	81.1(0.0,18.9)1.3
	0.01	97.7(0.8, 1.5)4.4	76.2(2.7,21.1)1.8	89.8(4.0, 6.2)1.9	62.3(0.0,37.7)0.89
	0.05	94.8(0.0, 5.2)3.3	61.9(0.9,37.2)1.6	92.5(0.8, 6.7)1.8	40.7(0.0,59.3)0.52
12/20/100	0	96.3(0.1, 3.6)1.3	85.4(0.4,14.2)1.2	91.6(0.4, 8.0)0.88	69.2(0.1,30.7)0.62
	0.01	92.9(0.0, 7.1)1.1	74.3(0.5,25.2)1.1	89.8(0.1,10.1)0.86	56.9(0.0,43.1)0.52
	0.05	84.0(0.0,16.0)0.84	38.2(0.0,61.8)1.0	86.5(0.0,13.5)0.82	31.2(0.0,68.8)0.34
12/20/200	0	95.0(4.7, 0.3)1.6	92.5(5.3, 2.2)1.3	89.4(8.3, 2.3)1.0	91.2(0.9, 7.9)0.91
	0.01	94.6(0.1, 5.3)1.2	78.8(0.8,20.4)1.1	93.2(0.2, 6.6)0.87	61.8(0.1,38.1)0.57
	0.05	82.6(0.0,17.4)0.84	41.6(0.1,58.3)1.0	87.9(0.0,12.1)0.83	31.8(0.0,68.2)0.35
12/40/100	0	94.7(5.0, 0.3)1.7	93.5(5.3, 1.2)1.3	88.9(9.3, 1.8)1.0	94.3(0.8, 4.9)0.95
	0.01	97.7(1.0, 1.3)1.4	88.7(2.6, 8.7)1.2	91.5(2.4, 6.1)0.92	78.9(0.0,21.1)0.71
	0.05	89.0(0.0,11.0)0.97	61.6(0.0,38.4)1.0	89.1(0.0,10.9)0.83	44.8(0.0,55.2)0.43
20/20/100	0	93.6(0.2, 6.2)0.91	88.4(0.4,11.2)0.87	79.6(0.2,20.2)0.68	69.9(0.0,30.1)0.54
	0.01	87.1(0.0,12.9)0.78	63.5(0.3,36.2)0.76	65.6(0.0,34.4)0.65	54.1(0.0,45.9)0.45
	0.05	67.1(0.0,32.9)0.56	21.3(0.0,78.7)0.74	34.0(0.0,66.0)0.62	23.4(0.0,76.6)0.29
20/20/200	0	92.4(7.3, 0.3)1.2	94.0(5.4, 0.6)0.92	87.9(10.2, 1.9)0.77	94.9(1.3, 3.8)0.76
	0.01	93.0(0.1, 6.9)0.84	76.5(0.8,22.7)0.78	73.4(0.3,26.3)0.66	62.4(0.0,37.6)0.51
	0.05	71.7(0.0,28.3)0.58	25.3(0.0,74.7)0.72	38.8(0.0,61.2)0.62	26.5(0.0,73.5)0.31
20/40/100	0	91.9(7.9, 0.2)1.2	92.8(6.6, 0.6)0.91	88.1(10.2, 1.7)0.78	93.7(2.0, 4.3)0.78
	0.01	97.2(0.4, 2.4)0.99	92.6(1.1, 6.3)0.83	86.3(0.9,12.8)0.70	80.0(0.1,19.9)0.61
	0.05	81.7(0.0,18.3)0.68	40.3(0.3,59.4)0.74	50.6(0.0,49.4)0.63	36.5(0.0,63.5)0.37
40/20/100	0	92.0(0.0, 8.0)0.62	77.0(0.0,23.0)0.58	73.7(0.0,26.3)0.48	63.2(0.0,36.8)0.44
	0.01	73.4(0.0,26.6)0.52	35.9(0.0,64.1)0.51	47.0(0.0,53.0)0.46	37.6(0.0,62.4)0.36
	0.05	34.5(0.0,65.5)0.36	2.60(0.0,97.4)0.48	13.0(0.0,87.0)0.43	8.20(0.0,91.8)0.23
40/20/200	0	87.5(12.5, 0.0)0.79	92.9(7.1, 0.0)0.61	87.0(12.3, 0.7)0.54	94.0(4.5, 1.5)0.55
	0.01	84.7(0.0,15.3)0.58	60.6(0.0,39.4)0.53	66.2(0.0,33.8)0.47	54.5(0.0,45.5)0.42
	0.05	40.0(0.0,60.0)0.38	5.50(0.0,94.5)0.48	14.9(0.0,85.1)0.43	9.10(0.0,90.9)0.25
40/40/100	0	87.8(12.1, 0.1)0.80	91.2(8.8, 0.0)0.60	85.2(14.4, 0.4)0.54	93.3(5.7, 1.0)0.56
	0.01	95.8(0.3, 3.9)0.67	89.8(0.3, 9.9)0.55	87.5(0.3,12.2)0.50	81.0(0.0,19.0)0.49
	0.05	58.0(0.0,42.0)0.46	17.6(0.0,82.4)0.49	31.1(0.0,68.9)0.44	21.6(0.0,78.4)0.31

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.43: Confidence interval for H_a with 'no' heterogeneity, control group disease rates $r_B = \{0.35, 0.45, 0.50, 0.55\}$ for MOVER, Q distribution, test-based, based on τ_c^2, nonparametric bootstrap based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ(truncated), and odds ratio $\psi = 0.7$

$k/N_j/m$	ρ	MOVER/τ_c^2 C(L,R)W	Q C(L,R)W	Test-Based C(L,R)W	NB C(L,R)W
4/20/100	0	98.4(1.6, 0.0)2.3	96.7(3.3, 0.0)1.5	95.9(4.1, 0.0)1.9	100(0.0, 0.0)0.28
	0.01	97.0(3.0, 0.0)2.5	94.3(5.7, 0.0)1.4	93.9(6.1, 0.0)1.9	100(0.0, 0.0)0.35
	0.05	97.4(2.6, 0.0)2.4	93.9(6.1, 0.0)1.4	93.2(6.8, 0.0)1.9	100(0.0, 0.0)0.33
4/20/200	0	98.2(1.8, 0.0)2.4	95.8(4.2, 0.0)1.5	95.6(4.4, 0.0)1.9	100(0.0, 0.0)0.31
	0.01	97.0(3.0, 0.0)2.5	95.0(5.0, 0.0)1.5	94.4(5.6, 0.0)1.9	100(0.0, 0.0)0.35
	0.05	97.2(2.8, 0.0)2.5	93.7(6.3, 0.0)1.4	92.4(7.6, 0.0)1.9	100(0.0, 0.0)0.36
4/40/100	0	98.1(1.9, 0.0)2.4	96.1(3.9, 0.0)1.5	95.6(4.4, 0.0)1.9	100(0.0, 0.0)0.31
	0.01	97.3(2.7, 0.0)2.5	94.5(5.5, 0.0)1.4	93.5(6.5, 0.0)1.9	100(0.0, 0.0)0.36
	0.05	97.0(3.0, 0.0)2.5	94.5(5.5, 0.0)1.4	93.5(6.5, 0.0)1.9	100(0.0, 0.0)0.33
12/20/100	0	99.1(0.9, 0.0)0.63	97.6(2.2, 0.2)1.0	98.4(1.6, 0.0)0.83	99.9(0.1, 0.0)0.25
	0.01	96.4(3.6, 0.0)0.68	94.6(5.3, 0.1)0.99	95.1(4.9, 0.0)0.83	99.8(0.2, 0.0)0.29
	0.05	96.4(3.6, 0.0)0.66	95.4(4.5, 0.1)0.99	95.6(4.4, 0.0)0.84	99.6(0.4, 0.0)0.28
12/20/200	0	99.3(0.7, 0.0)0.64	98.4(1.6, 0.0)1.0	98.6(1.4, 0.0)0.82	100(0.0, 0.0)0.23
	0.01	97.1(2.9, 0.0)0.71	95.0(5.0, 0.0)0.97	95.9(4.1, 0.0)0.84	99.7(0.3, 0.0)0.31
	0.05	97.3(2.7, 0.0)0.67	95.9(4.0, 0.1)0.99	96.3(3.7, 0.0)0.82	99.4(0.6, 0.0)0.29
12/40/100	0	98.0(2.0, 0.0)0.68	96.4(3.5, 0.1)1.0	96.6(3.4, 0.0)0.83	99.6(0.4, 0.0)0.26
	0.01	97.5(2.5, 0.0)0.70	95.2(4.6, 0.2)0.99	95.9(4.1, 0.0)0.83	99.4(0.6, 0.0)0.29
	0.05	97.2(2.8, 0.0)0.68	95.2(4.7, 0.1)0.99	95.8(4.2, 0.0)0.83	99.3(0.7, 0.0)0.29
20/20/100	0	98.1(1.9, 0.0)0.42	96.5(2.7, 0.8)0.78	97.7(2.3, 0.0)0.62	99.5(0.5, 0.0)0.20
	0.01	96.2(3.8, 0.0)0.48	93.7(5.9, 0.4)0.73	94.9(5.1, 0.0)0.62	98.9(1.1, 0.0)0.26
	0.05	96.8(3.2, 0.0)0.45	94.8(4.7, 0.5)0.73	95.7(4.3, 0.0)0.62	99.0(1.0, 0.0)0.25
20/20/200	0	98.8(1.2, 0.0)0.46	96.7(2.3, 1.0)0.77	98.2(1.8, 0.0)0.62	99.7(0.3, 0.0)0.21
	0.01	97.4(2.6, 0.0)0.47	95.6(3.9, 0.5)0.73	96.6(3.4, 0.0)0.63	99.8(0.2, 0.0)0.25
	0.05	97.7(2.3, 0.0)0.44	96.0(3.6, 0.4)0.74	97.2(2.8, 0.0)0.62	99.4(0.6, 0.0)0.23
20/40/100	0	97.8(2.2, 0.0)0.47	95.4(3.5, 1.1)0.76	97.0(3.0, 0.0)0.63	99.6(0.4, 0.0)0.22
	0.01	97.5(2.5, 0.0)0.48	95.8(3.9, 0.3)0.73	96.6(3.4, 0.0)0.62	99.7(0.3, 0.0)0.25
	0.05	96.4(3.6, 0.0)0.45	95.0(4.7, 0.3)0.74	95.7(4.3, 0.0)0.62	98.7(1.3, 0.0)0.24
40/20/100	0	98.7(1.3, 0.0)0.26	96.3(1.4, 2.3)0.50	98.7(1.3, 0.0)0.44	99.4(0.6, 0.0)0.14
	0.01	97.3(2.7, 0.0)0.30	95.0(3.9, 1.1)0.47	96.7(3.3, 0.0)0.43	99.0(1.0, 0.0)0.19
	0.05	96.8(3.2, 0.0)0.29	94.7(4.5, 0.8)0.47	96.2(3.8, 0.0)0.43	99.3(0.7, 0.0)0.20
40/20/200	0	98.7(1.3, 0.0)0.30	96.1(1.7, 2.2)0.50	98.6(1.4, 0.0)0.43	99.7(0.3, 0.0)0.15
	0.01	97.1(2.9, 0.0)0.30	94.8(4.0, 1.2)0.47	96.7(3.3, 0.0)0.43	98.8(1.2, 0.0)0.19
	0.05	97.5(2.5, 0.0)0.28	95.2(3.5, 1.3)0.47	96.7(3.3, 0.0)0.43	99.2(0.8, 0.0)0.18
40/40/100	0	98.6(1.4, 0.0)0.32	95.6(2.3, 2.1)0.49	98.3(1.7, 0.0)0.43	99.6(0.4, 0.0)0.16
	0.01	97.7(2.3, 0.0)0.33	96.0(2.9, 1.1)0.47	97.5(2.5, 0.0)0.43	99.2(0.8, 0.0)0.20
	0.05	97.2(2.8, 0.0)0.28	95.1(3.5, 1.4)0.47	97.0(3.0, 0.0)0.43	98.5(1.5, 0.0)0.18

