Statistical Tests for Bias in Meta-Analysis with Binary Outcomes

Dissertation

zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften
der Universität Dortmund

Dem Fachbereich Statistik
der Universität Dortmund
vorgelegt von

Guido Schwarzer

aus

Rheine in Westfalen

Freiburg 2003

Gutachter: Prof. Dr. Joachim Hartung
Prof. Dr. Martin Schumacher (Freiburg)
Prof. Dr. Wolfgang Urfer

Tag der mündlichen Prüfung: 8. März 2004

Für die wissenschaftliche Betreuung dieser Arbeit und seine stets konstruktiven Kommentare bedanke ich mich ganz herzlich bei Prof. Dr. Martin Schumacher, Abteilung Medizinische Biometrie und Statistik am Universitätsklinikum Freiburg. Mein besonderer Dank gilt auch Dr. Gerd Antes, der mich für die Mitarbeit am Deutschen Cochrane Zentrum in Freiburg gewinnen konnte. In meiner Promotion konnte ich auf viele bei der Auswertung von Cochrane Reviews gewonnene Erfahrungen zurückgreifen.

Bedanken möchte ich mich auch bei meinen Kolleginnen und Kollegen der Abteilung Medizinische Biometrie und Statistik sowie des Freiburger Zentrums für Datenanalyse und Modellbildung (insbesondere für ihre kritischen Kommentare anläßlich meiner beiden Probevorträge in der Vorbereitung des Promotionskolloquiums). Besonders danken möchte ich meinem langjährigen Zimmerkollegen Dr. Norbert Holländer für seine Kommentare zu meiner Arbeit und viele promotionsspezifische Tipps.

Ganz besonderer Dank gilt meiner Frau Silvia Strieker, die meine diversen Stimmungsschwankungen während der Erstellung dieser Arbeit ertragen mußte. Außerdem tendierte mein Anteil an der täglichen Hausarbeit während dieser Zeit gegen Null – sogenannter „Dissertation related houseworking bias“. Diese Biasquelle wird in Zukunft, ohne Einführung eines neuen Testverfahrens, wieder verschwinden.

Contents

Zusammenfassung **iii**

1 Introduction **1**

2 Analysis and meta-analysis of binary data **7**

2.1 Analysis of binary data 7

2.1.1 Measures of the treatment effect 8

2.1.2 Estimation in sparse data 9

2.2 Meta-analysis of binary data 11

2.2.1 Fixed effect model 11

2.2.2 Random effects model 15

2.3 Existing methods for the detection of bias in meta-analysis 16

2.3.1 Funnel plot 16

2.3.2 Radial plot 18

2.3.3 Rank correlation test 19

2.3.4 Linear regression test 21

2.3.5 Other methods for identifying bias in meta-analysis 21

2.4 A new test for bias in meta-analysis with binary outcome data 23

3 Design of simulation study **27**

3.1 Design of published simulation studies 27

3.2 Simulation design 34

4 Results **39**

4.1 Simulation results under null hypothesis 39

4.1.1 Simulation I 40

4.1.2 Simulation II 49

4.1.3 Explanation for inflation of type I error rates with increasing treatment effect . . . 56
4.2 Simulation results under alternatives . . . 60
4.2.1 Specifying critical values . . . 60
4.2.2 Simulation I . . . 64
4.2.3 Simulation II . . . 73
4.3 Summary of simulation results . . . 78

5 Discussion **81**

Bibliography **85**

A Tables of simulation results **95**
A.1 Publication bias absent . . . 95
A.2 Publication bias present . . . 120

Zusammenfassung

Randomisierte klinische Studien sind ein wichtiges Mittel für die Erkenntnisgewinnung in der medizinischen Forschung. Beispielsweise bilden ihre Ergebnisse die Grundlage für eine Evidenz-basierte Medizin (EbM), welche durch den gewissenhaften, ausdrücklichen und vernünftigen Gebrauch der gegenwärtig besten externen, wissenschaftlichen Evidenz für Entscheidungen in der medizinischen Versorgung individueller Patienten gekennzeichnet ist. Die Praxis der EbM bedeutet die Integration individueller klinischer Expertise mit der bestmöglichen externen Evidenz aus systematischer Forschung. Ein wesentliches Problem in der Praxis der EbM ist die Diskrepanz zwischen der Menge medizinischer Publikationen und der zur Verfügung stehenden Lesezeit. Zusammenfassungen der Primärliteratur in sogenannten Übersichtsarbeiten (Reviews) bieten somit die Möglichkeit, große Mengen an Literatur schnell erfaßbar zu machen.

Systematische Reviews randomisierter Studien werden weltweit als das höchste Evidenzlevel angesehen. Eine systematische Übersichtsarbeit ist dadurch gekennzeichnet, daß nach allen Studien in systematischer Weise gesucht und daß alle Studien, unabhängig vom Ergebnis, bei der Bewertung Berücksichtigung finden. Dies ist anders als bei klassischen Übersichtsarbeiten, in denen häufig nur selektiv über einige Studien berichtet wird, andere jedoch gar nicht, oder nur am Rande, erwähnt werden. In systematischen Reviews spielt die Meta-Analyse, d.h. die statistische Auswertung der Ergebnisse von mehreren Studien zur gleichen Fragestellung, eine zentrale Rolle. Insbesondere der in einer Meta-Analyse durch die gewichtete Zusammenfassung der einzelnen Studienergebnisse geschätzte Therapieeffekt wird häufig zur Beurteilung der therapeutischen Effektivität einer Behandlung herangezogen. Dementsprechend ist es wichtig, daß der geschätzte Therapieeffekt eine unverzerrte Schätzung ist.

Ein grundlegendes Problem systematischer Übersichtsarbeiten besteht darin, ob alle zu einer Fragestellung durchgeführten Studien identifiziert werden bzw. ob die identifizierten Studien zumindest repräsentativ für alle durchgeführten Studien sind. In den letzten zwei Jahrzehnten wurde für den Bereich der Medizin in einer Reihe empirischer Untersuchungen gezeigt, daß Ergebnisse von Studien die eine statistisch signifikante Überlegenheit einer Behandlung zeigen, (i) mit größerer Wahrscheinlich-

keit publiziert werden, (ii) früher publiziert werden, (iii) häufiger in englischsprachigen Journalen, insbesondere solchen mit einem hohen Impakt-Faktor, publiziert werden. Diese Mechanismen führen potentiell zu einer Überschätzung des Therapieeffekts, welche häufig unter dem Begriff Selektionsbias zusammengefaßt werden; Publikationsbias und Sprachbias sind Teilaspekte des Selektionsbias.
Neben Ansätzen zur Minimierung von Bias in systematischen Übersichtsarbeiten durch Initiativen zur prospektiven Registrierung randomisierter klinischer Studien und zur Einrichtung von umfassenden Literaturdatenbanken sowohl publizierter als auch unpublizierter Studien kommen statistischen Methoden zur Detektion von Selektionsbias eine wichtige Rolle zu.

In dieser Arbeit werden die Eigenschaften statistischer Tests auf Bias in Meta-Analysen mit binären Zielgrößen im Rahmen von Simulationen sowohl unter der Null-Hypothese als auch unter zwei verschiedenen Alternativen (mit moderatem und starkem Selektionsbias) untersucht. Insbesondere wird der Einfluß der Studiengröße der Einzelstudien, des zugrundeliegenden Therapieeffekts, der Ereignisrate in der Kontrollgruppe und des Ausmaßes der Heterogenität zwischen den Studien auf die Ergebnisse der Tests auf Bias bewertet.
In die Untersuchungen werden die beiden gebräuchlichsten Testverfahren auf Bias in Meta-Analysen sowie ein neu vorgeschlagener Test eingeschlossen. Ein vorhandener Test basiert auf der Rangkorrelation zwischen einer standardisierten Form des geschätzten Therapieeffekts (z.B. log odds ratio) und der geschätzten Varianz des Therapieeffekts (z.B. geschätzte Varianz des log odds ratios). Der zweite Test basiert auf einer linearen Regression des standardisierten geschätzten Therapieeffekts auf die Inverse der Wurzel der geschätzten Varianz. In beiden Testverfahren wird insbesondere die geschätzte asymptotische Varianz verwendet. Der neu vorgeschlagene Test auf Bias in Meta-Analysen stellt eine Modifikation des Rankkorrelationstests dar. Anstelle des standardisierten geschätzten Therapieeffekts und der geschätzten asymptotischen Varianz des Therapieeffekts werden Größen verwendet, die aus der nicht-zentralen hypergeometrischen Verteilung abgeleitet sind.
Bei der Festlegung des Simulationsdesigns wurde auf vorhandene empirische Informationen zurückgegriffen. Insgesamt wurden vier verschiedene Simulationsdesigns verwendet, die sich in der Art der Generierung der Studiengröße unterscheiden. Alle andere Faktoren, d.h. zugrundeliegender Therapieeffekt, Ereignisrate in der Kontrollgruppe und Ausmaß der Heterogenität zwischen den Studien, wurden in allen Simulationsdesigns gleich gewählt. In zwei Simulationsdesigns ist der Anteil der generierten Studien mit einer Gesamtanzahl von weniger als 100 Patienten mit 66% und 30% sehr hoch; desweiteren ist in diesen Designs der Anteil von generierten Studien mit einem Gesamtanzahl größer 500 mit 2% und 0% sehr gering. Bei den in

den Simulationen zugrundegelegten Kontroll-Ereignisraten von 0.1 und 0.3 gibt es in diesen beiden Simulationsdesigns dementsprechend eine große Anzahl von Studien mit nur wenigen Ereignissen pro Behandlungsgruppe, im Englischen häufig als "sparse (binary) data" bezeichnet. Im Gegensatz dazu liegt in einem weiteren Design der Anteil von Studien mit einer Gesamtanzahl von weniger als 100 Patienten bei 0% und von mehr als 500 Patienten bei 65%.

Die zentralen Ergebnisse der Simulationsstudien unter der Nullhypothese lassen sich folgendermaßen zusammenfassen:

- In den beiden Simulationsdesigns mit einem großen Anteil kleiner Studien ist für die beiden gebräuchlichen Testverfahren eine deutliche Inflation des Fehlers 1. Art mit zunehmendem Therapieeffekt zu erkennen; desweiteren ist mit zunehmendem Therapieeffekt eine deutliche Asymmetrie in der unteren und oberen Ablehnungsrate zu erkennen.
- Für den neu vorgeschlagenen Test ist in dieser Situation keine Inflation des Fehlers 1. Art zu erkennen. Lediglich für sehr starke Therapieeffekte, d.h. Odds ratio von 0.25, ist eine Asymmetrie in den beiden Ablehnungsraten zu erkennen. Diese Asymmetrie fällt deutlich geringer aus als für die beiden gebräuchlichen Tests.
- In den Simulationsdesigns mit einem großen Anteil kleiner Studien ist in Meta-Analysen mit 10 Studien, was bei medizinischen Fragestellungen einer typischen Anzahl von Studien entspricht, lediglich für den Test basierend auf der linearen Regression eine deutliche Inflation des Fehlers 1. Art zu erkennen.
- Im Simulationsdesign mit großem Anteil großer Studien liegen die Ergebnisse des linearen Regressionstests in der Nähe des vorgegebenen Signifikanzniveaus während der Rankkorrelationstest und der neue Test teilweise sehr konservative Ergebnisse haben.

In Simulationen unter der Alternative zeigen sich für alle Simulationsdesigns ähnliche Ergebnisse, die sich wie folgt zusammenfassen lassen:

- Die Power aller Testverfahren nimmt mit zunehmendem Therapieeffekt ab. Dieses Ergebnis läßt sich durch den mit zunehmendem Therapieeffekt ebenfalls steigenden Anteil publizierter Studien erklären.
- Die Power aller Testverfahren nimmt mit steigender Heterogenität zwischen den Studien ab, obwohl gleichzeitig der Bias zunimmt. Insgesamt zeigt sich

für den neu vorgeschlagenen Test ein stärkerer Powerverlust mit zunehmender Heterogenität.

- In Meta-Analysen mit 10 Studien ist die Power aller Verfahren sehr gering. Für moderaten Selektionsbias liegt die maximale Power bei 34.1% (linearer Regressionstest), 23.7% (Rankkorrelationstest) und 25.0% (neuer Test). Bei starkem Selektionsbias ist die maximale Power 49.1% (linearer Regressionstest), 33.8% (Rankkorrelationstest) und 34.3% (neuer Test).

Chapter 1

Introduction

The use of systematic reviews, i.e. reviews using explicit and rigorous methods to identify, critically appraise, and synthesise relevant studies, has rapidly grown in the medical field during the last two decades. Reasons for this development are the increase in the number of articles published in medical journals and advantages of systematic reviews as compared to traditional narrative reviews (Egger and Smith, 1997).

The statistical method of meta-analysis is used in systematic reviews to combine two or more individual study results (Fleiss, 1993). More specific, an overall treatment effect is estimated by calculating a weighted average of treatment estimates in individual studies; various methods of meta-analysis exist that differ mainly in the weighting scheme utilised. Meta-analysis provides a statistical method to evaluate the direction and size of the treatment effect as well as the question whether the treatment effect is consistent across studies. Furthermore, the precision of a treatment estimate can be improved when it is based on information from several studies. Accordingly, the power to detect a real treatment effect as statistically significant is increased by combining several study results.

Systematic reviews of randomised controlled trials are considered as the highest level of evidence to evaluate the effects of health-care interventions and play an important role in the practice of evidence-based medicine, i.e. the integration of individual clinical expertise with the best available external clinical evidence from systematic research (Sackett et al., 2000). Accordingly, focus of the Cochrane Collaboration, the largest organisation in the world engaged in the production and maintenance of systematic reviews of health-care interventions, is on systematic reviews of randomised controlled trials (Clarke and Langhorne, 2001). However, the reliability of systematic reviews of randomised controlled trials can be affected by various sources of bias.

Lack of methodological quality of included trials is one potential source of bias. A landmark publication showing this association is the paper by Schulz et al. (1995) assessing the methodological quality of 250 controlled trials from 33 meta-analyses contained in the Cochrane Pregnancy and Childbirth Database. They showed that estimates of treatment effects differ depending on the adequacy of concealment of treatment allocation and the adequacy of blinding. Odds ratios were exaggerated, on average, by 41% for inadequately concealed trials and by 30% for unclear concealed trials (i.e. trials that did not report or incompletely reported the method of concealment) in comparison to trials with adequate method of concealment. Odds ratios were exaggerated, on average by 17% in trials that were not double-blind. These results have been confirmed in other empirical studies (Jüni et al., 2001). An approach to tackle the problem of the quality of included trials in a systematic review consists in a thorough assessment of the quality of included trials and the conduct of subgroup analyses or meta-regression techniques (Thompson and Higgins, 2002) depending on the quality of trials. In this thesis, methodological problems resulting from differences in the quality of trials are not considered any further.

Another type of bias in meta-analysis emerges if the trials included in the systematic review do not constitute a representative sample of all conducted trials. This type of bias which is referred to as selection bias or merely bias in subsequent chapters of this thesis can be introduced in many ways during the location of trials for inclusion in a systematic review. Especially, publication bias, i.e. the selective publication of trials with significant results is a predominating source of bias (Begg and Berlin, 1988); empirical evidence of publication bias is described in detail in the next subsection. Other potential sources of bias include language bias (Egger et al., 1997), i.e. the selective publication of significant results in English language journals, and database bias (Egger and Smith, 1998) resulting from the publication of trials with significant results in high-ranking journals which are typically indexed in literature databases like MEDLINE or EMBASE. Compared to publication bias, the empirical evidence for these two sources of bias is not so strong. For example, contradictory results on the impact of language bias are available (Moher et al., 2000).

Empirical evidence of publication bias

The problem of publication bias was reported for the first time in a paper by Sterling (1959) indicating that published results of scientific investigations are not representative for results of all scientific studies. He found that 286 of 294 research reports (97.3%) published in four major psychology journals between 1955 and 1956 reported statistically significant results for the major scientific hypothesis. An update of this work covering reports published between 1986 and 1987 in the same psychology journals reported very similar results with 538 of 563 reports (95.6%) having statistically significant results (Sterling et al., 1995).

In the context of meta-analysis of clinical trials, publication bias was reported for the first time by Simes (1986). He identified all trials comparing combination therapy with alkylating agents as treatment for advanced ovarian cancer that either were listed in an international cancer trials registry or were published by using a computer search for the indexed literature contained in MEDLINE. Pooled analysis showed a statistically significant beneficial effect of the combination therapy in published trials (median survival ratio: 1.16, 95% confidence interval: 1.06 - 1.27, P-value: 0.02) but no significant effect in registered trials (median survival ratio: 1.05, 95% confidence interval: 0.98 - 1.12, P-value: 0.25).

Similar findings were reported in several other empirical studies (Berlin et al., 1989; Dickersin, 1990, 1997; Egger et al., 1997).

Recently, a survey was published evaluating 510 randomised controlled trials presented at annual meetings of the American Society of Clinical Oncology (ASCO) between 1989 and 1998 each with a total sample size greater than 200 (Krzyzanowska et al., 2003). The overall probability of publication within 5 years after presentation at the ASCO meeting was 74%. The probability of publication of trials with statistically significant results (i.e. P-value < 0.05) was 81% whereas this probability was only 68% in trials with non-significant results; this difference is highly statistically significant with P-value < 0.001.

How to tackle the problem of selection bias in meta-analysis

Different strategies are available to approach the problem of selection bias. At present, none of these strategies should be used exclusively but a combination of approaches is the preferred method.

Obviously, the prospective registration of all initiated randomised controlled trials would overcome this bias (Dickersin and Rennie, 2003); an example is Current Controlled Trials' Meta-Register, an international database combining registers of ongoing randomised controlled trials in all areas of health-care (`http://www.controlled-trials.com`). However, these type of compilations are still far from being complete (Tonks, 1999).

Another approach consists in an exhaustive search strategy including hand-searching of journals. The Cochrane Controlled Trials Register which is part of the Cochrane Library (`http://www.cochranelibrary.com`) is a bibliography of controlled trials identified by contributors to the Cochrane Collaboration and others by hand-searching the world's journals. The German Cochrane Centre in Freiburg, for example, is involved in this international effort by coordinating the hand-searching of several German language journals. The register includes many published reports not currently listed in literature databases like MEDLINE. Again, this register is still far from being complete.

A third approach, which actually is pursued in this thesis, is to tackle the problem of selection bias by developing and using statistical methods for identifying bias in meta-analysis.

Commonly used methods for the detection of bias in meta-analysis

A funnel plot (Light and Pillemer, 1984), i.e. a scatter plot of the estimated treatment effect and a measure of the precision of the treatment estimate, is commonly used to check informally the presence of bias in meta-analysis. The display looks like a funnel if neither selection bias nor excess between-trial heterogeneity exists showing decreasing fluttering with increasing precision. Asymmetry in the funnel plot is taken as an indication of bias in meta-analysis. As noted by Egger et al. (1997), other explanations for funnel plot asymmetry exist and they prefer the term "small study effects". Beside this informal graphical method, statistical tests for bias in meta-analysis have been proposed.

Two statistical tests for bias in meta-analysis, available in the statistical software package Stata (StataCorp, 2003), have been used in a number of medical applications: a rank correlation test (Begg and Mazumdar, 1994) and a test based on a linear regression of the standard normal deviate on precision (Egger et al., 1997). For both tests, the variance of the estimated treatment effect in each individual trial is of central importance.

Binary outcomes are typically used in medical applications with risk ratio or odds ratio as measure of the treatment effect. The logarithm of these relative effect measures is often taken to calculate an overall treatment effect (Fleiss, 1993). Accordingly, the asymptotic variance of the log risk ratio or log odds ratio is utilised both in the funnel plot and in tests for bias in meta-analysis.

So far, the statistical properties of the linear regression and rank correlation test in meta-analysis with binary outcome data have been examined only in few publications (Sterne et al., 2000; Macaskill et al., 2001; Schwarzer et al., 2002). Nevertheless, a common problem of these tests, i.e. an inflation of type I error rates with increasing treatment effect, was identified both in Macaskill et al. (2001) and Schwarzer et al. (2002); furthermore, this problem was mentioned with respect to large treatment effects in Sterne et al. (2000).

Outline

In this thesis, which is partly based on Schwarzer et al. (2002), the properties of the linear regression and rank correlation test which are the two most commonly used tests for bias in meta-analysis are investigated in detail. Furthermore, an alternative test for bias in meta-analysis with binary outcome data is introduced and the properties of this test are compared with those of existing tests both under the null hypothesis of no bias in meta-analysis and alternatives.

In Chapter 2, statistical methods for the analysis and meta-analysis of binary outcome data that are of relevance for this thesis are reviewed. The most common measures of the treatment effect are introduced and the problem of sparse binary data is discussed. In Section 2.2, statistical methods for meta-analysis of binary data are briefly described. Methods for the detection of bias in meta-analysis evaluated in this thesis are reviewed in Section 2.3. Finally, a new test for bias in meta-analysis with binary outcomes is introduced in Section 2.4.

The simulation design used in this thesis should be realistic and represent meta-analyses published in medical journals; this aspect is elaborated in Chapter 3. Empirical information coming from published simulation studies evaluating tests for bias in meta-analysis is reviewed in Section 3.1. The actual simulation design utilised in this thesis is described in Section 3.2.

In Chapter 4, results of various simulation studies are described. Results under the null hypothesis of no bias in meta-analysis are described in Section 4.1; results under the alternative of moderate and strong selection bias are given in Section 4.2. A listing of the main simulation results can be found in Section 4.3. Additional information on simulation results in tabular form is contained in Appendix A.

In Chapter 5, this thesis ends with a discussion of simulation results and an outlook on future research.

Chapter 2

Analysis and meta-analysis of binary data

In this chapter, only those statistical methods for the analysis and meta-analysis of binary outcome data are reviewed that are of relevance for this thesis. Statistical methods for the analysis of binary data such as tests for treatment differences, e.g., χ^2 test and Fisher's exact test, and logistic regression models are not discussed. Furthermore, other approaches to meta-analysis such as Bayesian meta-analysis (Carlin, 1992; Smith et al., 1995; Warn et al., 2002) and multivariate approaches to meta-analysis (van Houwelingen et al., 2002) are not covered in this chapter.

Lets assume that a meta-analysis of k randomised controlled trials with binary outcome data is conducted to compare a new intervention with a control intervention; the available information for each trial can then be summarised as depicted in Table 2.1. The most common measures of the treatment effect are introduced and the problem of sparse binary data is discussed in Section 2.1. In Section 2.2, statistical methods for the meta-analysis of binary data are reviewed. Methods for the detection of bias in meta-analysis evaluated in this thesis are described in Section 2.3. Finally, a new test for bias in meta-analysis with binary outcomes is introduced in Section 2.4.

2.1 Analysis of binary data

The result of each individual randomised controlled trial comparing a new intervention and a control treatment with respect to a binary outcome can be summarised as shown in Table 2.1. In this table, number of patients in the two treatment groups $a_i + b_i$ and $c_i + d_i$ are assumed to be fixed. Based on this assumption, the number of events in each treatment group is distributed according to a binomial distribution

	Event	No event	Group size
Intervention	a_i	b_i	$a_i + b_i$
Control	c_i	d_i	$c_i + d_i$
	$a_i + c_i$	$b_i + d_i$	n_i

Table 2.1: Summary data of trial i with binary outcome data, $i = 1, \ldots, k$.

$Bin(n, p)$ (Agresti, 1990, p. 38). More precisely, cell count $a_i \sim Bin(a_i + b_i, p_i^I)$ and cell count $c_i \sim Bin(c_i + d_i, p_i^C)$ where p_i^I and p_i^C denote the event probabilities for intervention and control group, respectively. The event probabilities p_i^I and p_i^C are also referred to as treatment and control event rate in this thesis. Estimates of these probabilities are given by the relative cell counts $\widehat{p}_i^I = a_i/(a_i + b_i)$ and $\widehat{p}_i^C = c_i/(c_i + d_i)$.

2.1.1 Measures of the treatment effect

The most commonly used measures of the treatment effect for binary outcome data are risk difference, risk ratio, and odds ratio. The risk difference is defined as the difference between the two event probabilities $p_i^I - p_i^C$. The risk ratio, often called the relative risk, is defined as $\phi_i = p_i^I / p_i^C$. Finally, the odds ratio is defined as $\psi_i = (p_i^I/(1 - p_i^I)) \,/\, (p_i^C/(1 - p_i^C))$.

Recommendations on the choice of measure of the treatment effect in a meta-analysis can be found in Deeks and Altman (2001). In systematic reviews, the odds ratio and risk ratio are typically used as summary measures. The main reason is that these relative effect measures are, on average, more stable across trials than the risk difference (Engels et al., 2000; Deeks, 2002). Accordingly, this thesis focuses on the odds ratio and risk ratio as measures of the treatment effect.

Estimates of odds ratio ψ_i and risk ratio ϕ_i are given by the following formulae:

$$\widehat{\psi}_i = \frac{a_i\, d_i}{b_i\, c_i}. \tag{2.1}$$

$$\widehat{\phi}_i = \frac{a_i}{a_i + b_i} \Bigg/ \frac{c_i}{c_i + d_i} \tag{2.2}$$

These estimators converge to a normal distribution, but, have a highly skewed distribution for small sample sizes. Therefore, the logarithm of the risk ratio and odds ratio are typically considered which converge much faster to the normal distribution (Agresti, 1990, p. 54).

Large sample variance estimates for the logarithm of the odds ratio and risk ratio are given by

$$\widehat{\mathrm{Var}}(\log \widehat{\psi}_i) = \frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i}. \tag{2.3}$$

$$\widehat{\mathrm{Var}}(\log \widehat{\phi}_i) = \frac{1}{a_i} + \frac{1}{c_i} - \frac{1}{a_i + b_i} - \frac{1}{c_i + d_i} \tag{2.4}$$

Estimates defined in equation (2.1) to (2.4) are utilised, for example, in the calculation of approximate confidence intervals for the respective measure of the treatment effect, and in a meta-analysis using the inverse variance method for pooling (see Section 2.2.1). An approximate 95% confidence interval for the odds ratio is given by

$$\exp\left(\log \widehat{\psi}_i \pm 1.96 \times \mathrm{S.E.}(\log \widehat{\psi}_i)\right) \tag{2.5}$$

with standard error $\mathrm{S.E.}(\log \widehat{\psi}_i) = \sqrt{\widehat{\mathrm{Var}}(\log \widehat{\psi}_i)}$ (Agresti, 1990, p. 55).

2.1.2 Estimation in sparse data

A two-by-two table is referred to as sparse if any of the cells counts is small (Agresti, 1990, p. 244). Sparse data occur either if the total sample size of a trial is small or if the event probability is very low or very high in trials with large sample sizes. In the context of systematic reviews, the first case is much more likely, i.e. sparse data resulting from small sample sizes.

In the extreme case, a two-by-two table has entries with zero cell counts. If the event probability is low, e.g., in a population with a small risk of the event of interest, cell count a_i or c_i may be zero. If one of these two cell counts is zero, both the estimated odds ratio and estimated risk ratio are either 0 or ∞. If both cell counts are zero, the two summary measures are undefined. In any case, variance estimates given in (2.3) and (2.4) are ∞ due to division by zero. For the odds ratio, the same statements are true if either cell count b_i or d_i is zero; the risk ratio has finite estimates in this case.

For the odds ratio, Gart and Zweifel (1967) showed that adding 0.5 to the cell counts (a_i, b_i, c_i, d_i) in Table 2.1 improved the estimators in (2.1) and (2.3) by reducing their bias:

$$\widehat{\psi}_i^{mod} = \frac{(a_i + 0.5)(d_i + 0.5)}{(b_i + 0.5)(c_i + 0.5)} \tag{2.6}$$

and

$$\widehat{\mathrm{Var}}(\log \widehat{\psi}_i^{mod}) = \frac{1}{a_i + 0.5} + \frac{1}{b_i + 0.5} + \frac{1}{c_i + 0.5} + \frac{1}{d_i + 0.5}. \tag{2.7}$$

This modification can be used in general, however, is typically only applied if any of the cell counts is zero. The estimated variance of the log odds ratio gets smaller if 0.5 is added to each cell of the two-by-two table since all four denominators in equation (2.3) are increased.

For the risk ratio, Pettigrew et al. (1986) showed that the estimators in (2.2) and (2.4) can be improved by reducing their bias using the following formulae:

$$\widehat{\phi}_i^{mod} = \frac{a_i + 0.5}{a_i + b_i + 0.5} \Big/ \frac{c_i + 0.5}{c_i + d_i + 0.5} \tag{2.8}$$

and

$$\widehat{\mathrm{Var}}(\log \widehat{\phi}_i^{mod}) = \frac{1}{a_i + 0.5} + \frac{1}{c_i + 0.5} - \frac{1}{a_i + b_i + 0.5} - \frac{1}{c_i + d_i + 0.5}. \tag{2.9}$$

Again, this modification can be used in general, however, is often only applied if any of the cell counts is zero. Adding 0.5 to the cell counts (a_i, b_i, c_i, d_i) in Table 2.1 results in different estimates; this modification is used, for example, in the Stata procedure `metan` (StataCorp, 2003) which provides methods for meta-analysis of binary outcome data.

Peto method

An alternative method for the estimation of the odds ratio was proposed by Yusuf et al. (1985). For this method, often referred to as the Peto method or the Yussuf and Peto method, no correction for zero cell counts is necessary. The method is based on the observed cell count a_i and the expected cell count $E(a_i|\ldots;\psi = 1)$ where '…' denotes the four fixed marginal totals. The Peto odds ratio is given by

$$\widehat{\psi}_i^* = \exp\left(\frac{a_i - \mathrm{E}(a_i|\ldots;\psi_i = 1)}{\mathrm{Var}(a_i|\ldots;\psi_i = 1)}\right) \tag{2.10}$$

where $\mathrm{Var}(a_i|\ldots;\psi_i = 1)$ is the hypergeometric variance of a_i. Formulae for the expected cell count and variance are given by

$$\mathrm{E}(a_i|\ldots;\psi_i = 1) = \frac{(a_i + b_i)(a_i + c_i)}{n_i}$$

and

$$\mathrm{Var}(a_i|\ldots;\psi_i = 1) = (a_i + b_i)(c_i + d_i)(a_i + c_i)(b_i + d_i)/(n_i^2(n_i - 1)).$$

An estimator of the variance of $\log \widehat{\psi}_i^*$ is given by

$$\widehat{\mathrm{Var}}(\log \widehat{\psi}_i^*) = 1/\mathrm{Var}(a_i|\ldots;\psi_i = 1). \tag{2.11}$$

Greenland and Salvan (1990) showed that the Peto method performs poor in unbalanced designs and in nearly balanced designs if the odds ratio differs substantially from 1.00. In the context of randomised controlled trials, the use of the Peto method is reasonable in most cases, i.e. if treatment groups are of comparable size (in the case of 1:1 randomisation) and treatment effects are moderate.

2.2 Meta-analysis of binary data

The statistical method of meta-analysis is used to combine two or more individual trial results. An overall treatment effect, e.g., a summary odds ratio, is estimated by calculating a weighted average of treatment estimates in individual trials. Meta-analysis provides a statistical method to evaluate the direction and size of the treatment effect as well as the question whether the treatment effect is consistent across trials. Furthermore, the precision of a treatment estimate can be improved when it is based on information from several trials.

Two general approaches to meta-analysis exist (Fleiss, 1993). In a fixed effect meta-analysis, it is assumed that the underlying true treatment effect is equal in all trials, e.g., for the odds ratio $\psi_1 = \ldots = \psi_k = \psi$. Accordingly, the estimated overall treatment effect is a typical treatment effect for all trials.

In a random effects meta-analysis, it is allowed that the true treatment effect may differ between trials. Due to the many differences between trials (e.g., trial populations, type of intervention), this seems to be a more appropriate modelling approach. However, it is assumed that individual treatment effects vary around an average overall treatment effect; typically effects are assumed to have a normal distribution, e.g., for the odds ratio as measure of the treatment effect: $\log \psi_i \sim N(\log \psi, \tau^2)$, $i = 1, \ldots, k$, with between-trial variance τ^2.

There is no consensus whether a fixed effect or random effects meta-analysis should be used in a systematic review (Thompson and Pocock, 1991). One strategy often applied is to compare the results of the two methods in a sensitivity analysis. For testing an overall treatment effect, an alternative approach was proposed by Hartung (1999). For binary outcomes, this approach performs better than commonly used tests for fixed effect and random effects model (Hartung and Knapp, 2001). Furthermore, it is not necessary to choose in advance between fixed effect and random effects model.

2.2.1 Fixed effect model

Under a fixed effect model, three approaches exist for the estimation of a pooled treatment effect in meta-analyses of binary data. The inverse variance and Mantel-

Haenszel method can be used for odds ratio, risk ratio and risk difference, whereas the Peto method was solely developed for the odds ratio.

Inverse variance method

The inverse variance method is a generic method for meta-analysis. In order to use this method, an estimate of the treatment effect and its variance is needed. This method can be used in very different settings, e.g., meta-analysis of continuous data, meta-analysis of survival data using the log hazard ratio as measure of the treatment effect (Parmar et al., 1998), and meta-analysis of prognostic factors (Altman, 2001; Taggart et al., 2001).

The pooled odds ratio using the inverse variance method is given by

$$\widehat{\psi}_{IV} = \exp\left(\frac{\sum_{i=1}^{k} w_i \cdot \log\widehat{\psi}_i}{\sum_{i=1}^{k} w_i}\right) \tag{2.12}$$

with weights $w_i = \widehat{\text{Var}}(\log\widehat{\psi}_i)^{-1} = (1/a_i\ +\ 1/b_i\ +\ 1/c_i\ +\ 1/d_i)^{-1}$.

The estimated variance of the logarithm of $\widehat{\psi}_{IV}$ is given by

$$\widehat{\text{Var}}(\log\widehat{\psi}_{IV}) = \frac{1}{\sum_{i=1}^{k} w_i}. \tag{2.13}$$

An approximate 95% confidence interval for the pooled odds ratio is given by

$$\exp\left(\log\widehat{\psi}_{IV} \pm 1.96 \times \text{S.E.}(\log\widehat{\psi}_{IV})\right) \tag{2.14}$$

with standard error $\text{S.E.}(\log\widehat{\psi}_{IV}) = \sqrt{\widehat{\text{Var}}(\log\widehat{\psi}_{IV})}$ (Fleiss, 1993).

For the risk ratio, both $\log\widehat{\psi}_i$ and $\widehat{\text{Var}}(\log\widehat{\psi}_i)$ are replaced by the respective quantities defined in Section 2.1.

It is well-known that the pooled estimate based on the inverse variance method is a biased estimate for odds ratio and risk ratio in sparse binary data (Breslow, 1981; Greenland and Robins, 1985).

In the case of sparse binary data, the procedure described in Section 2.1 can be utilised in different ways leading to three general meta-analytic strategies. Firstly, 0.5 can be added to all two-by-two tables (labelled "add all"). Secondly, 0.5 can be added to all two-by-two tables only in the case of zero cell counts in one or more trials (labelled "add all selective"). Thirdly, 0.5 can be added only to cell counts of corresponding two-by-two tables with zero cell counts (labelled "add selective"). In any case, trials with zero events in both treatment groups are typically excluded from the meta-analysis. The Stata procedure `metan` (StataCorp, 2003) uses the "add selective" approach. All three strategies can be utilised in Stata by using the procedure `meta` which requires variables containing the estimated treatment effect and its standard error.

Mantel-Haenszel method

Mantel and Haenszel (1959) proposed an estimator for the common odds ratio in a stratified case-control study; this method can also be used in a meta-analysis of randomised controlled trials. The pooled odds ratio is estimated by combining the individual odds ratio on the natural scale:

$$\widehat{\psi}_{MH} = \frac{\sum\limits_{i=1}^{k} w_i \cdot \widehat{\psi}_i}{\sum\limits_{i=1}^{k} w_i} \tag{2.15}$$

with weights $w_i = \dfrac{b_i c_i}{n_i}$.

An estimator of the variance of the logarithm of $\widehat{\psi}_{MH}$ that is robust both in sparse data and large strata models was introduced by Robins, Breslow, and Greenland (1986), see also Robins, Greenland, and Breslow (1986):

$$\widehat{\mathrm{Var}}(\log \widehat{\psi}_{MH}) = \frac{\sum\limits_{i=1}^{k} P_i R_i}{2\left(\sum\limits_{i=1}^{k} R_i\right)^2} + \frac{\sum\limits_{i=1}^{k} (P_i S_i + Q_i R_i)}{2\sum\limits_{i=1}^{k} R_i \sum\limits_{i=1}^{k} S_i} + \frac{\sum\limits_{i=1}^{k} Q_i S_i}{2\left(\sum\limits_{i=1}^{k} S_i\right)^2} \tag{2.16}$$

with $P_i = \dfrac{a_i + d_i}{n_i}$, $Q_i = \dfrac{b_i + c_i}{n_i}$, $R_i = \dfrac{a_i\, d_i}{n_i}$, and $S_i = \dfrac{b_i\, c_i}{n_i}$.

The Mantel-Haenszel method was extended to the risk ratio as measure of the treatment effect by Greenland and Robins (1985). The pooled risk ratio $\widehat{\phi}_{MH}$ is calculated by combining individual risk ratios $\widehat{\phi}_i$ on the natural scale using weights $w_i = \dfrac{(a_i + b_i)c_i}{n_i}$.

A robust estimator of the variance of the logarithm of $\widehat{\phi}_{MH}$ is given by the following formula (Greenland and Robins, 1985):

$$\widehat{\mathrm{Var}}(\log \widehat{\phi}_{MH}) = \frac{\displaystyle\sum_{i=1}^{k} \frac{(a_i+b_i)(c_i+d_i)(a_i+c_i) - a_i c_i n_i}{n_i^2}}{\displaystyle\sum_{i=1}^{k} \frac{a_i(c_i+d_i)}{n_i} \sum_{i=1}^{k} \frac{c_i(a_i+b_i)}{n_i}}. \tag{2.17}$$

A 95% confidence interval for $\widehat{\psi}_{MH}$ or $\widehat{\phi}_{MH}$ can be calculated by replacing $\log \widehat{\psi}_{IV}$ and S.E.$(\log \widehat{\psi}_{IV})$ in equation (2.14) with the respective quantities.

For the Mantel-Haenszel method, there is no need to add 0.5 to each cell of two-by-two tables with zero cell counts. Nevertheless, this modification is utilised in commonly used software, e.g., in the Stata procedure `metan` (StataCorp, 2003) and in Review Manager, the program for preparing and maintaining Cochrane reviews (The Cochrane Collaboration, 2003).

Especially for pooling of odds ratios, the Mantel-Haenszel approach has been recommended as method of choice in most situations (Emerson, 1994). Accordingly, the Mantel-Haenszel method is the default procedure in Review Manager (The Cochrane Collaboration, 2003) for meta-analysis of binary data.

Peto method

The Peto method can also be used to calculate a common odds ratio:

$$\widehat{\psi}_{Peto} = \exp\left(\frac{\sum_{i=1}^{k} w_i \cdot \log \widehat{\psi}_i^*}{\sum_{i=1}^{k} w_i} \right) \tag{2.18}$$

with $w_i = 1/\widehat{\mathrm{Var}}(\log \widehat{\psi}_i^*)$; $\widehat{\psi}_i^*$ and $\widehat{\mathrm{Var}}(\log \widehat{\psi}_i^*)$ as defined in (2.10) and (2.11), respectively.

The variance of the log odds ratio is given by

$$\widehat{\mathrm{Var}}(\log \widehat{\psi}_{Peto}) = \frac{1}{\sum_{i=1}^{k} \widehat{\mathrm{Var}}(\log \widehat{\psi}_i^*)} \tag{2.19}$$

Again, a 95% confidence interval for $\widehat{\psi}_{Peto}$ can be calculated by replacing $\log \widehat{\psi}_{IV}$ and S.E.$(\log \widehat{\psi}_{IV})$ in equation (2.14) with the respective quantities.

For the Peto method to combine individual log odds ratio, no correction for zero cell counts is necessary. The method performs well in meta-analysis with very sparse data (Deeks et al., 1998).

2.2.2 Random effects model

In a random effects model, the assumption of a constant treatment effect across trials is relaxed by allowing that individual treatment effects may differ from the pooled treatment effect according to a specific distribution; typically, according to a normal distribution with mean (logarithm of common odds/risk ratio) and variance τ^2, e.g., for the odds ratio:

$$\log \psi_i \sim N(\log \psi, \tau^2).$$

The between-trial variance τ^2 describes the extent of heterogeneity between individual trial results; other methods to quantify heterogeneity have been introduced by Higgins and Thompson (2002).

A number of methods are available to estimate the between-trial variance τ^2 (Thompson and Sharp, 1999). The most popular method is probably that by DerSimonian and Laird (1986) which is available in Review Manager (The Cochrane Collaboration, 2003) and in the Stata procedures `meta` and `metan` (StataCorp, 2003).

DerSimonian and Laird (1986) proposed a non-iterative, moment-based estimator for the between-trial variance τ^2

$$\widehat{\tau}^2 = \frac{Q-(k-1)}{\sum_{i=1}^{k} w_i - \frac{\sum_{i=1}^{k} w_i^2}{\sum_{i=1}^{k} w_i}}, \tag{2.20}$$

with heterogeneity statistic $Q = \sum_{i=1}^{k} w_i(\log \widehat{\psi}_i - \log \widehat{\psi}_{IV})^2$ and w_i as defined in equation (2.12). The estimator $\widehat{\tau}^2$ is set to zero if $Q < k-1$.

The pooled odds ratio in the random effects model is calculated by

$$\widehat{\psi}_{DL} = \exp\left(\frac{\sum_{i=1}^{k} w_i^* \cdot \log \widehat{\psi}_i}{\sum_{i=1}^{k} w_i^*}\right) \tag{2.21}$$

with weights $w_i^* = (\widehat{\mathrm{Var}}(\log \widehat{\psi}_i) + \widehat{\tau}^2)^{-1}$. A variance estimator of $\log \widehat{\psi}_{DL}$ and a 95% confidence interval for $\widehat{\psi}_{DL}$ can be derived from equation (2.13) and (2.14) by replacing w_i with w_i^*. If $\tau^2 = 0$, the DerSimonian and Laird estimator corresponds to the inverse variance method given in equation (2.12).

2.3 Existing methods for the detection of bias in meta-analysis

Graphical methods for the detection of bias are described in Section 2.3.1 and 2.3.2. The two most commonly used tests for bias in meta-analysis are described in Section 2.3.3 and 2.3.4; regardless of the method used to calculate the pooled odds ratio, only information from the inverse variance method, see equation (2.12), is utilised in these tests. Although, all methods are solely introduced for the odds ratio as measure of the treatment effect all statements apply also to the risk ratio as measure of the treatment effect.

2.3.1 Funnel plot

The most prominent graphical method to check informally the presence of bias in meta-analysis is a funnel plot which is a scatter plot of the estimated treatment effect on the horizontal axis and a measure of the precision of the treatment estimate on the vertical axis (Light and Pillemer, 1984). Guidelines on choice of axis have been suggested by Sterne and Egger (2001). They recommend to use relative effect measures on the horizontal axis, i.e.

$$x_i = \log \widehat{\psi}_i.$$

Furthermore, the use of the standard error is recommended for the vertical axis, i.e.

$$y_i = \text{S.E.}(\log \widehat{\psi}_i).$$

The display looks like a funnel if neither bias nor excess between-trial heterogeneity exists showing decreasing fluttering with increasing precision. This is illustrated by using an artificial example in Figure 2.1. Trial results from smaller, less precise trials at the bottom of the graph scatter more widely than larger trials. Overall, individual trial results differ from the pooled treatment effect, indicated by the dotted line, at random. A meta-analysis of the 15 trials using the inverse variance method results in a pooled odds ratio $\widehat{\psi}_{IV} = 0.78$ with 95% confidence interval $[0.51; 1.18]$, p-value $= 0.23$.

In Figure 2.2, it is assumed that all four trials with negative results for the new treatment as well as one small trial with a result very close to the null effect line did not get published. A meta-analysis of the remaining 10 published trials yields a pooled odds ratio $\widehat{\psi}_{IV} = 0.62$ with 95% confidence interval $[0.38; 1.01]$, p-value $=$ 0.055. Accordingly, the estimated odds ratio is exaggerated by about 20% if only

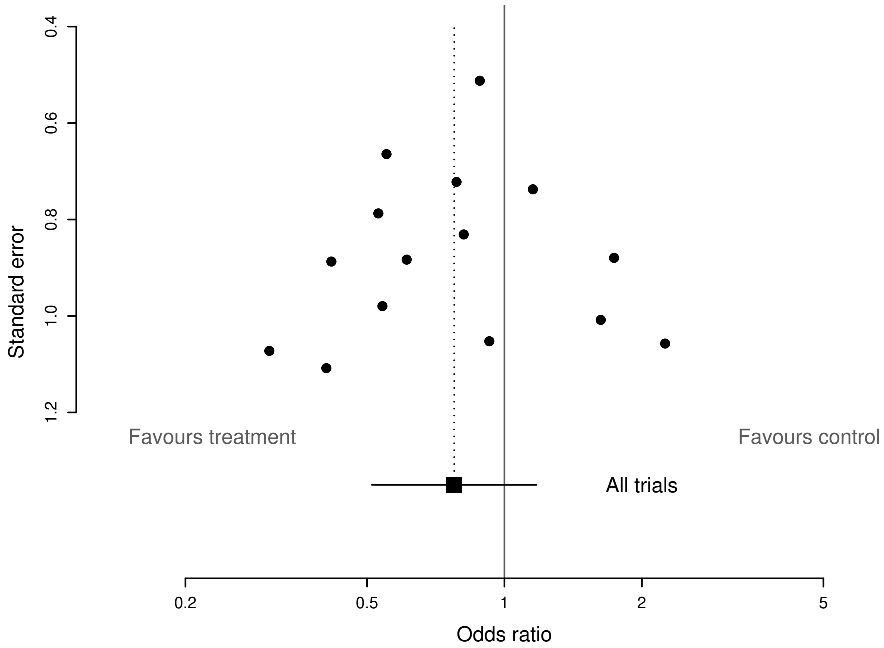


Figure 2.1: Funnel plot of an artificial example with 15 trials; pooled odds ratio and 95% confidence interval for inverse variance method included in the plot; pooled odds ratio is indicated by dotted line.

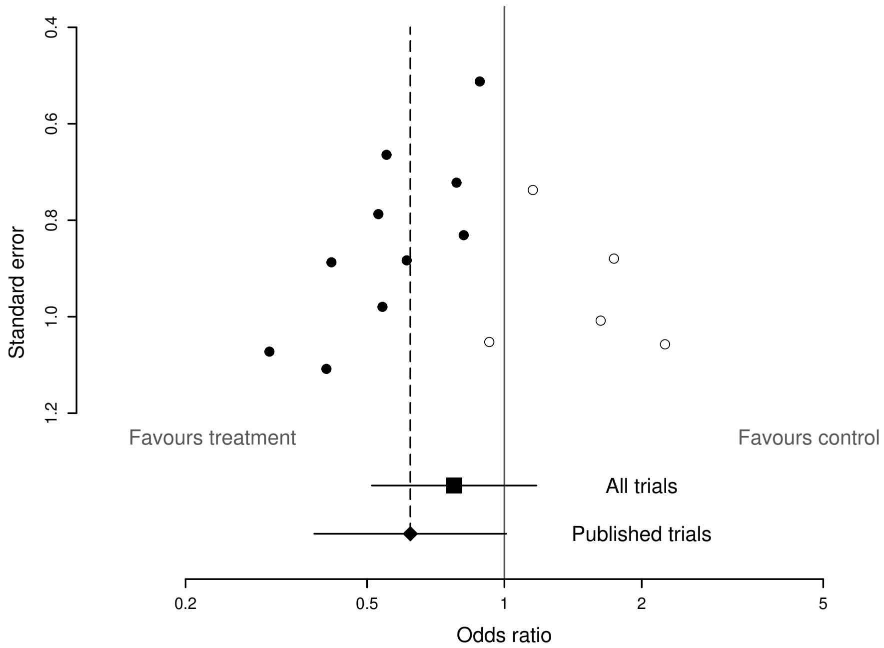


Figure 2.2: Funnel plot of an artificial example with 15 trials; published trials are plotted as filled points and unpublished trials as circles; pooled odds ratio of published trials and 95% confidence interval for inverse variance method included in the plot; pooled odds ratio of published trials is indicated by dashed line.

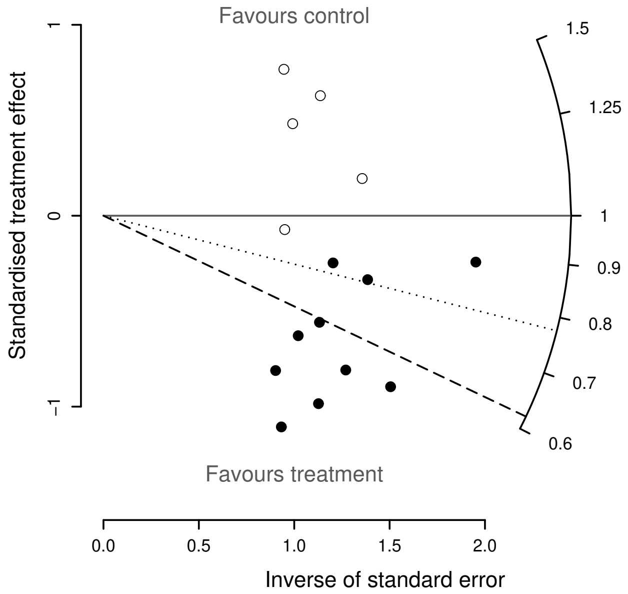


Figure 2.3: Radial plot of an artificial example with 15 trials; published trials are plotted as filled points and unpublished trials as circles; pooled odds ratio of all trials is indicated by dotted line, pooled odds ratio of published trials is indicated by dashed line.

published trials are considered. Furthermore, this meta-analysis almost reaches statistical significance using the conventional significance level of 0.05.

It is apparent from Figure 2.2 that results of published trials do not differ randomly from the pooled treatment effect which is indicated by the dashed line. The estimated odds ratio of small, less precise trials is below the pooled odds ratio whereas the estimated odds ratio in large trials is above the pooled odds ratio. Accordingly, asymmetry in the funnel plot is taken as an indication of bias in meta-analysis. The rank correlation test described in Section 2.3.3 relies heavily on this association.

2.3.2 Radial plot

Galbraith (1988a) introduced a graphical method in order to display point estimates with different standard errors. An additional paper (Galbraith, 1988b) is focused on medical applications and the use of the log odds ratio as measure of the treatment effect. If the number of estimates is large, a radial plot has certain advantages as compared to other displays like a forest plot which is commonly used for summarising the result of a meta-analysis.

For the odds ratio, a scatter plot of

$$x_i = 1/\text{S.E.}(\log \widehat{\psi}_i)$$

and

$$y_i = \log \widehat{\psi}_i/\text{S.E.}(\log \widehat{\psi}_i)$$

is called a radial plot and has under a fixed effect model the following properties (Galbraith, 1988a):

(a) $\text{Var}(y_i)$ is equal to 1,

(b) for each trial, the estimated log odds ratio, $\log \widehat{\psi}_i$, is equal to the slope of a line through the origin (0,0) and the point (x_i, y_i),

(c) points are close to zero on the x-axis for large $\text{S.E.}(\log \widehat{\psi}_i)$,

(d) logarithm of pooled odds ratio using the inverse variance method, $\log \widehat{\psi}_{IV}$, see (2.12), equals the slope β_1 in a linear regression going through the origin $y_i = \beta_1 \cdot x_i$.

Due to properties (b) and (d) a circular scale is typically displayed on the right-hand side of a radial plot making it easier to read off the estimated treatment effect. The values y_i are often called z-scores because they correspond to the statistic of a test that $\log \psi_i$ is different from zero. Sometimes, $y_i^* = (\log \widehat{\psi}_i - \log \widehat{\psi}_{IV})/\text{S.E.}(\log \widehat{\psi}_i)$ is plotted against x_i to get a better visual discrimination. In this case, the estimated overall effect coincides with the horizontal axis and departures from the overall effect are more apparent.

A radial plot of the artificial example is depicted in Figure 2.3. In this figure, the gap in the upper left corner is not as obvious as in the funnel plot. The linear regression test described in Section 2.3.4 is strongly connected to a radial plot.

2.3.3 Rank correlation test

An adjusted rank correlation test was proposed by Begg and Mazumdar (1994) to test for publication bias in meta-analysis; the power of the test was evaluated via simulations assuming a normal distribution for the estimated treatment effect. The test is based on the correlation between a standardised treatment effect and the within-trial variance; the inverse variance method is used in the construction of the test. For the odds ratio as measure of the treatment effect, the corresponding quantities are the variance of the log odds ratio, $\widehat{\mathrm{Var}}(\log \widehat{\psi}_i)$, and a standardised treatment effect

$$s(\widehat{\psi}_i) = \left(\log \widehat{\psi}_i - \log \widehat{\psi}_{IV}\right) \Big/ \sqrt{\widehat{\mathrm{Var}}(\log \widehat{\psi}_i) - \widehat{\mathrm{Var}}(\log \widehat{\psi}_{IV})}\ .$$

Begg and Mazumdar (1994) utilised Kendall's tau as correlation measure. Other rank correlation coefficients like Spearman's rho which is probably better known might be used alternatively (Begg, 1994).

Let x denote the number of pairs of trials with standardised effects and variances ranked in the same order, i.e.

$$\left(s(\widehat{\psi}_i) > s(\widehat{\psi}_j) \text{ and } \widehat{\mathrm{Var}}(\log \widehat{\psi}_i) > \widehat{\mathrm{Var}}(\log \widehat{\psi}_j)\right)$$

or

$$\left(s(\widehat{\psi}_i) < s(\widehat{\psi}_j) \text{ and } \widehat{\mathrm{Var}}(\log \widehat{\psi}_i) < \widehat{\mathrm{Var}}(\log \widehat{\psi}_j)\right),$$

for $i \neq j$. The number of pairs ranked in the opposite order are denoted by y. The normalised test statistic for the case that no ties neither within $s(\widehat{\psi}_i)$ nor $\widehat{\mathrm{Var}}(\log \widehat{\psi}_i)$ are present is

$$z = \frac{x - y}{\sqrt{k(k-1)(2k+5)/18}} \tag{2.22}$$

where k is the number of trials involved in the analysis.

In the case of ties, i.e. $s(\widehat{\psi}_i) = s(\widehat{\psi}_j)$ or $\widehat{\mathrm{Var}}(\log \widehat{\psi}_i) = \widehat{\mathrm{Var}}(\log \widehat{\psi}_j)$, the standard error given in the denominator in equation (2.22) requires some modification; a modified version for tied observations can be found in Kendall and Gibbons (1990, p. 66).

The test statistic z is asymptotically distributed according to a standard normal distribution under the null hypothesis of no bias in meta-analysis. The null hypothesis of no bias in meta-analysis is rejected at the significance level α if

$$|z| > z_{1-\alpha/2},$$

with $z_{1-\alpha/2}$ denoting the $(1 - \alpha/2)$ quantile of the standard normal distribution.

Alternatively to conducting a statistical test, an estimate of Kendall's tau with $100 \cdot (1 - \alpha)\%$ confidence interval could be reported in a systematic review. I am not aware of any publication reporting these quantities.

Sparse binary data

As described in Section 2.2.1, three meta-analytic strategies exist to estimate the log odds ratio and its variance in sparse binary data: add 0.5 to all two-by-two tables ("add all"), add 0.5 to all two-by-two tables only in the case of zero cell counts in one or more trials ("add all selective"), or add 0.5 only to cell counts of corresponding two-by-two tables with zero cell counts ("add selective"). Accordingly, any of these approaches can be used in the rank correlation test.

2.3.4 Linear regression test

The test proposed by Egger et al. (1997) for the detection of bias in meta-analyses is strongly connected to a radial plot. The test is based on a simple linear regression:

$$y_i = \beta_0 + \beta_1 \cdot x_i,$$

with

$$x_i = 1/\text{S.E.}(\log \widehat{\psi}_i)$$

and

$$y_i = \log \widehat{\psi}_i / \text{S.E.}(\log \widehat{\psi}_i).$$

In contrast to the radial plot, the regression line is not constrained to run through the origin. In fact, the test is constructed by testing for a non-zero intercept β_0 which is asymptotically distributed according to a t distribution with $k-2$ degrees of freedom under the null hypothesis of no bias in meta-analysis.

The approach is justified by the intuitive argument that, in the presence of publication bias, small trials with non-significant or negative results are less likely to get published. Thus, points close to zero on the x-axis do not scatter randomly around the overall effect resulting in a non-zero intercept which is a departure from property (d) of a radial plot described in Section 2.3.2.

The null hypothesis of no bias in meta-analysis is rejected at the significance level α if

$$\left| \widehat{\beta}_0 / \text{S.E.}(\widehat{\beta}_0) \right| > t_{k-2;1-\alpha/2}$$

with $t_{k-2;1-\alpha/2}$ denoting the $(1-\alpha/2)$ quantile of the t distribution with $k-2$ degrees of freedom. The test procedure is implicitly based on the assumption that linearity still holds in the presence of bias.

Sparse binary data

Again, as described in Section 2.2.1, three meta-analytic strategies exist to estimate the log odds ratio and its standard error in sparse binary data which accordingly can be used in the linear regression test.

2.3.5 Other methods for identifying bias in meta-analysis

Several other methods for the detection of bias in meta-analysis have been proposed in the literature, however, the uptake of these methods in systematic reviews is very small, at least in the medical field. This section contains a brief overview of some of these methods. A more comprehensive review can be found in Thornton and Lee

(2000) and (Sutton et al., 2000). The methods described in this section are not further considered in this thesis.

Rosenthal (1979) proposed a method for the estimation of the minimum number of unpublished studies N_{FS} with on average non-significant results that are necessary to turn the significant overall effect in a meta-analysis of k studies to non-significance; the number N_{FS} is called fail-safe N. The method is based on the combination of z-scores Z_i, $i = 1, \dots, k$, which in the context of randomised controlled trials usually come from the statistic of a test for treatment difference. The overall z-score Z is defined as $Z = \sum_{i=1}^{k} Z_i/\sqrt{k}$. The fail-safe N is calculated under the assumption that the average z-score of N_{FS} unpublished studies is zero: $\sum_{i=1}^{N_{FS}} Z_{k+i} = 0$. The decision on the existence of publication bias is based on an informal and subjective comparison of the number of studies included in the meta-analysis and the calculated fail-safe N. An extension of Rosenthal's method to meta-analysis of standardised mean differences was published by Orwin (1983).

Gleser and Olkin (1996) introduced two general models for the estimation of the number of unpublished studies. These models defined in terms of the p-values reported in the published studies are based on the following assumptions: the k published and N unpublished studies are mutually statistically independent; the p-value in each study is based on a continuous test statistic; the null hypothesis tested is true, i.e. the new intervention compared to control is ineffective. The first general model makes certain assumptions on the order of the published p-values, e.g., by assuming that studies with the k smallest p-values among the $k + N$ studies are published. The second general model assumes that the probability of publishing a study can be described as a function $w(p)$ of the underlying p-value; the function $w(\cdot)$ is called the selection function. In order to estimate the number of missing studies either the proportion of published studies has to be known or some information about the selection function $w(\cdot)$ must be present. The estimated number of missing trials may differ considerably from Rosenthal's fail safe N (Gleser and Olkin, 1996).

A third method to estimate the number of missing studies in meta-analysis is the non-parametric trim-and-fill method (Duval and Tweedie, 2000a,b). The method relies heavily on the symmetry of the funnel plot. It is assumed that studies with the most extreme negative treatment estimates have not been published. An iterative rank-based algorithm is used to estimate the number of missing studies. The method can also be used to adjust for bias in meta-analysis. It has been shown recently that this method spuriously adjusts for non-existing publication bias in the presence of heterogeneity (Terrin et al., 2003).

A different approach to the problem of bias in meta-analysis consists in attempts to model the selection process causing publication bias. Selection models using weighted distribution theory have been under development for this purpose for almost two decades. Weight functions were introduced into meta-analysis by Hedges (1984), following work published by Lane and Dunlap (1978). The weight function determines the probability that a trial result will be published; in most cases, it is assumed that the weight function is a function of the p-value of a trial. Either a specific functional form for the weight function is predetermined (Iyengar and Greenhouse, 1988) or estimated non-parametrically from the observed p-values (Hedges, 1992). Typically, trials with small p-values are assumed to be more likely to get published. Selection models have been criticised for making too simplistic assumptions with respect to the probability of publication; furthermore, these assumptions cannot be verified in the analysis (Sutton et al., 2000).

Selectivity models (Copas, 1999; Copas and Shi, 2000, 2001) have been introduced as an alternative to selection models. The probability of selection is modelled more realistically by considering both the size of the treatment effect and the precision (i.e. sample size). It is not possible to estimate all parameters in selectivity models and therefore it is suggested to use the method in a sensitivity analysis by considering different patterns of selection bias. So far, the properties of selectivity models have not been investigated in much detail. Furthermore, only few applications have been published probably due to the complexity of the method and a lack of user-friendly software to conduct the analyses.

2.4 A new test for bias in meta-analysis with binary outcome data

As described in Section 2.1, estimation of the variance of the log odds ratio is problematic in sparse binary data. Breslow (1981) has shown that the pooled odds ratio using the inverse variance method is a biased estimate in sparse binary data. In the context of tests for bias in meta-analysis, an inflation of type I error rates with increasing treatment effect has been shown in simulation studies for both the linear regression and rank correlation test (Macaskill et al., 2001; Schwarzer et al., 2002). It has been argued by Schwarzer et al. (2002) that this inflation is due to the use of the estimated variance of the log odds ratio. In Section 4.1.3 of this thesis, it is shown that an association between estimated treatment effect and its standard error is in fact a plausible explanation for the inflation of type I error rates. Accordingly, a statistical test for bias in meta-analysis with binary outcomes was sought that does not utilise the rather unreliable variance estimate of the log odds ratio given in equation (2.3) and (2.7).

The underlying idea of the rank correlation test based on the correlation between a standardised treatment effect and a variance estimate is quiet intuitive. Therefore, I thought about a statistical test similar to the rank correlation test which in particular is using a different variance estimate; this led to a test based on the non-central hypergeometric distribution.

Conditional on fixed marginal totals in Table 2.1, the probability of observing cell count a_i follows a non-central hypergeometric distribution:

$$\Pr^*(a_i|\ldots;\psi_i) = \frac{\binom{a_i+b_i}{a_i}\binom{c_i+d_i}{c_i}\psi_i^{a_i}}{\sum_{x\in U_i}\binom{a_i+b_i}{x}\binom{c_i+d_i}{a_i+c_i-x}\psi_i^{x}}$$

with '...' denoting the fixed marginal totals and U_i the range of possible values for a_i given the fixed marginal totals (Breslow and Day, 1980). Define $\Pr(a_i;\psi_i) := \Pr^*(a_i|\ldots;\psi_i)$ to simplify notation.

The mean and variance of the non-central hypergeometric distribution are given by

$$\mathrm{E}(a_i;\psi_i) = \sum_{x\in U_i} x\cdot\Pr(x;\psi_i)$$

and

$$\mathrm{Var}(a_i;\psi_i) = \sum_{x\in U_i}(x-\mathrm{E}(a_i;\psi_i))^2\cdot\Pr(x;\psi_i).$$

The calculation of the mean and variance of the non-central hypergeometric distribution for given ψ_i is cumbersome for large trial sizes due to the combinatorial numbers involved in the formulae. Several approximations to the conditional distribution of a_i exist which facilitate the calculation of mean and variance (Plackett, 1974, p. 37). A Fortran program to calculate mean and variance is available since about a decade (Liao, 1992). Recently, by the same author, another fast and stable algorithm was proposed to calculate these quantities (Liao and Rosen, 2001). An implementation of the algorithm is available for the statistical software package R (R Development Core Team, 2003) and is used in this thesis.

For a single study, the mean and variance of the non-central hypergeometric distribution are typically utilised in the estimation of the conditional maximum likelihood estimator of the odds ratio $\widehat{\psi}_i^{ML}$ satisfying the following equations:

$$a_i - \mathrm{E}(a_i;\widehat{\psi}_i^{ML}) = 0 \tag{2.23}$$

and

$$\widehat{\mathrm{Var}}(\log\widehat{\psi}_i^{ML}) \sim 1/\mathrm{Var}(a_i;\psi_i), \tag{2.24}$$

see Plackett (1974), p. 39, for further details.

The relationship between the variance of the log odds ratio and the variance of cell count a_i given in equation (2.24) is utilised in the construction of an alternative test. My idea is to replace the unreliable variance estimate $\widehat{\mathrm{Var}}(\log \widehat{\psi}_i)$ used in the rank correlation test by Begg and Mazumdar (1994) with the quantity

$$1/\mathrm{Var}(a_i; \widehat{\psi}_{MH}),$$

i.e. a variance estimate of the log odds ratio based on the non-central hypergeometric distribution with the Mantel-Haenszel odds ratio $\widehat{\psi}_{MH}$ defined in equation (2.15) plugged-in for the unknown odds ratio ψ_i. Thus, variance estimates for individual trials are derived from the non-central hypergeometric distribution with the Mantel-Haenszel odds ratio $\widehat{\psi}_{MH}$ and the marginal totals as parameters. Marginal totals are completely defined for example by using n_i, $a_i + b_i$ and $a_i + c_i$, i.e. total sample size, sample size in intervention group, and total number of events .

The use of the Mantel-Haenszel odds ratio as a plug-in estimate seems to be perfectly reasonable in a fixed effect model, whereas, the impact of using this method in a random effects model is rather unclear. This topic will be investigated in the simulation studies conducted in this thesis.

The variance estimate $\widehat{\mathrm{Var}}(\log \widehat{\psi}_i)$ is also utilised in the calculation of the second quantity considered in the rank correlation test by Begg and Mazumdar (1994), i.e. the standardised treatment effect. One option would be to solely replace the variance estimate $\widehat{\mathrm{Var}}(\log \widehat{\psi}_i)$ with $1/\mathrm{Var}(a_i; \widehat{\psi}_{MH})$ in the calculation of the standardised treatment effect. However, a different approach is used in the new test utilising the fact that under the non-central hypergeometric distribution an explicit functional relationship exists between cell count and log odds ratio (Plackett, 1974, p. 37). The second quantity considered in the new test is based on a standardised form of the difference between observed and expected cell count:

$$s(a_i; \widehat{\psi}_{MH}) = \frac{a_i - \mathrm{E}(a_i; \widehat{\psi}_{MH})}{\sqrt{\mathrm{Var}(a_i; \widehat{\psi}_{MH})}}.$$

Under the null hypothesis of no bias in meta-analysis, the observed cell count a_i should deviate randomly from the expected cell count. However, in the case of selection bias, observed cell counts of less precise trials are expected to be smaller than the expected cell count which corresponds to an exaggerated treatment effect, whereas observed cell counts of more precise trials are expected to be larger than the respective expected cell count.

Accordingly, an alternative test for the detection of bias in meta-analysis is proposed which is based on the correlation between the standardised cell count $s(a_i; \widehat{\psi}_{MH})$ and the variance estimate $1/\mathrm{Var}(a_i; \widehat{\psi}_{MH})$. These quantities are used in the rank correlation test instead of the factors proposed by Begg and Mazumdar (1994). Otherwise, the test is conducted essentially in the same way. Especially, I also used Kendall's tau as correlation measure instead of Spearman's rho. The main advantage of Kendall's tau is that its distribution approaches the normal distribution more rapidly under the null hypothesis of independence than the distribution of Spearman's rho (Conover, 1971, p. 249). This property is clearly an advantage in meta-analysis with a limited number of trials.

Let x denote the number of pairs of trials with standardised effects and variances ranked in the same order, i.e.

$$\left(s(a_i; \widehat{\psi}_{MH}) > s(a_j; \widehat{\psi}_{MH}) \text{ and } 1/\mathrm{Var}(a_i; \widehat{\psi}_{MH}) > 1/\mathrm{Var}(a_j; \widehat{\psi}_{MH})\right)$$

or

$$\left(s(a_i; \widehat{\psi}_{MH}) < s(a_j; \widehat{\psi}_{MH}) \text{ and } 1/\mathrm{Var}(a_i; \widehat{\psi}_{MH}) < 1/\mathrm{Var}(a_j; \widehat{\psi}_{MH})\right),$$

for $i \neq j$. The number of pairs ranked in the opposite order are denoted by y. The normalised test statistic for the case that no ties neither within $s(a_i; \widehat{\psi}_{MH})$ nor $1/\mathrm{Var}(a_i; \widehat{\psi}_{MH})$ are present is given in equation (2.22). A modified version for tied observations, i.e. $s(a_i; \widehat{\psi}_{MH}) = s(a_j; \widehat{\psi}_{MH})$ or $1/\mathrm{Var}(a_i; \widehat{\psi}_{MH}) = 1/\mathrm{Var}(a_j; \widehat{\psi}_{MH})$, can be found in Kendall and Gibbons (1990, p. 66).

The resulting test statistic z is asymptotically distributed according to a standard normal distribution under the null hypothesis of no bias in meta-analysis. Therefore, the null hypothesis of no bias in meta-analysis is rejected at the significance level α if

$$|z| > z_{1-\alpha/2},$$

with $z_{1-\alpha/2}$ denoting the $(1-\alpha/2)$ quantile of the standard normal distribution.

Chapter 3

Design of simulation study

Aim of the simulation study is to evaluate the properties of the new test (described in Section 2.4) both under the null hypothesis of no publication bias and different alternatives. Furthermore, the power of the new test is compared to the rank correlation test (Section 2.3.3) and linear regression test (Section 2.3.4).

The simulation design used in this thesis should be realistic and represent meta-analyses published in medical journals. Therefore, empirical information was considered in the decision on the simulation design. In Section 3.1, a review of published simulation studies evaluating tests for bias in meta-analysis with binary outcome is given; all publications used empirical data in an attempt to create realistic meta-analyses. The actual design used in this thesis is described in Section 3.2.

3.1 Design of published simulation studies

In the paper proposing the rank correlation test for bias in meta-analysis, results of a simulation study evaluating the power of this test were also published (Begg and Mazumdar, 1994). However, a normal distribution was assumed for the estimated treatment effect in the simulation study, even though all three examples, from medical research, were meta-analyses with binary outcome data. No simulation results were published in the paper introducing the linear regression test (Egger et al., 1997). In the discussion of this paper, conflicting results of simulations based on binary outcome data were reported (Irwig et al., 1998).

Since 1997, three simulation studies were published evaluating the linear regression and rank correlation test for bias in meta-analysis with binary outcome data (Sterne et al., 2000; Macaskill et al., 2001; Schwarzer et al., 2002). Recently, a simulation study evaluating the trim-and-fill method which was described briefly in Section 2.3.5 was published (Terrin et al., 2003); the design of this study can also be used

Reference	Empirical data	Number of trials	Treatment effect	Event rate	Sparse data	Alternative
Sterne et al. (2000)	Characteristics of 78 meta-analyses published between 1993 and 1997	5, 10, 20, 30	OR: $0.25, 0.50, 1.00$	Control group: $0.05, 0.10, 0.20$	add selective	Linear regression
Macaskill et al. (2001)	Characteristics of 70 meta-analyses published between 1990 and 1995	21	OR: $0.25, 0.50,$ $0.67, 1.00$	Control group: $U[0.10; 0.50]$	add all	Based on p-value
Schwarzer et al. (2002)	Empirical distr. of 1555 trials published between 1948 and 1998	10, 20, 50	OR/RR: $0.50, 0.67, 1.00$	Average rate: $0.10, 0.30$	add selective	—
Terrin et al. (2003)	Characteristics of 125 meta-analyses published between 1990 and 1996	10, 25	OR: $0.50, 0.80, 1.00$	Control group: $0.15, 0.30$	unclear	—

Table 3.1: Simulation designs (under homogeneity)

to evaluate the linear regression and rank correlation test. The key features of these four simulation studies for homogeneous meta-analysis are summarised in Table 3.1.

Use of empirical information

Sterne et al. (2000) defined four hypothetical meta-analyses containing 5, 10, 20, and 30 trials. Sample sizes of individual trials were generated according to findings in 78 meta-analyses based on at least five trials with binary outcomes published in four general medicine journals (*Annals of Internal Medicine, BMJ, JAMA, Lancet*) and four specialist journals (*American Journal of Cardiology, Cancer, Circulation, Obstetrics and Gynecology*) between 1993 and 1997.

Macaskill et al. (2001) used characteristics observed in 70 meta-analyses published in six general medicine journals (*Annals of Internal Medicine, Archives of Internal Medicine, BMJ, JAMA, Lancet, New England Journal of Medicine*) and one specialist journal (*Circulation*) between 1990 and 1995; this data as well as additional information from 45 meta-analyses published in the *Cochrane Database of Systematic Reviews* in 1994 was used before in an empirical study of the effect of the control event rate as predictor of treatment efficacy (Schmid et al., 1998). No information is given in Macaskill et al. (2001) how the empirical evidence was utilised in the simulations. For example, meta-analyses with 21 trials were generated in the simulation study, but, the median number of trials in published meta-analysis was only 11.5 (Schmid et al., 1998).

Schwarzer et al. (2002) utilised information from a hand-searching project at the German Cochrane Centre in Freiburg (Dr. Daniel Galandi, personal communication). The empirical distribution from 1555 primary randomised controlled trials published in eight German language general health care journals (*Deutsche Medizinische Wochenschrift, Journal of Molecular Medicine, Medizinische Klinik, Medizinische Welt, Münchener Medizinische Wochenschrift, Schweizer Medizinische Wochenschrift, Wiener Medizinische Wochenschrift, Zeitschrift für Allgemeinmedizin*) between 1948 and 1998 was used to generate sample sizes of individual trials in meta-analyses. The number of trials in meta-analysis was chosen as 10, 20, and 50, respectively.

The simulation study of Terrin et al. (2003) is based on data from an empirical study investigating the association of the choice of the measure of treatment effect and heterogeneity in 125 meta-analyses with binary endpoints (Engels et al., 2000). This set actually contains the 70 meta-analyses considered in the simulation study by Macaskill et al. (2001). Information from 10 additional meta-analyses published in 1996 in the seven medical journals mentioned above and from 45 meta-analyses published in the *Cochrane Database of Systematic Reviews* in 1994 were also used

Sample size	Scenario A ($k = 5$)	Scenario B ($k = 10$)	Scenario C ($k = 20$)	Scenario D ($k = 30$)
	Number of trials			
24–49	1	1	2	3
50–74	1	1	3	3
75–99	0	1	3	3
100–149	1	2	3	5
150–249	1	1	3	4
250–499	0	2	3	6
500–999	1	1	2	3
1000–4999	0	1	1	3

Table 3.2: Distribution of sample sizes (Sterne et al., 2000)

to set several parameters for the simulations: number of trials per meta-analysis, sample sizes of individual trials, mean and variance of control event rate, and the residual variance. The number of trials in meta-analysis was chosen as 10 and 25.

Generation of sample sizes

A peculiar problem in the design of a simulation study in a meta-analytic setting is to decide on how to generate the sample size of individual trials. In three publications (Sterne et al., 2000; Schwarzer et al., 2002; Terrin et al., 2003), empirical information was incorporated in this decision; details on available data are described in the last subsection entitled "Use of empirical information".

Total sample sizes used in the simulation study by Sterne et al. (2000) are given in Table 3.2. Macaskill et al. (2001) fixed the total sample size of individual trials in advance and used two scenarios (see Table 3.3). In Schwarzer et al. (2002), logarithm of total sample size for each trial was drawn from a normal distribution with mean 3.798 and variance 1.104. The sample size was rounded to the next even number to get treatment groups of equal size. Only trials with total sample size of at least 30 patients were considered in the analysis. Terrin et al. (2003) considered three different scenarios. The logarithm of total sample size for each trial was drawn from a uniform distribution on

- $\log 50$ to $\log 500$ (Scenario A),
- $\log 100$ to $\log 1500$ (Scenario B),
- $\log 100$ to $\log 10000$ (Scenario C).

Sample size per group	Scenario A ($k = 21$)	Scenario B ($k = 21$)
	Number of trials	
100 / 100	11	10
200 / 200	6	5
300 / 300	4	3
500 / 500	–	2
1000 / 1000	–	1

Table 3.3: Distribution of sample sizes (Macaskill et al., 2001)

A crude summary of generated sample sizes in the four simulation studies is given in Table 3.4. The percentage of small trials (total sample size less than 100) varies greatly between different simulation studies ranging from 0% to 66%. Within a publication, the largest variation in sample sizes is apparent in the simulation study by Terrin et al. (2003). Even though based on similar empirical evidence, sample sizes generated in the simulation studies by Macaskill et al. (2001) and Terrin et al. (2003) are rather different.

Overall, the whole range of generated sample sizes is covered by the two design utilised in Schwarzer et al. (2002) and Terrin et al. (2003).

Treatment effect

Schwarzer et al. (2002) considered both odds ratio and risk ratio as measures of treatment effect; the other publications did only use the odds ratio as measure of treatment effect (see Table 3.1). Values chosen for the odds ratio were similar between publications ranging from 0.25 to 1.00. Due to the symmetry of the odds ratio, all simulations were restricted to values of the odds ratio less equal one.

Event rate

The control event rate was specified in all but one publication. Following the work of Smith et al. (1995), the average event rate was fixed by Schwarzer et al. (2002); corresponding control event rates can be calculated from the average event rate and the odds ratio, e.g., for an average event rate of 0.1, the control event rate is 0.136 and 0.12 for an odds ratio of 0.5 and 0.67, respectively. Overall, control event rates of 0.1 to 0.3 seem to be reasonable values.

Reference	Scenario	Total sample size < 100	100 – 500	500 - 1000	> 1000
Sterne et al. (2000)	A	40%	40%	20%	0%
	B	30%	50%	10%	10%
	C	40%	45%	10%	5%
	D	30%	50%	10%	10%
Macaskill et al. (2001)	A	0%	81%	19%	0%
	B	0%	71%	24%	5%
Schwarzer et al. (2002)	–	66%	32%	2%	0%
Terrin et al. (2003)	A	30%	70%	0%	0%
	B	0%	60%	25%	15%
	C	0%	35%	15%	50%

Table 3.4: Generated sample sizes of individual trials

Handling of sparse binary data

In Section 2.2.1, three different meta-analytic strategies were described for sparse binary data: add 0.5 to all two-by-two tables ("add all"), add 0.5 to all two-by-two tables only in the case of zero cell counts in one or more trials ("add all selective"), or add 0.5 only to cell counts of corresponding two-by-two tables with zero cell counts ("add selective"). Two of these approaches were utilised separately in the published simulation studies. To my knowledge, no comparison of these strategies was published so far.