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.44: Confidence interval for H_a with 'low' heterogeneity, control group disease rates $r_B = \{0.35, 0.45, 0.50, 0.55\}$ for MOVER, Q distribution, test-based, based on τ_c^2, nonparametric bootstrap based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ(truncated), and odds ratio $\psi = 0.7$

$k/N_j/m$	ρ	MOVER/τ_c^2 C(L,R)W	Q C(L,R)W	Test-Based C(L,R)W	NB C(L,R)W
4/20/100	0	96.4(1.8, 1.8)2.9	96.9(3.1, 0.0)1.6	94.6(3.7, 1.7)1.8	56.2(0.0,43.8)0.46
	0.01	95.8(0.4, 3.8)2.8	95.8(4.2, 0.0)1.5	94.5(3.7, 1.8)1.9	54.1(0.0,45.9)0.45
	0.05	94.7(0.8, 4.5)2.5	95.9(4.1, 0.0)1.5	95.2(2.9, 1.9)1.9	46.5(0.0,53.5)0.36
4/20/200	0	93.4(4.6, 2.0)3.3	95.9(4.1, 0.0)1.7	91.2(7.2, 1.6)1.8	65.1(0.0,34.9)0.59
	0.01	95.3(1.6, 3.1)2.9	95.9(4.1, 0.0)1.5	94.6(4.1, 1.3)1.8	55.9(0.0,44.1)0.47
	0.05	96.4(1.0, 2.6)2.6	96.8(3.2, 0.0)1.5	94.9(2.6, 2.5)1.8	47.5(0.0,52.5)0.37
4/40/100	0	91.8(6.6, 1.6)3.4	95.5(4.5, 0.0)1.7	91.0(8.1, 0.9)1.8	67.6(0.0,32.4)0.62
	0.01	95.2(2.7, 2.1)3.0	95.1(4.9, 0.0)1.6	92.3(6.0, 1.7)1.8	57.5(0.0,42.5)0.51
	0.05	95.9(0.7, 3.4)2.6	96.5(3.5, 0.0)1.5	94.9(2.9, 2.2)1.9	49.3(0.0,50.7)0.39
12/20/100	0	95.1(1.6, 3.3)0.81	90.5(1.6, 7.9)1.1	97.6(1.3, 1.1)0.83	72.6(0.0,27.4)0.37
	0.01	94.0(1.2, 4.8)0.79	90.5(2.6, 6.9)0.99	96.0(2.0, 2.0)0.83	71.3(0.0,28.7)0.37
	0.05	92.8(0.7, 6.5)0.72	87.6(2.3,10.1)0.98	96.9(1.0, 2.1)0.83	64.5(0.2,35.3)0.33
12/20/200	0	84.8(14.5, 0.7)0.98	92.2(3.8, 4.0)1.1	93.5(5.3, 1.2)0.86	81.7(0.7,17.6)0.49
	0.01	96.5(0.9, 2.6)0.82	92.6(2.4, 5.0)1.0	95.7(1.5, 2.8)0.82	73.9(0.3,25.8)0.38
	0.05	92.1(0.4, 7.5)0.71	87.7(1.8,10.5)0.99	96.9(0.9, 2.2)0.83	63.8(0.1,36.1)0.32
12/40/100	0	83.4(16.0, 0.6)1.0	93.3(3.5, 3.2)1.1	94.2(4.7, 1.1)0.86	84.7(0.3,15.0)0.53
	0.01	95.3(3.2, 1.5)0.88	91.5(3.2, 5.3)1.0	95.2(3.1, 1.7)0.84	78.0(0.0,22.0)0.43
	0.05	95.7(0.4, 3.9)0.75	90.5(1.5, 8.0)0.99	97.5(0.5, 2.0)0.83	69.1(0.0,30.9)0.34
20/20/100	0	95.3(1.4, 3.3)0.57	87.8(1.3,10.9)0.76	97.1(1.0, 1.9)0.62	75.3(0.1,24.6)0.32
	0.01	94.8(0.8, 4.4)0.56	88.2(1.0,10.8)0.72	97.1(0.8, 2.1)0.62	75.8(0.0,24.2)0.33
	0.05	92.3(0.1, 7.6)0.49	83.9(1.4,14.7)0.72	96.9(0.5, 2.6)0.62	69.2(0.0,30.8)0.28
20/20/200	0	81.1(18.5, 0.4)0.71	92.7(3.2, 4.1)0.80	93.3(5.3, 1.4)0.65	88.3(0.9,10.8)0.44
	0.01	93.8(1.0, 5.2)0.59	89.5(1.6, 8.9)0.72	96.6(1.2, 2.2)0.62	78.9(0.3,20.8)0.35
	0.05	91.3(0.3, 8.4)0.49	83.4(1.3,15.3)0.72	97.0(0.6, 2.4)0.62	68.3(0.0,31.7)0.28
20/40/100	0	78.3(21.7, 0.0)0.74	92.9(3.5, 3.6)0.79	93.8(4.9, 1.3)0.65	90.2(0.9, 8.9)0.47
	0.01	94.4(3.6, 2.0)0.64	90.0(3.6, 6.4)0.73	94.3(3.1, 2.6)0.63	83.3(0.4,16.3)0.39
	0.05	91.9(0.6, 7.5)0.51	83.6(1.3,15.1)0.73	96.4(0.8, 2.8)0.62	69.0(0.1,30.9)0.29
40/20/100	0	94.8(1.0, 4.2)0.39	82.5(0.6,16.9)0.50	96.6(0.4, 3.0)0.44	77.4(0.1,22.5)0.27
	0.01	93.0(0.3, 6.7)0.36	81.2(0.3,18.5)0.47	96.9(0.1, 3.0)0.43	74.4(0.0,25.6)0.25
	0.05	86.3(0.0,13.7)0.31	72.8(0.3,26.9)0.47	96.3(0.1, 3.6)0.43	66.0(0.0,34.0)0.22
40/20/200	0	70.2(29.8, 0.0)0.50	93.1(2.8, 4.1)0.54	95.4(3.5, 1.1)0.46	92.2(1.1, 6.7)0.37
	0.01	93.3(0.6, 6.1)0.39	83.3(1.3,15.4)0.48	96.3(0.9, 2.8)0.44	78.5(0.2,21.3)0.28
	0.05	84.8(0.1,15.1)0.32	72.5(0.2,27.3)0.47	97.0(0.1, 2.9)0.43	65.0(0.0,35.0)0.22
40/40/100	0	67.6(32.4, 0.0)0.50	93.0(3.0, 4.0)0.54	94.2(4.3, 1.5)0.46	92.1(1.1, 6.8)0.38
	0.01	95.8(2.6, 1.6)0.45	89.6(1.6, 8.8)0.49	96.8(1.1, 2.1)0.45	86.4(0.3,13.3)0.33
	0.05	87.7(0.4,11.9)0.34	75.9(0.8,23.3)0.47	96.6(0.4, 3.0)0.43	70.8(0.2,29.0)0.24

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.45: Confidence interval for H_a with 'moderate' heterogeneity, control group disease rates $r_B = \{0.35, 0.45, 0.50, 0.55\}$ for MOVER, Q distribution, test-based, based on τ_c^2, nonparametric bootstrap based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ(truncated), and odds ratio $\psi = 0.7$

$k/N_j/m$	ρ	MOVER/τ_c^2 C(L,R)W	Q C(L,R)W	Test-Based C(L,R)W	NB C(L,R)W
4/20/100	0	95.6(1.4, 3.0)3.4	88.5(3.3, 8.2)1.7	92.5(4.0, 3.5)1.8	57.4(0.0,42.6)0.64
	0.01	95.6(0.3, 4.1)3.0	85.1(3.0,11.9)1.6	95.5(2.0, 2.5)1.9	47.5(0.0,52.5)0.51
	0.05	94.5(0.1, 5.4)2.7	83.7(1.6,14.7)1.5	95.6(1.1, 3.3)1.9	40.4(0.0,59.6)0.40
4/20/200	0	94.5(4.6, 0.9)4.2	91.2(5.4, 3.4)1.9	85.5(11.2, 3.3)1.9	74.0(0.0,26.0)0.93
	0.01	95.7(0.4, 3.9)3.2	85.4(3.8,10.8)1.6	92.4(3.5, 4.1)1.9	51.1(0.0,48.9)0.56
	0.05	95.4(0.3, 4.3)2.7	84.0(2.4,13.6)1.5	94.3(1.5, 4.2)1.8	39.2(0.0,60.8)0.39
4/40/100	0	95.0(4.0, 1.0)4.2	91.0(5.8, 3.2)1.9	86.6(10.6, 2.8)1.9	74.1(0.0,25.9)0.92
	0.01	96.6(1.7, 1.7)3.5	89.5(3.6, 6.9)1.7	91.1(4.9, 4.0)1.8	59.7(0.0,40.3)0.66
	0.05	94.5(0.2, 5.3)2.8	84.8(3.2,12.0)1.5	93.9(2.0, 4.1)1.9	42.9(0.0,57.1)0.45
12/20/100	0	95.4(0.2, 4.4)0.97	79.6(1.2,19.2)1.1	95.1(0.9, 4.0)0.85	70.4(0.1,29.5)0.50
	0.01	93.7(0.1, 6.2)0.88	71.7(1.6,26.7)1.0	94.1(0.9, 5.0)0.84	61.0(0.0,39.0)0.43
	0.05	86.0(0.2,13.8)0.73	60.9(0.5,38.6)0.99	93.1(0.2, 6.7)0.82	48.7(0.0,51.3)0.33
12/20/200	0	88.2(11.4, 0.4)1.3	90.3(5.5, 4.2)1.2	88.6(8.5, 2.9)0.93	89.5(0.9, 9.6)0.75
	0.01	94.7(0.9, 4.4)0.95	77.2(2.0,20.8)1.1	94.0(1.5, 4.5)0.85	69.4(0.0,30.6)0.49
	0.05	88.0(0.0,12.0)0.75	62.7(0.6,36.7)0.99	93.5(0.2, 6.3)0.83	50.8(0.0,49.2)0.35
12/40/100	0	89.4(10.2, 0.4)1.3	90.6(4.8, 4.6)1.2	90.1(8.1, 1.8)0.94	90.1(0.2, 9.7)0.76
	0.01	96.0(1.9, 2.1)1.1	83.2(3.1,13.7)1.1	92.7(2.8, 4.5)0.87	77.2(0.2,22.6)0.58
	0.05	90.7(0.1, 9.2)0.82	66.9(1.4,31.7)1.0	94.3(0.4, 5.3)0.84	55.6(0.0,44.4)0.40
20/20/100	0	94.6(0.1, 5.3)0.70	82.4(0.4,17.2)0.82	92.9(0.2, 6.9)0.64	72.4(0.0,27.6)0.44
	0.01	89.2(0.2,10.6)0.62	72.1(0.5,27.4)0.74	91.9(0.3, 7.8)0.63	62.5(0.0,37.5)0.38
	0.05	80.7(0.0,19.3)0.52	57.0(0.1,42.9)0.72	91.6(0.0, 8.4)0.62	48.7(0.0,51.3)0.31
20/20/200	0	84.7(15.0, 0.3)0.90	92.2(5.1, 2.7)0.90	90.5(7.9, 1.6)0.72	92.6(1.5, 5.9)0.65
	0.01	94.0(0.5, 5.5)0.68	76.3(1.5,22.2)0.74	92.8(0.7, 6.5)0.64	70.0(0.0,30.0)0.44
	0.05	80.0(0.0,20.0)0.52	53.4(0.2,46.4)0.72	91.1(0.0, 8.9)0.62	44.6(0.0,55.4)0.30
20/40/100	0	86.7(13.0, 0.3)0.91	92.8(4.6, 2.6)0.88	91.9(7.3, 0.8)0.72	93.2(1.4, 5.4)0.66
	0.01	94.8(2.5, 2.7)0.78	84.1(3.0,12.9)0.77	93.5(2.2, 4.3)0.67	79.9(0.3,19.8)0.52
	0.05	83.7(0.1,16.2)0.57	65.1(0.4,34.5)0.73	91.3(0.1, 8.6)0.62	55.4(0.0,44.6)0.34
40/20/100	0	90.8(0.0, 9.2)0.49	85.5(0.1,14.4)0.54	77.9(0.0,22.1)0.46	72.9(0.0,27.1)0.38
	0.01	82.6(0.0,17.4)0.43	65.0(0.2,34.8)0.49	65.8(0.0,34.2)0.45	58.1(0.0,41.9)0.32
	0.05	59.1(0.0,40.9)0.33	27.6(0.0,72.4)0.47	36.9(0.0,63.1)0.43	32.0(0.0,68.0)0.24
40/20/200	0	78.2(21.8, 0.0)0.60	94.7(5.1, 0.2)0.59	89.8(8.7, 1.5)0.51	95.5(2.1, 2.4)0.52
	0.01	90.9(0.0, 9.1)0.47	77.1(0.1,22.8)0.51	73.8(0.0,26.2)0.45	65.5(0.0,34.5)0.35
	0.05	61.4(0.1,38.5)0.34	30.1(0.1,69.8)0.47	40.0(0.1,59.9)0.43	33.3(0.0,66.7)0.24
40/40/100	0	72.6(27.4, 0.0)0.61	93.0(6.7, 0.3)0.58	88.2(11.0, 0.8)0.51	94.4(3.8, 1.8)0.53
	0.01	94.2(2.8, 3.0)0.53	91.1(1.4, 7.5)0.53	87.3(0.8,11.9)0.47	83.7(0.1,16.2)0.43
	0.05	75.5(0.0,24.5)0.39	46.0(0.0,54.0)0.48	52.0(0.0,48.0)0.43	44.7(0.0,55.3)0.28