Alternatives

All publications evaluated the properties of tests for bias in meta-analysis under the null hypothesis of no publication bias. Two publications also evaluated the properties under alternatives.
Sterne et al. (2000) used the following regression equation to generate bias:

$$\text{log OR} = \text{true log OR} + (\text{bias coefficient} \times \text{standard error of log OR}).$$

This equation corresponds to the alternative hypothesis tested by the linear regression test. Accordingly, using this mechanism to compare the linear regression test with other tests seems somewhat unfair.

Macaskill et al. (2001) used the approach by Begg and Mazumdar (1994) to evaluate the properties of the tests under alternative hypotheses. The probability of publication was defined as a smooth exponential function

$$\exp\left(-b \cdot p^a\right)$$

with p denoting the one-sided P-value of a test of treatment difference. Moderate selection bias was generated by setting $a = 3$ and $b = 4$, and strong publication bias was generated by setting $a = 1.5$ and $b = 4$.

Generating heterogeneous meta-analyses

The influence of heterogeneity in meta-analysis was considered in two simulation studies (Schwarzer et al., 2002; Terrin et al., 2003) using rather different methods.

The random effects model described in Section 2.2.2 was utilised in Schwarzer et al. (2002). For each study, the logarithm of treatment effect was generated according to a normal distribution, i.e. $\log\psi_i \sim N(\log\psi, \tau^2)$ for the odds ratio as measure of treatment effect. For the odds ratio, values of τ^2 between 0.05 and 0.23 were utilised; for the risk ratio, values of τ^2 between 0.09 and 0.19 were used.

Treatment and control event rates, p_i^I and p_i^C, were calculated by

$$\text{logit}(p_i^I) = \text{logit}(p_A) + \log(\psi_i)/2$$

and

$$\text{logit}(p_i^C) = \text{logit}(p_A) - \log(\psi_i)/2$$

with p_A denoting the average event probability (Smith et al., 1995); for the risk ratio as measure of treatment effect the function $\text{logit}(\cdot)$ is replaced by $\log(\cdot)$.

In Terrin et al. (2003) a model incorporating the influence of the control event rate p_i^C on the outcome was used. Both the control event rate and the logarithm of the odds ratio, conditioned on p_i^C, are assumed to follow a normal distribution:

$$p_i^C \sim N(p_C, \tau_C^2)$$

and

$$\log(\psi_i | p_i^C) \sim N(\log\psi + \beta(p_i^C - p_C), \tau_{res}^2)$$

with average control rate p_C, parameter β denoting the slope of the regression of $\log\psi_i$ on p_i^C, and residual variance from the regression τ_{res}^2. For τ_{res}^2, values of 0.01 and 0.15 were utilised.

3.2 Simulation design

Referring to the discussion of properties of published simulation studies in Section 3.1, the simulation design described below is used in this thesis. For each combination, 1000 meta-analyses were generated (see Table 3.5). All simulations were conducted using the statistical software package R, version 1.8.1 (R Development Core Team, 2003) on Gnu/Linux Compute-Servers (2×Xeon 1.8GHz, Debian woody) at the Freiburg Centre for Modelling and Data Analysis (`http://www.fdm.uni-freiburg.de`).

Number of trials per meta-analysis / Sample sizes of individual trials

As described in Section 3.1, the whole range of generated sample sizes is covered by the simulation designs utilised in Schwarzer et al. (2002) and Terrin et al. (2003). Furthermore, the following properties of the simulation designs used in Sterne et al. (2000) and Macaskill et al. (2001) justify to neglect these two designs. In Sterne et al. (2000), different mechanisms to generate sample sizes are used for different numbers of trials. Accordingly, it is not easy to judge whether differences in test results for different numbers of trials per meta-analysis are due to the number of trials per meta-analysis or the mechanism used to generate sample sizes. The design utilised in Macaskill et al. (2001) is restricted to meta-analyses with 21 trials, thus, it is not possible to evaluate the influence of different numbers of trials on the properties of tests for bias. It seems sensible to consider only the mechanisms used in Schwarzer et al. (2002) and Terrin et al. (2003) to generate sample sizes; in this thesis, the two designs are labelled "Simulation I" and "Simulation II", respectively.

Schwarzer et al. (2002) and Terrin et al. (2003) utilised mechanism to generate sample sizes of individual trials that can be used for any number of trials per meta-analysis. I decided to generate meta-analyses with 5, 10, 20, and 50 trials. On the one hand, these numbers correspond to the range of number of trials used in the four simulation studies. On the other hand, the reported median number of trials per meta-analysis published in medical journals varies between 10 trials with interquartile range from 6 to 19 (McAuley et al., 2000) and 13.5 trials (Jadad et al., 1998); reported median number of trials per meta-analysis in Cochrane reviews varies between 5 trials (Jadad et al., 1998) and 6 trials with interquartile range from 3 to 12 (Mallett and Clarke, 2002). Accordingly, meta-analyses with 5, 10, and 20 trials are representative for published meta-analyses and simulation results for meta-analyses with 50 trials give an indication for asymptotic behaviour of the statistical tests.

Treatment effect

Odds ratio as well as risk ratio are utilised as measures of treatment effect. The following values were chosen for these effect measures corresponding to a large variety of treatment effects: 0.25, 0.50, 0.67, and 1.00.

Control event rate

Event probabilities are defined by using the control event rate; values of 0.1 and 0.3 are used in the simulation study.

Handling of sparse binary data

For each generated meta-analysis, both the "add selective" and "add all" method described in Section 2.2.1 were applied to deal with sparse binary data using the formulae (2.6) to (2.9); results of the two approaches are compared.

Alternative

The alternative based on a linear regression equation used in Sterne et al. (2000) is not considered in this thesis. Rather the approach by Macaskill et al. (2001) following work by Begg and Mazumdar (1994) is used to evaluate the properties of the tests under alternative hypotheses.

Alternative hypotheses considered in the simulations depend on one-sided P-values of a test for treatment difference, e.g., for the odds ratio as measure of treatment effect: $p_i = \Phi\left(\log\widehat{\psi}_i/\text{S.E.}(\log\widehat{\psi}_i)\right)$, with Φ denoting the cumulative distribution function of the standard normal distribution. The probability of publication is given by a smooth exponential function $\exp(-b \cdot p_i^a)$. Moderate selection bias was generated by setting $a = 3$ and $b = 4$, and strong publication bias was generated by setting $a = 1.5$ and $b = 4$.

Figure 3.1 shows the differences in the magnitude of bias induced. For example, the probability of publication of a trial with a P-value of 0.2 is 97% and 70% in the situation with moderate and strong selection bias, respectively. For an odds ratio or risk ratio of 1.00, the area under the curve corresponds to the overall probability of publication which is approximately 36% for strong selection bias and 56% for moderate selection bias (Begg and Mazumdar, 1994).

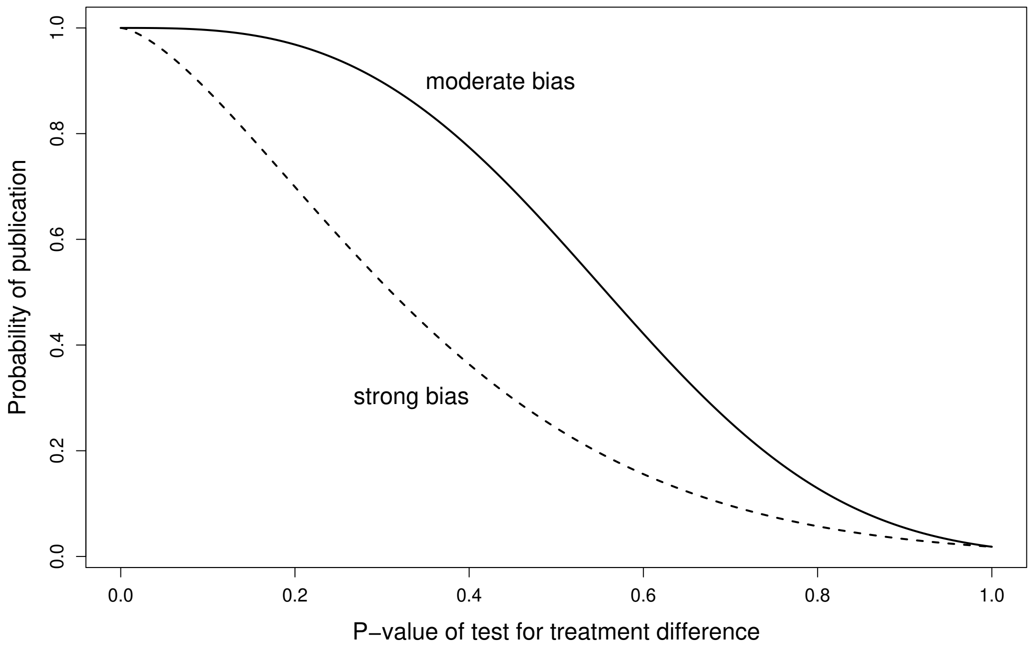


Figure 3.1: Selection mechanisms used in simulations referred to as "moderate" and "strong" selection bias in the text

Generating heterogeneous meta-analyses

The standard random effects model described in Section 2.2.2 is used as model to generate heterogeneous meta-analyses which is the model used in Schwarzer et al. (2002). The approach of Terrin et al. (2003) modelling the influence of the control event rate on the outcome is not pursued in this thesis. Values of 0.00, 0.05 and 0.25 are utilised for the between-trial variance τ^2.

For values of the between-trial variance τ^2 greater than 0.00, a very small probability exists to generate a control event rate larger than 1.00 for the risk ratio as measure of treatment effect; these probabilities were truncated at 0.9999.

	Simulation I[a]	Simulation II[b]
Number of trials	5, 10, 20, 50	
Total sample size (on log scale)	Normal distribution: $N(3.798, 1.104)$	Uniform distribution: A: $\log 50$ to $\log 500$ B: $\log 100$ to $\log 1500$ C: $\log 100$ to $\log 10000$
Treatment effect	Odds ratio ψ: 0.25, 0.50, 0.67, 1.00 Risk ratio ϕ: 0.25, 0.50, 0.67, 1.00	
Control event rate	p_C: 0.1, 0.3 $\Rightarrow \text{logit}(p_A) = \text{logit}(p_C) + 0.5 \cdot \log(\psi)$; $\log(p_A) = \log(p_C) + 0.5 \cdot \log(\phi)$	
Heterogeneity	$\text{logit}(p_i^I) = \text{logit}(p_A) + \log(\psi_i)/2$ $\text{logit}(p_i^C) = \text{logit}(p_A) - \log(\psi_i)/2$ with $\log(\psi_i) \sim N(\log\psi, \tau^2)$, $\tau^2 = 0, 0.05, 0.25$; (logit$(\cdot)$ replaced by $\log(\cdot)$ for risk ratio ϕ)	

Table 3.5: Generation of meta-analyses ($n = 1000$); sample sizes according to [a]Schwarzer et al. (2002), [b]Terrin et al. (2003)

Chapter 4

Results

In this chapter, results for the various simulation designs are described. In the text, figures are utilised to display simulation results. In Appendix A, additional information can be found in tabular form; references to these tables are included throughout in the text. Simulation results for meta-analysis with 10 trials are reported in detail since this number of trials is typically found in medical applications.

In Section 4.1, results under the null hypothesis of no bias in meta-analysis are described. For Simulation I, results are discussed in-depth for both odds ratio and risk ratio as measure of treatment effect. Furthermore, an explanation for the inflation of type I error rates with increasing treatment effect is given. In Section 4.2, results under the alternative of moderate and strong selection bias are described; again, details are reported for Simulation I. Finally, a short summary of simulation results is given in Section 4.3.

4.1 Simulation results under null hypothesis

A significance level of 10% is used throughout this thesis in all simulations evaluating statistical tests for bias in meta-analysis. Therefore, the proportion of significant findings which is used as an estimate of the type I error rate should match this value in simulations under the null hypothesis of no publication bias.

Since publication bias may either lead to an over- or underestimation of the treatment effect, tests for bias in meta-analysis are typically utilised as two-sided tests with corresponding lower and upper critical values. Under the null hypothesis of no bias in meta-analysis, the proportion of significant results below the lower critical value (lower tail) and above the upper critical value (upper tail) should be of similar size, i.e. each 5% for a nominal significance level of 10%.

For odds ratio and risk ratio less than 1.00, a test result below the lower critical value is indicative for an overestimation of the treatment effect which corresponds to the situation illustrated in Figure 2.2; a test result above the upper critical value is indicative for an underestimation of the treatment effect. For an odds ratio and risk ratio above 1.00, the relation between over-/underestimation and lower/upper tail is vice versa. Typically, an overestimation of the treatment effect is expected as a consequence of publication bias.

4.1.1 Simulation I

In Simulation I, sample sizes are generated according to Schwarzer et al. (2002). The proportion of small trials is greatest in this simulation design from all designs considered in this thesis (see Table 3.4).

In this section, results are reported in detail for both odds ratio and risk ratio as measures of treatment effect. In Section 4.1.2 describing results for Simulation II, only results for the odds ratio are reported in the text to economise on space.

Accuracy of estimated treatment effects

The distribution of the estimated treatment effects was examined in an attempt to validate the software code used in simulations. Furthermore, this information was utilised to compare results of different methods to estimate the overall treatment effect.

In Appendix A, median, 5% and 95% quantiles of the estimated treatment effect are listed for the Mantel-Haenszel method in Table A.1. For a control event rate of 0.3, the median of the estimated odds ratio and risk ratio is very close to the true value with a maximum difference of 0.01. For a control event rate of 0.1 and a treatment effect of 0.25, the median of the estimated odds ratio and risk ratio is always larger than the true treatment effect with values ranging from 0.27 to 0.28. For treatment effects of 0.50 to 1.00, the median of the estimated treatment effect is close to the true value. Overall, results are very similar for the odds ratio and risk ratio as measure of treatment effect.

For the inverse variance method using the “add selective” approach, median, 5% and 95% quantiles of the estimated treatment effect are listed in Table A.2. For a control event rate of 0.3 and a treatment effect of 0.25, the median of the estimated treatment effect is always larger than the true treatment effect with values ranging from 0.26 to 0.27 for the odds ratio and 0.27 to 0.29 for the risk ratio. Results are better for a treatment effect of 0.50, however, the estimated treatment effect is larger than the true treatment effect in most cases for the odds ratio (range: 0.50 to

0.53) and in all cases for the risk ratio (0.51 to 0.54). Results are worse for a control event rate of 0.1. All medians of estimated treatment effects are larger than the true value for treatment effects of 0.25, 0.50, and 0.67. The largest difference is seen for a true treatment effect of 0.25 with medians of estimated odds ratios ranging from 0.30 to 0.31 and from 0.30 to 0.32 for estimated risk ratios. In general, estimates are closer to the true value for the odds ratio as measure of treatment effect. Results are even worse for the inverse variance method using the "add all" approach which are listed in Table A.3.

The Peto method which may also be used for the estimation of the odds ratio performs poor for a true treatment effect of 0.25 with values ranging from 0.28 to 0.30; for odds ratios closer to 1.00 results of the Peto method are almost identical to those of the Mantel-Haenszel method (results not shown).

These results support the assertion that the Mantel-Haenszel method is the method of choice to estimate the treatment effect for sparse binary data. With respect to the two approaches for the inverse variance method, the "add selective" approach is the preferred method to estimate the treatment effect.

Odds ratio - estimated type I error rates

Results for a control event rate of 0.1 are given in Figure 4.1 and Table A.4 in Appendix A. Visual inspection of the figure clearly shows the inflation of type I error rates for the linear regression and rank correlation test; results of the two strategies to deal with sparse binary data ("add selective"/"add all") are very similar. Overall, the linear regression test shows the largest inflation of type I error rates. The new test holds the predefined significance level with a single exception (odds ratio of 0.25, large between-trial heterogeneity, and 50 trials in meta-analysis).

The linear regression test exceeds the predefined significance level of 10% in all configurations. The empirical size is ranging from 10.7% to 36.4% ("add selective") and 10.1% to 37.3% ("add all"). The empirical size is larger than 20% in 11 ("add selective") and 9 ("add all") of 48 configurations. In meta-analyses with 10 trials, the empirical size is ranging from 11.9% to 18.2% ("add selective") and 12.5% to 19.5% ("add all"). Even in meta-analysis with five trials, the empirical size is going up to 15.2% ("add selective") and 17.7% ("add all"). The type I error rate of the linear regression test is increasing with increasing number of trials per meta-analysis, increasing treatment effect, and increasing degree of between-trial heterogeneity.

The empirical size of the rank correlation test is ranging from 5.8% to 23.4% ("add selective") and 6.0% to 19.7% ("add all"); the type I error rate is larger than 20% only in 2 of 48 configurations. In meta-analyses with 10 trials, the empirical size is ranging from 7.1% to 14.8% ("add selective") and 7.1% to 15.1% ("add all"). The

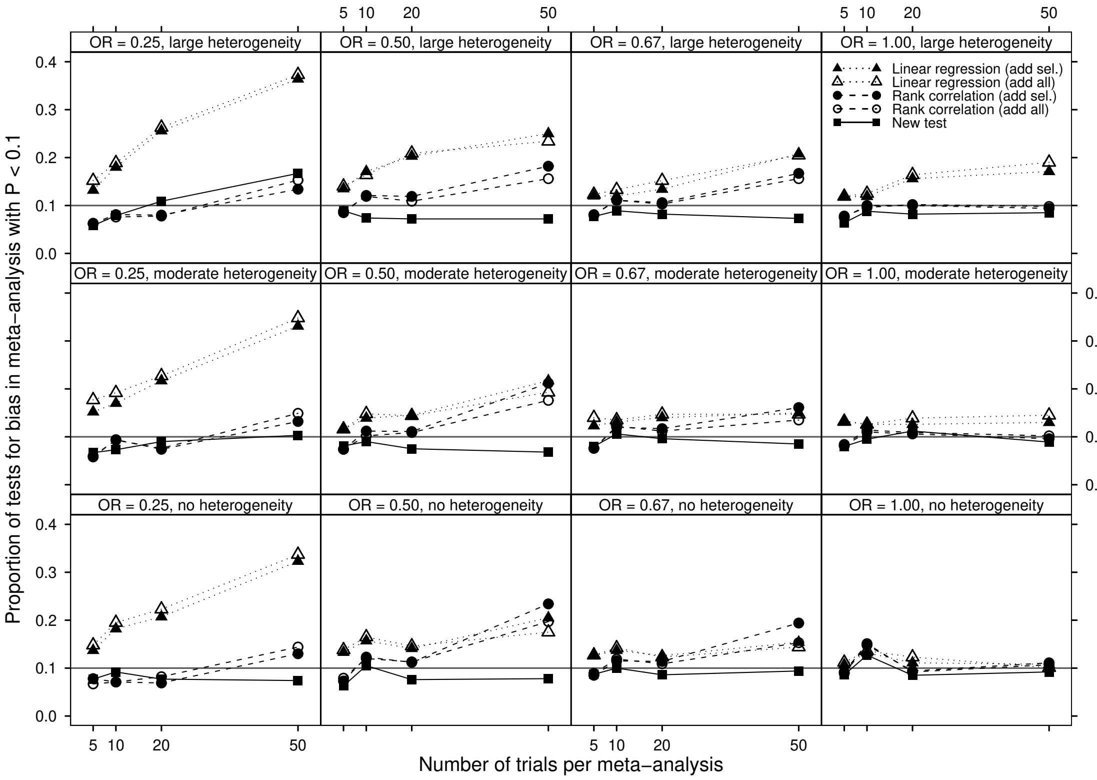


Figure 4.1: Simulation I: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.1; nominal significance level is 10%.

rank correlation test holds the predefined significance level in meta-analyses with 5 trials: empirical size is ranging from 5.8% to 8.9% ("add selective") and 6.0% to 9.2% ("add all"); some of these results are rather conservative. For an odds ratio less than 1.00, the type I error rate is increasing with increasing number of trials per meta-analysis. As compared to the linear regression test, the number of significant results of the rank correlation test is smaller for an odds ratio of 0.25 than for an odds ratio of 0.50; this observation is consistent across different degrees of heterogeneity.

Results for the linear regression and rank correlation test are similar for odds ratio values of 0.50 to 1.00 and no between-trial heterogeneity. For other configurations, results of the rank correlation test are below that of the linear regression test. Especially for an odds ratio of 0.25, results are markedly different.

The new test holds the predefined significance level in all but one situation. Only in meta-analyses with 50 trials, for an odds ratio of 0.25 and large between-trial heterogeneity the empirical size of 16.7% is clearly above the predefined significance level of 10%. In the remaining configurations, the empirical size is ranging from 5.8%

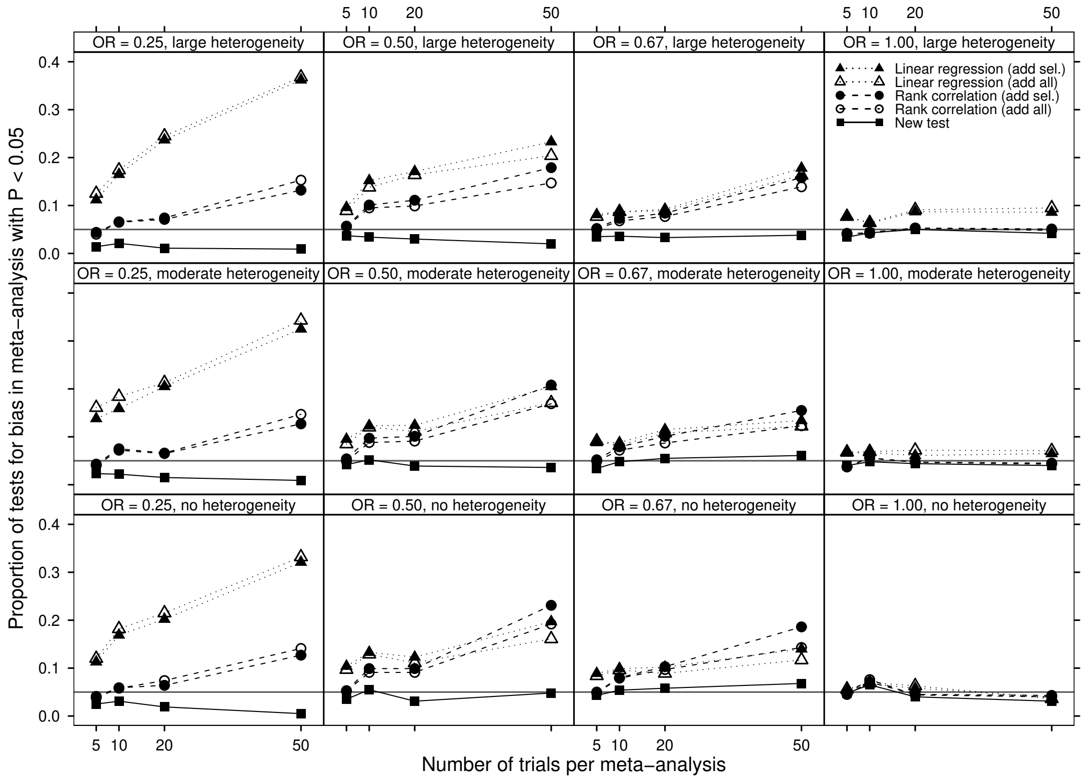


Figure 4.2: Simulation I: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of one-sided test for bias (lower tail); odds ratio as measure of treatment effect; control event rate of 0.1; nominal significance level is 5%.

to 12.7%. Results of the rank correlation test and the new test are very similar in meta-analyses with five trials. In meta-analyses with 10 trials, the empirical size is ranging from 7.3% to 12.7% which is in good agreement with the predefined significance level. As compared to the linear regression and rank correlation test, a tendency of anti-conservatism with increasing number of trials is apparent only in a single situation (odds ratio of 0.25 and large between-trial heterogeneity).

Figure 4.2 and 4.3, and Table A.5 show the proportion of significant results separately for lower and upper tail. A tendency of asymmetry in tail probabilities with increasing treatment effects is clearly apparent for both the linear regression and rank correlation test; both tests more often have significant results below the lower critical value corresponding to an overestimation of the treatment effect. Results for the new test are much more balanced with lower and upper tail probabilities close to the nominal significance level of 5%. However, for an odds ratio of 0.25, asymmetry in tail probabilities is apparent even for the new test. This asymmetry is in the opposite direction of the other two tests, i.e. a significant result for the new test is indicative more often for an underestimation of the treatment effect; this

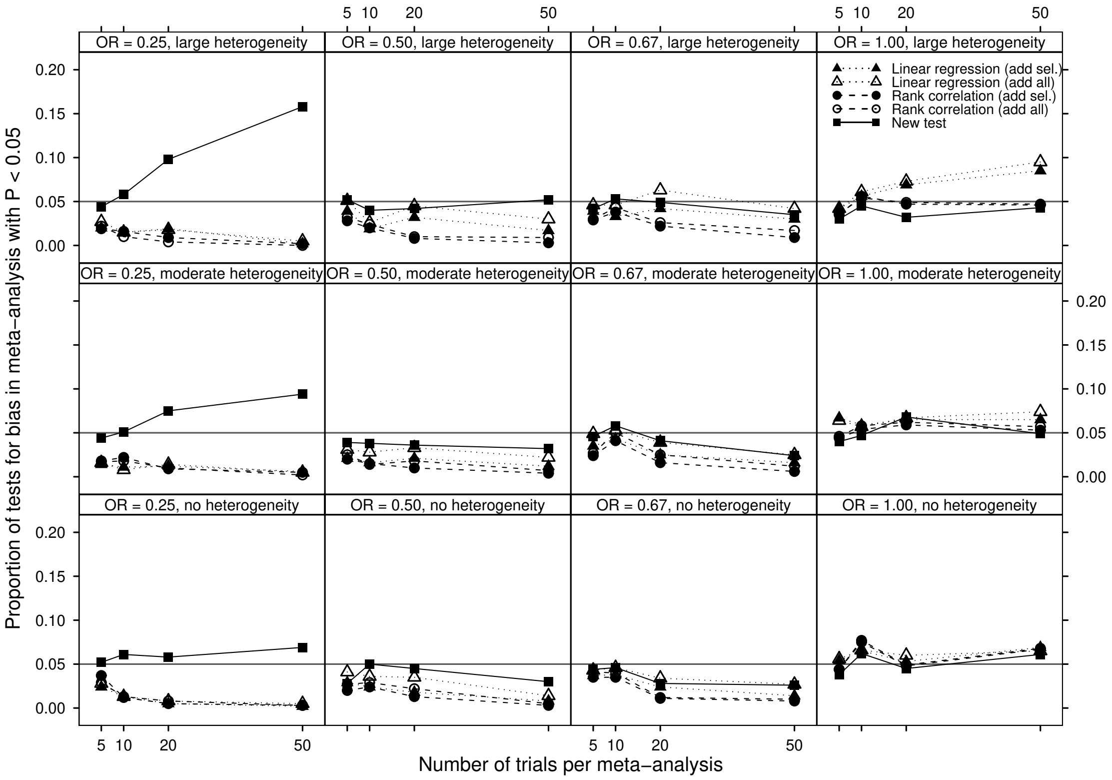


Figure 4.3: Simulation I: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of one-sided test for bias (upper tail); odds ratio as measure of treatment effect; control event rate of 0.1; nominal significance level is 5%.

observation is consistent across different degrees of heterogeneity.

Results for a control event rate of 0.3 are given in Figure 4.4 as well as Table A.6 and A.7 in Appendix A. Again, the inflation of type I error rates is clearly visible for the linear regression and rank correlation test. For large treatment effects, the percentage of significant tests for bias is even larger than for a control event rate of 0.1 with values as high as 48.6% and 58.8% for the linear regression and rank correlation test, respectively. In meta-analyses with 10 trials, the empirical size of the linear regression test is ranging from 11.3% to 23.6% ("add selective") and 11.6% to 19.5% ("add all"); empirical size of the rank correlation test is ranging from 10.2% to 18.4% ("add selective") and from 9.6% to 14.1% ("add all"), respectively. Asymmetry in tail probabilities with increasing treatment effects is certainly a problem for a control event rate of 0.3 for both the linear regression and rank correlation test (see Table A.7).

Two issues are rather different as compared to results of simulations with a control

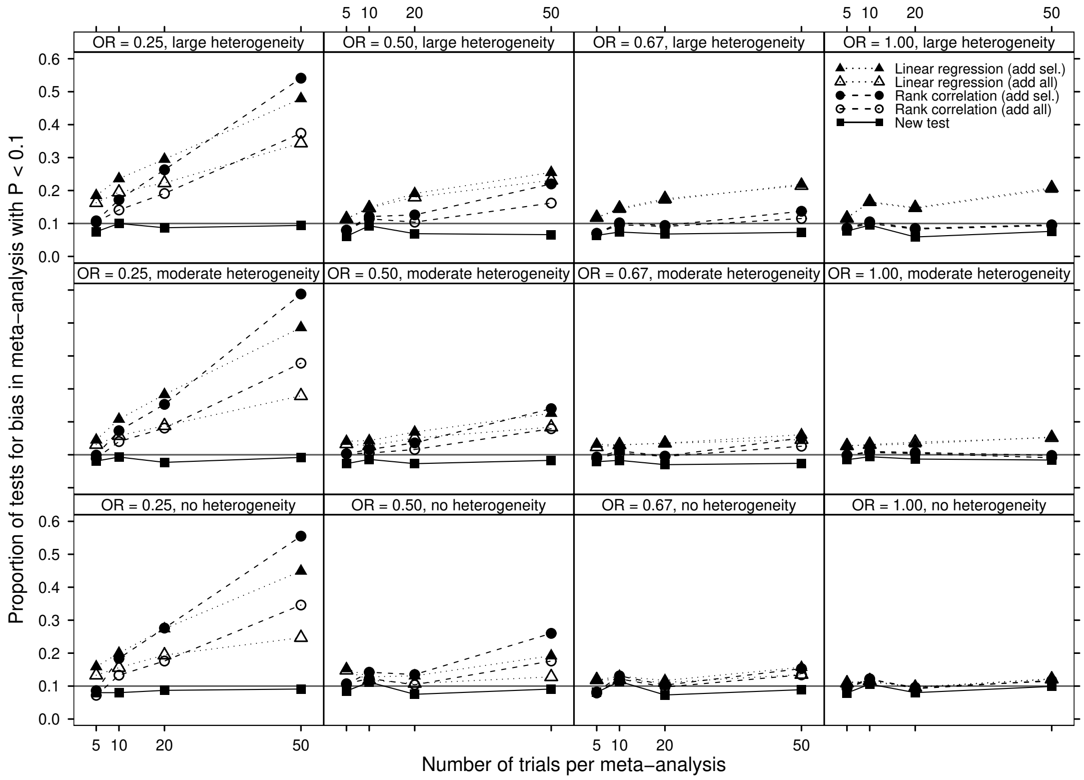


Figure 4.4: Simulation I: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.3; nominal significance level is 10%.

event rate of 0.1. Firstly, for an odds ratio of 0.25, the percentage of significant results of the rank correlation test is markedly increased. For a control event rate of 0.1, the maximum percentage of significant results for the rank correlation test is 13.4% ("add selective") and 15.3% ("add all"); for a control event rate of 0.3, the maximum percentage of significant test results is 58.8% ("add selective") and 37.8% ("add all"). Secondly, for both the linear regression and rank correlation test, results of the two strategies to deal with sparse binary data are different for odds ratios of 0.25 and 0.50; use of the "add all" approaches results in a much smaller number of significant test results. For example, for an odds ratio of 0.25, no between-trial heterogeneity, and 20 trials per meta-analysis, the number of significant results for the linear regression test is 27.4% using the "add selective" approach compared to 19.4% using the "add all" approach. For meta-analyses with 5 to 20 trials and values of the odds ratio from 0.50 to 1.00, the empirical size of the rank correlation using the "add all" method is ranging from 7.0% to 12.3% which is in good agreement with the predefined significance level. In this situation, the empirical size of the linear regression test using the "add all" method is ranging from 9.6% to 18.0%.

The new test holds the predefined significance level in all configurations with empirical size ranging from 5.9% to 11.4%; some of these results are rather conservative. In meta-analyses with 10 trials, the empirical size is ranging from 7.4% to 11.4%. Furthermore, tail probabilities are symmetric for the new test in almost all situations. Only for an odds ratio of 0.25 and large between-trial heterogeneity, lower and upper tail probabilities are different. In this case, the percentage of significant results above the upper critical value is 6.5% and 8.1% compared to a percentage of 2.2% and 1.3% of significant results below the lower critical value for meta-analyses with 20 and 50 trials, respectively. However, the asymmetry observed for the new test in this situation is by far not as dramatic as that of the linear regression and rank correlation test. For the linear regression test using the “add all” approach, the percentage of significant results below the lower critical value is 20.7% and 33.2% compared to a percentage of 1.6% and 1.1% of significant results above the upper critical value for meta-analyses with 20 and 50 trials, respectively. For the rank correlation test using the “add all” approach, the percentage of significant results above the upper critical value is only 0.3% and 0.0% for meta-analyses with 20 and 50 trials, respectively.

Risk ratio - estimated type I error rates

Results for the risk ratio as measure of treatment effect and a control event rate of 0.1 are given in Figure 4.5 and Table A.8 in Appendix A. At first glance, this figure looks very similar to Figure 4.1, however, there are some subtle but important differences.

Once more, the linear regression test exceeds the predefined significance level of 10% in all configurations. However, in this case, the percentage of significant test results is even more exaggerated with values ranging from 11.3% to 45.9% (“add selective”) and 11.9% to 47.2% (“add all”); the empirical size is larger than 20% in 15 (“add selective”) and 14 (“add all”) of 48 configurations. Simulation results in meta-analyses with 10 trials are comparable to results for the odds ratio as measure of treatment effect, i.e. only slightly larger; the empirical size is ranging from 11.4% to 21.7% (“add selective”) and 12.5% to 22.7% (“add all”). Again, the type I error rate of the linear regression test is increasing with increasing number of trials per meta-analysis, increasing treatment effect, and increasing degree of between-trial heterogeneity; results of the two strategies to deal with sparse binary data (“add selective”/“add all”) are very similar.

On the other hand, the percentage of significant results for the rank correlation test is smaller for the risk ratio as measure of treatment effect in almost all configurations where the risk ratio is less than 1.00. The empirical size of the rank correlation test

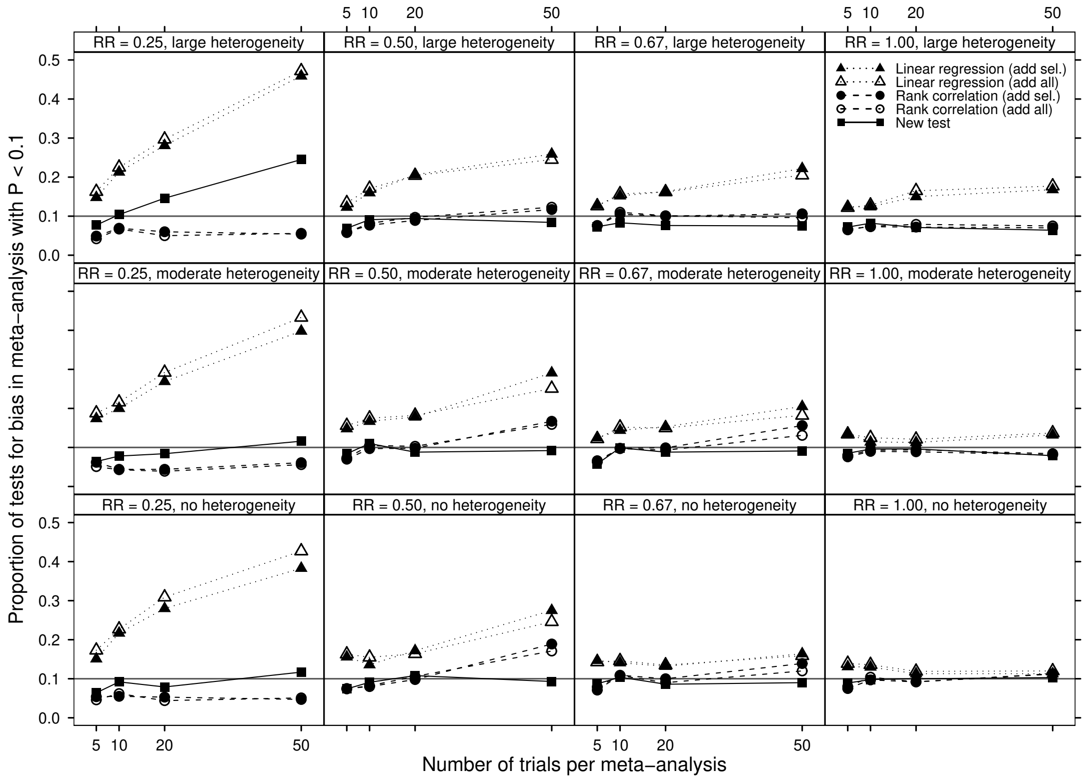


Figure 4.5: Simulation I: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; risk ratio as measure of treatment effect; control event rate of 0.1; nominal significance level is 10%.

is ranging from 4.4% to 18.9% ("add selective") and 3.9% to 17.1% ("add all"); some of these results are very conservative. In meta-analyses with 10 trials, the empirical size is ranging from only 4.4% to 10.9% ("add selective") and 4.3% to 11.0% ("add all"). Again, the number of significant results of the rank correlation test is smaller for a risk ratio of 0.25 than for a risk ratio of 0.50.

The new test holds the predefined significance level in all but two configuration. Only in meta-analyses with 20 and 50 trials, for a risk ratio of 0.25 and large between-trial heterogeneity the empirical size is markedly above the predefined significance level with values of 14.6% and 24.5%, respectively. In the remaining configurations, the empirical size is ranging from 5.8% to 11.7%. In meta-analyses with 10 trials, the empirical size is ranging from 7.8% to 11.0% which is in good agreement with the predefined significance level. A clear tendency of inflation of type I error rates with increasing number of trials is apparent only in a single situation (risk ratio of 0.25 and large between-trial heterogeneity).

Proportions of significant results separately for lower and upper tail are listed in Table A.8. For values of the risk ratio less than 1.00, asymmetry in tail probabilities with increasing treatment effects is clearly apparent for the linear regression test. In all configurations, the lower tail probability of the linear regression test is much larger than the upper tail probability. Even in meta-analysis with five trials and a moderate treatment effect, i.e. risk ratio of 0.67, the lower tail probability is about twice as large as the upper tail probability: 9.5% vs. 5.2% ("add selective") and 8.6% vs. 5.7% ("add all") for no between-trial heterogeneity, 9.2% vs. 3.3% ("add selective") and 8.4% vs. 3.8% ("add all") for moderate between-trial heterogeneity, 8.2% vs. 4.6% ("add selective") and 7.5% vs. 5.1% ("add all") for large between-trial heterogeneity.

Asymmetry in tail probabilities of the rank correlation test is not as dramatic as for the linear regression test which is primarily due to smaller values for the lower tail probability, upper tail probabilities are of comparable size. For example, for meta-analysis with 10 trials and a risk ratio of 0.50, the lower tail probability of the linear regression test using the "add all" approach ranges from 11.8% to 21.0% compared to percentages ranging from 4.3% to 7.8% for the rank correlation test; upper tail probabilities of the linear regression test range from 1.5% to 3.7% compared to percentages of comparable size ranging from 2.4% to 2.7% for the rank correlation test. Again, results for the new test are much more balanced with lower and upper tail probabilities close to the nominal significance level of 5%. For a risk ratio of 0.25, asymmetry in tail probabilities is clearly visible for the new test with upper tail probability as high as 24.4% (compared to a lower tail probability of 0.1%).

Results for a control event rate of 0.3 are given in Figure 4.6 and Table A.9 in Appendix A. For large treatment effects, inflation of type I error rates is much more pronounced than for a control event rate of 0.1; the largest type I error rate is 72.3% and 79.2% for the linear regression and rank correlation test, each for meta-analysis with 50 trials, a risk ratio of 0.25 and no between-trial heterogeneity. Again, the "add all" approach results in a much smaller number of significant test results for both the linear regression and rank correlation test. In meta-analyses with 10 trials, the empirical size of the linear regression test is ranging from 11.9% to 31.4% ("add selective") and 11.7% to 25.3% ("add all"); empirical size of the rank correlation test is ranging from 8.6% to 20.4% ("add selective") and from 8.7% to 17.0% ("add all"), respectively.

Asymmetry in tail probabilities with increasing treatment effects is certainly a problem for both the linear regression and rank correlation test. For the linear regression test and values of the risk ratio less than 1.00, the lower tail probability is ranging from 7.9% to 72.3% ("add selective") and 6.9% to 56.1% ("add all") compared to upper tail probabilities of only 0.0% to 4.7% ("add selective") and 1.0% to 5.7%

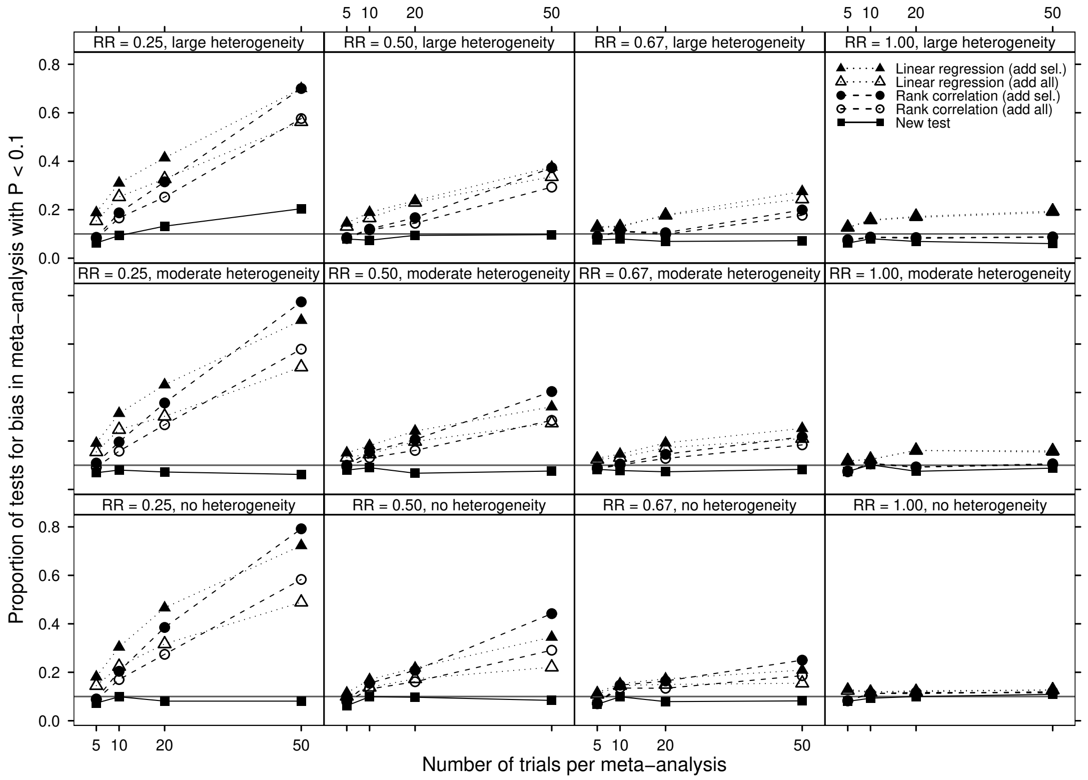


Figure 4.6: Simulation I: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; risk ratio as measure of treatment effect; control event rate of 0.3; nominal significance level is 10%.

("add all"); results for the rank correlation test are very similar. Results for the new test are much more balanced with lower and upper tail probabilities close to the nominal significance level of 5%. Only for the case of large between-trial heterogeneity and values of the risk ratio of 0.25 and 0.50, asymmetry in tail probabilities is apparent for the new test which is in the opposite direction of the other two tests. However, this asymmetry is not as dramatic as that of the linear regression and rank correlation test.

4.1.2 Simulation II

In Simulation II, three different mechanisms to generate sample sizes are considered (Terrin et al., 2003). A large proportion of sample sizes generated according to Scenario A is very small (see Table 3.4); furthermore, the range of total sample size is very narrow going from 50 to 500. From all designs considered in this thesis, the proportion of large trials is greatest in Scenario C with 50% of trials having a sample size larger than 1000. Sample sizes generated in Scenario B lie somewhere between these two extremes.

Results are only reported for the odds ratio as measure of treatment effect. In general, results for the risk ratio as measure of treatment effect are worse than results for the odds ratio.

Scenario A - estimated type I error rates

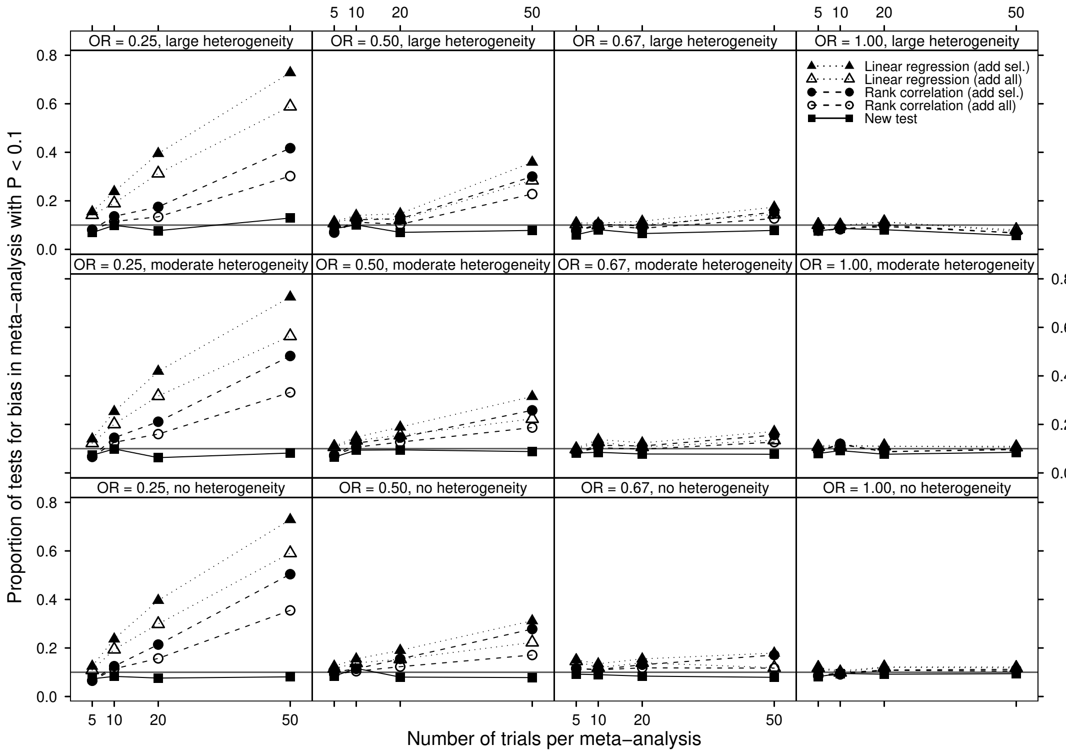
Median, 5% and 95% quantiles of Mantel-Haenszel estimates are listed in Table A.10. Results for the odds ratio as measure of treatment effect and control event rates of 0.1 and 0.3 are given in Figure 4.7 and Table A.11 and A.12; results for the risk ratio as measure of treatment effect are listed in Table A.13 and A.14.

For a control event rate of 0.1, the percentage of significant results is larger than those observed in Simulation I for both the linear regression and rank correlation test. The empirical size of the linear regression test is ranging from 7.4% to 72.9% ("add selective") and 8.0% to 59.1% ("add all"). Results of the rank correlation test are slightly better with values ranging from 6.6% to 50.4% ("add selective") and 6.5% to 35.5% ("add all"). For both tests the inflation of type I error rates is much smaller for the "add all" approach. For the new test, the percentage of significant results is ranging from 5.7% to 12.9%. The largest percentage is observed in meta-analyses with 50 trials, for an odds ratio of 0.25 and large between-trial heterogeneity.

In meta-analyses with 10 trials, the percentage of significant results is close to the nominal significance level for the rank correlation test using the "add all" approach with values ranging from 8.3% to 12.6% and the new test with values ranging from 8.1% to 11.6%; results for the linear regression test are not sufficient with values ranging from 9.9% to 20.1% for the "add all" approach (results for the "add selective" approach are even worse).

As described in the results section of Simulation I, tail probabilities are very unbalanced for the linear regression and rank correlation test for an odds ratio less than 1.00; this is also the case for results based on Simulation II, Scenario A. For the linear regression test and an odds ratio less than 1.00, the lower tail probability is ranging from 6.9% to 72.9% ("add selective") and 6.0% to 59.1% ("add all") compared to upper tail probabilities of 0.0% to 4.7% ("add selective") and 0.0% to 5.2% ("add all"); results for the rank correlation test are similar. For the new test and an odds ratio less than 1.00, the lower tail probability is ranging from 0.4% to 6.4% compared to upper tail probabilities of 2.4% to 12.5%; asymmetry in tail probabilities is only observed for an odds ratio of 0.25.

For a control event rate of 0.3, the empirical size of the linear regression test is ranging from 8.4% to 52.7% ("add selective") and 8.3% to 38.6% ("add all"). Results of the rank correlation test are better with values ranging from 6.5% to 43.8% ("add



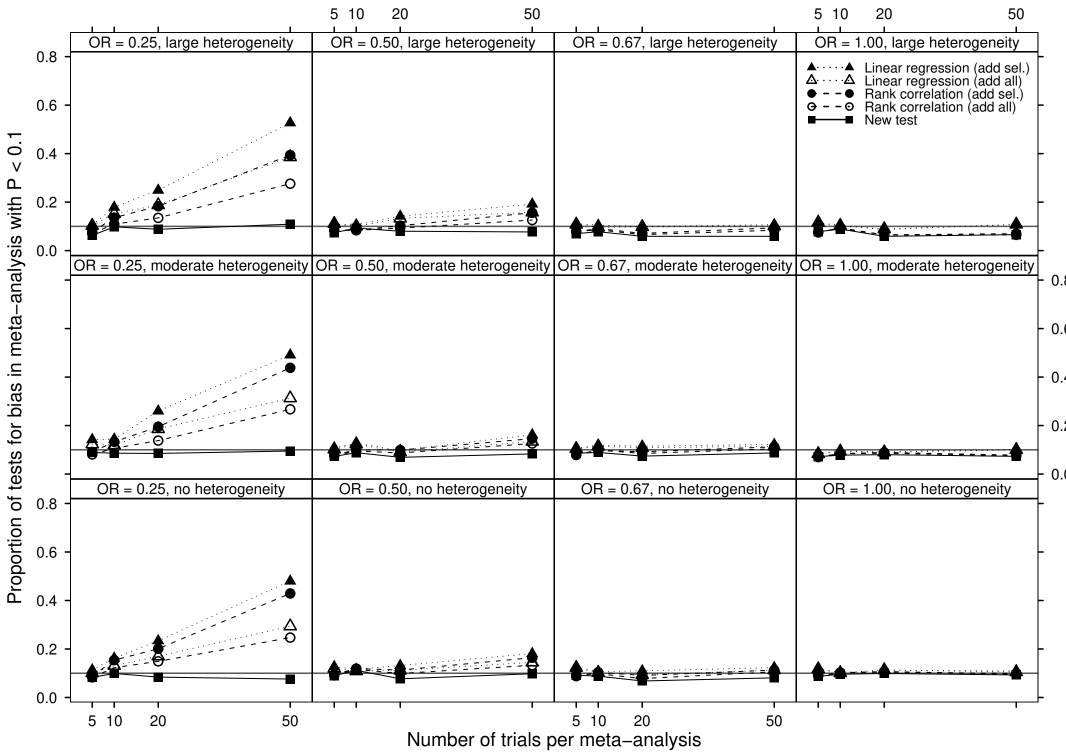


Figure 4.7: Simulation II, Scenario A: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.1 (upper figure) and 0.3 (lower figure); nominal significance level is 10%.

selective") and 6.5% to 27.6% ("add all"). For both tests the inflation of type I error rates is much smaller for the "add all" approach. For the new test, the percentage of significant results is ranging from 5.9% to 11.2%; some of these results are rather conservative.

In meta-analyses with 10 trials, the percentage of significant results is close to the nominal significance level for the rank correlation test using the "add all" approach with values ranging from 8.4% to 12.3% and the new test with values ranging from 7.8% to 11.2%. Results for the linear regression test are only sufficient for an odds ratio larger than 0.25 with values ranging from 9.3% to 12.3% for the "add all" approach; for an odds ratio of 0.25 the inflation of type I error rates is still quiet large with values from 14.4% to 17.9% ("add selective") and 11.9% to 15.2% ("add all").

Again, asymmetry in tail probabilities is clearly visible for both the linear regression and rank correlation test and to a much lesser extent for the new test.

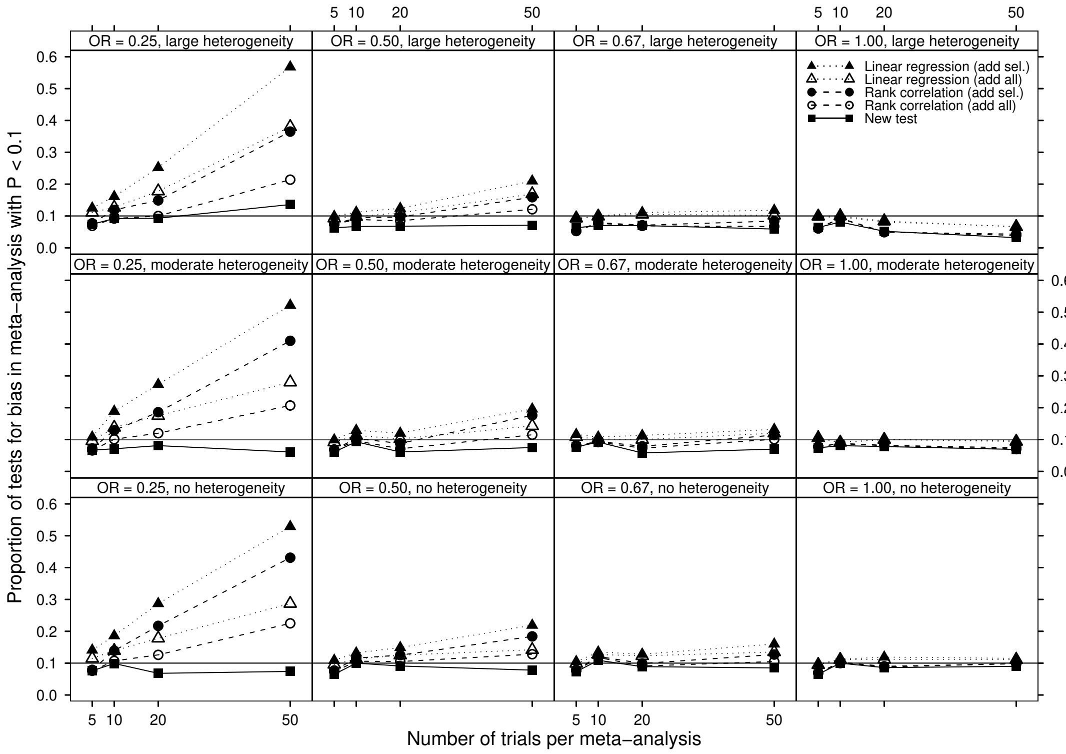
Scenario B - estimated type I error rates

Median, 5% and 95% quantiles of Mantel-Haenszel estimates are listed in Table A.15. Results for the odds ratio as measure of treatment effect and control event rates of 0.1 and 0.3 are given in Figure 4.8 and Table A.16 and A.17; results for the risk ratio as measure of treatment effect are listed in Table A.18 and A.19.

For a control event rate of 0.1, an inflation of type I error rates is apparent for the linear regression and rank correlation test for an odds ratio of 0.25. Furthermore, inflation of type I error rates is observed for an odds ratio of 0.50 and meta-analyses with 50 trials. Otherwise, empirical sizes do not exceed the predefined significance level too far; use of the "add all" approach results in a smaller number of significant results. For an odds ratio of 0.67, the largest type I error rate of the linear regression test is 16.0% ("add selective") and 13.5% ("add all"), both observed in meta-analyses with 50 trials.

In meta-analyses with 10 trials, the percentage of significant results is ranging from 9.1% to 18.9% ("add selective") and 9.1% to 13.9% ("add all") for the linear regression test, from 7.9% to 13.9% ("add selective") and 7.7% to 11.8% ("add all") for the rank correlation test, and from 6.7% to 10.9% for the new test. These results are sufficiently close to the nominal significance level with exception of results for the linear regression test using the "add selective" approach.

For an odds ratio of 1.00 and 20 or 50 trials per meta-analysis, a tendency of decreasing empirical size with increasing heterogeneity is apparent for both the rank correlation and new test. The percentage of significant results of the new test is ranging from 8.6% to 9.0% in meta-analysis without between-trial heterogeneity;



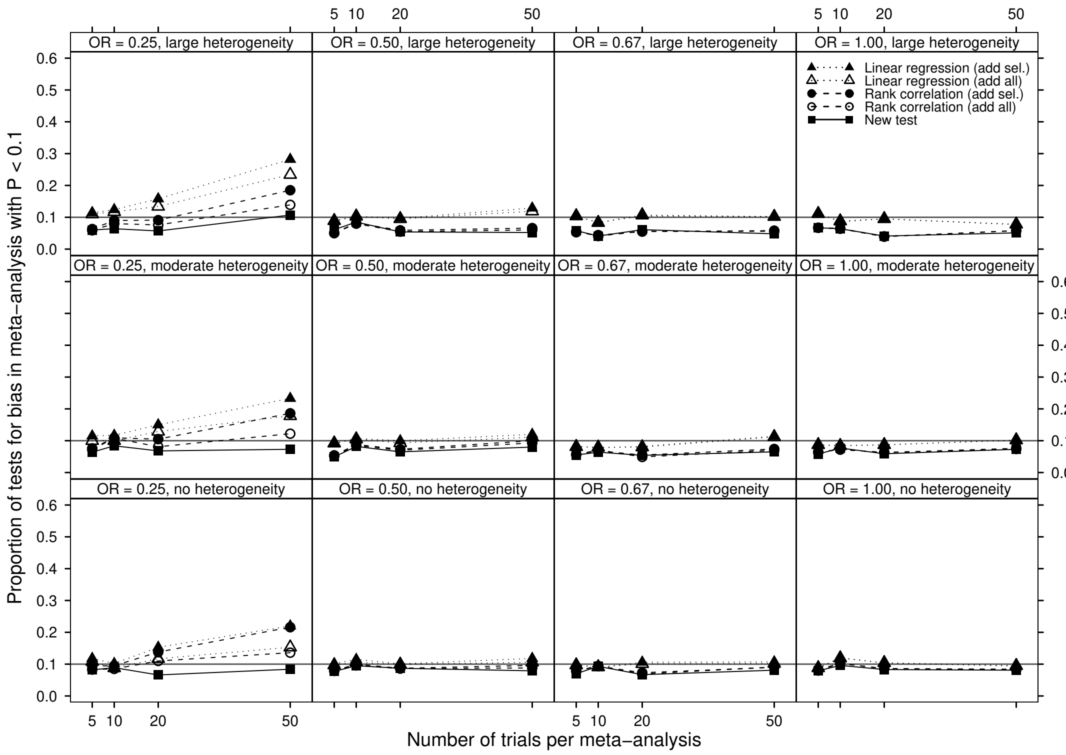


Figure 4.8: Simulation II, Scenario B: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.1 (upper figure) and 0.3 (lower figure); nominal significance level is 10%.

this percentage reduces to values between 3.2% and 5.2% in meta-analyses with large between-trial heterogeneity yielding rather conservative test results. For the new test, numbers of significant results in meta-analyses with 20 or 50 trials being well below the nominal significance level are also observed for an odds ratio of 0.50 and 0.67.

For a control event rate of 0.3, an inflation of type I error rates is only apparent for an odds ratio of 0.25, for both the linear regression and rank correlation test. Otherwise, empirical sizes do not exceed the predefined significance level too far; use of the "add all" approach results in a smaller number of significant results. In fact, both the rank correlation and new test have conservative results in the situation of large between-trial heterogeneity and odds ratios between 0.50 and 1.00. For an odds ratio of 0.67, the percentage of significant results for the rank correlation test is ranging from 5.3% to 8.4% ("add selective") and 5.3% to 7.7% ("add all") and from 5.9% to 7.0% for the new test. In this situation, results for the linear regression test are much closer to the nominal significant level, e.g., ranging from 9.3% to 10.5% using the "add all" approach.

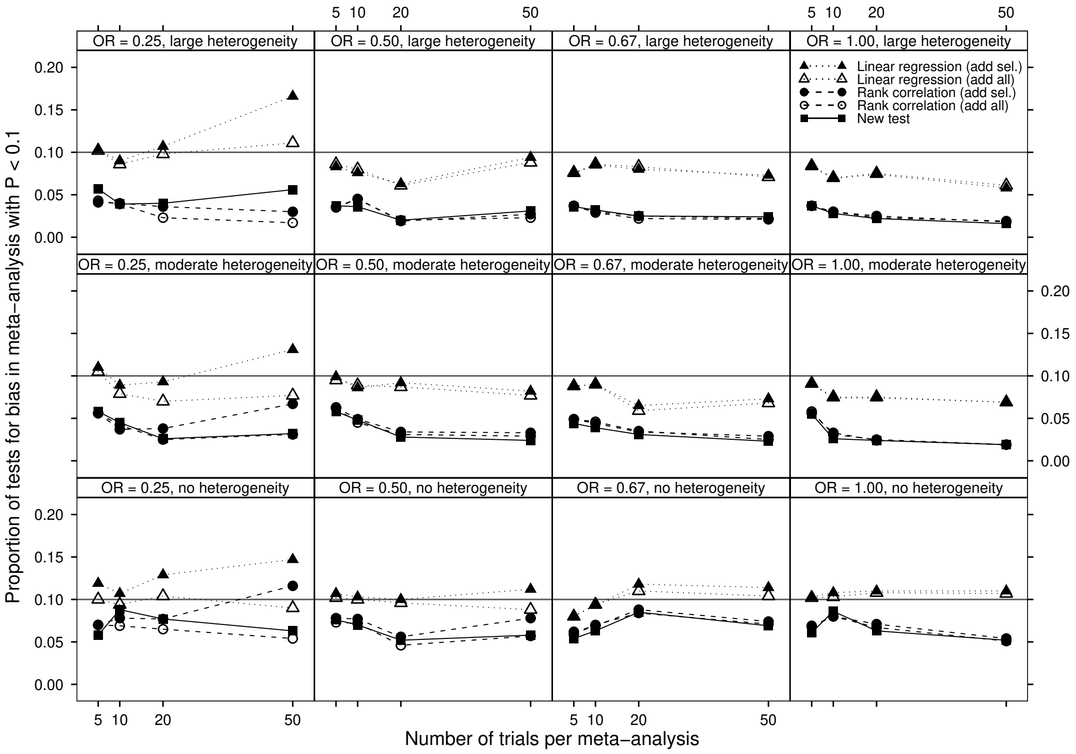
In meta-analyses with 10 trials, the percentage of significant results is ranging from 7.8% to 12.4% ("add selective") and 7.9% to 11.9% ("add all") for the linear regression test, from 4.4% to 10.9% ("add selective") and 4.1% to 10.8% ("add all") for the rank correlation test, and from 4.1% to 9.6% for the new test.

Asymmetry in tail probabilities exists for both control event rates of 0.1 and 0.3 for the linear regression and rank correlation test and to a much smaller extent for the new test.

Scenario C - estimated type I error rates

Median, 5% and 95% quantiles of Mantel-Haenszel estimates are listed in Table A.20. Results for the odds ratio as measure of treatment effect and control event rates of 0.1 and 0.3 are given in Figure 4.9 and Table A.21 and A.22; results for the risk ratio as measure of treatment effect are listed in Table A.23 and A.24. In Scenario C with a large proportion of trials with large samples sizes, findings are very different from simulation results discussed so far.

For a control event rate of 0.1, an inflation of type I error rates is only apparent for the linear regression test using the "add selective" approach in a single situation, i.e. in meta-analyses with 50, for an odds ratio of 0.25 and large between-trial heterogeneity. Otherwise, results for the linear regression test are close to the nominal significance level; especially results for the "add all" approach are good. Overall, the percentage of significant results is too small for both the rank correlation and new test. The percentage of significant results of the rank correlation test ranges



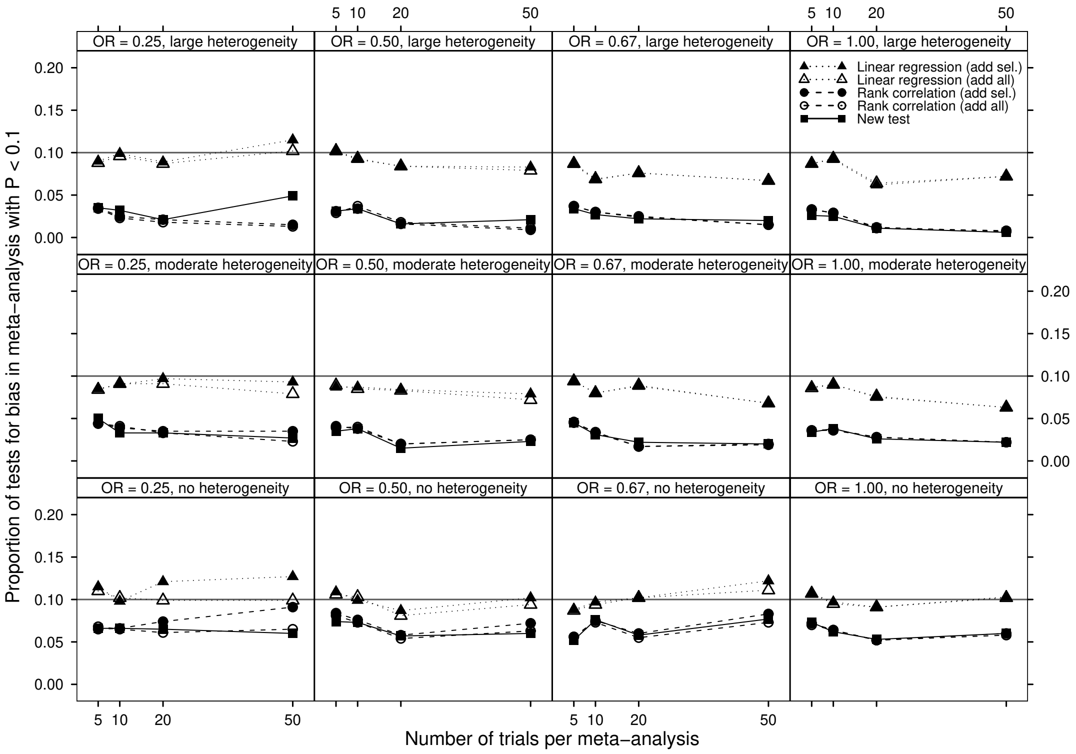


Figure 4.9: Simulation II, Scenario C: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.1 (upper figure) and 0.3 (lower figure); nominal significance level is 10%.

from 1.8% to 11.6% ("add selective") and from 1.7% to 8.4% ("add all"); the empirical size is above the nominal significance level in a single case. Results for the new test are ranging from 1.6% to 8.8%, i.e. all results are well below the predefined significance level.

Results for a control event rate of 0.3 are very similar to those reported in the last paragraph. Overall, the linear regression test shows the best simulation results and both rank correlation and new test yield conservative results.

4.1.3 Explanation for inflation of type I error rates with increasing treatment effect

An inflation of type I error rates with increasing treatment effect has been shown for the design of Simulation I as well as Simulation II, Scenario A and B. For example in Simulation I, for a true odds ratio of 0.25, a control event rate of 0.3, and no between-trial heterogeneity, type I error rates of the linear regression and rank correlation are as high as 48.6% and 58.8%, respectively, whereas these rates are close to 10% for an odds ratio of 0.67 (see Figure 4.4). The examination of the association between the estimated log odds ratio and its standard error in these two situations yields an explanation for the inflation of type I error rates.

In Figure 4.10, image plots of 10000 trials are displayed for different total sample sizes with the estimated odds ratio plotted on the x-axis and the standard error of the estimated log odds ratio plotted on the y-axis, i.e. the plot can be seen as an analogue to a funnel plot. The following parameters were kept fix in the four sub-figures: true odds ratio of 0.25, control event rate of 0.3, and no between-trial heterogeneity. The only parameter that changes in the sub-figures is the total sample size ranging from 30 to 500. These sample sizes were selected on purpose; in Simulation I, the minimum total sample sizes is 30 and only a small percentage of 2% of the generated trials has a sample size larger than 500. In Simulation II, Scenario A, total sample size is ranging from 50 to 500.

The density of points is highest for rectangles plotted in black, this density is decreasing by moving from black to light grey. For a total sample size of 30, large variability is apparent both in the estimated odds ratio as well as the standard error. Despite this large variability, an association between estimated odds ratio and standard error can be seen at first glance. The probability of observing large standard errors, e.g., larger than 1.1, is larger in trials with an estimated odds ratio less than the true odds ratio of 0.25. On the other hand, the probability of observing small standard errors, e.g., smaller than 1.0, is much larger in trials with an estimated odds ratio above 0.25. In trials with a total sample size of 50 or 100, the association between estimated odds ratio and standard error of the log odds ratio is more apparent. Even in trials with a total sample size of 500 an association between these two quantities is observed.

Each sub-figure gives an indication on the bivariate distribution of estimated odds ratio and standard error only for a single sample size. An image plot for a meta-analysis of trials with different sample sizes could be constructed by putting the four sub-figure on top of each other weighted according to the probability of generating a trial of the corresponding sample size.

An assumption of the linear regression and rank correlation test is that the estimated treatment effect (e.g., log odds ratio) and its standard error are independent under the null hypothesis of no bias in meta-analysis. This assumption is obviously violated for the situation depicted in Figure 4.10.

For a true odds ratio of 0.67, the association between estimated odds ratio and standard error of estimated log odds ratio is much less pronounced (see Figure 4.11). Accordingly, the inflation of type I error rates is much smaller.

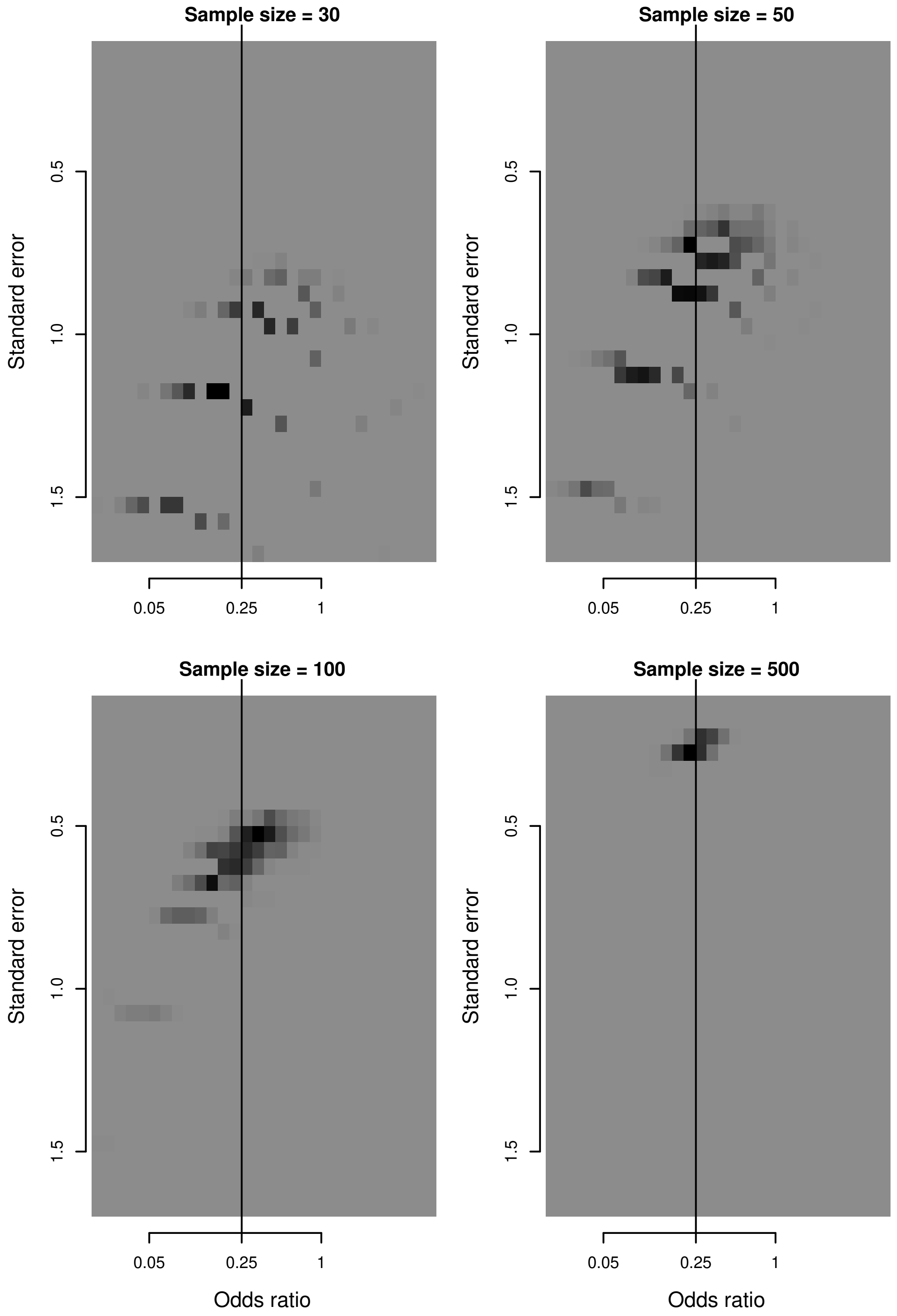


Figure 4.10: Simulation I: image plot of estimated odds ratios and standard error of estimated log odds ratio, i.e. an analogue to a funnel plot, for 10000 trials generated under null hypothesis; true odds ratio of 0.25 indicated by vertical line; control event rate of 0.3; no between-trial heterogeneity.

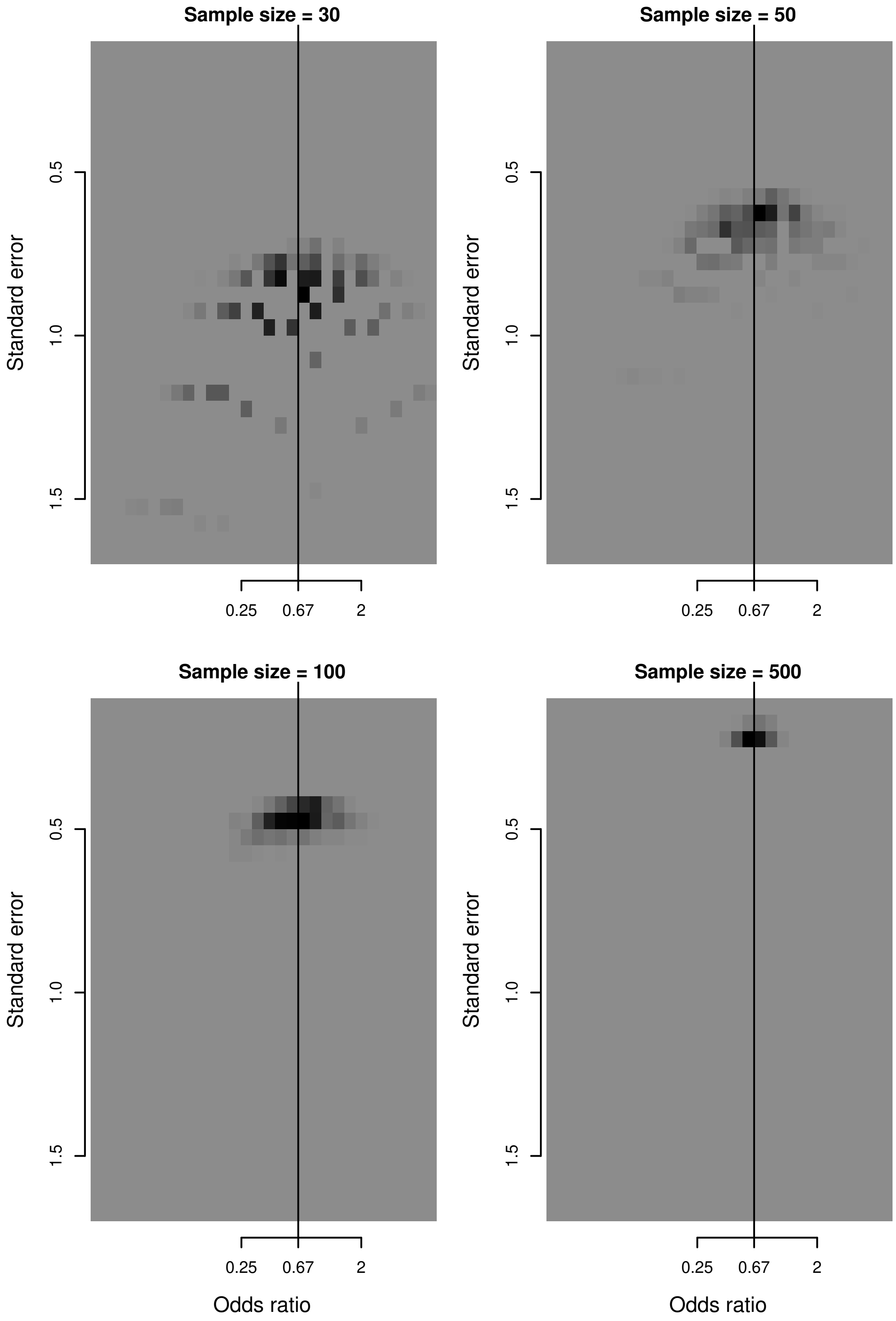


Figure 4.11: Simulation I: image plot of estimated odds ratios and standard error of estimated log odds ratio, i.e. an analogue to a funnel plot, for 10000 trials generated under null hypothesis; true odds ratio of 0.67 indicated by vertical line; control event rate of 0.3; no between-trial heterogeneity.