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Table 6.46: Confidence interval for H_a with 'high' heterogeneity, control group disease rates $r_B = \{0.35, 0.45, 0.50, 0.55\}$ for MOVER, Q distribution, test-based, based on τ_c^2, nonparametric bootstrap based on 1000 simulations of k trials each with N_j clusters per group of size m, intracluster correlation ρ(truncated), and odds ratio $\psi = 0.7$

$k/N_j/m$	ρ	MOVER/τ_c^2 C(L,R)W	Q C(L,R)W	Test-Based C(L,R)W	NB C(L,R)W
4/20/100	0	97.0(0.4, 2.6)3.7	76.0(2.9,21.1)1.8	92.3(3.1, 4.6)1.9	55.2(0.0,44.8)0.74
	0.01	95.3(0.2, 4.5)3.3	67.7(1.6,30.7)1.6	94.3(1.0, 4.7)1.9	47.5(0.0,52.5)0.61
	0.05	93.7(0.1, 6.2)2.7	58.1(2.1,39.8)1.5	93.5(0.7, 5.8)1.8	32.9(0.0,67.1)0.42
4/20/200	0	96.4(2.7, 0.9)4.8	88.6(4.2, 7.2)2.1	86.1(10.7, 3.2)2.0	79.2(0.0,20.8)1.2
	0.01	96.2(0.3, 3.5)3.5	71.3(2.4,26.3)1.7	93.8(2.0, 4.2)1.9	51.0(0.0,49.0)0.67
	0.05	93.3(0.0, 6.7)2.6	56.7(0.5,42.8)1.5	92.8(0.4, 6.8)1.9	30.2(0.0,69.8)0.38
4/40/100	0	96.6(3.0, 0.4)4.8	86.8(5.0, 8.2)2.1	87.0(9.6, 3.4)2.0	78.5(0.0,21.5)1.2
	0.01	96.8(0.9, 2.3)3.9	77.0(4.5,18.5)1.8	91.6(5.1, 3.3)1.9	62.0(0.0,38.0)0.83
	0.05	94.6(0.1, 5.3)3.0	64.0(1.7,34.3)1.6	92.8(1.0, 6.2)1.8	39.1(0.0,60.9)0.49
12/20/100	0	97.3(0.1, 2.6)1.2	86.4(1.2,12.4)1.2	91.9(0.7, 7.4)0.89	72.4(0.0,27.6)0.65
	0.01	92.4(0.0, 7.6)0.98	73.8(0.6,25.6)1.1	89.1(0.3,10.6)0.85	56.1(0.0,43.9)0.51
	0.05	82.2(0.0,17.8)0.81	46.2(0.4,53.4)0.99	88.7(0.0,11.3)0.83	39.9(0.0,60.1)0.39
12/20/200	0	93.4(6.3, 0.3)1.5	92.9(4.2, 2.9)1.3	89.7(8.7, 1.6)1.0	92.2(0.7, 7.1)0.92
	0.01	94.0(0.3, 5.7)1.1	80.3(1.3,18.4)1.1	90.5(0.8, 8.7)0.88	66.6(0.1,33.3)0.60
	0.05	82.1(0.0,17.9)0.80	47.4(0.1,52.5)0.99	89.2(0.0,10.8)0.83	40.1(0.0,59.9)0.38
12/40/100	0	92.2(7.7, 0.1)1.5	91.7(6.0, 2.3)1.3	88.0(10.7, 1.3)1.0	93.1(0.6, 6.3)0.93
	0.01	96.3(0.6, 3.1)1.2	87.3(2.4,10.3)1.1	92.8(2.3, 4.9)0.92	78.0(0.1,21.9)0.72
	0.05	88.1(0.0,11.9)0.89	63.0(0.2,36.8)1.0	89.9(0.0,10.1)0.84	46.8(0.0,53.2)0.45
20/20/100	0	96.5(0.1, 3.4)0.82	92.1(0.3, 7.6)0.86	82.5(0.1,17.4)0.68	70.8(0.0,29.2)0.57
	0.01	86.3(0.0,13.7)0.70	62.7(0.2,37.1)0.75	63.3(0.1,36.6)0.65	53.1(0.0,46.9)0.46
	0.05	66.9(0.0,33.1)0.54	26.0(0.0,74.0)0.72	39.1(0.0,60.9)0.62	28.0(0.0,72.0)0.32
20/20/200	0	91.3(8.7, 0.0)1.0	94.5(5.2, 0.3)0.92	90.1(8.8, 1.1)0.77	95.1(1.6, 3.3)0.77
	0.01	92.2(0.1, 7.7)0.77	79.2(0.2,20.6)0.78	74.9(0.2,24.9)0.67	64.7(0.0,35.3)0.51
	0.05	68.3(0.0,31.7)0.56	30.1(0.1,69.8)0.72	42.3(0.1,57.6)0.63	33.1(0.0,66.9)0.34
20/40/100	0	88.8(11.1, 0.1)1.0	91.0(8.5, 0.5)0.91	85.0(13.1, 1.9)0.78	93.5(2.9, 3.6)0.79
	0.01	97.1(0.6, 2.3)0.87	92.2(2.3, 5.5)0.83	86.2(1.5,12.3)0.70	78.9(0.2,20.9)0.62
	0.05	76.5(0.0,23.5)0.62	45.5(0.1,54.4)0.74	53.8(0.1,46.1)0.64	42.9(0.0,57.1)0.39
40/20/100	0	91.8(0.0, 8.2)0.56	79.5(0.0,20.5)0.58	76.7(0.0,23.3)0.48	66.7(0.0,33.3)0.45
	0.01	71.2(0.0,28.8)0.49	41.0(0.0,59.0)0.51	50.1(0.0,49.9)0.46	42.6(0.0,57.4)0.38
	0.05	36.3(0.0,63.7)0.36	7.70(0.0,92.3)0.47	19.9(0.0,80.1)0.44	14.8(0.0,85.2)0.26
40/20/200	0	84.9(15.0, 0.1)0.68	92.7(7.2, 0.1)0.61	87.0(12.4, 0.6)0.54	94.7(4.2, 1.1)0.57
	0.01	86.1(0.0,13.9)0.54	62.9(0.1,37.0)0.53	68.3(0.0,31.7)0.47	58.2(0.0,41.8)0.43
	0.05	36.2(0.0,63.8)0.37	8.00(0.0,92.0)0.47	19.3(0.0,80.7)0.43	13.9(0.0,86.1)0.27
40/40/100	0	83.8(16.2, 0.0)0.69	91.0(9.0, 0.0)0.60	85.0(14.9, 0.1)0.55	94.3(5.2, 0.5)0.58
	0.01	95.5(0.2, 4.3)0.59	89.3(0.5,10.2)0.55	87.6(0.3,12.1)0.50	80.8(0.0,19.2)0.49
	0.05	54.1(0.0,45.9)0.43	20.2(0.0,79.8)0.49	33.2(0.0,66.8)0.45	25.8(0.0,74.2)0.32

C means the probability coverage. L and R denote tail errors from left and right in percent, respectively. W denotes the average interval width.

Chapter 7

Meta-analysis of practice-based secondary prevention programs for patients with heart disease risk factors

7.1 Introduction

The aim of this chapter is to illustrate the analytic methods described in a meta-analysis of four cluster randomization trials. These trials were conducted to compare two or more interventions to reduce coronary heart disease (CHD) risk factors in primary care, where the unit of randomization in each trial was at the practice level. The trials include the Assessment of Implementation Strategies Trial (ASSIST) (Moher et al., 2001), the Diabetes Care from Diagnosis study (Woodcock et al., 1999), the Hypertension Decision Support study (Montgomery et al., 2000) and the Southampton Heart Integrated Care Project (SHIP) (Jolly et al., 1999). The outline of this chapter is as follows: the four trials are described in Section 7.2, and the methods of analysis and the results described in Sections 7.3 and 7.4, respectively. Results are summarized and compared to related

meta-analyses in Section 7.5.

7.2 Aspects of Study Data

Table 7.1: Description of studies.

Study	Type of patients	Intervention	Type of study design
ASSIST (Moher et al., 2001)	CHD	Recall to a general practitioner	stratified
Diabetes Care from Diagnosis (Woodcock et al., 1999)	Diabetes	Trained practitioners and nurses	stratified
Hypertension Decision Support (Montgomery et al., 2000)	Hypertension	Decision support system + risk chart	completely
SHIP (Jolly et al., 1999)	MI+Angina	Specialist cardiac liaison nurses	stratified

The four cluster randomization trials were included in a meta-analysis of intracluster correlation coefficients involving 31 primary care cluster randomization trials (Adams et al., 2004). All trials enrolled 10 or more practices and were conducted in English-speaking countries or Northern Europe. The four studies were selected for the purposes of conducting a meta-analysis to investigate secondary prevention programs for patients with heart disease risk factors using the presence or absence of hypertension at one year as the endpoint. Hypertension is defined as having systolic blood pressure exceeding 140 mm Hg, or having diastolic blood pressure exceeding 90 mm Hg (Chobanian et al., 2003). These four trials were conducted in England with individual patient data available with different study designs. For illustrative purposes, they were assumed completely randomized.

In ASSIST, the study objective was to compare three different interventions of care delivery for secondary prevention of coronary heart disease (CHD) in primary care: audit and feedback; recall to a general practitioner; and recall to a nurse clinic (Moher et al., 2001). For our analyses, attention was limited to two intervention groups: audit and feedback (control) vs. recall to a general practitioner (experimental). For the control group, practices provided usual care, while for the experimental group, each practice developed a disease register and recall system for regular review.

The Diabetes Care from Diagnosis study investigated the effects of patient-centered training for general practitioners and nurses who cared for Type 2 diabetes patients who were diagnosed over a 1 year period (Woodcock et al., 1999). General practitioners and nurses in the experimental group received training sessions to recognize and practice skills of patient-centered consulting, which were not provided to general practitioners and nurses in the control group. At 6 and 12 months into patient recruitment, the nurses in the experimental group met with the trainer for group support, and reviewed recruitment with the research team. The nurses in the control group discussed the recruitment and the use of British Diabetic Association materials with the research team.

The Hypertension Decision Support study was designed to evaluate a computer-based clinical decision support system for patients with high blood pressure (Montgomery et al., 2000). Medical practices were randomly assigned to a computer-based clinical decision support system plus cardiovascular risk chart (which gives identical information about risk); risk chart alone; or usual care (no information given about cardiovascular risk). Again, attention was limited to two intervention groups: computer-based clinical decision support system plus cardiovascular risk chart vs. usual care.