4.2 Simulation results under alternatives

I decided to evaluate the power of the linear regression and rank correlation test only for the "add all" approach since this method consistently shows better results under the null hypothesis of no bias in meta-analysis. Furthermore, only results for the odds ratio as measure of treatment effect are reported.

4.2.1 Specifying critical values

As shown in Section 4.1, the linear regression and rank correlation test are too liberal in many situations whereas the new test holds the predefined significance level in almost all situations. In order to allow a fair comparison of the power of the three test procedures, an adjustment for this imbalance is needed.

For both the linear regression and rank correlation test, the distribution of the test statistic is shifted to the left under the null hypothesis of no bias for true odds ratios less than 1.00 (Schwarzer et al., 2002). Accordingly, the empirical 5% and 95% quantiles of the test statistics are smaller than the respective critical values used in these tests.

My idea is to use empirical 5% and 95% quantiles of the test statistic derived from simulations under the null hypothesis as rejection limits in simulations under the alternative of bias in meta-analysis. However, two questions remain: should these empirical quantiles be used for all tests, i.e. linear regression test, rank correlation test, and new test, and for which simulation designs?

In Figure 4.12, empirical 5% and 95% quantiles of the test statistic of the rank correlation test (for simulations under the null hypothesis of no bias according to the design of Simulation I and a control event rate of 0.3) are plotted for all combinations of treatment effect, number of trials per meta-analysis, and between-trial heterogeneity. The empirical 5% quantile is smaller than the original critical value of -1.64 in almost all situations with a true odds ratio of 0.25 and 0.50; only in meta-analyses with 5 trials, for an odds ratio of 0.25, and no between-trial heterogeneity the empirical quantile is larger than -1.64. For an odds ratio of 0.67, the empirical 5% quantile is smaller than the original critical value in most situations. Finally, most empirical 5% and 95% quantiles are close to the original critical values of -1.64 and 1.64, for an odds ratio of 1.00. In four of nine simulations with 5 trials per meta-analysis, the 5% and 95% quantile are within the range of the original critical values; these rather conservative test results are typical for rank tests with a small number of observations.

Bias as generated in simulations under the alternative (see Section 3.2 for details) will result in a shift of test statistics to more negative values for all tests considered

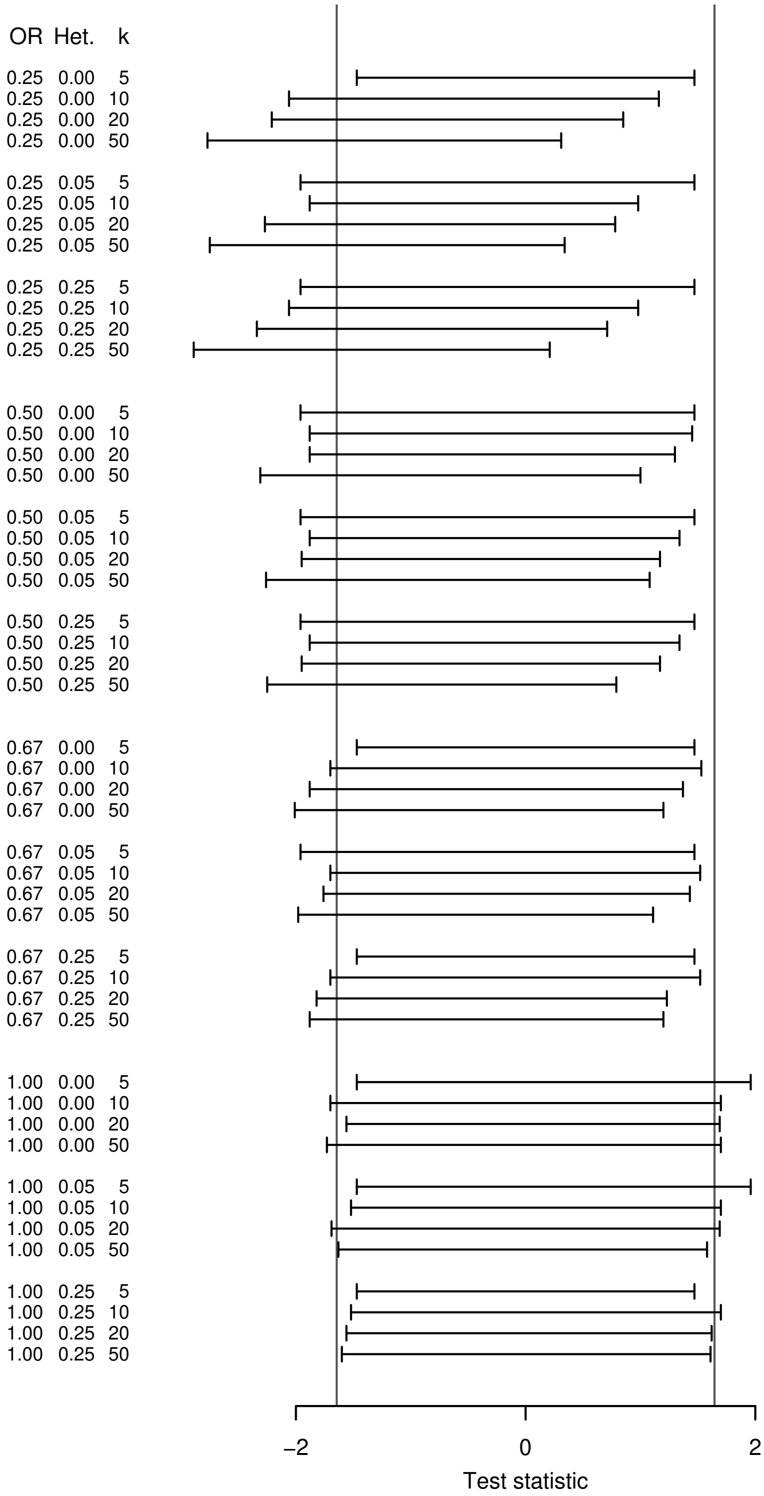


Figure 4.12: Simulation I: empirical 5% and 95% quantiles of the test statistic of the rank correlation test using "add all" approach derived from simulations under the null hypothesis of no bias in meta-analysis; odds ratio as measure of treatment effect; control event rate of 0.3.

in this thesis. Accordingly, the use of critical values based on empirical quantiles of simulations under the null-hypothesis will lead to a smaller number of significant results for the rank correlation test in situations with true odds ratio less than 1.00. Similar results to Figure 4.12 were observed for the linear regression test (results not shown).

Empirical 5% and 95% quantiles of the test statistic of the new test under the null-hypothesis of no bias are given in Figure 4.13. The empirical 5% quantile is almost always larger than the original critical value of -1.64. Thus, the use of critical values based on empirical quantiles would yield a larger number of significant results for the new test.

Overall, for simulations based on the design of Simulation I and a control event rate of 0.3, the following strategy seems sensible:

- Linear regression and rank correlation test:
  - use of empirical 5% and 95% quantiles of test statistics derived from simulations under the null hypothesis of no bias as rejection limits, for true treatment effects of 0.25, 0.50, and 0.67 (labelled "adjusted critical values"),
  - use of original critical values, for a true treatment effect of 1.00.
- New test: use of the original critical values.

Similar considerations as described above in the discussion of Figure 4.12 and 4.13 resulted in the use of this strategy in simulations based on Simulation I (control event rate of 0.1) and Simulation II, Scenario A, too.

For simulations based on Simulation II, Scenario B the following strategy was utilised:

- Linear regression and rank correlation test:
  - use of adjusted critical values, for a true treatment effect of 0.25,
  - use of original critical values, for true treatment effects of 0.50, 0.67, and 1.00.
- New test: use of the original critical values.

For simulations based on Simulation II, Scenario C the original critical values were utilised for all three tests.

Both the adjusted and original critical values are reported in the corresponding tables in Appendix A.

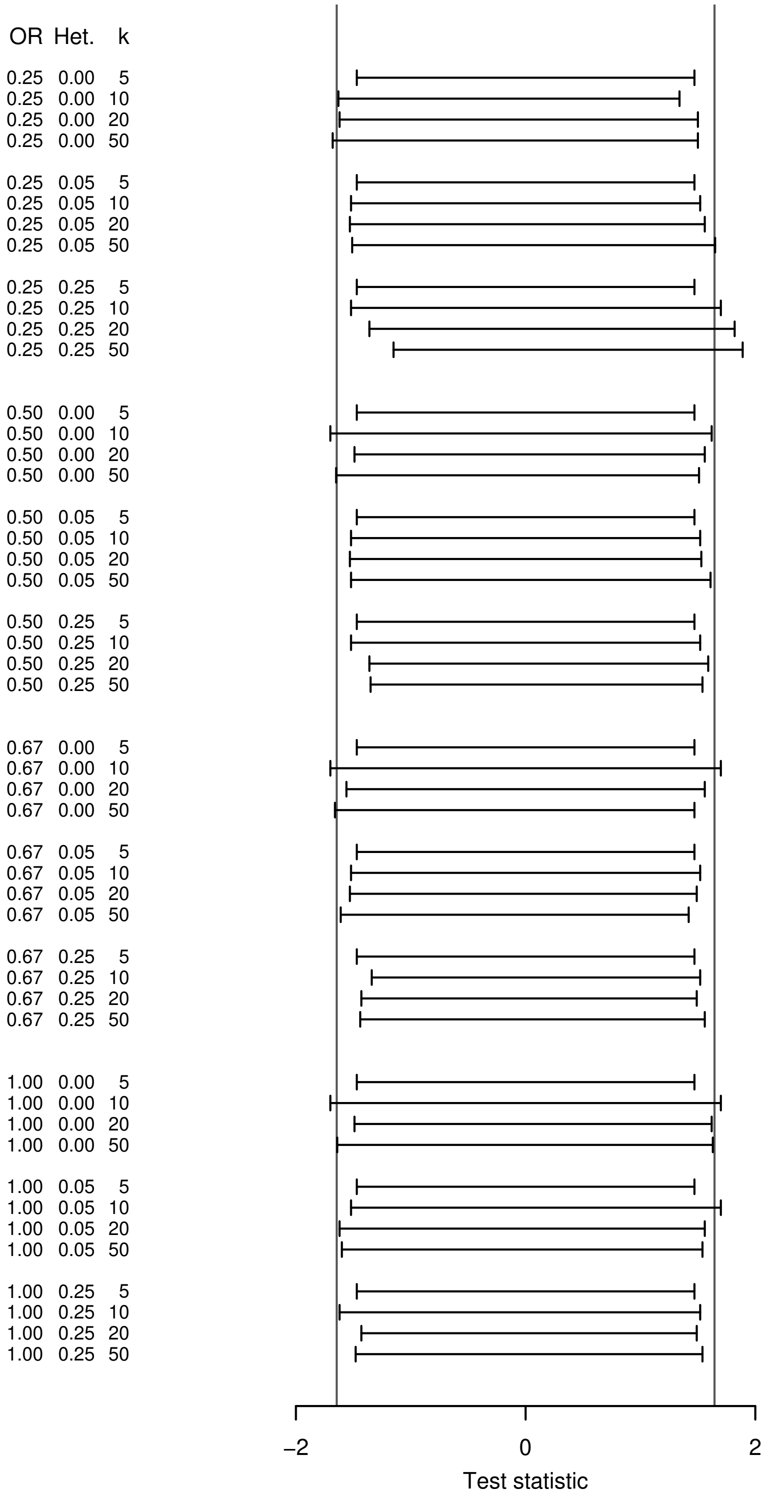


Figure 4.13: Simulation I: empirical 5% and 95% quantiles of the test statistic of the new test derived from simulations under the null hypothesis of no bias in meta-analysis; odds ratio as measure of treatment effect; control event rate of 0.3.

4.2.2 Simulation I

Figure 4.14 displays the proportion of meta-analyses with significant results for the three tests for bias; meta-analyses were generated according to moderate selection bias. Table A.25 and A.26 contain the corresponding numbers for control event rates of 0.1 and 0.3, respectively, as well as information on the probability of publication (column "% selected") and the median, 5% and 95% quantile of Mantel-Haenszel estimates of the treatment effect.

Moderate bias - decreasing power with increasing treatment effect

The power of all tests is decreasing with increasing treatment effect for both control event rates. This result is not surprising since the probability of publication is increasing with increasing treatment effect (Macaskill et al., 2001). For a control event rate of 0.1, the probability of publication is ranging from 55.6% to 62.5% for an odds ratio of 1.00, this proportion is increasing to values ranging from 89.3% to 100% for an odds ratio of 0.25. For a control event rate of 0.3, the probability of publication is similar for an odds ratio of 1.00 and even larger for an odds ratio of 0.25 with values ranging from 95.2% to 100%. In general, for odds ratios less than 1.00, the proportion of published trials is larger for a control event rate of 0.3 than for a rate of 0.1. Thus, the bias induced in simulations with moderate selection bias is larger for an event rate of 0.1.

In accordance with findings described in the last paragraph, Mantel-Haenszel estimates are much closer to the true values for large treatment effects. For a true odds ratio of 0.25, the median of Mantel-Haenszel estimates lies between 0.23 and 0.25 (control event rate of 0.1) and 0.24 and 0.25 (control event rate of 0.3). Thus, for an odds ratio of 0.25 there is hardly any bias left in meta-analyses generated according to moderate selection bias. On the other hand, the bias induced is quiet severe for an odds ratio of 1.00. The median of Mantel-Haenszel estimates is ranging from 0.59 and 0.70 for a control event rate of 0.1 and from 0.64 to 0.79 for an event rate of 0.3. In general, estimates are less biased in meta-analyses with a control event rate of 0.3 as compared to results for a rate of 0.1.

Moderate bias - decreasing power with increasing heterogeneity

The power of all tests is decreasing with increasing heterogeneity whereas the amount of bias is increasing at the same time. For example, in meta-analyses with 50 trials, a control event rate of 0.1, and an odds ratio of 1.00, the power of the linear regression test is decreasing from 79.6% in meta-analyses without heterogeneity to 72.1% in meta-analyses with moderate heterogeneity to 58.5% in meta-analyses with large heterogeneity which is in contrast to the increase in bias in the Mantel-Haenszel estimates with medians ranging from 0.68 to 0.70 (no heterogeneity), 0.66 to 0.68 (moderate heterogeneity), and 0.59 to 0.60 (large heterogeneity). This observation which is even more pronounced for a control event rate of 0.3 is rather surprising.

A weak tendency for decreasing probability of published trials with increasing heterogeneity is apparent in Table A.25 and A.26. This tendency can not fully explain the decrease in power. In the example given in the last paragraph, the probability of publication is only slightly decreasing from 56.8% in meta-analyses without between-trial heterogeneity, to 56.2% in meta-analyses with moderate between-trial heterogeneity, and finally to 55.6% in meta-analyses with large between-trial heterogeneity. Furthermore, a huge decrease in power with increasing heterogeneity is also observed in situations where the proportion of published trials is constant. For example, in meta-analyses with 50 trials, a control event rate of 0.3, and an odds ratio of 1.00, the power of the linear regression test is decreasing from 72.6% (no between-trial heterogeneity) to 55.8% (moderate between-trial) even though the probability of publication is 56.2% for both degrees of heterogeneity.

An explanation for this result will be given below after discussing the results for strong selection bias.

Moderate bias - power comparison

For an odds ratio of 1.00, the new test has the lowest power of all three tests for both control event rates of 0.1 and 0.3 (see Figure 4.14). The power of the linear regression and rank correlation test is of comparable size in most situations. In meta-analyses with 5 to 20 trials, the power of the linear regression test is always slightly larger than the power of the rank correlation test. This small difference can be explained by the small inflation of type I error rates observed for the linear regression test (see Figures 4.1 and 4.4). The power of all tests is decreasing with increasing degree of heterogeneity. This tendency is more markedly in the new test procedure. In meta-analyses with 10 trials, the power of all tests to detect bias is rather small; the maximum power is 29.3% for the linear regression test, 23.3% for the rank correlation test, and 19.0% for the new test. These values are observed in meta-analyses with a control event rate of 0.1 and no between-trial heterogeneity; the median of Mantel-Haenszel estimates is 0.70 in this situation which corresponds to a seriously biased estimate.

For odds ratios of 0.50 and 0.67, the power of the three tests is of comparable size in the situation of no or moderate between-trial heterogeneity. For large between-trial heterogeneity, the power of the new test is smaller than the power of the other tests. In meta-analyses with 10 trials, the maximum power of the three tests is below

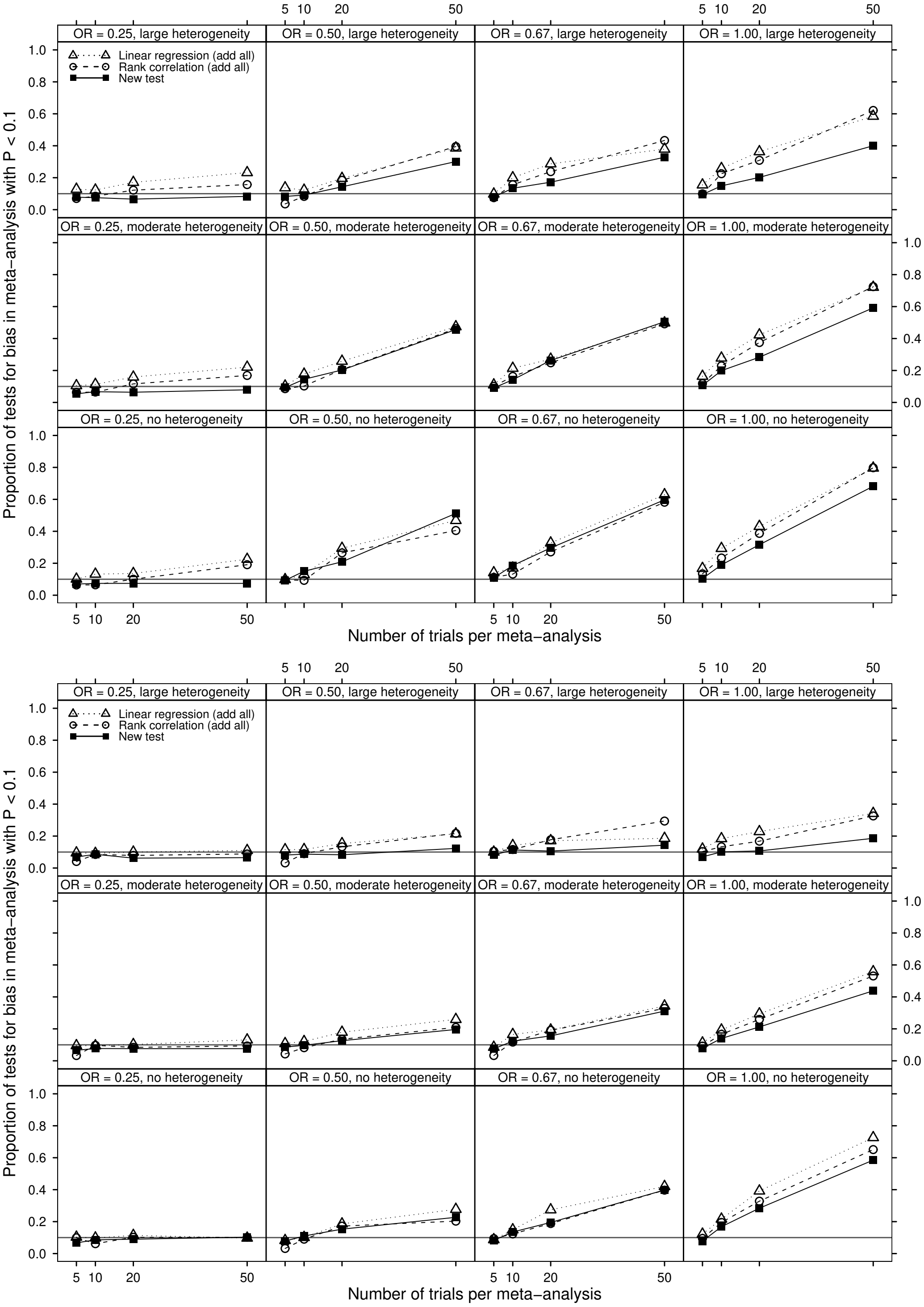


Figure 4.14: Simulation I: results under alternative hypothesis of moderate selection bias ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.1 (upper figure) and 0.3 (lower figure); nominal significance level is 10%.

25% for an odds ratio of 0.67 and below 20% for an odds ratio of 0.50 even though the bias in these situation is quiet large with the median of the Mantel-Haenszel estimate ranging from 0.47 to 0.61 (for an odds ratio of 0.67) and from 0.39 to 0.47 (for an odds ratio of 0.50).

For an odds ratio of 0.25, the power of the new test is close to the predefined significance level of 10% in all situations. For the linear regression and rank correlation test, the percentage of significant results is slightly increasing with the number of trials per meta-analysis for a control event rate of 0.1; the largest power is observed in meta-analyses with 50 trials with values of 23.2% for the linear regression test and 19.1% for the rank correlation test. For an control event rate of 0.3, the power of the linear regression and rank correlation test is close to the predefined significance level of 10%.

Strong selection bias - decreasing power with increasing treatment effect

Results for the strong selection mechanism are given in Figure 4.15 and in Table A.27 and A.28.

Similar to results for moderate selection bias, the power of all tests is decreasing with increasing treatment effect for both control event rates which can be explained by the increase in the proportion of published trials with increasing treatment effect. The largest probability of publication for an odds ratio of 1.00 is 41.7% which is even markedly smaller than the smallest probability of publication for an odds ratio of 0.25 which is 74.1%. Again, for odds ratios less than 1.00, the proportion of published trials is larger for a control event rate of 0.3 than for a rate of 0.1 indicating that the bias induced by the strong selection mechanism is larger for an event rate of 0.1. Furthermore, as expected, the probabilities of publication are smaller and the Mantel-Haenszel estimates are more biased for the strong selection mechanism than for the moderate selection mechanism.

Strong bias - decreasing power with increasing heterogeneity

Again, the power of all tests is decreasing with increasing heterogeneity and the amount of bias is increasing at the same time. In this case, a decrease in the probability of publication is no explanation for the power loss. Actually, for an odds ratio of 1.00, the proportion of published trials is increasing with heterogeneity. For instance, the proportion of published trials is increasing from approximately 35.5% to 38% for a control event rate of 0.1 whereas the power of all tests is clearly decreasing.

The increase in the proportion of published trials with increasing heterogeneity for an odds ratio of 1.00 and strong selection mechanism is not astonishing. The proportion

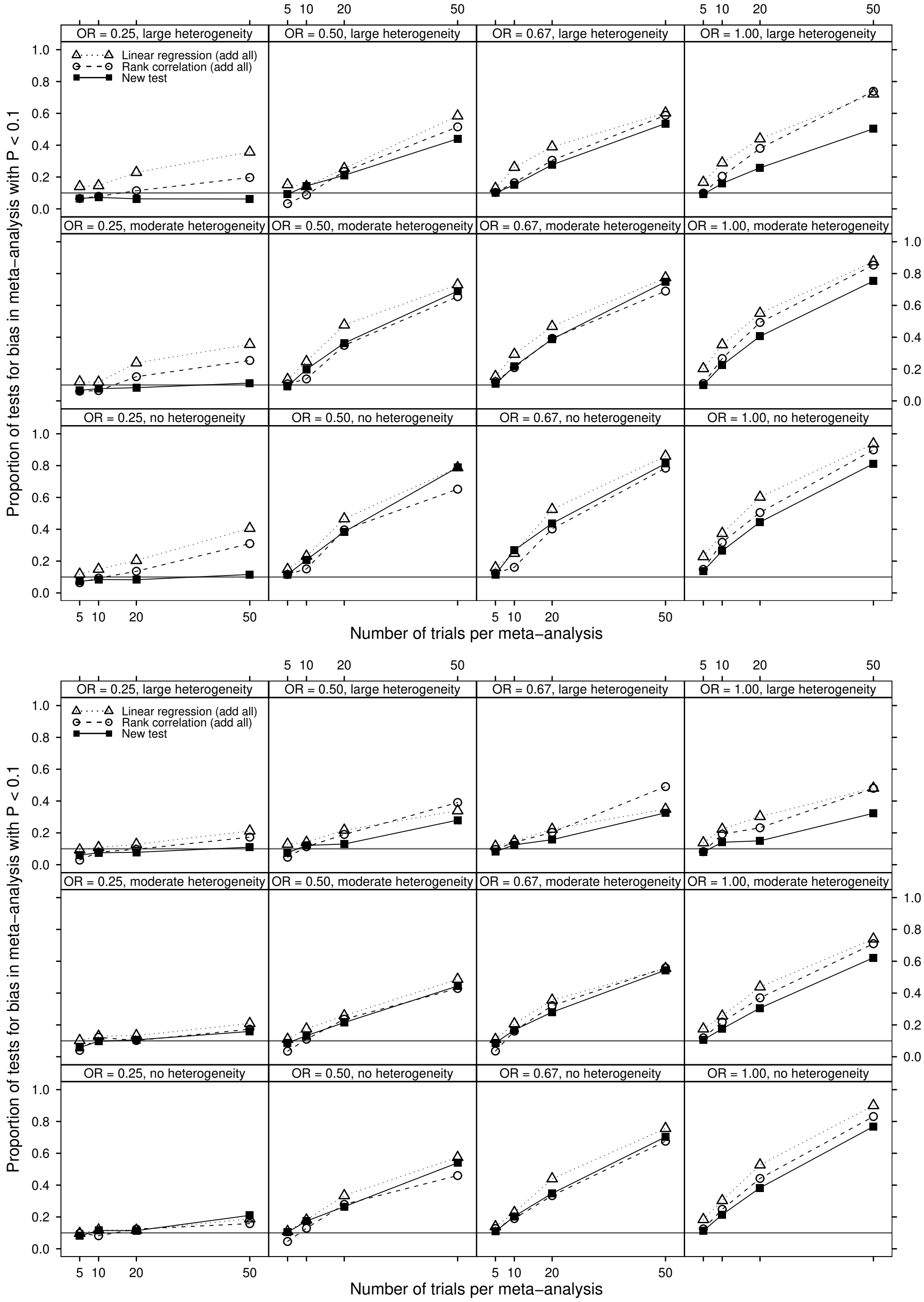


Figure 4.15: Simulation I: results under alternative hypothesis of strong selection bias ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.1 (upper figure) and 0.3 (lower figure); nominal significance level is 10%.

of trials with extreme treatment effects (both in positive and negative direction) is increasing with increasing heterogeneity. Thus, the proportion of extreme, i.e. very small or large, p-values is increasing. In fact, in the limiting case of infinite between-trial heterogeneity, one half of the trials will have a true odds ratio far below 1.00 (these trials will be published) and the other half will have a true odds ratio far above 1.00 (these trials will not be published). Accordingly, the proportion of published trials is decreasing with increasing heterogeneity in all situations where the proportion of published trials is above 50% in meta-analyses without between-trial heterogeneity, i.e. the majority of cases considered in this thesis. On the other hand, the proportion of published trials is increasing with increasing heterogeneity in all situations where the proportion of published trials is below 50% in meta-analyses without between-trial heterogeneity, i.e. in this thesis only for an odds ratio of 1.00 and strong selection bias.

Explanation for decreasing power with increasing heterogeneity

The decrease in power with increasing heterogeneity and increasing bias is explained by the use of Figure 4.16 and 4.17. In these figures, funnel plots of randomly selected meta-analyses with 50 trials generated according to the strong selection mechanism are displayed for a true odds ratio of 1.00 and control event rates of 0.1 and 0.3. Funnel plots look similar for moderate selection bias.

No between-trial heterogeneity was induced in meta-analyses shown in Figure 4.16. The Mantel-Haenszel odds ratio as well as the p-value of the rank correlation test using the "add all" approach is reported for each sub-figure. For an event rate of 0.1, the rank correlation test is significant at the 10% level in all 12 meta-analyses; in 7 cases the p-value is even smaller than 0.01. Mantel-Haenszel estimates are ranging from 0.54 to 0.67 indicating substantial bias. For an event rate of 0.3, a significant test result is observed in 11 meta-analyses; in 4 cases the p-value is smaller than 0.01. Mantel-Haenszel estimates are ranging from 0.68 to 0.78 indicating substantial bias (even though less biased than for an event rate of 0.1). For a control event rates of 0.1, asymmetry in funnel plots is clearly visible in most cases. For an event rate of 0.3, asymmetry in funnel plots is not that obvious as the standard error does not vary as much in this case.

As described above, the proportion of trials with extreme treatment effects (both in positive and negative direction) is increasing with increasing heterogeneity. Large, precise trials with odds ratios substantially greater than 1.00 will not get published due to a large p-value for the test of treatment difference. On the other hand, large, precise trials with odds ratio (substantially) smaller than 1.00 and small p-values for the test of treatment difference will get published. This will lead to more biased

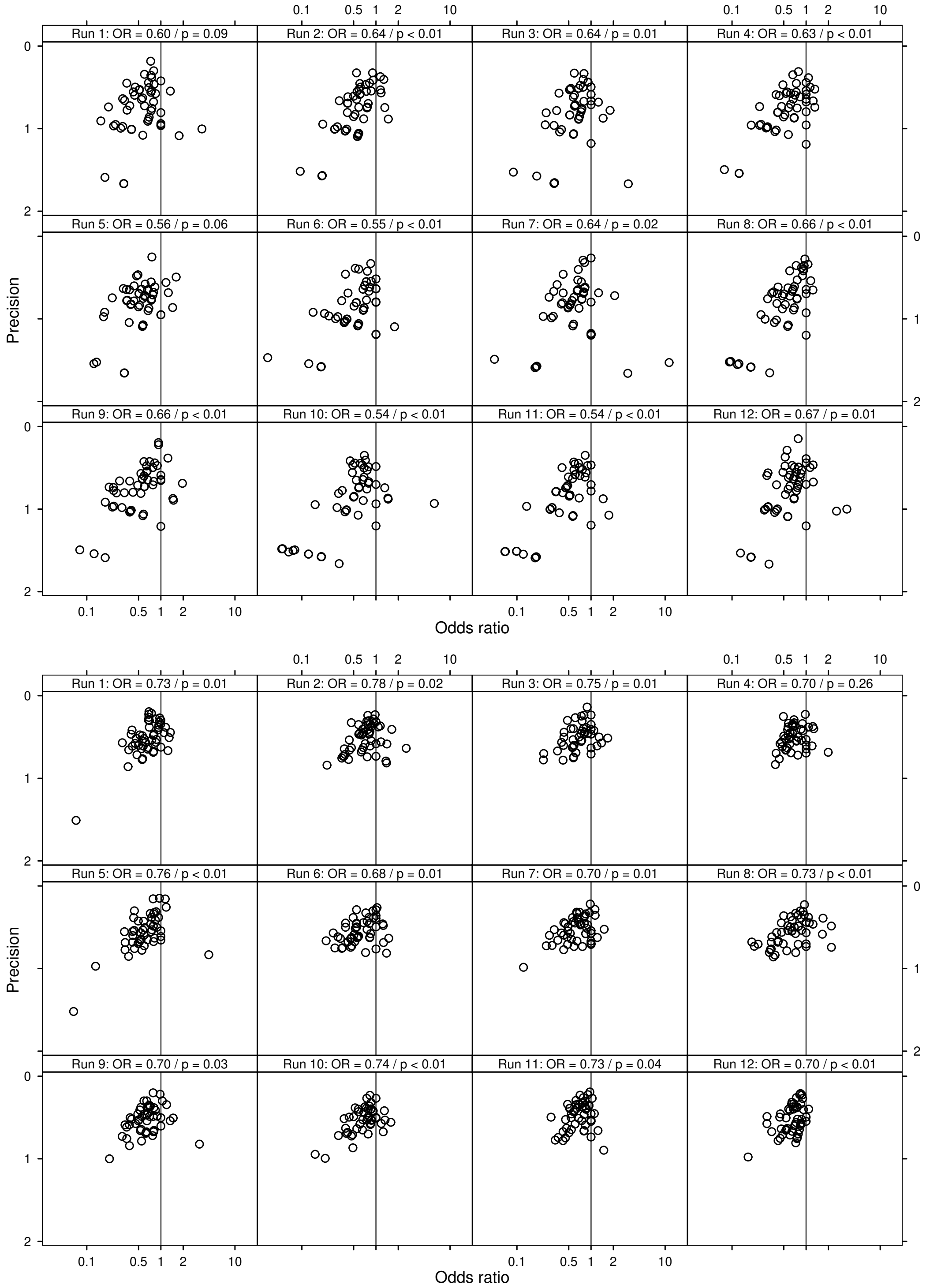


Figure 4.16: Simulation I: funnel plots of 12 randomly selected meta-analyses with 50 trials generated under strong selection bias; true odds ratio of 1.00; control event rate of 0.1 (upper figure) and 0.3 (lower figure); no between-trial heterogeneity; Mantel-Haenszel estimate and p-value of rank correlation test reported for each sub-figure.

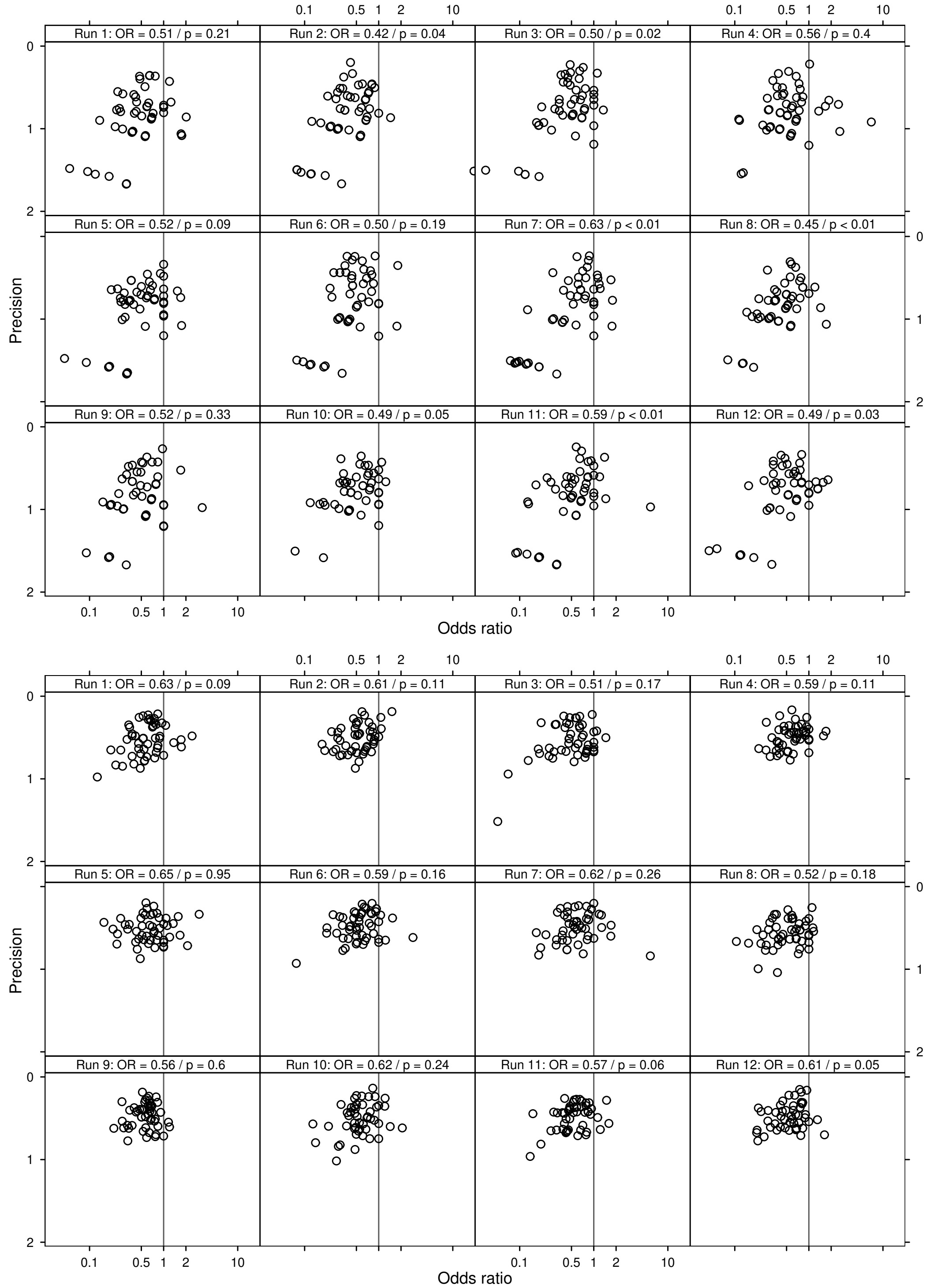


Figure 4.17: Simulation I: funnel plots of 12 randomly selected meta-analyses with 50 trials generated under strong selection bias; true odds ratio of 1.00; control event rate of 0.1 (upper figure) and 0.3 (lower figure); large between-trial heterogeneity; Mantel-Haenszel estimate and p-value of rank correlation test reported for each sub-figure.

treatment estimates in heterogeneous meta-analysis than in meta-analysis without between-trial heterogeneity.

Large between-trial heterogeneity was induced in meta-analyses plotted in Figure 4.17. For an event rate of 0.1, the rank correlation test is significant at the 10% level in 8 meta-analyses; in 3 cases the p-value is smaller than 0.01. Compared to results for no between-trial heterogeneity, more variability is present in the upper part of the funnel plots, i.e. large, more precise trials do scatter more due to between-trial heterogeneity. For an event rate of 0.3, this tendency is much more pronounced. The rank correlation test is significant at the 10% level in only 3 of 12 meta-analyses; the p-value is smaller than 0.01 in none of these cases. Asymmetry in funnel plots is not apparent in most plots; in fact, it seems as if points scatter randomly in horizontal direction.

Strong bias - power comparison

Once more, for an odds ratio of 1.00, the new test has the lowest power and the power of the linear regression and rank correlation test is of comparable size. In meta-analyses with 5 to 20 trials, the power of the linear regression test is always slightly larger than the power of the rank correlation test which can be explained by the inflation of type I error rates for the linear regression test (see Figures 4.1 and 4.4). While the power of all tests is decreasing with increasing degree of heterogeneity, this tendency is more pronounced for the new test. In meta-analyses with 10 trials, the power of all tests to detect bias is rather small with a maximum power of 37.4% (linear regression test), 31.8% (rank correlation test), and 26.9% (new test). All values are observed in meta-analyses with a control event rate of 0.1 and no between-trial heterogeneity. In this situation, the median of Mantel-Haenszel estimates is ranging from 0.60 to 0.62 indicating severe bias.

For odds ratios of 0.50 and 0.67, the power of the three tests is of comparable size in the situation of no or moderate between-trial heterogeneity. In meta-analyses with 5 to 20 trials, the power of the linear regression test is slightly larger than the power of the other tests. In meta-analyses with 10 trials, the maximum power observed for the three tests is below 30% for an odds ratio of 0.67 and below 25% for an odds ratio of 0.50. Nevertheless, severe bias is apparent in the Mantel-Haenszel estimates listed in Table A.27 and A.28.

For an odds ratio of 0.25 and a control event rate of 0.1, the power of the new test is close to the predefined significance level of 10% in all situations. The linear regression test has the largest power in this case. For an event rate of 0.3, the power of all tests is close to the predefined significance level of 10%, for meta-analyses with 5 to 20 trials.

4.2.3 Simulation II

In order to economise on space, results for the design of Simulation II are only described in the text for the case of strong selection bias; for moderate selection mechanism, the power of the tests is consistently smaller, otherwise, results are of a similar nature.

Scenario A

Results for strong selection bias are given in Figure 4.18 as well as Table A.29 and A.30 for control event rates of 0.1 and 0.3, respectively. The results are very similar to those reported for Simulation I.

The amount of bias is decreasing with increasing treatment effect; the power of all tests is decreasing at the same time. For example, the median of Mantel-Haenszel estimates ranges from 0.56 to 0.78 and probability of publication ranging from 35.7% to 42.6% for a true odds ratio of 1.00; for a true odds ratio of 0.25 the median of Mantel-Haenszel estimates ranges from 0.22 to 0.25 with corresponding probability of publication ranging from 83.3% to 100%.

The amount of bias is clearly increasing with increasing heterogeneity for true odds ratios of 0.50, 0.67, and 1.00; for a true odds ratio of 0.25, only a slight and not consistent increase is observed. The power of all tests is decreasing with increasing heterogeneity. This tendency is more pronounced for the new test, the power of the new test is clearly smaller for large between-trial heterogeneity. For no between-trial heterogeneity, the power of all tests is comparable.

In meta-analyses with 10 trials, the maximum power is below 30% for the linear regression test and below 25% for the rank correlation and new test.

Scenario B

Results are given in Figure 4.19 and Table A.31 and A.32 for Scenario B. Again, the power of all tests is decreasing with increasing treatment effect due to the fact that bias in meta-analysis is decreasing and correspondingly probability of publication is increasing. Furthermore, the power of all tests is decreasing with increasing heterogeneity which is more pronounced for the new test.

For large between-trial heterogeneity and a control event rate of 0.1, the power of the linear regression and rank correlation test is apparently larger for a true odds ratio of 0.50 than for an odds ratio of 0.67 in meta-analyses with 20 and 50 trials even though bias is decreasing at the same time. This increase can be explained by the larger percentage of significant results under the null hypothesis for an odds

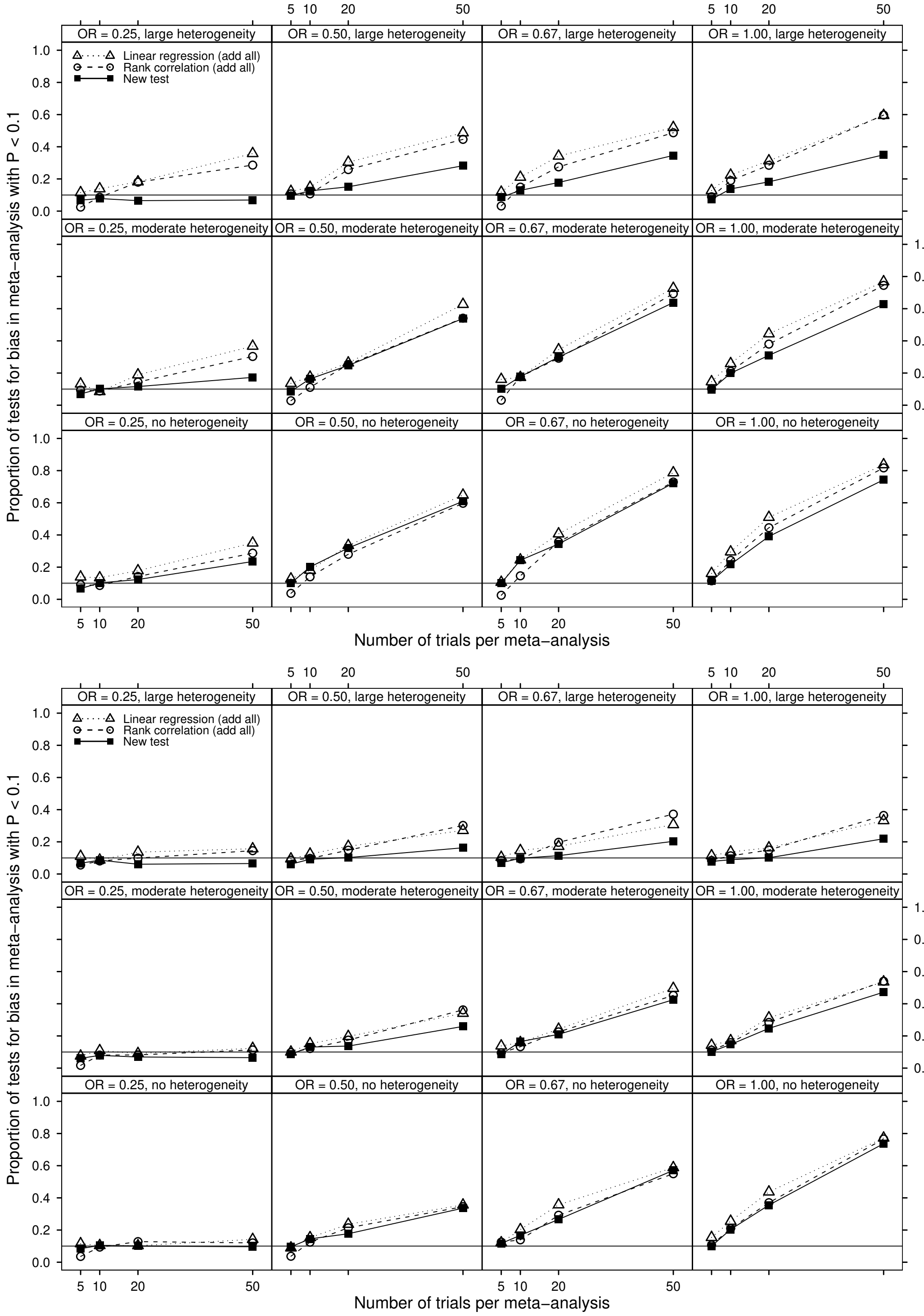


Figure 4.18: Simulation II, Scenario A: results under alternative hypothesis of strong selection bias ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.1 (upper figure) and 0.3 (lower figure); nominal significance level is 10%.

ratio of 0.50 as compared to an odds ratio of 0.67 for the linear regression and rank correlation test (see Figure 4.8 and Table A.16).
The power of the new test is always less equal than the power of the other two tests. In the case of no between-trial heterogeneity, the power of the linear regression and rank correlation test is clearly larger than the power of the new test for a true odds ratio of 0.5. This can be explained by differences in the percentage of significant results under the null hypothesis of no bias (see Figure 4.8 and Table A.16, again). In meta-analyses with 10 trials, the maximum power for all tests is observed for a true odds ratio of 1.00 and no between-trial heterogeneity. In this situation, the median of Mantel-Haenszel estimates is 0.79, however, the power of all tests to detect this apparent bias is rather small with values of 30.4% (linear regression test), 25.6% (rank correlation test), and 23.8% (new test).

Scenario C

The most diverse power results are present for Scenario C (see Figure 4.20, Table A.33 and A.34). For all three test procedures, the largest power of all simulation designs is observed for Simulation C. In meta-analyses with 50 trials and no between-trial heterogeneity, the maximum power is 99.0% for the linear regression test, 95.4% for the rank correlation test, and 95.3% for the new test; power results are very similar for control event rates of 0.1 and 0.3. On the other hand, the decrease in power with increasing heterogeneity for an odds ratio of 1.00 is most dramatic for Scenario C. For example, the power of the linear regression test is decreasing from 99.0% in meta-analyses of 50 trials with no between-trial heterogeneity to merely 18.3% in meta-analyses of 50 trials with large between-trial heterogeneity. For all three tests, the smallest power in meta-analyses with 50 trials and large between-trial heterogeneity exists in Scenario C.
For a true odds ratio of 1.00, the power of the linear regression test is much larger than the power of the rank correlation and new test for meta-analyses with 5 to 20 trials. This can be explained by differences in the percentage of significant results under the null hypothesis of no bias (see Figure 4.9, Table A.21 and A.22). Results for the linear regression test under the null hypothesis are close to the predefined significant level of 10% whereas both rank correlation and new test are rather conservative.
In meta-analyses with 10 trials, the maximum power for all tests is observed for a true odds ratio of 1.00, a control event rate of 0.3, and no between-trial heterogeneity: 49.1% (linear regression test), 33.8% (rank correlation test), and 34.3% (new test). The power of the linear regression test is substantial larger than in other simulation designs, whereas the gain in power in not that markedly for the rank correlation and new test.

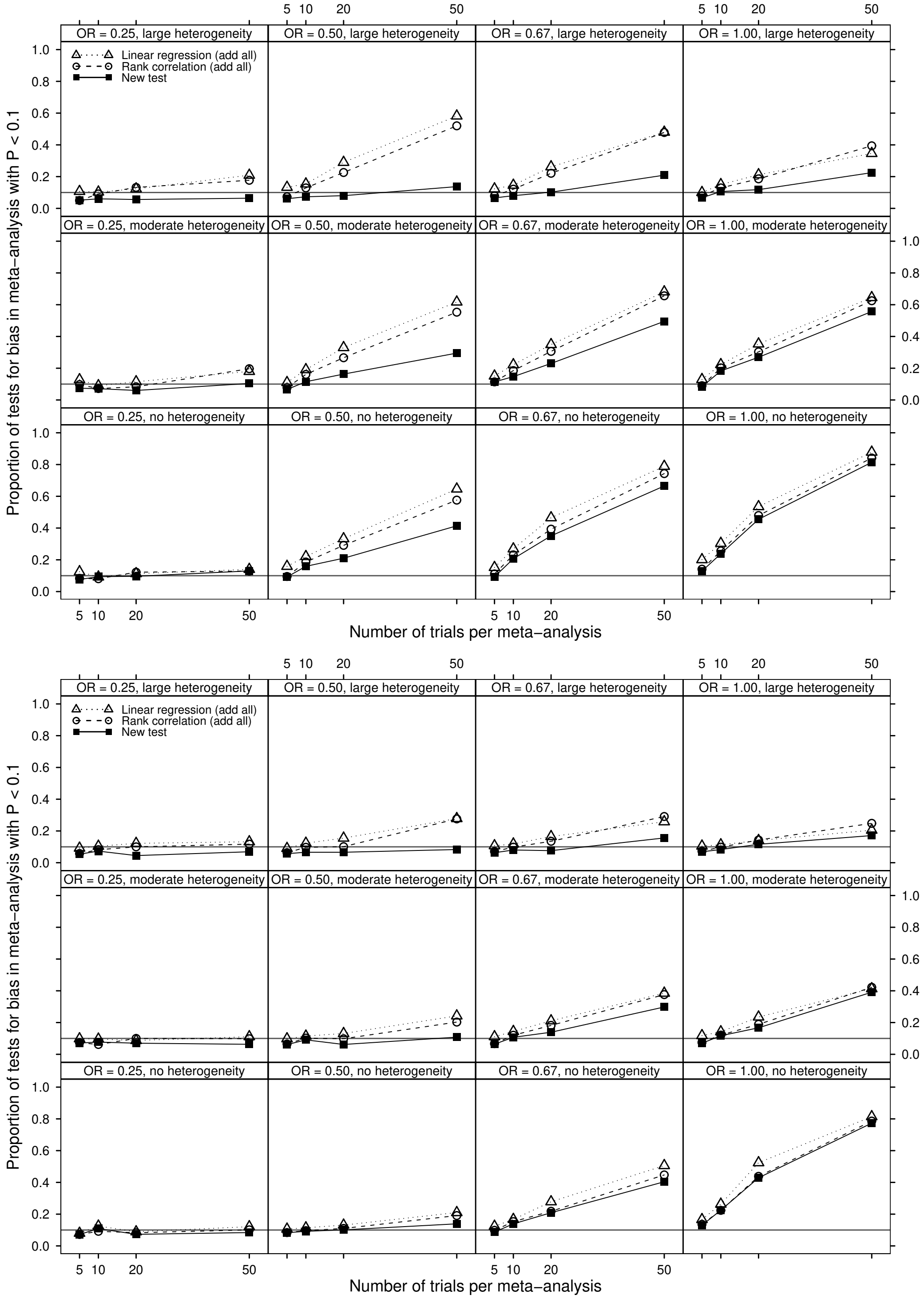


Figure 4.19: Simulation II, Scenario B: results under alternative hypothesis of strong selection bias ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.1 (upper figure) and 0.3 (lower figure); nominal significance level is 10%.

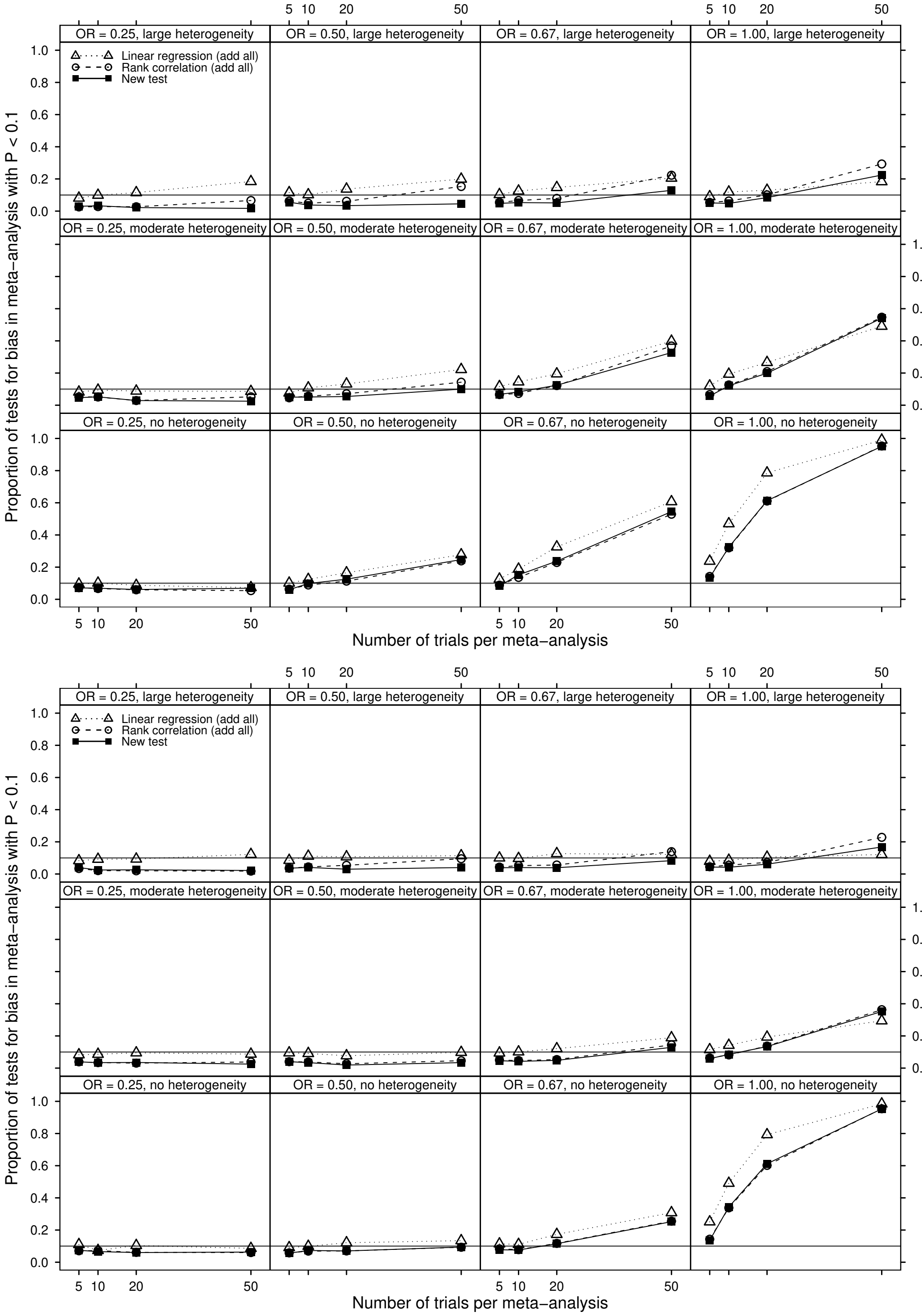


Figure 4.20: Simulation II, Scenario C: results under alternative hypothesis of strong selection bias ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.1 (upper figure) and 0.3 (lower figure); nominal significance level is 10%.

4.3 Summary of simulation results

In this chapter, type I error rate and power of tests for bias in meta-analysis are evaluated in four different simulation designs labelled Simulation I and Simulation II, Scenario A to C. These simulation designs differ only in the mechanism utilised to generate sample sizes of individual trials, all other factors in simulations are kept fixed.

Results under the null hypothesis

In simulation designs with a large proportion of small trials, i.e. Simulation I and Simulation II, Scenario A, main results under the null hypothesis can be summarised as follows:

- Linear regression and rank correlation test show a clear tendency of inflation of type I error rates with increasing treatment effect. Furthermore, asymmetry in tail probabilities with increasing treatment effect is clearly apparent for both the linear regression and rank correlation test.
- For the new test, no such tendency is present. Inflation of type I error rates is only observed in a single situation: meta-analyses with 50 trials for an odds ratio of 0.25, a control event rate of 0.1, and large between-trial heterogeneity. Asymmetry in tail probabilities is apparent for an odds ratio of 0.25 with increasing heterogeneity; the upper tail probability is increased as compared to the other two tests.
- In general, results for the risk ratio as measure of treatment effect are worse than results for the odds ratio for all tests.
- Overall, type I error rates are less inflated for both linear regression and rank correlation test using the “add all” approach.
- In meta-analyses with 10 trials, type I error rates are sufficiently close to the predefined significance level of 10% for both the rank correlation test using the “add all” approach (maximum value: 15.1%) and the new test (maximum value: 12.7%); for the linear regression test the maximum value is 20.1%.

Main results for simulations based on Simulation II, Scenario B can be subsumed in the following way:

- Linear regression and rank correlation test show a clear tendency of inflation of type I error rates with increasing treatment effect. Furthermore, asymmetry in tail probabilities with increasing treatment effect is clearly apparent for both the linear regression and rank correlation test.

- For the new test, no such tendency is present. Asymmetry in tail probabilities is apparent for an odds ratio of 0.25 with increasing heterogeneity.
- Many results for rank correlation and new test are conservative in the situation of large between-trial heterogeneity with empirical sizes as low as 3.9% (rank correlation test) and 3.2% (new test).
- In general, results for the risk ratio as measure of treatment effect are worse than results for the odds ratio for all tests.
- Overall, type I error rates are less inflated for both linear regression and rank correlation test using the "add all" approach.
- In meta-analyses with 10 trials, type I error rates are ranging from 7.9% to 13.9% and from 4.1% to 11.8% for linear regression and rank correlation test both using "add all" approach and from 4.1% to 10.9% for the new test.

Main results for simulations based on Simulation II, Scenario C are given in the following list:

- Results for the linear regression test are close to the predefined significance level of 10%.
- The percentage of significant results is too small for both the rank correlation and new test, i.e. these tests are rather conservative. Some results are very conservative with type I error rates as low as 0.7% and 0.6% for rank correlation and new test, respectively.
- The proportion of significant results is decreasing with increasing heterogeneity for both the rank correlation and new test.
- In meta-analyses with 10 trials, type I error rates are ranging from 6.9% to 10.3% and from 2.3% to 8.0% for linear regression and rank correlation test both using the "add all" approach and from 2.5% to 8.8% for the new test.

Results under alternatives

Main results under the alternative of publication bias are similar for different study designs and can be summarised as follows:

- Power of all tests is decreasing with increasing treatment effect. This result is not surprising since the probability of publication of a trial is increasing with treatment effect.

- Power of all tests is decreasing with increasing heterogeneity whereas the amount of bias is increasing at the same time. Overall, the loss in power is more pronounced for the new test.

- In meta-analyses with 10 trials and moderate selection bias, power estimates are ranging from

  - 9.2% to 29.3% (linear regression test),
    6.2% to 23.3% (rank correlation test),
    6.7% to 19.9% (new test) for Simulation I
  - 7.6% to 21.9% (linear regression test),
    5.7% to 20.0% (rank correlation test),
    6.4% to 17.7% (new test) for Simulation II, Scenario A
  - 8.0% to 24.2% (linear regression test),
    6.2% to 20.6% (rank correlation test),
    5.4% to 19.0% (new test) for Simulation II, Scenario B
  - 8.4% to 34.1% (linear regression test),
    2.5% to 23.7% (rank correlation test),
    2.8% to 25.0% (new test) for Simulation II, Scenario C.

- In meta-analyses with 10 trials and strong selection bias, power estimates are ranging from

  - 10.8% to 37.4% (linear regression test),
    6.5% to 31.8% (rank correlation test),
    7.2% to 26.9% (new test) for Simulation I
  - 8.5% to 29.5% (linear regression test),
    8.1% to 24.2% (rank correlation test),
    7.8% to 24.3% (new test) for Simulation II, Scenario A
  - 9.0% to 30.4% (linear regression test),
    6.2% to 25.6% (rank correlation test),
    6.0% to 23.8% (new test) for Simulation II, Scenario B
  - 7.3% to 49.1% (linear regression test),
    2.1% to 33.8% (rank correlation test),
    2.5% to 34.3% (new test) for Simulation II, Scenario C.

Chapter 5

Discussion

In this thesis, I investigate the properties of statistical tests for bias in meta-analysis via simulations both under the null hypothesis of no publication bias and different alternatives. Two commonly used tests, i.e. a rank correlation test (Begg and Mazumdar, 1994) and a linear regression test (Egger et al., 1997), as well as a newly proposed test are considered in the simulations. The new test is a modification of the rank correlation test by Begg and Mazumdar (1994) using different factors; the main difference consists in the use of a different estimate for the within-trial variance. In the new test, the variance estimate is derived from the conditional distribution of the observed cell count given the marginal totals which follows a non-central hypergeometric distribution.

Empirical information is considered in the decision on the simulation design in an attempt to create realistic meta-analyses. A total of four different simulation designs is used which differ only in the mechanism utilised to generate sample sizes of individual trials; other factors (number of trials per meta-analysis, true treatment effect, control event rate, between-trial variance, amount of bias) are equal in all designs. The proportion of small trials is very large in two simulation designs (labelled Simulation I and Simulation II, Scenario A) with 66% and 30% of trials having a total sample size less than 100, respectively. On the other hand, one simulation design is used with a large proportion of large trials (labelled Simulation II, Scenario C); 50% of trials have a sample size larger than 1000 in this simulation design. Sample sizes generated in the fourth simulation design (labelled Simulation II, Scenario B) lie somewhere between these extremes. Accordingly, the amount of sparse data is very different in these simulation designs.

Different degrees of between-trial heterogeneity are considered in simulations both under the null hypothesis and alternatives. The influence of heterogeneity on results of the linear regression and rank correlation test under the null hypothesis has been investigated before (Schwarzer et al., 2002). In this thesis, the influence of heterogeneity on the power of these tests (i.e. under alternatives) is to my knowledge investigated for the first time. Furthermore, the evaluation of the influence of different strategies to deal with sparse binary data on the results of the linear regression and rank correlation test is also novel.

A clear tendency of inflation of type I error rates with increasing treatment effect is apparent for linear regression and rank correlation test in simulations with a large amount of sparse binary data (Simulation I and Simulation II, Scenario A). Furthermore, a notable tendency of asymmetry in tail probabilities with increasing treatment effects is observed in this situation; both tests more often have significant results below the lower critical value which corresponds to an overestimation of the treatment effect. Results for the new test are much better in this situation; an inflation of type I error rates is only observed in meta-analyses with 50 trials in the situation of an odds ratio or risk ratio of 0.25, large between-trial heterogeneity, and a control event rate of 0.1. Moreover, a tendency of asymmetry in tail probabilities is only apparent for an odds ratio or risk ratio of 0.25.

In Section 4.1.3, it is shown that an association between estimated treatment effect, i.e. log odds ratio, and its standard error is a plausible explanation for the inflation of type I error rates with increasing treatment effect observed in the linear regression and rank correlation test. Estimates of the odds ratio and the variance of the log odds ratio are based explicitly on the four cell counts in the corresponding two-by-two table. As compared to the rank correlation test, variance estimates used in the new test depend on the following quantities: Mantel-Haenszel odds ratio, total sample size, sample size in intervention group, and total number of events. The Mantel-Haenszel odds ratio is kept fixed for each trial, thus, variance estimates of individual trials are essentially defined by the total sample size and the margins of the corresponding two-by-two tables.

Another potential modification of the rank correlation test which is not considered in this thesis consists in the use of smoothed variance estimates which have been utilised in simulation studies in the context of random-effects meta-regression (Berkey et al., 1995; Knapp and Hartung, 2003). Smoothed variance estimates are used to reduce the correlation between estimated treatment effect and its variance estimate. Evaluation of the usefulness of smoothed variance estimates in tests for bias in meta-analysis is an interesting topic for future research.

Two approaches to deal with sparse binary data are evaluated in simulations: either 0.5 is added to all two-by-two tables ("add all") or 0.5 is added only to cell counts of corresponding two-by-two tables with zero cell counts ("add selective"). Overall, type I error rates are less inflated for both linear regression and rank correlation test using the "add all" approach. Due to this result, the use of the "add all" approach is suggested for linear regression and rank correlation test.

McAuley et al. (2000) report a median number of 10 trials per meta-analysis published in medical journals with interquartile range from 6 to 19; the median number of trials per meta-analysis in Cochrane reviews is 6 trials with interquartile range from 3 to 12 according to Mallett and Clarke (2002). Therefore, simulation results for meta-analysis with 10 trials can be regarded as results for a typical number of trials per meta-analysis found in medical applications. In this situation, the inflation of type I error rates is not that dramatic for the rank correlation test using the "add all" approach with a maximum percentage of significant results of 15.1%. For the linear regression test using the "add all" approach this probability is still 20.1%. Accordingly, the application of the rank correlation test using the "add all" approach seems justified in meta-analyses with 10 trials.

Simulations under the alternative of bias in meta-analysis reveal two main results. Firstly, the power of all tests is decreasing with increasing treatment effect which is not a surprising result since the probability of publication of a trial is increasing and correspondingly the amount of bias is decreasing with treatment effect. Secondly, the power of all tests is decreasing with increasing heterogeneity whereas the amount of bias is increasing at the same time; overall, loss in power is more pronounced for the new test. An explanation for this contradictory result is given in Section 4.2.2. Essentially, the problem is that the basic assumption of a funnel plot that all trials come from the same underlying population is violated in heterogeneous meta-analyses. Accordingly, it was suggested in a recent paper that the funnel plot is inappropriate for heterogeneous meta-analyses (Terrin et al., 2003). One should be aware that there is a high chance that the treatment effect estimated in a meta-analysis with substantial heterogeneity might be biased even though tests for bias in meta-analysis have a non-significant result.

However, the power of all tests is not sufficient even in homogeneous meta-analyses in clinical important situations. In simulations with strong selection bias, the maximum power in meta-analysis with 10 trials is 49.1% for the linear regression test and merely 33.8% and 34.3% for the rank correlation and new test, respectively. Similar findings have been reported in other publications using different simulation designs (Macaskill et al., 2001; Sterne et al., 2000). Due to this low power, one may argue that approaches to quantify the extent of bias should be used instead of statistical tests for bias.

A straightforward approach for the rank correlation test is to report Kendall's tau with corresponding confidence interval instead of the p-value. The uncertainty on the extent of selection bias is reflected in the width of the confidence interval. A disadvantage of this approach is that Kendall's tau does not reveal any information on the actual extent of bias in the treatment effect.

One method to quantify the bias and ultimately adjust for bias in meta-analysis are selection models (Terrin et al., 2003). These models using weighted distribution theory have been under development for almost two decades; weight functions were introduced into meta-analysis by Hedges (1984). Selection models have been criticised for making too simplistic assumptions with respect to the probability of publication; furthermore, these assumptions cannot be verified in the analysis (Sutton et al., 2000).

A more promising approach to quantify bias in meta-analysis are selectivity models (Copas, 1999; Copas and Shi, 2000; Shi and Copas, 2002). Compared to selection models, the probability of publication is modelled more realistically by considering both the size of the treatment effect and the precision (i.e. sample size). It is not possible to estimate all parameters in selectivity models and therefore it is suggested to use the method in a sensitivity analysis by considering different patterns of publication bias. So far, the properties of selectivity models have not been investigated in much detail. Furthermore, only few applications have been published probably due to the complexity of the method and a lack of user-friendly software to conduct the analyses. Future research should focus on the evaluation of the properties of selectivity models both in simulations and applications.

Bibliography

Agresti, A. (1990). *Categorical Data Analysis.* New York: Wiley.

Altman, D. G. (2001). Systematic reviews of evaluations of prognostic variables. *British Medical Journal 323*, 224–228.

Begg, C. B. (1994). Publication bias. In H. Cooper and L. V. Hedges (Eds.), *The Handbook of Research Synthesis*, pp. 399–409. Newbury Park, CA: Russell Sage Foundation.

Begg, C. B. and J. A. Berlin (1988). Publication bias: A problem in interpreting medical data. *Journal of the Royal Statistical Society, Series A 151*, 419–445. (C/R: p445-463).

Begg, C. B. and M. Mazumdar (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics 50*, 1088–1101.

Berkey, C. S., D. C. Hoaglin, F. Mosteller, and G. A. Colditz (1995). A random-effects regression model for meta-analysis. *Statistics in Medicine 14*, 395–411.

Berlin, J. A., C. B. Begg, and T. A. Louis (1989). An assessment of publication bias using a sample of published clinical trials. *Journal of the American Statistical Association 84*, 381–392.

Breslow, N. and N. Day (1980). *Statistical methods in cancer research. Volume I - The analysis of case-control studies.* Number 32. Lyon: IARC Sci Pub.

Breslow, N. E. (1981). Odds ratio estimators when the data are sparse. *Biometrika 68*, 73–84.

Carlin, J. B. (1992). Meta-analysis for 2×2 tables: A Bayesian approach. *Statistics in Medicine 11*, 141–158.

Clarke, M. and P. Langhorne (2001). Revisiting the Cochrane Collaboration. *British Medical Journal 323*, 821.

Conover, W. J. (1971). *Practical Nonparametric Statistics.* New York: Wiley.

Copas, J. (1999). What works?: Selectivity models and meta-analysis. *Journal of the Royal Statistical Society, Series A 162*, 95–109.

Copas, J. and J. Q. Shi (2000). Meta-analysis, funnel plots and sensitivity analysis. *Biostatistics 1*, 247–262.

Copas, J. B. and J. Q. Shi (2001). A sensitivity analysis for publication bias in systematic reviews. *Statistical Methods in Medical Research 10*, 251–265.

Deeks, J. J. (2002). Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine 21*, 1575–1600.

Deeks, J. J. and D. G. Altman (2001). Effect measures for meta-analysis of trials with binary outcome. In M. Egger, G. D. Smith, and D. G. Altman (Eds.), *Systematic reviews in health care: Meta-analysis in context* (2 ed.)., pp. 313–335. London: BMJ Publishing Group.

Deeks, J. J., M. J. Bradburn, W. Bilker, R. Localio, and J. Berlin (1998). Much ado about nothing: Meta-analysis for rare events [abstract]. 6th Cochrane Colloquium, Baltimore, MD.

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

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Journal of the American Medical Association 263*, 1385–1389.

Dickersin, K. (1997). How important is publication bias? A synthesis of available data. *AIDS Education and Prevention 9*, Supplement A, 15–21.

Dickersin, K. and D. Rennie (2003). Registering clinical trials. *Journal of the American Medical Association 290*, 516–523.

Duval, S. and R. Tweedie (2000a). A nonparametric "Trim and Fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association 95*, 89–98.

Duval, S. and R. Tweedie (2000b). Trim and Fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics 56*, 455–463.

Egger, M. and G. D. Smith (1997). Meta-analysis: Potentials and promise. *British Medical Journal 315*, 1371–1374.

Egger, M. and G. D. Smith (1998). Meta-analysis: Bias in location and selection of studies. *British Medical Journal 316*, 61–66.

Egger, M., G. D. Smith, M. Schneider, and C. Minder (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal 315*, 629–634.

Egger, M., T. Zellweger-Zähner, M. Schneider, C. Junker, C. Lengeler, and G. Antes (1997). Language bias in randomised controlled trials published in English and German. *The Lancet 350*, 326–329.

Emerson, J. D. (1994). Combining estimates of the odds ratio: The state of the art. *Statistical Methods in Medical Research 3*, 157–178.

Engels, E. A., C. H. Schmid, N. Terrin, I. Olkin, and J. Lau (2000). Heterogeneity and statistical significance in meta-analysis: An empirical study of 125 meta-analyses. *Statistics in Medicine 19*, 1707–1728.

Fleiss, J. L. (1993). The statistical basis of meta-analysis. *Statistical Methods in Medical Research 2*, 121–145.

Galbraith, R. F. (1988a). Graphical display of estimates having differing standard errors. *Technometrics 30*, 271–281.

Galbraith, R. F. (1988b). A note on graphical presentation of estimated odds ratios from several clinical trials. *Statistics in Medicine 7*, 889–894.

Gart, J. J. and J. R. Zweifel (1967). On the bias of various estimators of the logit and its variance with application to quantal bioassay. *Biometrika 54*, 181–187.

Gleser, L. J. and I. Olkin (1996). Models for estimating the number of unpublished studies. *Statistics in Medicine 15*, 2493–2507.

Greenland, S. and J. M. Robins (1985). Estimation of a common effect parameter from sparse follow-up data. *Biometrics 41*, 55–68. (C/R: V45 p1323–1324).

Greenland, S. and A. Salvan (1990). Bias in the one-step method for pooling study results. *Statistics in Medicine 9*, 247–252.

Hartung, J. (1999). An alternative method for meta-analysis. *Biometrical Journal 41*, 901–916.

Hartung, J. and G. Knapp (2001). A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine 20*, 3875–3889.

Hedges, L. V. (1984). Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics 9*, 61–85.

Hedges, L. V. (1992). Modeling publication selection effects in meta-analysis. *Statistical Science 7*, 246–255.

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

Irwig, L., P. Macaskill, G. Berry, and P. Glasziou (1998). Bias in meta-analysis detected by a simple, graphical test. Graphical test is itself biased. *British Medical Journal 316*, 470; author reply 470–1.

Iyengar, S. and J. B. Greenhouse (1988). Selection models and the file drawer problem. *Statistical Science 3*, 109–117.

Jadad, A. R., D. J. Cook, A. Jones, T. P. Klassen, P. Tugwell, M. Moher, and D. Moher (1998). Methodology and reports of systematic reviews and meta-analyses: A comparison of Cochrane reviews with articles published in paper-based journals. *Journal of the American Medical Association 280*, 278–80.

Jüni, P., D. G. Altman, and M. Egger (2001). Assessing the quality of controlled clinical trials. *British Medical Journal 323*, 42–46.

Kendall, M. and J. D. Gibbons (1990). *Rank Correlation Methods* (Fifth ed.). London: Edward Arnold.

Knapp, G. and J. Hartung (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine 22*, 2693–2710.

Krzyzanowska, M., M. Pintilie, and I. Tannock (2003). Factors associated with failure to publish large randomized trials presented at an oncology meeting. *Journal of the American Medical Association 290*, 495–501.

Lane, D. M. and W. P. Dunlap (1978). Estimating effect size: Bias resulting from the significance criterion in editorial decisions. *British Journal of Mathematical and Statistical Psychology 31*, 107–112.

Liao, J. (1992). An algorithm for the mean and variance of the noncentral hypergeometric distribution. *Biometrics 48*, 889–892.

Liao, J. G. and O. Rosen (2001). Fast and stable algorithms for computing and sampling from the noncentral hypergeometric distribution. *The American Statistician 55*, 366–369.

Light, R. J. and D. B. Pillemer (1984). *Summing up. The science of reviewing research.* London: Harvard University Press.

Macaskill, P., S. D. Walter, and L. Irwig (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine 20*, 641–654.

Mallett, S. and M. Clarke (2002). The typical Cochrane review. How many trials? How many participants? *International Journal of Technology Assessment in Health Care 18*, 820–831.

Mantel, N. and W. Haenszel (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute 22*, 719–748.

McAuley, L., B. Pham, P. Tugwell, and D. Moher (2000). Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *The Lancet 356*, 1228–31.

Moher, D., B. Pham, T. P. Klassen, K. F. Schulz, J. A. Berlin, A. R. Jadad, and A. Liberati (2000). What contributions do languages other than English make on the results of meta-analyses? *Journal of Clinical Epidemiology 53*, 964–972.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics 8*, 157–159.

Parmar, M. K. B., V. Torri, and L. Stewart (1998). Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Statistics in Medicine 17*, 2815–2834.

Pettigrew, H. M., J. J. Gart, and D. G. Thomas (1986). The bias and higher cumulants of the logarithm of a binomial variate. *Biometrika 73*, 425–435.

Plackett, R. L. (1974). *The Analysis of Categorical Data.* London: Charles Griffin & Company Ltd.

R Development Core Team (2003). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-00-3.

Robins, J., N. E. Breslow, and S. Greenland (1986). Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting models. *Biometrics 42*, 311–323.

Robins, J., S. Greenland, and N. E. Breslow (1986). A general estimator for the variance of the Mantel-Haenszel odds ratio. *American Journal of Epidemiology 124*, 719–723.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin 86*, 638–641.

Sackett, D. L., S. Straus, S. Richardson, W. Rosenberg, and R. B. Haynes (2000). *Evidence-Based Medicine: How to Practice and Teach EBM (2nd edition)*. London: Churchill Livingstone.

Schmid, C. H., J. Lau, M. W. McIntosh, and J. C. Cappelleri (1998). An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Statistics in Medicine 17*, 1923–1942.

Schulz, K. F., I. Chalmers, R. J. Hayes, and D. G. Altman (1995). Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association 273*, 408–412.

Schwarzer, G., G. Antes, and M. Schumacher (2002). Inflation of type I error rate in two statistical tests for the detection of publication bias in meta-analyses with binary outcomes. *Statistics in Medicine 21*, 2465–2477.

Shi, J. Q. and J. Copas (2002). Publication bias and meta-analysis for 2×2 tables: an average Markov chain Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B 64*, 221–236.

Simes, R. J. (1986). Publication bias: The case for an international registry of clinical trials. *Journal of Clinical Oncology 4*, 1529–1541.

Smith, T. C., D. J. Spiegelhalter, and A. Thomas (1995). Bayesian approaches to random-effects meta-analysis: A comparative study. *Statistics in Medicine 14*, 2685–2699.

StataCorp (2003). *Stata Statistical Software: Release 8.0.* College Station, TX: StataCorp LP.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance – or vice versa. *Journal of the American Statistical Association 54*, 30–34. (Comment: V54 p593).

Sterling, T. D., W. L. Rosenbaum, and J. J. Weinkam (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician 49*, 108–112.

Sterne, J. A. C. and M. Egger (2001). Funnel plots for detecting bias in meta-analysis: Guideline on choice of axis. *Journal of Clinical Epidemiology 54*, 1046–1055.