The SHIP was designed to assess the effectiveness of a program for coordinating and supporting follow-up care in general practice among myocardial infraction (MI) or angina

patients discharged from hospitals as compared to the usual care without any such support (Jolly et al., 1999). More specifically, the experimental group was led by specialist cardiac liaison nurses who contacted practices at the time of discharge to discuss future care and to book the first follow-up visit.

Table 7.1 describes the study design for the four cluster randomization trials included in the meta-analysis. Baseline characteristics of patients by intervention group for each trial are given in Table 7.2. Most of the patients had either CHD or hypertension, except for the one trial with diabetes patients. The number of practices was similar across the intervention groups but somewhat different across trials. Mean age of the patients ranged from 57 to 70 years. Specifically, the Diabetes Care from Diagnosis trial had the lowest mean age and the Hypertension Decision Support trial had the highest mean age. Men constituted more than 50% of the subjects in the ASSIST and the SHIP, while women constituted more than 50% of the subjects in the other two trials.

7.3 Method of analysis

We performed analyses by using SAS 9.2 (SAS Institute, Inc, Cary, NC). Given the individual patient data, the analysis for each trial was conducted using generalized estimating equation (GEE) to account for clustering by incorporating robust standard errors using an exchangeable working correlation matrix for patients within the same primary care practice. The methods of GEE were carried out with the procedure PROC GENMOD. The summary odds ratios were computed by using both the fixed and random effects models. The intracluster correlation coefficients for the four trials were calculated using the 'analysis of variance' method (Donner and Klar, 2000, p9) with the procedure PROC GLM, where negative values were truncated to zero. The approaches to addressing heterogeneity included the adjusted Q statistic, the heterogeneity variance estimators with

Table 7.2: Baseline characteristics of patients in intervention groups for each trial included in the meta-analysis. Age is in years.

	Experimental	**Control**
ASSIST (Moher et al., 2001)		
Practices	7	7
N	682	559
Mean age (SD)	66.2(5.4)	66.4(5.6)
Men (%)	457(67)	373(67)
Diabetes Care from Diagnosis (Woodcock et al., 1999)		
Practices	20	20
N	142	108
Mean age (SD)	57.9(9.6)	57.3(9.6)
Men (%)	59(42)	43(40)
Hypertension Decision Support (Montgomery et al., 2000)		
Practices	10	7
N	229	157
Mean age (SD)	70.6(5.5)	70.5(5.3)
Men (%)	106(46)	80(51)
SHIP (Jolly et al., 1999)		
Practices	33	34
N	277	320
Mean age (SD)	63.2(10.1)	64.1(10.3)
Men (%)	189(68)	237(74)

corresponding confidence intervals and measures of heterogeneity, also with corresponding confidence intervals (see Chapters 2-4). The complete list of the approaches is given in Table 5.1. The DerSimonian and Laird estimator was used to estimate τ_c^2 unless otherwise specified. Although some approaches performed better than others based on the simulation results, all the approaches were applied to the meta-analysis for the purpose of illustration.

As part of a planned sensitivity analyses, the summary odds ratios, the adjusted Q statistic and the adjusted I^2 were recalculated for the meta-analysis excluding the Diabetes Care intervention from the Diagnosis trial, which enrolled different type of patients as compared to the other three trials.

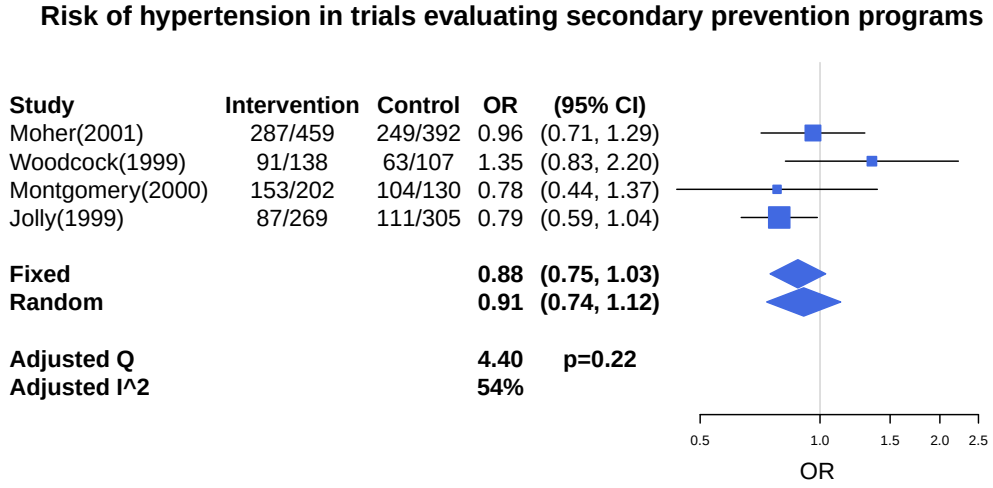

Figure 7.1: Forest plot for the meta-analysis of practice-based secondary prevention programs for patients with coronary heart disease risk factors.

7.4 Results

Three of the four trials reported a risk reduction for hypertension (experimental vs control); however, the reduction in each trial was not statistically significant as the corresponding confidence intervals included 1.0 (Figure 7.1). The summary odds ratios for hypertension for the fixed and random effects model were 0.88 (95%CI, 0.75-1.03) and 0.91 (95%CI, 0.74-1.12), respectively, combining data from all four trials. The adjusted Q statistic was 5.43 ($p = 0.14$) and the adjusted I^2 was 54%. The intracluster correlation cofficients were 0.003, -0.024, 0.013 and -0.028, corresponding to the ASSIST, the Diabetes Care from Diagnosis study, the Hypertension Decision Support study and the SHIP, respectively. The two negative values were truncated to zero.

The heterogeneity variance estimates ranged from $\hat{\tau}^2_{c.ML} = 0$ to $\hat{\tau}^2_{c.MV} = 0.033$ (Table 7.3), indicating a 'small' degree of heterogeneity. Although the point estimates of the overall intervention effect did not differ greatly across the eight values of τ^2_c, it is noted that the

slight differences among the eight estimators led to fairly large differences among the interval widths. This suggests that the choice of the estimator could lead to conflicting conclusions regarding the true overall intervention effect. However, in this case, all eight 95% confidence intervals for the overall intervention effect included 1, indicating that the reduction in the risk of having hypertension is not significant. The confidence intervals for τ_c^2 obtained with the various approaches discussed earlier are also given in Table 7.4. All the confidence intervals except the Sidik-Jonkman confidence interval includes zero, suggesting that heterogeneity in the meta-analysis is not significant. It is also noted that the precision tends to vary greatly across the different approaches.

Finally, the measures of heterogeneity were $H_a = 1.47$, $R_a = 1.43$ and $I_a^2 = 54\%$, indicating a 'moderate' degree of heterogeneity based on the guideline described in Table 4.4. However, the 'moderate' degree of heterogeneity was not considered statistically significant given that all the confidence intervals for H_a^2 included 1 (for a 'no' degree of heterogeneity) (Table 7.5).

For the meta-analysis of three cluster randomization trials omitting the Diabetes Care from Diagnosis study, the adjusted Q statistic was 1.17 with $p = 0.14$ and the adjusted I^2 was 0%. As a result, both summary odds ratios were identical, given by 0.84 (95%CI, 0.71-0.99), suggesting a statistically significant positive effect of secondary prevention programs on reducing hypertension.

7.5 Summary

Overall, the secondary prevention programs showed a risk reduction of approximately 9%-12% in hypertension for patients with CHD risk factors at 1 year follow-up but the reduction is not statistically significant. Clark et al. (2005) came to a different conclusion

Table 7.3: Heterogeneity variance estimators and random effects summary odds ratios .

Estimator	$\hat{\tau}_c^2$	OR(95% CI)
Variance component (VC)	0.020	0.91(0.82, 1.28)
DerSimonian and Laird (DL)	0.015	0.91(0.82, 1.25)
Two-step DL (DLVC)	0.016	0.91(0.82, 1.26)
Two-step DL (DL2)	0.016	0.91(0.82, 1.26)
Model error variance (MV)	0.033	0.92(0.80, 1.33)
Improved model error variance (MVVC)	0.019	0.91(0.82, 1.27)
Maximum likelihood (ML)	0.000	0.88(0.85, 1.17)
Restricted maximum likelihood (REML)	0.013	0.91(0.83, 1.24)

Table 7.4: Point estimates and confidence intervals for τ_c^2.

Confidence interval	$\hat{\tau}_c^2$	(95%CI)
Q profile	$\hat{\tau}^2_{c.DL} = 0.015$	(0.00, 0.85)
Biggerstaff-Tweedie	$\hat{\tau}^2_{c.DL} = 0.015$	(0.00, 1.31)
Profile likelihood (ML)	$\hat{\tau}^2_{c.ML} = 0.000$	(0.00, 0.17)
Profile likelihood (REML)	$\hat{\tau}^2_{c.RE} = 0.013$	(0.00, 1.47)
Wald-type (ML)	$\hat{\tau}^2_{c.ML} = 0.000$	(0.00, 0.03)
Wald-type (REML)	$\hat{\tau}^2_{c.RE} = 0.013$	(0.00, 0.08)
Sidik-Jonkman	$\hat{\tau}^2_{c.MV} = 0.033$	(0.01, 0.46)
Nonparametric bootstraps	$\hat{\tau}^2_{c.DL} = 0.015$	(0.00, 0.06)

that secondary prevention programs generally had a significant positive effect in processes of care for all-cause mortality. Several factors may explain the difference. First, only four trials were included in the meta-analysis, while the smallest meta-analysis reported by Clark et al. (2005) included data from eight trials with the secondary prevention programs solely exercise-based. Furthermore, the values of I^2 reported in the meta-analyses by Clark et al. (2005) were mostly zero as compared to 54% with our meta-analysis. Most of all, our meta-analysis included the trial with diabetes patients, which are a different type of patient, and had the opposite direction in intervention effect compared to the other trials. The inclusion of this trial may be the source of heterogeneity for the results of the sensitivity analysis excluding the trial with diabtetes patients showed that the adjusted I^2 statistic was reduced to zero and the risk reduction for hypertension became statistically significant. For illustrative purposes, the main analysis focused on the meta-anlaysis with all four trials.

Table 7.5: Confidence intervals for H_a.

Confidence interval for $H_a = 1.47$	(95%CI)
MOVER	(1.00, 5.84)
Based on distribution of Q_a	(0.00, 2.15)
Test-based	(0.79, 2.21)
Based on τ_c^2	(1.00, 5.31)
Nonparametric bootstraps	(1.00, 1.86)

The adjusted Q statistic was not statistically significant, indicating no significant differences among the four estimated odds ratios. It is noted that the highest power of the adjusted Q statistic for a meta-analysis with $k = 4$ is generally less than 30% (Table 6.3-6.6) unless the heterogeneity is considerably large ($I_a^2 \geq 78\%$).

All the heterogeneity variance estimators suggest a 'small' degree of heterogeneity but at a level considered not statistically significant with a 95% confidence interval including 0 (for a 'no' degree of heterogeneity). The results show that the overall intervention effect is relatively insensitive to changes in τ_c^2 but these small changes may lead to fairly large difference in confidence interval widths for the overall intervention effect (Sidik and Jonkman, 2007; Viechtbauer, 2007*a*). In our example, all the heterogeneity variance estimators came to a similar conclusion due to small values, but the two-step estimators and the REML estimator performed better than the other estimators with relatively low bias for $k = 4$. As for the confidence interval approaches for τ_c^2, the Q profile confidence interval generally performed relatively well, while other confidence intervals were either overly liberal (Sidik-Jonkman and bootstraps) or overly conservative (Biggerstaff-Tweedie, profile likelihood and Wald-type) (Table 6.19-6.22).

Finally, the measures of heterogeneity indicated that there was a 'moderate' degree of heterogeneity but which was not statistically significant, with a 95% confidence interval including 1.0 (for a 'no' degree of heterogeneity). Our simulation results in Chapter 6

suggested that caution must be taken when interpreting the measures of heterogeneity for a small number of trials ($k = 4$), which may result in a relatively large bias of approximately 0.15 (Table 6.35 to 6.38). Similarly, the test-based confidence interval and the confidence interval based on the distribution of the Q statistic performed relatively well for $k = 4$ based on our simulation results shown in the previous chapter (Table 6.43-6.46).

Overall, there is no substantial heterogeneity was found applying the proposed approaches.

Chapter 8

Conclusions

8.1 Introduction

The primary objective of this thesis was to to develop and evaluation methods that identify and quantify heterogeneity in a meta-analysis of cluster randomization trials assuming a fixed effects model. The possible approaches included the adjusted Q statistic, heterogeneity variance estimators with their corresponding confidence intervals and finally measures of heterogeneity with their corresponding confidence intervals. The discussion was limited to completely randomized cluster randomized trials having binary outcomes measured by the odds ratio comparing an experimental to a control intervention. The aim of this final chapter is to summarize main results in Section 8.2, identify potential limitations and propose areas of future research in Section 8.3.

8.2 Summary

8.2.1 Key findings

The different forms of the Q statistic were compared for testing heterogeneity in a meta-analysis of cluster randomization trials: the unadjusted Q statistic, the adjusted Q statistic with truncated ANOVA-based ρ, and the adjusted Q statistic omitting truncation. It was clearly seen that the unadjusted Q statistic resulted in severely inflated Type I error for clustered data. For example, the observed Type I error was close to 100% at $\rho = 0.05$ for all the parameter combinations with $k \geq 12$. In contrast, the adjusted Q statistic with truncated ρ showed generally satisfactory Type I error, except for $\rho = 0$ where the Type I error was overly conservative. On the other hand, the adjusted Q statistic omitting truncation of ρ maintained Type I error at nominal level throughout all parameter combinations.

Although the adjusted Q statistic is simple to calculate and has satisfactory Type I error, its power raises a concern because it depends heavily on the number of trials. In this case, power analysis before conducting the meta-analysis of cluster randomization trials is useful to ensure the validity of the test. Our results showed that the power calculated using the derived formula was similar to the power obtained from the simulation for all the parameter combinations investigated. An increase in the power of the adjusted Q statistic may be obtained by increasing the number of trials, overall sample size per trial (i.e. $n \times m$), or degree of heterogeneity. However, the power reduces dramatically for a small increase in the values of the intracluster correlation coefficient. In addition, for a fixed sample size, the power of the adjusted Q statistic is greater for a large number of small clusters than for a small number of large clusters. Based on the results, a meta-analysis with at least 12 trials for $\rho = 0$ and 40 trials for $\rho = 0.01$ is sufficiently large to detect 'high' heterogeneity in order to achieve a desired power of approximately 80%.

Eight heterogeneity variance estimators adjusted for clustering were compared, including the four noniterative estimators VC, DL, MV and MVVC; the two two-step estimators DLVC and DL2 and finally the two iterative estimators ML and REML. The simulation results indicated that the MVVC estimator for 'no' to 'low' heterogeneity and the DLVC estimator for 'moderate' to 'high' heterogeneity had the lowest bias as compared to other estimators, followed by the ML and REML estimators. These results are consistent with the conclusion presented by Viechtbauer (2007*a*) and also complement their findings, which focused on the meta-analysis of individually randomized trials, with no consideration of the two-step estimators.

We also compared the eight confidence intervals for the adjusted heterogeneity variance estimators. These included the Q profile (QP), Biggerstaff-Tweedie (BT), ML profile likelihood (pML), REML profile likelihood (pRE), ML Wald-type (wML), REML Wald-type (wRE), Sidik and Jonkman (SJ) and nonparametric bootstraps (NB) confidence intervals, where only the Sidik and Jonkman confidence interval had a closed-form solution. The simulation results showed that the Q profile confidence interval had relatively satisfactory performance in terms of coverage, tail errors and interval width at least for 'low' to 'high' heterogeneity with small meta-analyses of large trials. According to Viechtbauer (2007*a*) based on the meta-analysis of individually randomized trials, the large coverage above the nominal for the Q profile confidence interval at $\tau_c^2 = 0$ may be resulted of having the asymptotic distribution used to construct the confidence interval other than the expected chi-square distribution with one degree of freedom. Other confidence interval approaches were either overly conservative (BT, pML, pRE, wML, wRE) or overly liberal (SJ and NB).

The adjusted measures of heterogeneity were H_a, R_a and I_a^2. The simulation results showed that the adjusted statistics were generally an accurate indicator of the degree of heterogeneity with a relatively large number of trials. However, the adjusted statistics had a relatively large bias of 0.15 when the number of trials was small, compromising

their interpretation.

It was also useful to compare the confidence intervals of the H_a statistic including the MOVER, the confidence interval based on the Q distribution, the test-based confidence interval, the confidence interval based on τ_c^2 and the nonparametric bootstraps confidence interval. According to the simulation results, it appears that the within study variance (the denominator for MOVER) had little impact on the MOVER given that the within study variance was relatively small as compared to the between study variance (the numerator for MOVER). Also, given that the Q profile approach was used for constructing the confidence interval for the denominator of MOVER and the confidence interval based on τ_c^2, they had similar performance. For 'low' to 'high' heterogeneity, the MOVER consistently maintained a nominal coverage level for small meta-analyses of large trials. Nevertheless, the confidence interval based on the Q distribution is preferred for 'no' heterogeneity.

8.2.2 Recommendations

It is apparent that the adjusted Q statistic is a reasonable choice for testing the heterogeneity of intervention effects obtained from cluster randomization trials given that the unadjusted Q statistic produces highly inflated Type I errors. However, it is known that the power of the adjusted Q statistic depends on the number of trials. This was shown by a derived algebraic formula for its power. As for the heterogeneity variance estimators, the REML estimator with consistently relatively low bias is recommended. Although this procedure requires an iterative scheme, several noniterative approaches are also available that show reasonable performance: the MVVC estimator for 'no' to 'low' and the two-step estimator DLVC for 'moderate' to 'high' heterogeneity. The REML profile likelihood can be used to construct the confidence interval of the REML estimator and the Q profile can be used to construct confidence intervals for the two-step estimator DLVC or the

MVVC estimator given relatively small meta-analyses with large trials. The measures of heterogeneity appear to be a consistent indicator of the degree of heterogeneity when the number of trials is relatively large. However, caution must be taken in interpreting the results with a small number of trials. In this case, confidence interval construction may be informative. For 'no' heterogeneity, the confidence interval based on the Q distribution is recommended. Otherwise, the MOVER approach, which had the similar performance as the profile Q confidence interval, is a reasonable choice to construct the confidence interval for H_a or I_a^2 given relatively small meta-analyses with large trials.

8.2.3 Practical issues

Meta-analysts must often select methods based on the form of the available data. According to Whitehead (2002), available data may be classified in three forms for individually randomized trials. A similar analogy will be applied to extend the forms of available data in the context of cluster randomized trials.

First, an estimate of the intervention effect for each trial and its corresponding standard error are the minimum information needed to apply the proposed approaches to assessing heterogeneity in the meta-analysis of cluster randomization trials. When the adjusted variance is not provided, the intraclass correlation coefficient and an average cluster size for each trial are needed to compute the inflation factor (IF) to account for clustering. According to Ivers et al. (2011), only 18% of the 300 manuscripts that they reviewed reported an estimated intracluster correlation coefficient. In this case, the missing intraclass correlation coefficient may be imputed by a common intraclass correlation coefficient extracted from other published papers reporting similar trials, although the risks of bias using this strategy is well-known. Also, in this form, the meta-analysis is limited to combine the studies with the same type of effect measures. For instance, a study with the mean difference as the measure of effect cannot be used in the meta-analysis limited

to binary outcomes.

The second form of available data consists of summary statistics for each intervention group, enabling a choice to be made between several different measures of the intervention effect. For binary data, one way is to record the number of events (A_{ijk}) and the cluster size for each cluster (m_{ijl}), which are sufficient for computing the odds ratio and the standard error (Equation 2.4) for each trial. Another approach is to record the disease rates (P_{ij}) for the control and experimental group, the average cluster size per intervention (m_{ij}) and the intraclass correlation coefficient. A summary of this notation can be found in Table 2.2.

The third form consists of individual patient data, allowing any measures of the intervention effect and method of estimation. In addition, if all the studies provide individual patient data, a more thorough analysis can be undertaken by employing a statistical modeling approach.

8.3 Limitations and future research

First, the focus of the thesis is limited to the meta-analysis of completely randomized cluster randomization trials. In practice, meta-analysts may encounter the challenge of combining cluster randomization trials using different designs, such as the stratified design or the matched-pair design. The approaches described here may be easily extended to meta-analyses of stratified cluster randomization trials by treating each stratum as a separate trial in the meta-analysis. As for the matched-pair designs, it will not be feasible to routinely calculate the intracluster correlation coefficient where the between-cluster variation is confounded with the intervention effect (Klar and Donner, 1997). Thus, one could conduct the meta-analyses separately for the completely randomized and for the matched-pair designs using standard techniques (Donner and Klar, 2002). Alternatively,

there is the option of ignoring the stratification for the stratified design or breaking the matches for the matched-pair design (Donner et al., 2007). However, this may lead to a loss in power if the stratification/matching is effective.

Second, the approaches presented here were developed specifically for the case of two intervention groups. However, many trials contain more than two intervention groups. For instance, two of the four cluster randomization trials in our example (ASSIST and SHIP) had three intervention groups but the third intervention group was discarded for the proposes of analysis. However, the approaches discussed may usefully be extended to incorporate the meta-analysis of cluster randomization trials with more than two intervention groups. A list of approaches for including multiple intervention groups from a given trial may be found in the Cochrane Handbook for Systematic Reviews of Interventions (Higgins and Green, 2008, Chapter 16).

Third, the binary outcome is assumed approximately normal on the log odds ratio scale. We would therefore expect that the conclusions for the binary outcomes found in this thesis might also be applied to normally distributed continuous outcome data, that yield standardized mean differences. However, the performance of these methods have not yet been determined for outcomes which do not follow approximately a normal distribution. Further study would be required to draw firm conclusions regarding continuous outcomes in the meta-analysis of cluster randomization trials, or to make recommendations for non-normally distributed effect measures.

Fourth, the proposed approaches are based on the assumption of fixed within study variances, where sampling errors in these variances are ignored. This issue arises particularly when the trials are small. Otherwise, it appears that this assumption would have little impact on the results (Bohning et al., 2002; Hardy and Thompson, 1996). For instance, the simulation results were similar when treating the within study variance as fixed when

applying the MOVER approach, as compared to the confidence interval based on τ_c^2 approach where the estimated within study variance was used. Also assuming a constant within study variance among trials in a meta-analysis tends to overestimate the statistical power of the adjusted Q statistic with varying within study variances. Similar results were found in a simulation study by Hardy and Thompson (1998).

Fifth, the assumption of a common intracluster correlation coefficient was used across all trials considered, by averaging the estimates as computed from the separate trials. This approach becomes less efficient when there is a substantial difference among the estimates of intracluster correlation coefficient across trials. In this case, a separate estimate of intracluster correlation coefficient for each trial may be used (Donner et al., 2001).

Sixth, the simulation study is limited to the data generated under a fixed effects model. Some researchers (Sidik and Jonkman, 2005; Viechtbauer, 2007*a*) generated data under the random effects assumption to allow more variability in intervention effects. It is expected that the decision to model as fixed effects in our study may result in greater statistical power; whereas the random effects model tends to lead to a loss in power.

Seventh, the simulation results presented are necessarily limited in scope in order to understand the performance of the approaches under simple scenarios.

For example, attention was restricted to an equal number of clusters with an equal number of subjects per intervention group. However, there is often considerable variation in both the number of clusters and cluster sizes in practice. An equal number of clusters per intervention group generally leads to an increase in efficiency as compared to unequal allocation (Donner and Klar, 2000, p.59). In the case of unequal cluster sizes, if the cluster size is replaced by its average, a slight underestimation in power would be expected. Another option is to use a more conservative approach by replacing the cluster size by its

maximum to provide some protection for statistical power (Donner and Klar, 2000, p.57).