Sterne, J. A. C., D. Gavaghan, and M. Egger (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology 53*, 1119–1129.

Sutton, A. J., F. Song, S. M. Gilbody, and K. R. Abrams (2000). Modelling publication bias in meta-analysis: A review. *Statistical Methods in Medical Research 9*, 421–445.

Taggart, D. P., R. D'Amico, and D. G. Altman (2001). Effect of arterial revascularisation on survival: A systematic review of studies comparing bilateral and single internal mammary arteries. *The Lancet 358*, 870–875.

Terrin, N., C. H. Schmid, J. Lau, and I. Olkin (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine 22*, 2113–2126.

The Cochrane Collaboration (2003). *Review Manager (RevMan) [Computer program]. Version 4.2 for Windows.* Oxford, England.

Thompson, S. G. and J. P. T. Higgins (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine 21*, 1559–1573.

Thompson, S. G. and S. J. Pocock (1991). Can meta-analyses be trusted? *The Lancet 338*, 1127–1130.

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

Thornton, A. and P. Lee (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology 53*, 207–216.

Tonks, A. (1999). Registering clinical trials. *British Medical Journal 319*, 1565–1568.

van Houwelingen, H. C., L. R. Arends, and T. Stijnen (2002). Tutorial in biostatistics – Advanced methods in meta-analysis: Multivariate approach and meta-regression. *Statistics in Medicine 21*, 589–624.

Warn, D. E., S. G. Thompson, and D. J. Spiegelhalter (2002). Bayesian random effects meta-analysis of trials with binary outcomes: Methods for the absolute risk difference and relative risk scales. *Statistics in Medicine 21*, 1601–1623.

Yusuf, S., R. Peto, J. Lewis, R. Collins, and P. Sleight (1985). Beta blockade during and after myocardial infarction: An overview of the randomized trials. *Progress in Cardiovascular Diseases 27*, 335–371.

Appendix

Appendix A

Tables of simulation results

A.1 Publication bias absent

The following abbreviations are used in tables listed in this section.

Given values:

τ^2	Between-trial variance
ψ	Underlying odds ratio
ϕ	Underlying risk ratio
k	Number of trials in meta-analysis
p_c	Control event rate

Estimated values:

$\widehat{\psi}_{MH}$	Mantel-Haenszel odds ratio
$\widehat{\phi}_{MH}$	Mantel-Haenszel risk ratio
$\widehat{\psi}_{IV}$	Pooled odds ratio based on inverse variance method
$\widehat{\phi}_{IV}$	Pooled risk ratio based on inverse variance method

τ^2	ψ / ϕ	k	$\widehat{\psi}_{MH}$ ($p_c = 0.1$) 50% [5%; 95%]	$\widehat{\psi}_{MH}$ ($p_c = 0.3$) 50% [5%; 95%]	$\widehat{\phi}_{MH}$ ($p_c = 0.1$) 50% [5%; 95%]	$\widehat{\phi}_{MH}$ ($p_c = 0.3$) 50% [5%; 95%]
0.00	0.25	5	0.27 [0.14; 0.51]	0.25 [0.16; 0.38]	0.27 [0.15; 0.49]	0.25 [0.16; 0.38]
0.00	0.25	10	0.27 [0.17; 0.42]	0.25 [0.19; 0.34]	0.27 [0.18; 0.42]	0.25 [0.19; 0.33]
0.00	0.25	20	0.27 [0.20; 0.37]	0.25 [0.20; 0.31]	0.27 [0.20; 0.37]	0.25 [0.20; 0.30]
0.00	0.25	50	0.27 [0.22; 0.33]	0.25 [0.22; 0.28]	0.27 [0.23; 0.33]	0.25 [0.22; 0.28]
0.00	0.50	5	0.51 [0.29; 0.86]	0.49 [0.35; 0.71]	0.51 [0.29; 0.84]	0.49 [0.36; 0.68]
0.00	0.50	10	0.51 [0.33; 0.73]	0.50 [0.39; 0.65]	0.51 [0.35; 0.74]	0.50 [0.41; 0.60]
0.00	0.50	20	0.51 [0.39; 0.66]	0.50 [0.42; 0.60]	0.50 [0.39; 0.66]	0.50 [0.44; 0.58]
0.00	0.50	50	0.51 [0.43; 0.59]	0.50 [0.45; 0.56]	0.51 [0.43; 0.59]	0.50 [0.46; 0.55]
0.00	0.67	5	0.67 [0.38; 1.15]	0.66 [0.47; 0.94]	0.67 [0.40; 1.10]	0.66 [0.50; 0.86]
0.00	0.67	10	0.66 [0.46; 0.97]	0.67 [0.52; 0.85]	0.67 [0.48; 0.94]	0.67 [0.55; 0.82]
0.00	0.67	20	0.68 [0.52; 0.87]	0.67 [0.57; 0.80]	0.68 [0.53; 0.85]	0.67 [0.59; 0.76]
0.00	0.67	50	0.67 [0.57; 0.80]	0.67 [0.61; 0.74]	0.67 [0.58; 0.78]	0.67 [0.62; 0.73]
0.00	1.00	5	1.00 [0.58; 1.63]	1.00 [0.72; 1.38]	1.00 [0.65; 1.53]	1.00 [0.79; 1.25]
0.00	1.00	10	1.00 [0.71; 1.41]	1.01 [0.80; 1.27]	1.00 [0.75; 1.35]	0.99 [0.86; 1.16]
0.00	1.00	20	1.01 [0.79; 1.28]	1.00 [0.85; 1.16]	1.00 [0.81; 1.23]	1.00 [0.89; 1.12]
0.00	1.00	50	1.00 [0.87; 1.15]	1.00 [0.91; 1.11]	1.00 [0.88; 1.14]	1.00 [0.94; 1.07]
0.05	0.25	5	0.27 [0.14; 0.54]	0.25 [0.15; 0.39]	0.27 [0.14; 0.52]	0.25 [0.15; 0.38]
0.05	0.25	10	0.27 [0.17; 0.44]	0.25 [0.18; 0.35]	0.27 [0.17; 0.43]	0.25 [0.18; 0.34]
0.05	0.25	20	0.27 [0.20; 0.37]	0.25 [0.20; 0.31]	0.27 [0.20; 0.37]	0.25 [0.20; 0.32]
0.05	0.25	50	0.27 [0.22; 0.33]	0.25 [0.22; 0.29]	0.28 [0.23; 0.33]	0.25 [0.22; 0.29]
0.05	0.50	5	0.51 [0.27; 0.92]	0.50 [0.31; 0.74]	0.50 [0.28; 0.87]	0.49 [0.34; 0.70]
0.05	0.50	10	0.50 [0.32; 0.75]	0.50 [0.38; 0.67]	0.52 [0.33; 0.76]	0.50 [0.38; 0.63]
0.05	0.50	20	0.51 [0.38; 0.70]	0.50 [0.41; 0.62]	0.51 [0.38; 0.67]	0.50 [0.42; 0.61]
0.05	0.50	50	0.51 [0.42; 0.61]	0.50 [0.44; 0.57]	0.51 [0.43; 0.60]	0.50 [0.45; 0.56]
0.05	0.67	5	0.68 [0.39; 1.20]	0.67 [0.44; 0.99]	0.67 [0.38; 1.09]	0.66 [0.47; 0.91]
0.05	0.67	10	0.68 [0.44; 1.00]	0.67 [0.51; 0.90]	0.67 [0.47; 0.97]	0.66 [0.53; 0.84]
0.05	0.67	20	0.67 [0.50; 0.89]	0.67 [0.55; 0.81]	0.68 [0.52; 0.87]	0.67 [0.56; 0.79]
0.05	0.67	50	0.67 [0.57; 0.80]	0.67 [0.59; 0.76]	0.67 [0.57; 0.79]	0.67 [0.60; 0.75]
0.05	1.00	5	1.00 [0.60; 1.71]	1.00 [0.69; 1.50]	1.00 [0.60; 1.63]	1.00 [0.74; 1.36]
0.05	1.00	10	1.00 [0.69; 1.46]	1.00 [0.75; 1.29]	1.00 [0.73; 1.39]	0.99 [0.81; 1.24]
0.05	1.00	20	0.99 [0.77; 1.29]	1.01 [0.83; 1.19]	1.00 [0.79; 1.26]	0.99 [0.86; 1.16]
0.05	1.00	50	1.00 [0.85; 1.18]	1.00 [0.88; 1.13]	1.00 [0.85; 1.17]	1.00 [0.91; 1.11]
0.25	0.25	5	0.27 [0.12; 0.57]	0.26 [0.14; 0.46]	0.28 [0.13; 0.56]	0.25 [0.14; 0.44]
0.25	0.25	10	0.28 [0.16; 0.48]	0.25 [0.17; 0.39]	0.27 [0.16; 0.45]	0.25 [0.16; 0.37]
0.25	0.25	20	0.27 [0.19; 0.40]	0.25 [0.19; 0.35]	0.28 [0.19; 0.40]	0.25 [0.18; 0.33]
0.25	0.25	50	0.27 [0.22; 0.35]	0.26 [0.21; 0.31]	0.27 [0.21; 0.34]	0.25 [0.20; 0.31]
0.25	0.50	5	0.52 [0.23; 1.05]	0.50 [0.28; 0.90]	0.50 [0.25; 1.00]	0.51 [0.28; 0.86]
0.25	0.50	10	0.51 [0.31; 0.84]	0.50 [0.34; 0.76]	0.50 [0.31; 0.81]	0.50 [0.34; 0.74]
0.25	0.50	20	0.51 [0.35; 0.74]	0.51 [0.37; 0.71]	0.50 [0.36; 0.72]	0.50 [0.37; 0.66]
0.25	0.50	50	0.51 [0.41; 0.62]	0.51 [0.42; 0.61]	0.51 [0.40; 0.63]	0.50 [0.41; 0.61]
0.25	0.67	5	0.68 [0.34; 1.32]	0.67 [0.38; 1.16]	0.66 [0.33; 1.31]	0.67 [0.39; 1.09]
0.25	0.67	10	0.68 [0.41; 1.07]	0.67 [0.46; 1.00]	0.68 [0.43; 1.10]	0.66 [0.46; 0.98]
0.25	0.67	20	0.68 [0.47; 0.96]	0.67 [0.51; 0.92]	0.67 [0.47; 0.96]	0.67 [0.51; 0.88]
0.25	0.67	50	0.67 [0.55; 0.84]	0.67 [0.55; 0.81]	0.67 [0.55; 0.84]	0.67 [0.55; 0.81]
0.25	1.00	5	1.00 [0.49; 1.93]	1.00 [0.58; 1.70]	0.97 [0.51; 1.89]	1.00 [0.62; 1.68]
0.25	1.00	10	1.02 [0.63; 1.65]	1.00 [0.68; 1.53]	1.00 [0.65; 1.58]	0.99 [0.70; 1.43]
0.25	1.00	20	1.03 [0.71; 1.43]	1.01 [0.77; 1.35]	0.99 [0.73; 1.35]	1.00 [0.78; 1.30]
0.25	1.00	50	0.99 [0.79; 1.24]	0.99 [0.82; 1.20]	1.00 [0.82; 1.24]	1.00 [0.83; 1.19]

Table A.1: Simulation I: results under null hypothesis ($n = 1000$); median, 5% and 95% quantiles of empirical distribution of estimated treatment effect based on Mantel-Haenszel method; both odds ratio ψ and risk ratio ϕ as measures of treatment effect.

τ^2	ψ / ϕ	k	$\widehat{\psi}_{IV}$ ($p_c = 0.1$) 50% [5%; 95%]	$\widehat{\psi}_{IV}$ ($p_c = 0.3$) 50% [5%; 95%]	$\widehat{\phi}_{IV}$ ($p_c = 0.1$) 50% [5%; 95%]	$\widehat{\phi}_{IV}$ ($p_c = 0.3$) 50% [5%; 95%]
0.00	0.25	5	0.30 [0.15; 0.55]	0.26 [0.17; 0.39]	0.30 [0.17; 0.54]	0.27 [0.17; 0.40]
0.00	0.25	10	0.30 [0.19; 0.45]	0.26 [0.20; 0.35]	0.30 [0.20; 0.46]	0.27 [0.21; 0.35]
0.00	0.25	20	0.30 [0.22; 0.40]	0.26 [0.22; 0.32]	0.31 [0.23; 0.42]	0.27 [0.23; 0.33]
0.00	0.25	50	0.30 [0.25; 0.36]	0.26 [0.23; 0.30]	0.31 [0.26; 0.37]	0.28 [0.24; 0.31]
0.00	0.50	5	0.53 [0.31; 0.88]	0.50 [0.36; 0.71]	0.54 [0.33; 0.88]	0.51 [0.38; 0.69]
0.00	0.50	10	0.53 [0.35; 0.75]	0.51 [0.40; 0.66]	0.54 [0.38; 0.76]	0.52 [0.43; 0.62]
0.00	0.50	20	0.53 [0.41; 0.68]	0.50 [0.43; 0.60]	0.54 [0.42; 0.68]	0.52 [0.45; 0.60]
0.00	0.50	50	0.53 [0.46; 0.62]	0.51 [0.45; 0.57]	0.54 [0.46; 0.62]	0.52 [0.48; 0.57]
0.00	0.67	5	0.68 [0.40; 1.13]	0.67 [0.47; 0.94]	0.69 [0.42; 1.10]	0.67 [0.52; 0.87]
0.00	0.67	10	0.68 [0.49; 0.98]	0.67 [0.53; 0.85]	0.69 [0.50; 0.95]	0.69 [0.57; 0.83]
0.00	0.67	20	0.70 [0.54; 0.88]	0.67 [0.57; 0.80]	0.70 [0.55; 0.87]	0.68 [0.60; 0.77]
0.00	0.67	50	0.69 [0.60; 0.81]	0.67 [0.61; 0.75]	0.70 [0.61; 0.80]	0.68 [0.63; 0.74]
0.00	1.00	5	1.00 [0.59; 1.59]	1.00 [0.72; 1.37]	1.00 [0.68; 1.49]	0.99 [0.79; 1.24]
0.00	1.00	10	1.01 [0.72; 1.39]	1.01 [0.80; 1.27]	1.00 [0.76; 1.32]	0.99 [0.86; 1.15]
0.00	1.00	20	1.00 [0.81; 1.27]	1.00 [0.85; 1.16]	1.00 [0.82; 1.21]	1.00 [0.90; 1.11]
0.00	1.00	50	1.00 [0.87; 1.14]	1.00 [0.91; 1.11]	1.00 [0.89; 1.13]	1.00 [0.94; 1.07]
0.05	0.25	5	0.30 [0.15; 0.56]	0.27 [0.16; 0.42]	0.30 [0.16; 0.57]	0.27 [0.17; 0.40]
0.05	0.25	10	0.30 [0.19; 0.49]	0.27 [0.19; 0.36]	0.31 [0.20; 0.46]	0.28 [0.20; 0.37]
0.05	0.25	20	0.30 [0.22; 0.41]	0.26 [0.21; 0.33]	0.31 [0.23; 0.41]	0.28 [0.23; 0.35]
0.05	0.25	50	0.30 [0.25; 0.36]	0.26 [0.23; 0.31]	0.31 [0.26; 0.38]	0.28 [0.24; 0.32]
0.05	0.50	5	0.53 [0.29; 0.93]	0.50 [0.32; 0.76]	0.53 [0.31; 0.90]	0.51 [0.37; 0.72]
0.05	0.50	10	0.53 [0.35; 0.78]	0.51 [0.38; 0.68]	0.55 [0.36; 0.79]	0.52 [0.40; 0.66]
0.05	0.50	20	0.54 [0.41; 0.72]	0.51 [0.41; 0.63]	0.54 [0.41; 0.70]	0.52 [0.44; 0.63]
0.05	0.50	50	0.54 [0.45; 0.64]	0.51 [0.45; 0.58]	0.55 [0.46; 0.64]	0.52 [0.47; 0.58]
0.05	0.67	5	0.70 [0.41; 1.21]	0.68 [0.45; 0.99]	0.69 [0.41; 1.09]	0.68 [0.49; 0.93]
0.05	0.67	10	0.69 [0.46; 1.00]	0.68 [0.51; 0.90]	0.70 [0.49; 0.98]	0.68 [0.54; 0.86]
0.05	0.67	20	0.68 [0.53; 0.90]	0.67 [0.56; 0.82]	0.70 [0.55; 0.89]	0.69 [0.58; 0.81]
0.05	0.67	50	0.69 [0.59; 0.82]	0.68 [0.60; 0.77]	0.70 [0.60; 0.81]	0.69 [0.62; 0.76]
0.05	1.00	5	1.00 [0.61; 1.65]	1.00 [0.68; 1.48]	1.01 [0.62; 1.56]	1.00 [0.74; 1.34]
0.05	1.00	10	1.01 [0.69; 1.44]	1.00 [0.75; 1.29]	1.00 [0.74; 1.36]	1.00 [0.82; 1.22]
0.05	1.00	20	1.00 [0.77; 1.28]	1.01 [0.83; 1.20]	1.00 [0.80; 1.24]	0.99 [0.86; 1.15]
0.05	1.00	50	1.00 [0.86; 1.17]	1.00 [0.89; 1.13]	1.00 [0.86; 1.16]	1.00 [0.91; 1.10]
0.25	0.25	5	0.30 [0.14; 0.62]	0.27 [0.15; 0.49]	0.32 [0.16; 0.62]	0.28 [0.16; 0.49]
0.25	0.25	10	0.31 [0.19; 0.53]	0.27 [0.18; 0.42]	0.32 [0.19; 0.51]	0.29 [0.19; 0.43]
0.25	0.25	20	0.31 [0.22; 0.44]	0.27 [0.20; 0.37]	0.32 [0.22; 0.46]	0.29 [0.21; 0.39]
0.25	0.25	50	0.31 [0.25; 0.39]	0.27 [0.22; 0.33]	0.32 [0.25; 0.40]	0.29 [0.24; 0.36]
0.25	0.50	5	0.55 [0.25; 1.06]	0.50 [0.30; 0.92]	0.54 [0.28; 1.02]	0.54 [0.31; 0.88]
0.25	0.50	10	0.55 [0.34; 0.88]	0.51 [0.34; 0.77]	0.54 [0.35; 0.85]	0.54 [0.37; 0.77]
0.25	0.50	20	0.55 [0.38; 0.77]	0.53 [0.38; 0.72]	0.55 [0.41; 0.77]	0.54 [0.40; 0.70]
0.25	0.50	50	0.54 [0.44; 0.67]	0.52 [0.43; 0.63]	0.56 [0.45; 0.68]	0.54 [0.45; 0.65]
0.25	0.67	5	0.70 [0.37; 1.34]	0.68 [0.39; 1.17]	0.69 [0.37; 1.29]	0.70 [0.43; 1.09]
0.25	0.67	10	0.70 [0.43; 1.09]	0.68 [0.47; 1.01]	0.70 [0.47; 1.10]	0.69 [0.49; 0.99]
0.25	0.67	20	0.70 [0.49; 0.98]	0.68 [0.52; 0.93]	0.71 [0.51; 0.99]	0.70 [0.54; 0.91]
0.25	0.67	50	0.70 [0.58; 0.86]	0.68 [0.56; 0.82]	0.71 [0.58; 0.87]	0.70 [0.59; 0.83]
0.25	1.00	5	1.01 [0.51; 1.87]	1.00 [0.59; 1.69]	0.98 [0.55; 1.80]	1.00 [0.63; 1.62]
0.25	1.00	10	1.02 [0.64; 1.61]	1.00 [0.68; 1.52]	1.00 [0.67; 1.54]	0.99 [0.71; 1.40]
0.25	1.00	20	1.02 [0.72; 1.42]	1.02 [0.77; 1.35]	1.00 [0.74; 1.33]	1.01 [0.80; 1.28]
0.25	1.00	50	1.00 [0.81; 1.23]	1.00 [0.82; 1.21]	1.00 [0.83; 1.24]	1.00 [0.85; 1.17]

Table A.2: Simulation I: results under null hypothesis ($n = 1000$); median, 5% and 95% quantiles of empirical distribution of estimated treatment effect based on inverse variance method using "add selective" approach; both odds ratio ψ and risk ratio ϕ as measures of treatment effect.

τ^2	ψ / ϕ	k	$\widehat{\psi}_{IV}$ ($p_c = 0.1$) 50% [5%; 95%]	$\widehat{\psi}_{IV}$ ($p_c = 0.3$) 50% [5%; 95%]	$\widehat{\phi}_{IV}$ ($p_c = 0.1$) 50% [5%; 95%]	$\widehat{\phi}_{IV}$ ($p_c = 0.3$) 50% [5%; 95%]
0.00	0.25	5	0.33 [0.18; 0.59]	0.27 [0.19; 0.41]	0.35 [0.20; 0.58]	0.29 [0.20; 0.43]
0.00	0.25	10	0.34 [0.22; 0.49]	0.28 [0.21; 0.36]	0.34 [0.23; 0.51]	0.29 [0.23; 0.38]
0.00	0.25	20	0.34 [0.25; 0.44]	0.27 [0.23; 0.34]	0.35 [0.26; 0.46]	0.29 [0.25; 0.35]
0.00	0.25	50	0.33 [0.28; 0.39]	0.28 [0.25; 0.31]	0.35 [0.30; 0.41]	0.29 [0.26; 0.33]
0.00	0.50	5	0.56 [0.35; 0.89]	0.51 [0.37; 0.72]	0.57 [0.37; 0.90]	0.52 [0.40; 0.70]
0.00	0.50	10	0.56 [0.39; 0.77]	0.52 [0.41; 0.67]	0.58 [0.42; 0.78]	0.53 [0.44; 0.64]
0.00	0.50	20	0.56 [0.44; 0.70]	0.51 [0.44; 0.61]	0.57 [0.45; 0.71]	0.53 [0.47; 0.61]
0.00	0.50	50	0.56 [0.49; 0.65]	0.52 [0.46; 0.57]	0.57 [0.50; 0.65]	0.53 [0.49; 0.58]
0.00	0.67	5	0.71 [0.44; 1.11]	0.68 [0.49; 0.94]	0.72 [0.46; 1.09]	0.68 [0.54; 0.88]
0.00	0.67	10	0.70 [0.51; 0.98]	0.68 [0.54; 0.86]	0.71 [0.54; 0.95]	0.69 [0.58; 0.84]
0.00	0.67	20	0.71 [0.56; 0.89]	0.68 [0.58; 0.81]	0.72 [0.58; 0.88]	0.69 [0.61; 0.78]
0.00	0.67	50	0.71 [0.62; 0.83]	0.68 [0.62; 0.75]	0.72 [0.63; 0.82]	0.69 [0.64; 0.75]
0.00	1.00	5	1.00 [0.62; 1.53]	1.00 [0.73; 1.36]	1.00 [0.70; 1.44]	0.99 [0.80; 1.23]
0.00	1.00	10	1.01 [0.74; 1.36]	1.01 [0.81; 1.26]	1.00 [0.78; 1.29]	0.99 [0.86; 1.14]
0.00	1.00	20	1.00 [0.82; 1.25]	1.00 [0.85; 1.16]	1.00 [0.83; 1.20]	1.00 [0.90; 1.11]
0.00	1.00	50	1.00 [0.88; 1.14]	1.00 [0.91; 1.10]	1.00 [0.90; 1.12]	1.00 [0.94; 1.07]
0.05	0.25	5	0.33 [0.18; 0.60]	0.28 [0.18; 0.43]	0.34 [0.18; 0.61]	0.30 [0.19; 0.43]
0.05	0.25	10	0.34 [0.23; 0.53]	0.28 [0.20; 0.38]	0.35 [0.23; 0.51]	0.30 [0.22; 0.39]
0.05	0.25	20	0.33 [0.25; 0.44]	0.28 [0.22; 0.34]	0.35 [0.26; 0.45]	0.30 [0.25; 0.37]
0.05	0.25	50	0.34 [0.28; 0.40]	0.28 [0.24; 0.32]	0.35 [0.29; 0.42]	0.30 [0.26; 0.34]
0.05	0.50	5	0.56 [0.33; 0.93]	0.51 [0.34; 0.76]	0.56 [0.36; 0.91]	0.53 [0.38; 0.73]
0.05	0.50	10	0.55 [0.38; 0.79]	0.52 [0.39; 0.69]	0.58 [0.40; 0.82]	0.53 [0.42; 0.67]
0.05	0.50	20	0.56 [0.43; 0.74]	0.52 [0.43; 0.64]	0.57 [0.45; 0.72]	0.54 [0.45; 0.64]
0.05	0.50	50	0.56 [0.48; 0.66]	0.52 [0.46; 0.59]	0.58 [0.49; 0.67]	0.54 [0.48; 0.59]
0.05	0.67	5	0.72 [0.44; 1.18]	0.68 [0.46; 0.99]	0.71 [0.45; 1.08]	0.69 [0.50; 0.93]
0.05	0.67	10	0.71 [0.49; 1.00]	0.69 [0.52; 0.90]	0.72 [0.52; 0.98]	0.69 [0.55; 0.86]
0.05	0.67	20	0.70 [0.55; 0.91]	0.68 [0.56; 0.83]	0.72 [0.58; 0.90]	0.69 [0.59; 0.82]
0.05	0.67	50	0.71 [0.62; 0.83]	0.68 [0.60; 0.77]	0.72 [0.63; 0.83]	0.70 [0.63; 0.77]
0.05	1.00	5	1.00 [0.63; 1.58]	1.00 [0.69; 1.47]	1.00 [0.65; 1.51]	1.00 [0.75; 1.33]
0.05	1.00	10	1.01 [0.71; 1.40]	1.00 [0.76; 1.28]	1.00 [0.76; 1.32]	1.00 [0.83; 1.21]
0.05	1.00	20	0.99 [0.79; 1.26]	1.01 [0.83; 1.19]	1.00 [0.82; 1.22]	0.99 [0.86; 1.15]
0.05	1.00	50	1.00 [0.87; 1.16]	1.00 [0.89; 1.13]	1.00 [0.87; 1.14]	1.00 [0.91; 1.10]
0.25	0.25	5	0.34 [0.17; 0.65]	0.29 [0.16; 0.50]	0.36 [0.18; 0.66]	0.30 [0.18; 0.50]
0.25	0.25	10	0.35 [0.22; 0.56]	0.29 [0.19; 0.43]	0.36 [0.23; 0.55]	0.30 [0.21; 0.44]
0.25	0.25	20	0.35 [0.25; 0.47]	0.28 [0.21; 0.38]	0.36 [0.26; 0.50]	0.31 [0.23; 0.40]
0.25	0.25	50	0.34 [0.28; 0.42]	0.29 [0.24; 0.35]	0.36 [0.29; 0.44]	0.31 [0.26; 0.37]
0.25	0.50	5	0.58 [0.29; 1.06]	0.52 [0.31; 0.93]	0.57 [0.31; 1.02]	0.55 [0.33; 0.88]
0.25	0.50	10	0.57 [0.37; 0.88]	0.52 [0.36; 0.78]	0.57 [0.38; 0.86]	0.55 [0.38; 0.77]
0.25	0.50	20	0.57 [0.41; 0.78]	0.54 [0.39; 0.73]	0.58 [0.44; 0.78]	0.55 [0.42; 0.70]
0.25	0.50	50	0.57 [0.47; 0.68]	0.53 [0.44; 0.63]	0.58 [0.48; 0.70]	0.55 [0.47; 0.66]
0.25	0.67	5	0.72 [0.39; 1.30]	0.69 [0.40; 1.16]	0.72 [0.41; 1.25]	0.71 [0.44; 1.09]
0.25	0.67	10	0.72 [0.47; 1.07]	0.68 [0.48; 1.00]	0.72 [0.50; 1.09]	0.70 [0.50; 0.99]
0.25	0.67	20	0.72 [0.52; 0.98]	0.69 [0.53; 0.93]	0.73 [0.54; 0.99]	0.70 [0.55; 0.90]
0.25	0.67	50	0.72 [0.60; 0.87]	0.69 [0.57; 0.82]	0.72 [0.60; 0.88]	0.70 [0.60; 0.83]
0.25	1.00	5	1.01 [0.54; 1.78]	1.00 [0.60; 1.66]	0.98 [0.57; 1.73]	1.00 [0.64; 1.60]
0.25	1.00	10	1.02 [0.66; 1.56]	1.00 [0.69; 1.50]	1.00 [0.68; 1.48]	0.99 [0.72; 1.39]
0.25	1.00	20	1.02 [0.74; 1.38]	1.01 [0.77; 1.34]	1.00 [0.76; 1.29]	1.00 [0.80; 1.27]
0.25	1.00	50	1.00 [0.82; 1.21]	1.00 [0.83; 1.20]	1.00 [0.84; 1.22]	1.00 [0.85; 1.17]

Table A.3: Simulation I: results under null hypothesis ($n = 1000$); median, 5% and 95% quantiles of empirical distribution of estimated treatment effect based on inverse variance method using "add all" approach; both odds ratio ψ and risk ratio ϕ as measures of treatment effect.

τ^2	ψ	k	Linear regression test "add selective"	Linear regression test "add all"	Rank correlation test "add selective"	Rank correlation test "add all"	New test
0.00	0.25	5	13.7%	14.8%	7.8%	6.7%	7.7%
0.00	0.25	10	18.2%	19.5%	7.1%	7.1%	9.2%
0.00	0.25	20	**20.7%**	**22.3%**	6.9%	8.2%	7.7%
0.00	0.25	50	**32.3%**	**33.7%**	13.0%	14.4%	7.4%
0.00	0.50	5	13.3%	13.8%	7.3%	7.9%	6.3%
0.00	0.50	10	15.7%	16.5%	12.3%	12.0%	10.5%
0.00	0.50	20	14.1%	14.6%	11.2%	11.3%	7.6%
0.00	0.50	50	**20.5%**	17.5%	**23.4%**	19.7%	7.8%
0.00	0.67	5	12.6%	12.7%	8.5%	8.9%	8.7%
0.00	0.67	10	13.7%	14.2%	11.5%	11.9%	10.0%
0.00	0.67	20	12.6%	12.3%	11.4%	10.9%	8.6%
0.00	0.67	50	15.4%	14.4%	19.4%	15.3%	9.4%
0.00	1.00	5	10.7%	11.2%	8.9%	9.2%	8.6%
0.00	1.00	10	13.2%	13.6%	14.8%	15.1%	12.7%
0.00	1.00	20	11.1%	12.3%	9.4%	9.2%	8.5%
0.00	1.00	50	10.8%	10.1%	11.1%	10.7%	9.2%
0.05	0.25	5	15.2%	17.7%	5.8%	6.0%	6.7%
0.05	0.25	10	17.0%	19.2%	9.4%	9.3%	7.3%
0.05	0.25	20	**21.7%**	**22.7%**	7.4%	7.6%	9.0%
0.05	0.25	50	**33.1%**	**34.8%**	13.2%	14.9%	10.3%
0.05	0.50	5	11.6%	11.6%	7.4%	7.4%	8.1%
0.05	0.50	10	13.9%	14.8%	11.2%	10.2%	9.0%
0.05	0.50	20	14.5%	14.4%	11.1%	10.9%	7.5%
0.05	0.50	50	**21.7%**	19.3%	**21.2%**	17.6%	6.8%
0.05	0.67	5	12.3%	14.0%	7.6%	7.6%	8.0%
0.05	0.67	10	13.1%	13.5%	12.0%	12.2%	10.6%
0.05	0.67	20	14.0%	14.7%	11.7%	11.2%	9.6%
0.05	0.67	50	15.0%	14.7%	16.1%	13.5%	8.5%
0.05	1.00	5	13.3%	13.2%	8.1%	8.4%	8.0%
0.05	1.00	10	12.6%	12.6%	11.0%	11.4%	9.5%
0.05	1.00	20	12.6%	13.9%	10.6%	10.9%	11.2%
0.05	1.00	50	13.0%	14.5%	9.6%	10.2%	8.9%
0.25	0.25	5	13.2%	15.2%	6.3%	6.2%	5.8%
0.25	0.25	10	18.0%	18.9%	8.1%	7.6%	7.9%
0.25	0.25	20	**25.6%**	**26.3%**	8.0%	7.8%	10.9%
0.25	0.25	50	**36.4%**	**37.3%**	13.4%	15.3%	16.7%
0.25	0.50	5	13.5%	14.0%	8.5%	8.8%	8.9%
0.25	0.50	10	17.1%	16.4%	12.1%	11.9%	7.4%
0.25	0.50	20	**20.3%**	**20.9%**	11.9%	10.9%	7.2%
0.25	0.50	50	**25.0%**	**23.4%**	18.2%	15.6%	7.2%
0.25	0.67	5	12.0%	12.3%	8.1%	8.0%	7.7%
0.25	0.67	10	12.0%	13.3%	11.1%	11.1%	8.9%
0.25	0.67	20	13.4%	15.2%	10.6%	10.3%	8.2%
0.25	0.67	50	**20.8%**	**20.5%**	16.7%	15.6%	7.3%
0.25	1.00	5	11.8%	11.9%	7.5%	7.8%	6.4%
0.25	1.00	10	11.9%	12.5%	10.0%	9.6%	8.8%
0.25	1.00	20	15.6%	16.4%	10.0%	10.2%	8.2%
0.25	1.00	50	17.1%	19.0%	9.4%	9.8%	8.5%

Table A.4: Simulation I: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio ψ as measure of treatment effect; control event rate equals 0.1; nominal significance level is 10%; error rates above 20% in bold font.

τ^2	ψ	k	Linear regression test		Rank correlation test		New
			"add selective"	"add all"	"add selective"	"add all"	test
0.00	0.25	5	11.3% / 2.4%	12.0% / 2.8%	4.1% / 3.7%	3.8% / 2.9%	2.5% / 5.2%
0.00	0.25	10	16.9% / 1.3%	18.2% / 1.3%	5.9% / 1.2%	5.8% / 1.3%	3.1% / 6.1%
0.00	0.25	20	**20.2%** / 0.5%	**21.5%** / 0.8%	6.4% / 0.5%	7.4% / 0.8%	1.9% / 5.8%
0.00	0.25	50	**32.1%** / 0.2%	**33.2%** / 0.5%	12.7% / 0.3%	14.1% / 0.3%	0.5% / 6.9%
0.00	0.50	5	10.4% / 2.9%	9.7% / 4.1%	5.3% / 2.0%	5.2% / 2.7%	3.5% / 2.8%
0.00	0.50	10	13.3% / 2.4%	12.9% / 3.6%	9.9% / 2.4%	9.1% / 2.9%	5.5% / 5.0%
0.00	0.50	20	12.3% / 1.8%	11.1% / 3.5%	9.9% / 1.3%	9.1% / 2.2%	3.1% / 4.5%
0.00	0.50	50	19.7% / 0.8%	16.1% / 1.4%	**23.1%** / 0.3%	19.2% / 0.5%	4.8% / 3.0%
0.00	0.67	5	8.9% / 3.7%	8.4% / 4.3%	5.0% / 3.5%	4.8% / 4.1%	4.3% / 4.4%
0.00	0.67	10	9.9% / 3.8%	9.6% / 4.6%	8.0% / 3.5%	7.9% / 4.0%	5.4% / 4.6%
0.00	0.67	20	10.2% / 2.4%	8.9% / 3.4%	10.3% / 1.1%	9.7% / 1.2%	5.8% / 2.8%
0.00	0.67	50	14.0% / 1.4%	11.7% / 2.7%	18.6% / 0.8%	14.3% / 1.0%	6.8% / 2.6%
0.00	1.00	5	5.3% / 5.4%	5.6% / 5.6%	4.5% / 4.4%	4.8% / 4.4%	4.8% / 3.8%
0.00	1.00	10	6.7% / 6.5%	7.0% / 6.6%	7.1% / 7.7%	7.6% / 7.5%	6.5% / 6.2%
0.00	1.00	20	5.8% / 5.3%	6.3% / 6.0%	4.5% / 4.9%	4.4% / 4.8%	4.0% / 4.5%
0.00	1.00	50	3.9% / 6.9%	3.6% / 6.5%	4.3% / 6.8%	4.0% / 6.7%	3.1% / 6.1%
0.05	0.25	5	13.8% / 1.4%	16.1% / 1.6%	4.0% / 1.8%	4.3% / 1.7%	2.3% / 4.4%
0.05	0.25	10	15.9% / 1.1%	18.4% / 0.8%	7.2% / 2.2%	7.5% / 1.8%	2.2% / 5.1%
0.05	0.25	20	**20.5%** / 1.2%	**21.3%** / 1.4%	6.5% / 0.9%	6.6% / 1.0%	1.5% / 7.5%
0.05	0.25	50	**32.5%** / 0.6%	**34.3%** / 0.5%	12.7% / 0.5%	14.7% / 0.2%	0.9% / 9.4%
0.05	0.50	5	9.5% / 2.1%	8.5% / 3.1%	5.4% / 2.0%	4.9% / 2.5%	4.2% / 3.9%
0.05	0.50	10	12.3% / 1.6%	12.0% / 2.8%	9.7% / 1.5%	8.8% / 1.4%	5.2% / 3.8%
0.05	0.50	20	12.4% / 2.1%	11.1% / 3.3%	10.1% / 1.0%	9.1% / 1.8%	3.9% / 3.6%
0.05	0.50	50	**20.5%** / 1.2%	17.1% / 2.2%	**20.8%** / 0.4%	16.9% / 0.7%	3.6% / 3.2%
0.05	0.67	5	8.8% / 3.5%	9.1% / 4.9%	5.2% / 2.4%	4.9% / 2.7%	3.4% / 4.6%
0.05	0.67	10	8.8% / 4.3%	8.2% / 5.3%	7.9% / 4.1%	7.2% / 5.0%	4.8% / 5.8%
0.05	0.67	20	11.5% / 2.5%	10.8% / 3.9%	10.1% / 1.6%	8.7% / 2.5%	5.5% / 4.1%
0.05	0.67	50	13.4% / 1.6%	12.2% / 2.5%	15.5% / 0.6%	12.3% / 1.2%	6.1% / 2.4%
0.05	1.00	5	6.6% / 6.7%	6.8% / 6.4%	3.7% / 4.4%	3.8% / 4.6%	4.0% / 4.0%
0.05	1.00	10	6.7% / 5.9%	6.9% / 5.7%	5.6% / 5.4%	5.6% / 5.8%	4.8% / 4.7%
0.05	1.00	20	6.1% / 6.5%	7.2% / 6.7%	4.7% / 5.9%	4.7% / 6.2%	4.4% / 6.8%
0.05	1.00	50	6.5% / 6.5%	7.1% / 7.4%	4.3% / 5.3%	4.5% / 5.7%	4.0% / 4.9%
0.25	0.25	5	11.2% / 2.0%	12.5% / 2.7%	4.4% / 1.9%	4.0% / 2.2%	1.4% / 4.4%
0.25	0.25	10	16.5% / 1.5%	17.4% / 1.5%	6.5% / 1.6%	6.6% / 1.0%	2.1% / 5.8%
0.25	0.25	20	**23.7%** / 1.9%	**24.5%** / 1.8%	7.1% / 0.9%	7.4% / 0.4%	1.1% / 9.8%
0.25	0.25	50	**36.2%** / 0.2%	**36.8%** / 0.5%	13.2% / 0.2%	15.3% / 0.0%	0.9% / 15.8%
0.25	0.50	5	9.6% / 3.9%	8.9% / 5.1%	5.7% / 2.8%	5.6% / 3.2%	3.7% / 5.2%
0.25	0.50	10	15.2% / 1.9%	13.8% / 2.6%	10.1% / 2.0%	9.5% / 2.4%	3.4% / 4.0%
0.25	0.50	20	17.1% / 3.2%	16.4% / 4.5%	11.1% / 0.8%	9.9% / 1.0%	3.0% / 4.2%
0.25	0.50	50	**23.3%** / 1.7%	**20.4%** / 3.0%	17.9% / 0.3%	14.7% / 0.9%	2.0% / 5.2%
0.25	0.67	5	8.1% / 3.9%	7.7% / 4.6%	5.2% / 2.9%	5.0% / 3.0%	3.5% / 4.2%
0.25	0.67	10	8.7% / 3.3%	8.7% / 4.6%	7.3% / 3.8%	6.8% / 4.3%	3.6% / 5.3%
0.25	0.67	20	9.2% / 4.2%	8.9% / 6.3%	8.4% / 2.2%	7.7% / 2.6%	3.3% / 4.9%
0.25	0.67	50	17.8% / 3.0%	16.3% / 4.2%	15.8% / 0.9%	13.9% / 1.7%	3.8% / 3.5%
0.25	1.00	5	7.6% / 4.2%	7.7% / 4.2%	4.0% / 3.5%	4.2% / 3.6%	3.4% / 3.0%
0.25	1.00	10	6.3% / 5.6%	6.4% / 6.1%	4.4% / 5.6%	4.2% / 5.4%	4.3% / 4.5%
0.25	1.00	20	8.7% / 6.9%	9.1% / 7.3%	5.3% / 4.7%	5.3% / 4.9%	5.0% / 3.2%
0.25	1.00	50	8.6% / 8.5%	9.5% / 9.5%	4.8% / 4.6%	5.1% / 4.7%	4.2% / 4.3%

Table A.5: Simulation I: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of one-sided test for bias (lower / upper tail); odds ratio ψ as measure of treatment effect; control event rate equals 0.1; nominal significance level is 5% for each tail; error rates above 20% in bold font.