The conclusion drawn here based on simulation results for trials with 20 or 40 clusters may not apply to trials with a fairly small number of clusters (10 or less) since the large sample approximation underlying these approaches may be questionable (Donner and Klar, 2000, p.100). Further study may be helpful to broaden our findings to more general settings by generating data under the random effects assumptions and considering unequal allocation with unbalanced cluster size.

Eighth, we recognize that the focus of this thesis has been on analytic methods used to identify the degree of heterogeneity. Thus, when substantial heterogeneity is detected, further study would be required to apply subgroup analysis or meta-regression to investigate the source of heterogeneity in a meta-analysis of cluster randomization trials (Higgins et al., 2002*b*; Rotondi and Khobzia, 2010).

Finally, the discussion is limited to the meta-analysis of only cluster randomization trials. However, meta-analysts may encounter the challenge of combining the results from both individually randomized and cluster randomized trials. In this case, the proposed approaches can be easily applied by setting the values of intracluster correlation coefficient equal to zero for the individually randomized trials (Darlington and Donner, 2007).

Appendix A

Derivation of Q statistic

Let $\theta_1, \ldots, \theta_k$ be the log odd ratios from k trials which are considered to be a random sample from a normal distribution of trials with mean θ and within study variance $\sigma_j^2 = w_j^{-1}$ for trial j, $j = 1, \ldots, k$. Then the likelihood function is

$$L(\theta) = \prod_{j=1}^{k} \left(\frac{w_j}{2\pi}\right)^{1/2} e^{-(1/2)w_j(\theta_j - \theta)^2}$$

and the log likelihood function is

$$lnL(\theta) = -\frac{k}{2} ln 2\pi + \frac{1}{2}\sum_{j=1}^{k} lnw_j - \frac{1}{2}\sum_{j=1}^{k} w_j(\theta_j - \theta)^2$$

The Q statistic (Cochran, 1954) tests the null hypothesis: $H_o : \theta_1 = \theta_2 = \ldots = \theta_k = \theta$ versus the alternative H_A: at least one trial had a truly different intervention effect as compared to the other trials, where θ denotes the common intervention effect. To calculate the likelihood ratio test (LRT) denoted by $\lambda(\theta)$, the maximum likelihood estimator (MLE) of θ under H_o and H_A must be determined by solving

$$\frac{dlnL(\theta)}{d\theta} = \sum_{j=1}^{k} w_j(\theta_j - \theta) = 0$$

Therefore, the MLE under H_o is θ_o and the MLE under H_A is

$$\hat{\theta} = \frac{\sum_{j=1}^{k} w_j \hat{\theta}_j}{\sum_{j=1}^{k} w_j} \tag{A.1}$$

Then, the likelihood ratio test is

$$\begin{aligned}
\lambda(\theta) &= \frac{L(\theta_o|\theta)}{L(\hat{\theta}|\theta)} \\
&= \frac{\prod_{j=1}^{k} \left(\frac{w_j}{2\pi}\right)^{1/2} e^{-(1/2)w_j(\theta_j-\theta_o)^2}}{\prod_{j=1}^{k} \left(\frac{w_j}{2\pi}\right)^{1/2} e^{-(1/2)w_j(\theta_j-\hat{\theta})^2}} \\
&= e^{-\frac{1}{2}\sum_{j=1}^{k} w_j\{(\theta_j-\theta_o)^2-(\theta_j-\hat{\theta})^2\}}
\end{aligned}$$

The likelihood ratio test can be further simplified by noting that

$$\begin{aligned}
& \sum_{j=1}^{k} w_j\{(\theta_j-\theta_o)^2-(\theta_j-\hat{\theta})^2\} \\
= & \sum_{j=1}^{k} w_j(\hat{\theta}-\theta_o)\{(\theta_j-\theta_o)+(\theta_j-\hat{\theta})\} \\
= & \sum_{j=1}^{k} w_j(\theta_j-\theta_o)\frac{\sum_{j=1}^{k} w_j\theta_j - \theta_o\sum_{j=1}^{k} w_j}{\sum_{j=1}^{k} w_j} + (\hat{\theta}-\theta_o)\frac{\sum_{j=1}^{k} w_j\theta_j \sum_{j=1}^{k} w_j - \sum_{j=1}^{k} w_j \sum_{j=1}^{k} w_j\theta_j}{\sum_{j=1}^{k} w_j} \\
= & \frac{\{\sum_{j=1}^{k} w_j(\theta_j-\theta_o)\}^2}{\sum_{j=1}^{k} w_j} \\
= & \frac{(\sum_{j=1}^{k} w_j\theta_j)^2 - 2\sum_{j=1}^{k} w_j\theta_j \sum_{j=1}^{k} w_j\theta_o + (\sum_{j=1}^{k} w_j\theta_o)^2}{\sum_{j=1}^{k} w_j} \\
= & \frac{(\sum_{j=1}^{k} w_j\hat{\theta})^2 - 2\sum_{j=1}^{k} w_j\hat{\theta} \sum_{j=1}^{k} w_j\theta_o + (\sum_{j=1}^{k} w_j\theta_o)^2}{\sum_{j=1}^{k} w_j} \\
= & \frac{(\sum_{j=1}^{k} w_j)^2(\hat{\theta}^2 - 2\hat{\theta}\theta_o + \theta_o^2)}{\sum_{j=1}^{k} w_j} = \sum_{j=1}^{k} w_j(\theta_o-\hat{\theta})^2
\end{aligned}$$

Therefore, the likelihood ratio is given by

$$\lambda(\theta) = e^{-\frac{1}{2}\sum_{j=1}^{k} w_j(\theta_o - \hat{\theta})^2}$$

Given one free parameter under H_o and k free parameters under H_A, the degrees of freedom are $k-1$. The null hypothesis is rejected if $-2log\lambda(\theta) = \sum_{j=1}^{k} w_j(\theta_o - \hat{\theta})^2 \geq \chi_{k-1,\alpha}$, where $\hat{\theta}$ is the MLE under H_A as given in Equation A.1. Under H_o, $-2log\lambda(\theta) = \sum_{j=1}^{k} w_j(\theta_j - \hat{\theta})^2$, known as the Q statistic asymptotically follows a chi square distribution with $k-1$ degrees of freedom (see Casella and Berger (2002, Theorem 10.3.3)).

Appendix B

Intracluster correlation coefficient (ANOVA estimator)

A suitable and convenient estimator of ρ is the analysis of variance (ANOVA) estimator(Snedecor and Cochran, 1980). This method was originally used for continuous data but is also suitable for binary data (Fleiss, 1981). Let S_{cj} and S_{wj} be the unbiased variance estimators between and within clusters in trial j, respectively. Given $S_{cj} = MSW$, $S_{wj} = (MSC_j - MSW_j)/m_{oj}$ and $\bar{m}_{Aij} = \sum_{l=1}^{n_{ij}} m_{ijl}^2/M_{ij}$, the analysis of variance estimator of intracluster correlation coefficient in trial j is given by

$$\hat{\rho}_j = \frac{S_{cj}}{S_{cj} + S_{wj}} = \frac{MSC_j - MSW_j}{MSC_j + (m_{oj} - 1)MSW_j} \tag{B.1}$$

where

$$MSC_j = \sum_{i=1}^{2} \sum_{l=1}^{n_{ij}} m_{ijl}(\hat{P}_{ijl} - \hat{P}_{ij})^2/(N_j - 2)$$

$$MSW_j = \sum_{i=1}^{2} \sum_{l=1}^{n_{ij}} m_{ijl}\hat{P}_{ijl}(1 - \hat{P}_{ijl})/(M_j - N_j)$$

and

$$m_{oj} = \left[M_j - \sum_{i=1}^{2} \bar{m}_{Aij} \right] / (N_j - 2)$$

where MSC_j and MSW_j are the pooled mean square errors between and within clusters in trial j, respectively. The analysis of variance (ANOVA) estimator is a consistent but not unbiased estimate of the intracluster correlation coefficient. It can also result in a negative value, indicating greater variation among individuals in the same cluster than among different clusters. In practice, negative values are generally regarded as implausible for cluster randomization trials; therefore, a negative estimated value of ρ_j is customary set equal to zero.

Appendix C

Variance component approach

Given the unweighted mean $\bar{\theta} = \sum_{j=1}^{k} \hat{\theta}_j / k$, the usual sample variance of $\hat{\theta}_j$ may be expressed as

$$S_\theta^2 = \frac{1}{k-1} \sum_{j=1}^{k} (\hat{\theta}_j - \bar{\theta})^2$$

Then the expected value of S_θ^2 in terms of variance components is

$$\begin{aligned} E[S_\theta^2] &= \frac{1}{k-1}[\sum_{j=1}^{k} E(\theta_j - \theta)^2 - \sum_{j=1}^{k} E(\bar{\theta} - \theta)^2] \\ &= \frac{1}{k-1}[\sum_{j=1}^{k} var(\theta_j) - k var(\bar{\theta})] \\ &= \frac{1}{k-1}[\sum_{j=1}^{k} (\sigma_{jc}^2 + \tau^2) - \frac{\sum_{j=1}^{k} (\sigma_{jc}^2 + \tau^2)}{k}] \\ &= \frac{\sum_{j=1}^{k} \sigma_{jc}^2}{k} + \tau^2 \end{aligned}$$

Then σ_{jc}^2 may be estimated using $\hat{\sigma}_{jc}^2$ in equation (2.4).

Appendix D

ML approach

Under a random effects model, the marginal distribution of the estimated intervention effect $\hat{\theta}_j$ for a meta-analysis of k cluster randomized trials follows a normal distribution with mean θ and variance $\hat{\sigma}^2_{jc} + \tau^2_c$. Given $\hat{w}^*_{jc} = 1/(\hat{\sigma}^2_{jc} + \tau^2_c)$, the likelihood function is given by

$$L(\theta, \tau^2_c) = \prod_{j=1}^{k} \left(\frac{\hat{w}^*_{jc}}{2\pi} \right)^{1/2} e^{-(1/2)\hat{w}^*_{jc}(\hat{\theta}_j - \theta)^2}$$

and the log likelihood function is

$$lnL(\theta, \tau^2_c) \quad = \quad -\frac{k}{2} ln 2\pi + \frac{1}{2} \sum_{j=1}^{k} ln\hat{w}^*_{jc} - \frac{1}{2} \sum_{j=1}^{k} \hat{w}^*_{jc}(\hat{\theta}_j - \theta)^2$$

First, by setting the first derivative of the log likelihood function $lnL(\theta, \tau^2_c)$ to zero in respect to θ, the maximum likelihood estimate of θ is given by

$$\hat{\theta}_c \quad = \quad \frac{\sum_{j=1}^{k} \hat{w}^*_{jc} \hat{\theta}_j}{\sum_{j=1}^{k} \hat{w}^*_{jc}} \tag{D.1}$$

Next, the first derivative of $lnL(\theta, \tau^2_c)$ in respect to τ^2_c is

$$\frac{dlnL(\theta)}{d\tau_c^2} = -\frac{1}{2}\sum_{j=1}^{k}\hat{w}_{jc}^{*} + \frac{1}{2}\sum_{j=1}^{k}(\hat{w}_{jc}^{*})^2(\hat{\theta}_j - \theta)^2 \quad \text{(D.2)}$$

Then, by setting (D.2) to zero and substituting θ by $\hat{\theta}_c$ in (D.1), we have

$$\sum_{j=1}^{k}(\hat{w}_{jc}^{*})^2(\hat{\sigma}_{jc}^2 + \tau_c^2) = \sum_{j=1}^{k}(\hat{w}_{jc}^{*})^2(\hat{\theta}_j - \hat{\theta}_c)^2 \quad \text{(D.3)}$$

By rearranging equation (D.3), the maximum likelihood estimate of τ_c^2 is obtained as

$$\hat{\tau}_{c.ML}^2 = \frac{\sum_{j=1}^{k}\hat{w}_{jc.ML}^2\{(\hat{\theta}_j - \hat{\theta}_{c.ML})^2 - \hat{\sigma}_{jc}^2\}}{\sum_{j=1}^{k}\hat{w}_{jc.ML}^2} \quad \text{(D.4)}$$

where $\hat{\theta}_{c.ML} = \sum_{j=1}^{k}\hat{w}_{jc.ML}\hat{\theta}_j / \sum_{j=1}^{k}\hat{w}_{jc.ML}$ and $\hat{w}_{jc.ML} = 1/(\hat{\sigma}_{jc}^2 + \hat{\tau}_{c.ML}^2)$. The estimate of $\tau_{c.ML}^2$ is obtained iteratively with an initial value of $\hat{\tau}_{c.ML}^2 = 0$. At each iteration, a positive value of $\hat{\tau}_{c.ML}^2$ is assured by setting the negative value equal to zero until convergence is reached.

Furthermore, the Wald-type confidence intervals for $\hat{\tau}_{c.ML}^2$ are constructed from the sampling variance calculated from the inverse of the Fisher information. The Fisher information I_{ML} can be computed by taking the negative of the second derivative of $lnL(\theta, \tau_c^2)$ in respect to τ_c^2 as follows:

$$I_{ML} = -\frac{d^2 lnL(\theta)}{(d\tau_c^2)^2} = -\frac{1}{2}\sum_{j=1}^{k}(\hat{w}_{jc}^*)^2 + \sum_{j=1}^{k}(\hat{w}_{jc}^*)^3(\hat{\theta}_j - \hat{\theta}_c)^2$$
$$= \frac{1}{2}\sum_{j=1}^{k}(\hat{w}_{jc}^*)^2 \quad (D.5)$$

which can be simplified by replacing the second term in equation (D.5) with the equality found in equation (D.3). Further, replacing $\hat{w}_{jc}^*$ in equation (D.5) by $\hat{w}_{jc.ML}$, the sampling variance for $\hat{\tau}_{c.ML}^2$ (i.e. the inverse of I_{ML}) is then obtained by

$$\hat{var}(\hat{\tau}_{c.ML}^2) = 2\left(\sum_{j=1}^{k}\hat{w}_{jc.ML}^2\right)^{-1}$$

The 95 percent Wald-type confidence interval is calculated as $\hat{\tau}_{c.ML}^2 \pm 1.96\sqrt{\hat{var}(\hat{\tau}_{c.ML}^2)}$

Appendix E

REML approach

For the restricted maximum likelihood estimator, the log likelihood function

$$lnL_R(\theta) = -\frac{k}{2}ln2\pi + \frac{1}{2}\sum_{j=1}^{k} ln\hat{w}^*_{jc} - \frac{1}{2}ln\sum_{j=1}^{k}\hat{w}^*_{jc} - \frac{1}{2}\sum_{j=1}^{k}\hat{w}^*_{jc}(\hat{\theta}_j - \theta)^2$$

The restricted maximum likelihood estimate of θ remains at the same as the maximum likelihood estimate of θ in D.1. The first derivative of $lnL_R(\theta)$ in respect to τ_c^2 is given by

$$\frac{dlnL_R(\theta)}{d\tau_c^2} = -\frac{1}{2}\sum_{j=1}^{k}\hat{w}^*_{jc} + \frac{1}{2}\frac{\sum_{j=1}^{k}(\hat{w}^*_{jc})^2}{\sum_{j=1}^{k}\hat{w}^*_{jc}} + \frac{1}{2}\sum_{j=1}^{k}(\hat{w}^*_{jc})^2(\hat{\theta}_j - \theta)^2 \qquad \text{(E.1)}$$

By setting equation (E.1) to zero and substituting θ by $\hat{\theta}^*_c$ in equation (D.1), we have

$$\sum_{j=1}^{k}(\hat{w}^*_{jc})^2(\hat{\sigma}^2_{jc} + \tau_c^2) = \frac{\sum_{j=1}^{k}(\hat{w}^*_{jc})^2}{\sum_{j=1}^{k}\hat{w}^*_{jc}} + \sum_{j=1}^{k}(\hat{w}^*_{jc})^2(\hat{\theta}_j - \hat{\theta}_c)^2 \qquad \text{(E.2)}$$

By rearranging equation (E.2), the restricted maximum likelihood estimate of τ_c^2 is

obtained as

$$\hat{\tau}^2_{c.RE} = \frac{\sum_{j=1}^{k} \hat{w}^2_{jc.RE}\{(\hat{\theta}_j - \hat{\theta}_{c.RE})^2 + 1/\sum_{j=1}^{k} \hat{w}_{jc.RE} - \hat{\sigma}^2_{jc}\}}{\sum_{j=1}^{k} \hat{w}^2_{jc.RE}} \quad \text{(E.3)}$$

where $\hat{w}_{jc.RE} = 1/(\hat{\sigma}^2_{jc} + \hat{\tau}^2_{c.RE})$ and $\hat{\theta}_{c.RE} = \sum_{j=1}^{k} \hat{w}_{jc.RE}\hat{\theta}_j / \sum_{j=1}^{k} \hat{w}_{jc.RE}$. $\hat{\tau}^2_{c.RE}$ is obtained iteratively with an initial value of zero. At each iteration, a negative value is truncated at zero until convergence is reached.

Furthermore, we can calculate the Fisher information to construct the Wald-type confidence intervals. The Fisher information I_{RE} for the REML estimate can be obtained by taking the negative of the second derivative of $lnL_R(\theta)$ in respect to τ^2_c after simplifying using the equality in equation (D.3), it is given by

$$\begin{aligned} I_{RE} = -\frac{d^2 lnL(\theta)}{(d\tau^2_c)^2} &= -\frac{1}{2}\sum_{j=1}^{k}(\hat{w}^*_{jc})^2 + \frac{\sum_{j=1}^{k}(\hat{w}^*_{jc})^3}{\sum_{j=1}^{k}\hat{w}^*_{jc}} - \frac{1}{2}\left(\frac{\sum_{j=1}^{k}(\hat{w}^*_{jc})^2}{\sum_{j=1}^{k}\hat{w}^*_{jc}}\right)^2 + \sum_{j=1}^{k}(\hat{w}^*_{jc})^3(\hat{\theta}_j - \hat{\theta}_c)^2 \\ &= \frac{1}{2}\sum_{j=1}^{k}(\hat{w}^*_{jc})^2 + \frac{\sum_{j=1}^{k}(\hat{w}^*_{jc})^3}{\sum_{j=1}^{k}\hat{w}^*_{jc}} - \frac{1}{2}\left(\frac{\sum_{j=1}^{k}(\hat{w}^*_{jc})^2}{\sum_{j=1}^{k}\hat{w}^*_{jc}}\right)^2 \quad \text{(E.4)} \end{aligned}$$

Replacing $\hat{w}^*_{jc}$ in equation (E.4) by $\hat{w}_{jc.RE}$, the sampling variance is then the inverse of the Fisher information, given by

$$\hat{var}(\hat{\tau}^2_{c.RE}) = 2\left(\sum_{j=1}^{k}\hat{w}^2_{jc.RE} - 2\frac{\sum_{j=1}^{k}\hat{w}^3_{jc.RE}}{\sum_{j=1}^{k}\hat{w}_{jc.RE}} + \left(\frac{\sum_{j=1}^{k}\hat{w}^2_{jc.RE}}{\sum_{j=1}^{k}\hat{w}_{jc.RE}}\right)^2\right)^{-1}$$

The 95 per cent Wald-type confidence interval is given by $\hat{\tau}^2_{c.RE} \pm 1.96\sqrt{\hat{var}(\hat{\tau}^2_{c.RE})}$.

Bibliography

Abramowitz, M. and Stegun, I. (1965), *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, Dover Publications: New York.

Adams, G., Gulliford, M., Ukoumunne, O., Eldridge, S., Chinn, S. and Campbell, M. (2004), 'Patterns of intra-cluster correlation from primary care research to inform study design and analysis.', *Journal of Clinical Epidemiology* **57**(8), 785–94.

Altman, D., Schulz, K., Moher, D., Egger, M., Davidoff, F., Elbourne, D., Gotzsche, P. and Lang, T. (2001), 'The revised consort statement for reporting randomized trials: explanation and elaboration', *Annals of Internal Medicine* **134**(8), 663–694.

Bass, M., McWhinney, I. and Donner, A. (1986), 'Do family physicians need medical assistants to detect and manage hypertension', *Canadian Medical Association Journal* **134**(11), 1247–1255.

Biggerstaff, B. and Jackson, D. (2008), 'The exact distribution of cochran's heterogeneity statistic in one-way random effects meta-analysis', *Statistics in Medicine* **27**, 6093–6110.

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.

Bland, J. (2004), 'Cluster randomised trials in the medical literature: two bibliometric surveys', *BMC Medical Research Methodology* **4**, 21–27.

Bland, J. and Altman, D. (2000), 'The odds ratio', *British Medical Journal* **320**, 1468.

Bohning, 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), 445–57.

Bradley, J. (1978), 'Robustness?', *British Journal of Mathematical and Statistical Psychology* **31**, 144–152.

Brockwell, S. and Gordon, I. (2001), 'A comparison of statistical methods for meta-analysis', *Statistics in Medicine* **20**(6), 825–840.

Burton, A., Altman, D. G., Royston, P. and Holder, R. L. (2006), 'The design of simulation studies in medical statistics', *Statistics in Medicine* **25**(24), 4279–4292.

Campbell, M., Fayers, P. and Grimshaw, J. (2005), 'Determinants of the intracluster correlation coefficient in cluster randomized trials: the case of implementation research', *Clinical Trials* **2**, 99–107.

Casella, G. and Berger, R. (2002), *Statistical Inference*, 2nd edn, Duxbury: North Scituate, MA.

Chobanian, A., Bakris, G., Black, H., Cushman, W., Green, L., Izzo, J. J., Jones, D., Materson, B., Oparil, S., Wright, J. J. and Roccella, E. (2003), 'The seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure: the jnc 7 report', *JAMA* **289**(19), 2560–72.

Clark, A., Hartling, L., Vandermeer, B. and McAlister, F. A. (2005), 'Meta-analysis: Secondary prevention programs for patients with coronary artery disease', *Annals of Internal Medicine* **143**(9), 659–672.

Cochran, W. (1954), 'The combination of estimates from different experiments', *Biometrics* **10**, 101–129.

Cohen, J. (1992), 'A power primer', *The Psychological Bulletin* **112**, 155–159.

Darlington, G. and Donner, A. (2007), 'Meta-analysis of community-based cluster randomization trials with binary outcomes', *Clinical Trials* **4**, 491–498.

DerSimonian, R. and Kacker, R. (2007), 'Random-effects model for meta-analysis of clinical trials: An update', *Contemporary Clinical Trials* **28**, 105–114.

DerSimonian, R. and Laird, N. (1986), 'Meta-analysis in clinical trials', *Controlled Clinical Trials* **7**, 177–188.

Dobson, A. (2002), *An Introduction to Generalized Linear Models*, Chapman & Hall/CRC Press.

Donner, A., Brown, K. and Brasher, P. (1990), 'A methodological review of non-therapeutic intervention trials employing cluster randomization, 1979-1989', *International Journal of Epidemiology* **19**, 795–800.

Donner, A. and Donald, A. (1987*a*), 'Analysis of data arising from a stratified design with cluster as unit if randomization', *Statistics in Medicine* **6**, 43–52.

Donner, A. and Donald, A. (1987*b*), 'The statistical analysis of multiple binary measurements', *Journal of Clinical Epidemiology* **41**(9), 899–905.

Donner, A. and Klar, N. (1996), 'Statistical considerations in the design and analysis of community intervention trials', *Journal of Clinical Epidemiology* **49**(4), 435–439.

Donner, A. and Klar, N. (2000), *Design and Analysis of Cluster Randomization Trials in Health Research*, Edward Arnold Publishers Ltd.

Donner, A. and Klar, N. (2002), 'Issues in the meta-analysis of cluster randomized trials', *Statistics in Medicine* **21**, 2971–80.

Donner, A., Piaggio, G. and Villar, J. (2001), 'Statistical methods for the meta-analysis of cluster randomization trials', *Statistical Methods in Medical Research* **10**, 325–338.

Donner, A., Piaggio, G. and Villar, J. (2003), 'Meta-analyses of cluster randomization trials: power considerations', *Evaluation & the Health Professions* **26**, 340–351.

Donner, A., Taljaard, M. and Klar, N. (2007), 'The merits of breaking the matches: a cautionary tale', *Statistics in Medicine* **26**, 2036–51.

Donner, A. and Zou, G. (2010), 'Closed-form confidence intervals for functions of the normal mean and standard deviation', *Statistical Methods Medical Research* pp. 1–13.

Eldridge, S., Ashby, D., Feder, G., Rudnicka, A. and Ukoumunne, O. (2004), 'Lessons for cluster randomised trials in the 21st century: a systematic review of trials in primary care', *Clinical Trials* **1**(1), 80–90.

Fawzi, W., Chalmers, T., Herrera, M. and Mosteller, F. (1993), 'Vitamin A supplementation and child-mortality - a meta-analysis', *Journal of the American Statistical Association* **269**, 898–903.

Fleiss, J. (1981), *Statistical Methods for Rates and Proportions*, 2nd edn, Wiley, New York.

Hardy, R. and Thompson, S. (1996), 'A likelihood approach to meta-analysis with random effects', *Statistics in Medicine* **15**(6), 619–629.

Hardy, R. and Thompson, S. (1998), 'Detecting and describing heterogeneity in meta-analysis', *Statistics in Medicine* **17**, 841–856.

Harville, D. (1977), 'Maximum likelihood approaches to variance component estimation and to related problems', *Journal of the American Statistical Association* **72**(358), 320–338.

Hedges, L. (1983), 'A random effects model for effect sizes', *The Psychological Bulletin* **93**, 388–395.

Hedges, L. and Olkin, I. (1985), *Statistical method for meta-analysis*, Academic Press.

Hedges, L. and Pigott, T. (2001), 'The power of statistical tests in meta-analysis', *Psychological Methods* **6**(3), 203–217.

Higgins, J. (2008), 'Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified', *International Journal of Epidemiology* **37**(5), 1158–1160.

Higgins, J. and Green, S. (2008), 'Cochrane handbook for systematic reviews of interventions', *Version 5.0.1* .

Higgins, J. and Thompson, S. (2002), 'Quantifying heterogeneity in a meta-analysis', *Statistics in Medicine* **21**(11), 1539–1558.

Higgins, J., Thompson, S., Deeks, J. and Altman, D. (2002*a*), 'Measuring inconsistency in meta-analyses', *British Medical Journal* **327**, 557–560.

Higgins, J., Thompson, S., Deeks, J. and Altman, D. (2002*b*), 'Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice', *Journal of Health Services Research & Policy* **7**(1), 51–61.

Huedo-Medina, T., Snchez-Meca, J., Marn-Martnez, F. and Botella, J. (2006), 'Assessing heterogeneity in meta-analysis: Q statistic or I^2 index?', *Psychological Methods* **11**, 193–206.

Ivers, N., Taljaard, M., Dixon, S., Bennett, C., McRae, A., Taleban, J., Skea, Z., Brehaut, J., Boruch, R., Eccles, M., Grimshaw, J., Weijer, C., Zwarenstein, M. and Donner, A. (2011), 'Impact of consort extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000-8', *British Medical Journal* **343**, 1–14.

Jackson, D. (2006), 'The power of standard test for the presence of heterogeneity in meta-analysis', *Statistics in Medicine* **25**, 2688–2699.

Jolly, K., Bradley, F., Sharp, S., Smith, H., Thompson, S., Kinmonth, A. and Mant, D. (1999), 'Randomised controlled trial of follow up care in general practice of patients with myocardial infarction and angina: final results of the southampton heart integrated care project (ship)', *British Medical Journal* **318**, 706–711.

Klar, N. and Darlington, G. (2004), 'Methods for modelling change in cluster randomization trials', *Statistics in Medicine* **23**, 2341–2357.

Klar, N. and Donner, A. (1997), 'The merits of matching in community intervention trials: a cautionary tale', *Statistics in Medicine* **16**, 1753–64.

Laopaiboon, M. (2003), 'Meta-analyses involving cluster randomization trials: A review of published literature in health care', *Statistical Methods in Medical Research* **12**(6), 515–530.

Lunn, A. and Davies, S. (1998), 'A note on generating correlated binary variables', *Biometrika* **85**, 487–490.

Mittlböck, M. and Heinzl, H. (2006), 'A simulation study comparing properties of heterogeneity measures in meta-analyses', *Statistics in Medicine* **25**, 4321–4333.

Moher, M., Yudkin, P., Wright, L., Turner, R., Fuller, A., Schofield, T. and Mant, D. (2001), 'Cluster randomised controlled trial to compare three methods of promoting secondary prevention of coronary heart disease in primary care', *British Medical Journal* **322**, 1338–1342.

Montgomery, A., Fahey, T., Peters, T., MacIntosh, C. and Sharp, D. (2000), 'Evaluation of computer based clinical decision support system and risk chart for management of hypertension in primary care: randomised controlled trial', *British Medical Journal* **320**, 686–690.

Montgomery, D. (2000), *Design and Analysis of Experiments*, 5nd edn, Wiley.

Morris, C. (1983), 'Parametric empirical Bayes inference: Theory and applications (C/R: P55-65)', *Journal of the American Statistical Association* **78**, 47–55.

Murray, D., Clark, M. and Wagenaar, A. (2000), 'Intraclass correlations from a community-based alcohol prevention study: the effect of repeat observations on the same communities', *Journal of studies on alcohol* **61**, 881–890.

Murray, D., Hannan, P., Wolfinger, R., Baker, W. and Dwyer, J. (1998), 'Analysis of data from group-randomized trials with repeat observations on the same groups', *Statistics in Medicine* **17**, 1581–600.

Paul, S. and Donner, A. (1989), 'A comparison of tests of homogeneity of odds ratios in k 2x2 tables', *Statistics in Medicine* **8**, 1455–1468.

Rao, J. and Scott, A. (1992), 'A simple method for the analysis of clustered binary data', *Biometrics* **48**, 577–585.

Raudenbush, S. and Bryk, A. (1985), 'Empirical bayes meta-analysis', *Journal of Educational Statistics* **10**, 75–98.

Rücker, G., Schwarzer, G., Carpenter, J. and Schumacher, M. (2008), 'Undue reliance on I^2 in assessing heterogeneity may mislead', *BMC Medical Research Methodology* **8**, 79.

RevMan (2008), 'Review manager', `http://www.cc-ims.net/RevMan/RevMan5/`.

Rotondi, M. and Khobzia, N. (2010), 'Vitamin A supplementation and neonatal mortality in the developing world: a meta-regression of cluster-randomized trials', *Bulletin World Health Organization* **88**(9), 697–702.

Schlattmann, P. (2009), *Medical Applications of Finite Mixture Models*, New York, NY: Springer.

Schmidt, F. (1992), 'What do data really mean - research findings, metaanalysis, and cumulative knowledge in psychology', *American Psychologist* **47**, 1173–1181.

Schulze, R. (2007), 'Current methods for meta-analysis', *Zeitschrift fur Psychologie / Journal of Psychology* **215**, 90–103.

Schuster, H. and Metzger, W. (2010), *Biometrics: Methods, Applications and Analyses*, Nova Science Pub Inc.

Sidik, K. and Jonkman, J. (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. (2006), 'Robust variance estimation for random effects meta-analysis', *Computational Statistics and Data Analysis* **50**(12), 3681–3701.

Sidik, K. and Jonkman, J. (2007), 'A comparison of heterogeneity variance estimators in combining results of studies', *Statistics in Medicine* **26**(9), 1964–1981.

Snedecor, G. and Cochran, W. (1980), *Statistical Methods*, 7th edn, Iowa State University Press, Ames.

Sommer, A., Djunaedi, E., Loeden, A., Tareotjo, I., West, K., Tilden, R. and Mele, L. (1986), 'Impact of vitamin-A supplementation on childhood mortality - a randomized controlled community trial', *Lancet* **1**, 1169–1173.

Song, J. (2004), 'Adjusted homogeneity tests of odds ratios when data are clustered', *Pharmaceutical Statistics* **3**(2), 81–87.

Swallow, W. and Monahan, J. (1984), 'Monte Carlo comparison of ANOVA, MIVQUE, REML, and ML estimators of variance components', *Technometrics* **26**, 47–57.

Takkouche, B., CadarsoSurez, 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.

Thompson, S. and Sharp, S. (1999), 'Explaining heterogeneity in meta-analysis: A comparison of methods', *Statistics in Medicine* **18**, 2693–2708.

Valentine, J., Pigott, T. and Rothstein, H. (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.

Viechtbauer, W. (2007*a*), 'Confidence intervals for the amount of heterogeneity in meta-analysis', *Statistics in Medicine* **26**(1), 37–52.

Viechtbauer, W. (2007*b*), 'Hypothesis tests for population heterogeneity in meta-analysis', *British Journal of Mathematical & Statistical Psychology* **60**, 29–60.

Villar, J., Mackey, M., Carroli, G. and Donner, A. (2001), 'Meta-analyses in systematic reviews of randomized controlled trials in perinatal medicine: Comparison of fixed and random effects models', *Statistics in Medicine* **20**(23), 3635–3647.

Wade, D. (1999), 'Randomized controlled trials - a gold standard?', *Clinical Rehabilitation* **13**(6), 453–455.

Whitehead, A. (2002), *Meta-Analysis of controlled clinical trials*, John Wiley & Sons.

Woodcock, A., Kinmonth, A., Campbell, M., Griffin, S. and Spiegal, N. (1999), 'Diabetes care from diagnosis: Effects of training in patient-centred care on beliefs, attitudes and behaviour of primary care professionals', *Patient Education and Counseling* **37**(1), 65–79.

Woolf, B. (1955), 'On estimating the relation between blood group and disease', *Annals of Human Genetics* **19**(4), 251–3.

Zou, G. (2008), 'On the estimation of additive interaction using the four-by-two table and beyond', *American Journal of Epidemiology* **168**, 212–224.

Zou, G., Huang, W. and Zhang, X. (2009), 'A note on confidence interval estimation for a linear function of binomial proportions', *Computational Statistics & Data Analysis* **53**(4), 1080–1085.

Vita

NAME: Shun Fu Chen

PLACE OF BIRTH: Taipei, Taiwan

EDUCATION

Department of Mathematics and Statistics
McGill University, Montreal, Quebec
1998-2002 B.Sc. Mathematics and Computer Science

Department of Statistics and Actuarial Science
University of Waterloo, Waterloo, Ontario
2002-2004 M.Sc. Biostatistics

Department of Epidemiology and Biostatistics
University of Western Ontario, London, Ontario
2006-2012 Ph.D Biostatistics

WORK EXPERIENCE:

Junior IT Network Administrator
Canadian Space Agency
St-Hubert, Quebec, 2000 (8 Month Internship)

Statistical Consultant
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario, 2003-2004

Biostatistician
Division of Epidemiology and Biostatistics
Montreal General Hospital
Montreal, Quebec, 2004-2006

TEACHING EXPERIENCE:

Teaching Assistant
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario, 2003-2004

Teaching Assistant
Department of Epidemiology and Biostatistic
University of Western Ontario
London, Ontario, 2008-2011

PUBLICATIONS:

Karp I, Chen SF, Pilote L. Sex differences in the effectiveness of statins after myocardial infarction. Canadian Medical Association Journal 2007; 176(3):333-8.

Keyhan G, Chen SF, Pilote L. Angiotensin-converting enzyme inhibitors and survival in women and men with heart failure. European Journal of Heart Failure 2007; 9(6-7):594-601.

Keyhan G, Chen SF, Pilote L. The effectiveness of β-blockers in women with congestive heart failure. Journal of General Internal Medicine 2007; 22(7):955-61.

Dasgupta K, O'Loughlin J, Chen S, Karp I, Paradis G, Tremblay J, Hamet P, Pilote L. Emergence of sex differences in prevalence of high systolic blood pressure: analysis of a longitudinal adolescent cohort. Circulation 2006; 114(24):2663-70.