τ^2	ψ	k	Linear regression test "add selective"	Linear regression test "add all"	Rank correlation test "add selective"	Rank correlation test "add all"	New test
0.00	0.25	5	15.9%	13.3%	8.6%	7.2%	8.0%
0.00	0.25	10	**20.0%**	15.6%	18.4%	13.3%	8.0%
0.00	0.25	20	**27.4%**	19.4%	**27.6%**	17.6%	8.7%
0.00	0.25	50	**44.9%**	**24.7%**	**55.5%**	**34.6%**	9.1%
0.00	0.50	5	15.3%	14.8%	10.5%	10.7%	8.5%
0.00	0.50	10	12.8%	11.8%	14.2%	12.3%	11.2%
0.00	0.50	20	13.1%	10.8%	13.5%	10.4%	7.5%
0.00	0.50	50	19.2%	12.8%	**26.0%**	17.6%	9.1%
0.00	0.67	5	11.8%	11.9%	8.2%	7.9%	8.1%
0.00	0.67	10	13.0%	11.9%	13.1%	12.2%	11.4%
0.00	0.67	20	11.7%	11.1%	10.3%	9.7%	7.3%
0.00	0.67	50	15.7%	13.5%	15.3%	13.4%	8.9%
0.00	1.00	5	11.2%	10.9%	9.7%	9.3%	7.9%
0.00	1.00	10	11.3%	11.6%	12.2%	12.0%	10.6%
0.00	1.00	20	9.7%	9.6%	9.2%	9.6%	8.0%
0.00	1.00	50	12.2%	12.1%	11.7%	11.5%	9.9%
0.05	0.25	5	14.5%	13.1%	9.9%	8.9%	8.1%
0.05	0.25	10	**20.8%**	15.8%	17.3%	14.0%	9.3%
0.05	0.25	20	**28.3%**	18.8%	**25.3%**	18.1%	7.7%
0.05	0.25	50	**48.6%**	**27.9%**	**58.8%**	**37.8%**	9.2%
0.05	0.50	5	14.1%	13.3%	10.3%	10.3%	7.3%
0.05	0.50	10	14.3%	12.7%	11.6%	10.5%	8.6%
0.05	0.50	20	16.9%	15.1%	13.6%	11.6%	7.3%
0.05	0.50	50	**22.6%**	18.4%	**24.0%**	17.9%	8.3%
0.05	0.67	5	13.2%	12.4%	9.3%	8.9%	7.9%
0.05	0.67	10	13.0%	13.0%	11.2%	10.5%	8.3%
0.05	0.67	20	13.5%	13.5%	9.6%	9.5%	7.0%
0.05	0.67	50	16.1%	14.6%	15.1%	12.6%	7.4%
0.05	1.00	5	12.8%	12.7%	9.8%	9.9%	8.5%
0.05	1.00	10	12.9%	13.1%	11.0%	10.7%	9.4%
0.05	1.00	20	13.3%	13.8%	10.7%	10.4%	8.7%
0.05	1.00	50	15.4%	15.2%	9.8%	9.1%	8.4%
0.25	0.25	5	18.5%	16.3%	10.8%	10.4%	7.5%
0.25	0.25	10	**23.6%**	19.5%	17.2%	14.1%	10.0%
0.25	0.25	20	**29.5%**	**22.3%**	**26.3%**	19.1%	8.7%
0.25	0.25	50	**47.9%**	**34.4%**	**54.1%**	**37.4%**	9.4%
0.25	0.50	5	11.4%	11.4%	8.0%	7.9%	6.1%
0.25	0.50	10	14.9%	14.7%	12.1%	11.4%	9.3%
0.25	0.50	20	19.1%	18.0%	12.6%	10.4%	6.9%
0.25	0.50	50	**25.5%**	**23.1%**	**22.0%**	16.2%	6.6%
0.25	0.67	5	11.8%	11.8%	7.0%	7.0%	6.4%
0.25	0.67	10	14.3%	14.6%	10.2%	9.6%	7.4%
0.25	0.67	20	17.1%	17.5%	9.4%	9.1%	6.8%
0.25	0.67	50	**21.9%**	**21.5%**	13.7%	11.5%	7.3%
0.25	1.00	5	11.8%	11.6%	8.6%	8.4%	7.7%
0.25	1.00	10	16.4%	16.6%	10.5%	10.0%	9.5%
0.25	1.00	20	14.6%	14.8%	8.5%	8.3%	5.9%
0.25	1.00	50	**20.4%**	**20.8%**	9.6%	9.4%	7.6%

Table A.6: Simulation I: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio ψ as measure of treatment effect; control event rate equals 0.3; nominal significance level is 10%; error rates above 20% in bold font.

τ^2	ψ	k	Linear regression test "add selective"	Linear regression test "add all"	Rank correlation test "add selective"	Rank correlation test "add all"	New test
0.00	0.25	5	13.7% / 2.2%	10.2% / 3.1%	6.9% / 1.7%	4.9% / 2.3%	4.7% / 3.3%
0.00	0.25	10	19.0% / 1.0%	13.6% / 2.0%	17.7% / 0.7%	12.2% / 1.1%	5.0% / 3.0%
0.00	0.25	20	**26.6%** / 0.8%	17.7% / 1.7%	**27.3%** / 0.3%	17.0% / 0.6%	4.8% / 3.9%
0.00	0.25	50	**44.7%** / 0.2%	**24.1%** / 0.6%	**55.5%** / 0.0%	**34.5%** / 0.1%	5.5% / 3.6%
0.00	0.50	5	9.2% / 6.1%	8.1% / 6.7%	6.7% / 3.8%	6.0% / 4.7%	4.1% / 4.4%
0.00	0.50	10	9.7% / 3.1%	8.0% / 3.8%	11.0% / 3.2%	8.8% / 3.5%	6.3% / 4.9%
0.00	0.50	20	11.7% / 1.4%	8.1% / 2.7%	12.6% / 0.9%	8.9% / 1.5%	3.5% / 4.0%
0.00	0.50	50	18.3% / 0.9%	10.4% / 2.4%	**25.4%** / 0.6%	16.7% / 0.9%	5.3% / 3.8%
0.00	0.67	5	6.8% / 5.0%	6.4% / 5.5%	5.1% / 3.1%	4.7% / 3.2%	4.0% / 4.1%
0.00	0.67	10	8.8% / 4.2%	6.9% / 5.0%	8.6% / 4.5%	7.2% / 5.0%	5.7% / 5.7%
0.00	0.67	20	8.1% / 3.6%	6.6% / 4.5%	8.0% / 2.3%	7.1% / 2.6%	4.1% / 3.2%
0.00	0.67	50	12.7% / 3.0%	9.9% / 3.6%	13.6% / 1.7%	10.8% / 2.6%	5.2% / 3.7%
0.00	1.00	5	4.9% / 6.3%	4.8% / 6.1%	4.4% / 5.3%	4.0% / 5.3%	3.4% / 4.5%
0.00	1.00	10	5.2% / 6.1%	5.2% / 6.4%	6.1% / 6.1%	5.8% / 6.2%	5.2% / 5.4%
0.00	1.00	20	4.5% / 5.2%	4.4% / 5.2%	4.0% / 5.2%	4.0% / 5.6%	3.3% / 4.7%
0.00	1.00	50	6.0% / 6.2%	5.9% / 6.2%	5.7% / 6.0%	5.8% / 5.7%	5.0% / 4.9%
0.05	0.25	5	12.6% / 1.9%	10.4% / 2.7%	8.6% / 1.3%	7.0% / 1.9%	4.8% / 3.3%
0.05	0.25	10	**20.1%** / 0.7%	13.4% / 2.4%	16.7% / 0.6%	12.0% / 2.0%	4.6% / 4.7%
0.05	0.25	20	**27.9%** / 0.4%	16.8% / 2.0%	**25.2%** / 0.1%	17.8% / 0.3%	3.5% / 4.2%
0.05	0.25	50	**48.3%** / 0.3%	**27.3%** / 0.6%	**58.8%** / 0.0%	**37.5%** / 0.3%	4.1% / 5.1%
0.05	0.50	5	10.1% / 4.0%	8.9% / 4.4%	7.2% / 3.1%	6.9% / 3.4%	3.6% / 3.7%
0.05	0.50	10	11.5% / 2.8%	8.9% / 3.8%	9.2% / 2.4%	7.5% / 3.0%	4.6% / 4.0%
0.05	0.50	20	14.2% / 2.7%	11.5% / 3.6%	12.1% / 1.5%	9.3% / 2.3%	3.6% / 3.7%
0.05	0.50	50	**20.4%** / 2.2%	14.1% / 4.3%	**23.2%** / 0.8%	16.3% / 1.6%	3.7% / 4.6%
0.05	0.67	5	8.5% / 4.7%	7.3% / 5.1%	6.1% / 3.2%	5.8% / 3.1%	4.5% / 3.4%
0.05	0.67	10	8.2% / 4.8%	7.5% / 5.5%	7.9% / 3.3%	6.5% / 4.0%	4.3% / 4.0%
0.05	0.67	20	9.0% / 4.5%	8.2% / 5.3%	7.1% / 2.5%	6.3% / 3.2%	3.8% / 3.2%
0.05	0.67	50	13.6% / 2.5%	11.4% / 3.2%	14.0% / 1.1%	11.2% / 1.4%	4.5% / 2.9%
0.05	1.00	5	6.5% / 6.3%	6.4% / 6.3%	4.3% / 5.5%	4.2% / 5.7%	3.6% / 4.9%
0.05	1.00	10	6.1% / 6.8%	6.3% / 6.8%	4.6% / 6.4%	4.5% / 6.2%	3.9% / 5.5%
0.05	1.00	20	6.6% / 6.7%	6.8% / 7.0%	5.4% / 5.3%	5.2% / 5.2%	4.5% / 4.2%
0.05	1.00	50	7.5% / 7.9%	7.7% / 7.5%	5.1% / 4.7%	4.9% / 4.2%	4.5% / 3.9%
0.25	0.25	5	16.3% / 2.2%	13.1% / 3.2%	9.0% / 1.8%	8.1% / 2.3%	3.3% / 4.2%
0.25	0.25	10	**22.0%** / 1.6%	17.5% / 2.0%	16.5% / 0.7%	13.0% / 1.1%	3.9% / 6.1%
0.25	0.25	20	**28.8%** / 0.7%	**20.7%** / 1.6%	**26.3%** / 0.0%	18.8% / 0.3%	2.2% / 6.5%
0.25	0.25	50	**47.6%** / 0.3%	**33.2%** / 1.2%	**54.1%** / 0.0%	**37.4%** / 0.0%	1.3% / 8.1%
0.25	0.50	5	7.7% / 3.7%	7.2% / 4.2%	5.7% / 2.3%	5.5% / 2.4%	3.0% / 3.1%
0.25	0.50	10	10.8% / 4.1%	9.9% / 4.8%	9.7% / 2.4%	8.8% / 2.6%	4.5% / 4.8%
0.25	0.50	20	15.6% / 3.5%	13.3% / 4.7%	11.5% / 1.1%	9.0% / 1.4%	2.9% / 4.0%
0.25	0.50	50	**23.0%** / 2.5%	19.5% / 3.6%	**21.8%** / 0.2%	15.7% / 0.5%	2.7% / 3.9%
0.25	0.67	5	7.4% / 4.4%	7.1% / 4.7%	4.8% / 2.2%	4.8% / 2.2%	4.0% / 2.4%
0.25	0.67	10	9.7% / 4.6%	9.1% / 5.5%	7.2% / 3.0%	6.3% / 3.3%	2.9% / 4.5%
0.25	0.67	20	12.8% / 4.3%	12.0% / 5.5%	7.7% / 1.7%	7.2% / 1.9%	3.1% / 3.7%
0.25	0.67	50	16.5% / 5.4%	14.8% / 6.7%	12.3% / 1.4%	9.5% / 2.0%	2.9% / 4.4%
0.25	1.00	5	5.1% / 6.7%	5.0% / 6.6%	4.4% / 4.2%	4.4% / 4.0%	3.8% / 3.9%
0.25	1.00	10	9.1% / 7.3%	9.3% / 7.3%	4.9% / 5.6%	4.7% / 5.3%	4.8% / 4.7%
0.25	1.00	20	7.9% / 6.7%	8.0% / 6.8%	3.7% / 4.8%	3.6% / 4.7%	2.8% / 3.1%
0.25	1.00	50	10.6% / 9.8%	10.8% / 10.0%	4.8% / 4.8%	4.6% / 4.8%	3.7% / 3.9%

Table A.7: Simulation I: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of one-sided test for bias (lower / upper tail); odds ratio ψ as measure of treatment effect; control event rate equals 0.3; nominal significance level is 5% for each tail; error rates above 20% in bold font.

τ^2	ϕ	k	Linear regression test "add selective"	Linear regression test "add all"	Rank correlation test "add selective"	Rank correlation test "add all"	New test
0.00	0.25	5	13.5% / 1.6%	15.4% / 1.9%	2.8% / 2.6%	2.4% / 2.2%	1.1% / 5.3%
0.00	0.25	10	**20.5%** / 1.2%	**21.5%** / 1.2%	2.7% / 2.8%	3.2% / 3.0%	3.0% / 6.2%
0.00	0.25	20	**27.1%** / 0.9%	**30.1%** / 0.8%	3.5% / 1.8%	3.4% / 1.0%	1.8% / 6.1%
0.00	0.25	50	**38.1%** / 0.2%	**42.3%** / 0.4%	4.2% / 0.5%	4.8% / 0.3%	0.6% / 11.1%
0.00	0.50	5	13.7% / 1.9%	13.3% / 3.0%	5.9% / 1.6%	5.4% / 2.0%	4.6% / 2.8%
0.00	0.50	10	11.8% / 1.8%	11.8% / 3.7%	5.9% / 2.1%	5.5% / 2.7%	4.4% / 4.7%
0.00	0.50	20	15.8% / 1.4%	14.1% / 2.3%	8.6% / 1.2%	9.2% / 1.4%	7.1% / 3.7%
0.00	0.50	50	**27.3%** / 0.2%	**23.2%** / 1.4%	18.8% / 0.1%	17.0% / 0.1%	6.0% / 3.3%
0.00	0.67	5	9.5% / 5.2%	8.6% / 5.7%	4.6% / 2.5%	4.9% / 3.0%	4.4% / 4.3%
0.00	0.67	10	9.9% / 4.2%	9.4% / 5.2%	7.5% / 3.4%	7.0% / 3.8%	5.3% / 5.1%
0.00	0.67	20	10.8% / 2.4%	10.1% / 3.4%	8.9% / 1.1%	7.5% / 1.5%	5.4% / 3.2%
0.00	0.67	50	15.1% / 1.3%	13.0% / 2.9%	13.6% / 0.3%	11.0% / 1.0%	6.5% / 2.5%
0.00	1.00	5	6.2% / 6.9%	6.6% / 7.3%	3.8% / 3.7%	4.0% / 4.0%	4.5% / 4.4%
0.00	1.00	10	6.6% / 6.4%	6.8% / 6.8%	5.0% / 4.7%	5.3% / 5.1%	5.2% / 4.6%
0.00	1.00	20	5.5% / 5.8%	5.7% / 6.2%	4.1% / 5.1%	4.2% / 5.0%	4.8% / 5.2%
0.00	1.00	50	5.9% / 5.4%	6.4% / 5.6%	6.5% / 4.7%	6.6% / 4.7%	5.8% / 4.5%
0.05	0.25	5	14.3% / 3.1%	16.0% / 2.8%	2.2% / 4.0%	1.9% / 3.2%	2.0% / 4.4%
0.05	0.25	10	19.2% / 0.8%	**21.0%** / 0.6%	2.4% / 2.0%	2.6% / 1.7%	2.1% / 5.7%
0.05	0.25	20	**26.1%** / 0.8%	**28.6%** / 0.6%	2.6% / 1.8%	2.8% / 1.1%	0.9% / 7.5%
0.05	0.25	50	**39.6%** / 0.2%	**43.0%** / 0.3%	5.0% / 1.1%	5.0% / 0.6%	0.4% / 11.2%
0.05	0.50	5	12.0% / 2.8%	11.5% / 4.1%	4.6% / 2.4%	4.6% / 2.7%	4.3% / 4.1%
0.05	0.50	10	14.8% / 1.9%	14.6% / 2.9%	8.1% / 1.6%	7.8% / 2.7%	6.3% / 4.7%
0.05	0.50	20	16.7% / 1.1%	16.1% / 2.1%	8.8% / 1.0%	8.8% / 1.5%	4.9% / 3.9%
0.05	0.50	50	**28.4%** / 0.7%	**24.2%** / 0.9%	16.5% / 0.2%	15.6% / 0.3%	5.6% / 3.6%
0.05	0.67	5	9.2% / 3.3%	8.4% / 3.8%	3.7% / 2.6%	3.9% / 2.7%	2.8% / 3.0%
0.05	0.67	10	11.0% / 3.4%	10.8% / 4.4%	7.3% / 2.5%	6.6% / 3.1%	4.6% / 5.3%
0.05	0.67	20	13.0% / 2.3%	11.5% / 3.5%	8.8% / 1.1%	7.8% / 1.5%	5.5% / 3.3%
0.05	0.67	50	19.0% / 1.5%	15.5% / 2.7%	15.1% / 0.5%	12.1% / 1.0%	5.7% / 3.4%
0.05	1.00	5	6.4% / 7.1%	6.6% / 6.8%	3.5% / 4.1%	3.9% / 4.1%	3.8% / 4.7%
0.05	1.00	10	5.2% / 6.2%	5.4% / 7.1%	4.2% / 4.8%	4.2% / 5.1%	5.6% / 4.0%
0.05	1.00	20	5.5% / 5.8%	5.8% / 6.3%	4.6% / 4.3%	4.8% / 4.6%	5.2% / 4.4%
0.05	1.00	50	7.4% / 5.8%	7.8% / 5.9%	4.3% / 3.8%	4.7% / 3.7%	4.3% / 3.6%
0.25	0.25	5	12.7% / 2.1%	13.8% / 2.5%	2.1% / 2.8%	1.9% / 2.3%	1.8% / 5.9%
0.25	0.25	10	19.5% / 1.8%	**21.0%** / 1.5%	3.8% / 3.1%	4.3% / 2.4%	0.9% / 9.5%
0.25	0.25	20	**26.8%** / 1.3%	**28.6%** / 1.1%	3.9% / 2.1%	3.7% / 1.3%	0.7% / 13.9%
0.25	0.25	50	**45.1%** / 0.8%	**47.0%** / 0.2%	4.1% / 1.3%	5.0% / 0.6%	0.1% / **24.4%**
0.25	0.50	5	9.9% / 2.4%	9.9% / 3.5%	4.0% / 1.9%	3.3% / 2.5%	3.4% / 3.5%
0.25	0.50	10	14.3% / 1.7%	14.4% / 2.7%	6.5% / 1.2%	6.6% / 1.6%	4.1% / 5.0%
0.25	0.50	20	19.0% / 1.6%	18.4% / 2.0%	8.2% / 0.7%	8.8% / 0.9%	4.2% / 5.2%
0.25	0.50	50	**25.1%** / 0.8%	**23.1%** / 1.4%	11.5% / 0.2%	11.9% / 0.4%	2.1% / 6.3%
0.25	0.67	5	8.2% / 4.6%	7.5% / 5.1%	5.2% / 2.4%	4.9% / 2.7%	3.3% / 4.0%
0.25	0.67	10	12.2% / 3.0%	11.6% / 4.1%	7.7% / 2.8%	7.8% / 3.2%	4.7% / 3.6%
0.25	0.67	20	13.6% / 2.7%	12.4% / 3.8%	9.0% / 1.0%	8.7% / 1.4%	4.6% / 3.0%
0.25	0.67	50	19.8% / 2.4%	17.6% / 2.9%	10.3% / 0.3%	9.0% / 0.6%	3.2% / 4.3%
0.25	1.00	5	6.9% / 5.5%	6.9% / 5.3%	3.2% / 3.5%	3.1% / 3.4%	2.9% / 4.3%
0.25	1.00	10	5.9% / 6.5%	6.2% / 6.6%	3.8% / 3.5%	3.8% / 3.6%	4.2% / 4.0%
0.25	1.00	20	7.4% / 7.6%	8.5% / 8.0%	3.7% / 3.5%	3.9% / 4.0%	3.5% / 3.6%
0.25	1.00	50	9.7% / 7.1%	10.0% / 7.7%	3.5% / 3.5%	3.7% / 3.8%	2.9% / 3.5%

Table A.8: Simulation I: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of one-sided test for bias (lower / upper tail); risk ratio ϕ as measure of treatment effect; control event rate equals 0.1; nominal significance level is 5% for each tail; error rates above 20% in bold font.

τ^2	ϕ	k	Linear regression test		Rank correlation test		New
			"add selective"	"add all"	"add selective"	"add all"	test
0.00	0.25	5	16.8% / 1.3%	12.6% / 1.9%	7.6% / 1.4%	6.7% / 2.0%	3.4% / 3.9%
0.00	0.25	10	**30.4%** / 0.0%	**21.8%** / 0.6%	**20.2%** / 0.2%	16.4% / 0.6%	5.8% / 4.1%
0.00	0.25	20	**46.5%** / 0.1%	**31.2%** / 0.5%	**38.5%** / 0.0%	**27.1%** / 0.3%	4.2% / 3.9%
0.00	0.25	50	**72.3%** / 0.0%	**48.8%** / 0.1%	**79.2%** / 0.0%	**58.3%** / 0.0%	4.6% / 3.5%
0.00	0.50	5	9.3% / 2.2%	7.6% / 2.5%	6.5% / 1.9%	5.3% / 2.5%	3.2% / 3.0%
0.00	0.50	10	16.0% / 1.0%	12.2% / 1.6%	14.3% / 1.2%	11.3% / 1.8%	5.1% / 4.8%
0.00	0.50	20	**20.5%** / 1.3%	15.3% / 2.0%	19.8% / 1.0%	14.6% / 1.5%	4.0% / 5.7%
0.00	0.50	50	**34.4%** / 0.1%	**21.2%** / 0.9%	**44.2%** / 0.0%	**29.0%** / 0.1%	3.8% / 4.6%
0.00	0.67	5	7.9% / 3.7%	6.9% / 4.0%	4.6% / 2.4%	4.5% / 2.8%	3.0% / 3.8%
0.00	0.67	10	12.2% / 3.1%	10.4% / 4.1%	12.0% / 2.7%	10.6% / 2.9%	5.9% / 4.0%
0.00	0.67	20	15.2% / 2.1%	12.1% / 2.9%	15.5% / 0.9%	11.9% / 1.5%	5.0% / 2.9%
0.00	0.67	50	**20.3%** / 0.6%	14.1% / 1.4%	**24.4%** / 0.6%	17.9% / 0.7%	5.2% / 3.0%
0.00	1.00	5	6.5% / 6.5%	6.2% / 6.3%	4.7% / 3.6%	4.7% / 3.5%	4.4% / 3.6%
0.00	1.00	10	6.2% / 5.7%	6.1% / 5.6%	6.1% / 5.1%	6.0% / 5.3%	4.4% / 4.9%
0.00	1.00	20	6.2% / 6.2%	6.2% / 5.8%	5.6% / 6.1%	5.3% / 5.9%	4.8% / 5.2%
0.00	1.00	50	6.2% / 6.5%	6.2% / 6.5%	5.5% / 6.0%	5.9% / 6.1%	5.4% / 5.5%
0.05	0.25	5	17.8% / 1.3%	13.8% / 1.7%	9.5% / 1.4%	7.2% / 1.9%	2.1% / 4.8%
0.05	0.25	10	**30.5%** / 0.9%	**23.4%** / 1.3%	18.8% / 0.8%	14.7% / 1.1%	3.4% / 4.6%
0.05	0.25	20	**42.8%** / 0.4%	**29.6%** / 0.6%	**35.6%** / 0.1%	**26.4%** / 0.3%	2.3% / 4.9%
0.05	0.25	50	**69.8%** / 0.0%	**50.4%** / 0.1%	**77.4%** / 0.0%	**57.9%** / 0.0%	1.2% / 5.0%
0.05	0.50	5	12.0% / 3.1%	9.8% / 3.4%	8.7% / 1.3%	7.7% / 1.7%	4.5% / 3.5%
0.05	0.50	10	16.1% / 2.0%	12.2% / 2.4%	13.3% / 2.2%	10.4% / 2.7%	4.0% / 5.0%
0.05	0.50	20	**22.7%** / 1.4%	17.5% / 2.3%	**20.3%** / 0.3%	15.4% / 0.7%	3.7% / 3.0%
0.05	0.50	50	**33.7%** / 0.4%	**25.8%** / 1.6%	**40.3%** / 0.1%	**28.2%** / 0.3%	2.4% / 5.2%
0.05	0.67	5	9.6% / 3.3%	8.5% / 3.9%	6.3% / 2.7%	5.5% / 3.0%	4.2% / 4.1%
0.05	0.67	10	11.8% / 2.8%	9.7% / 3.5%	9.0% / 1.5%	8.3% / 1.8%	3.6% / 4.2%
0.05	0.67	20	16.7% / 2.5%	14.1% / 3.1%	13.3% / 1.2%	11.3% / 1.5%	3.4% / 3.9%
0.05	0.67	50	**22.7%** / 2.5%	17.6% / 3.4%	**21.2%** / 0.5%	17.5% / 0.9%	3.3% / 5.0%
0.05	1.00	5	5.6% / 6.2%	5.7% / 6.1%	4.2% / 3.4%	4.2% / 3.1%	3.8% / 3.6%
0.05	1.00	10	5.3% / 7.1%	5.3% / 7.1%	5.2% / 5.3%	5.3% / 5.5%	4.8% / 5.4%
0.05	1.00	20	7.6% / 8.5%	7.5% / 8.6%	4.9% / 4.3%	4.8% / 4.5%	3.4% / 4.1%
0.05	1.00	50	6.2% / 9.2%	6.5% / 9.4%	3.9% / 6.4%	4.0% / 6.6%	4.1% / 4.7%
0.25	0.25	5	17.8% / 1.0%	14.0% / 1.4%	7.8% / 0.8%	6.9% / 1.2%	2.0% / 4.3%
0.25	0.25	10	**30.4%** / 0.6%	**24.1%** / 1.2%	18.4% / 0.3%	16.0% / 0.6%	1.9% / 7.4%
0.25	0.25	20	**41.0%** / 0.4%	**31.7%** / 1.0%	**31.5%** / 0.0%	**25.1%** / 0.1%	1.7% / 11.5%
0.25	0.25	50	**70.0%** / 0.0%	**56.1%** / 0.2%	**70.0%** / 0.0%	**57.6%** / 0.0%	0.5% / 19.9%
0.25	0.50	5	11.2% / 3.3%	9.4% / 3.7%	6.5% / 2.0%	6.1% / 2.0%	3.6% / 4.3%
0.25	0.50	10	16.0% / 2.9%	13.3% / 3.3%	10.3% / 1.7%	9.6% / 2.1%	1.9% / 5.5%
0.25	0.50	20	**21.4%** / 2.3%	19.8% / 3.2%	16.4% / 0.3%	13.7% / 0.7%	1.5% / 7.9%
0.25	0.50	50	**36.4%** / 1.2%	**31.1%** / 2.5%	**37.2%** / 0.1%	**29.2%** / 0.1%	1.0% / 8.7%
0.25	0.67	5	8.9% / 4.1%	8.4% / 4.3%	6.6% / 2.3%	6.2% / 2.4%	3.4% / 4.1%
0.25	0.67	10	9.4% / 3.8%	8.5% / 4.4%	9.0% / 2.1%	8.9% / 2.2%	3.4% / 4.5%
0.25	0.67	20	13.1% / 4.7%	12.0% / 5.7%	9.8% / 0.7%	8.4% / 1.4%	2.0% / 4.9%
0.25	0.67	50	**23.3%** / 4.2%	19.6% / 4.8%	19.3% / 0.6%	17.0% / 0.7%	1.0% / 6.2%
0.25	1.00	5	5.9% / 6.6%	5.8% / 6.9%	3.0% / 4.2%	3.2% / 4.4%	2.7% / 3.5%
0.25	1.00	10	8.2% / 7.4%	8.0% / 7.8%	4.4% / 4.2%	4.6% / 4.1%	3.8% / 4.2%
0.25	1.00	20	8.4% / 8.4%	8.7% / 8.5%	4.5% / 3.7%	4.7% / 3.8%	2.8% / 4.1%
0.25	1.00	50	9.2% / 9.7%	9.6% / 9.8%	4.1% / 4.7%	4.1% / 4.5%	3.0% / 3.0%

Table A.9: Simulation I: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of one-sided test for bias (lower / upper tail); risk ratio ϕ as measure of treatment effect; control event rate equals 0.3; nominal significance level is 5% for each tail; error rates above 20% in bold font.

τ^2	ψ / ϕ	k	$\widehat{\psi}_{MH}$ ($p_c = 0.1$) 50% [5%; 95%]	$\widehat{\psi}_{MH}$ ($p_c = 0.3$) 50% [5%; 95%]	$\widehat{\phi}_{MH}$ ($p_c = 0.1$) 50% [5%; 95%]	$\widehat{\phi}_{MH}$ ($p_c = 0.3$) 50% [5%; 95%]
0.00	0.25	5	0.26 [0.15; 0.40]	0.25 [0.19; 0.34]	0.26 [0.15; 0.42]	0.25 [0.18; 0.33]
0.00	0.25	10	0.25 [0.17; 0.36]	0.25 [0.20; 0.31]	0.25 [0.18; 0.36]	0.25 [0.20; 0.30]
0.00	0.25	20	0.26 [0.20; 0.33]	0.25 [0.21; 0.29]	0.26 [0.20; 0.33]	0.25 [0.22; 0.29]
0.00	0.25	50	0.26 [0.22; 0.30]	0.25 [0.23; 0.28]	0.26 [0.22; 0.30]	0.25 [0.23; 0.27]
0.00	0.50	5	0.51 [0.33; 0.78]	0.51 [0.38; 0.65]	0.51 [0.32; 0.74]	0.50 [0.40; 0.62]
0.00	0.50	10	0.50 [0.36; 0.69]	0.50 [0.42; 0.60]	0.50 [0.36; 0.66]	0.50 [0.43; 0.58]
0.00	0.50	20	0.50 [0.41; 0.61]	0.50 [0.44; 0.57]	0.49 [0.41; 0.60]	0.50 [0.45; 0.55]
0.00	0.50	50	0.50 [0.43; 0.57]	0.50 [0.46; 0.54]	0.50 [0.45; 0.56]	0.50 [0.47; 0.53]
0.00	0.67	5	0.67 [0.44; 1.00]	0.67 [0.53; 0.88]	0.67 [0.46; 0.95]	0.68 [0.56; 0.82]
0.00	0.67	10	0.67 [0.50; 0.87]	0.67 [0.56; 0.80]	0.67 [0.51; 0.87]	0.67 [0.59; 0.76]
0.00	0.67	20	0.67 [0.55; 0.81]	0.67 [0.59; 0.75]	0.67 [0.55; 0.80]	0.67 [0.61; 0.73]
0.00	0.67	50	0.67 [0.59; 0.75]	0.67 [0.62; 0.72]	0.67 [0.60; 0.75]	0.67 [0.63; 0.71]
0.00	1.00	5	1.02 [0.69; 1.42]	1.00 [0.78; 1.28]	1.00 [0.71; 1.37]	1.00 [0.84; 1.18]
0.00	1.00	10	1.00 [0.79; 1.29]	1.00 [0.85; 1.19]	1.00 [0.80; 1.27]	1.00 [0.89; 1.11]
0.00	1.00	20	1.00 [0.83; 1.18]	1.00 [0.90; 1.12]	1.00 [0.86; 1.18]	1.00 [0.93; 1.08]
0.00	1.00	50	1.00 [0.89; 1.12]	1.00 [0.93; 1.08]	1.00 [0.90; 1.10]	1.00 [0.95; 1.05]
0.05	0.25	5	0.25 [0.14; 0.43]	0.25 [0.17; 0.36]	0.26 [0.14; 0.41]	0.25 [0.17; 0.35]
0.05	0.25	10	0.26 [0.17; 0.37]	0.25 [0.19; 0.32]	0.26 [0.18; 0.37]	0.25 [0.19; 0.32]
0.05	0.25	20	0.25 [0.19; 0.34]	0.25 [0.21; 0.30]	0.26 [0.20; 0.33]	0.25 [0.21; 0.29]
0.05	0.25	50	0.26 [0.22; 0.30]	0.25 [0.22; 0.28]	0.26 [0.22; 0.30]	0.25 [0.23; 0.28]
0.05	0.50	5	0.50 [0.32; 0.82]	0.50 [0.37; 0.70]	0.50 [0.31; 0.78]	0.49 [0.37; 0.65]
0.05	0.50	10	0.50 [0.36; 0.68]	0.50 [0.39; 0.62]	0.50 [0.35; 0.68]	0.50 [0.41; 0.61]
0.05	0.50	20	0.50 [0.39; 0.62]	0.50 [0.43; 0.58]	0.50 [0.40; 0.62]	0.50 [0.43; 0.58]
0.05	0.50	50	0.50 [0.43; 0.58]	0.50 [0.46; 0.56]	0.50 [0.44; 0.57]	0.50 [0.45; 0.55]
0.05	0.67	5	0.66 [0.42; 1.02]	0.67 [0.49; 0.93]	0.67 [0.43; 1.00]	0.67 [0.51; 0.87]
0.05	0.67	10	0.66 [0.47; 0.91]	0.67 [0.52; 0.86]	0.66 [0.49; 0.89]	0.67 [0.55; 0.81]
0.05	0.67	20	0.67 [0.54; 0.83]	0.67 [0.57; 0.78]	0.67 [0.54; 0.81]	0.67 [0.58; 0.77]
0.05	0.67	50	0.67 [0.58; 0.76]	0.67 [0.61; 0.74]	0.67 [0.59; 0.76]	0.67 [0.62; 0.73]
0.05	1.00	5	1.00 [0.67; 1.52]	1.00 [0.74; 1.37]	1.00 [0.70; 1.47]	1.00 [0.78; 1.31]
0.05	1.00	10	1.00 [0.75; 1.33]	0.99 [0.80; 1.25]	1.00 [0.78; 1.30]	1.00 [0.84; 1.18]
0.05	1.00	20	0.99 [0.82; 1.21]	1.00 [0.86; 1.15]	1.01 [0.83; 1.22]	1.00 [0.88; 1.12]
0.05	1.00	50	1.00 [0.88; 1.13]	1.01 [0.91; 1.11]	1.00 [0.89; 1.12]	1.00 [0.92; 1.08]
0.25	0.25	5	0.25 [0.13; 0.47]	0.26 [0.15; 0.42]	0.25 [0.13; 0.47]	0.25 [0.15; 0.41]
0.25	0.25	10	0.26 [0.16; 0.40]	0.25 [0.18; 0.37]	0.26 [0.16; 0.39]	0.25 [0.17; 0.36]
0.25	0.25	20	0.26 [0.18; 0.35]	0.26 [0.20; 0.33]	0.26 [0.18; 0.36]	0.25 [0.19; 0.33]
0.25	0.25	50	0.26 [0.21; 0.31]	0.26 [0.22; 0.30]	0.26 [0.21; 0.32]	0.25 [0.21; 0.30]
0.25	0.50	5	0.50 [0.27; 0.90]	0.52 [0.31; 0.83]	0.50 [0.27; 0.91]	0.50 [0.33; 0.80]
0.25	0.50	10	0.50 [0.34; 0.76]	0.50 [0.35; 0.72]	0.49 [0.33; 0.75]	0.51 [0.36; 0.68]
0.25	0.50	20	0.51 [0.37; 0.68]	0.51 [0.39; 0.65]	0.50 [0.37; 0.66]	0.50 [0.39; 0.64]
0.25	0.50	50	0.51 [0.42; 0.61]	0.51 [0.43; 0.59]	0.50 [0.42; 0.60]	0.50 [0.43; 0.58]
0.25	0.67	5	0.65 [0.37; 1.16]	0.69 [0.43; 1.11]	0.67 [0.38; 1.14]	0.67 [0.41; 1.06]
0.25	0.67	10	0.67 [0.44; 1.03]	0.68 [0.49; 0.94]	0.66 [0.46; 1.00]	0.66 [0.47; 0.94]
0.25	0.67	20	0.67 [0.50; 0.88]	0.68 [0.53; 0.86]	0.67 [0.50; 0.89]	0.67 [0.53; 0.83]
0.25	0.67	50	0.67 [0.56; 0.81]	0.68 [0.59; 0.79]	0.67 [0.56; 0.79]	0.67 [0.58; 0.78]
0.25	1.00	5	1.04 [0.60; 1.76]	0.98 [0.62; 1.64]	1.02 [0.61; 1.65]	1.00 [0.62; 1.57]
0.25	1.00	10	1.00 [0.70; 1.45]	1.00 [0.72; 1.38]	1.00 [0.67; 1.44]	1.01 [0.73; 1.39]
0.25	1.00	20	1.01 [0.76; 1.30]	0.99 [0.79; 1.26]	1.00 [0.76; 1.31]	1.01 [0.80; 1.27]
0.25	1.00	50	1.00 [0.84; 1.18]	1.01 [0.87; 1.18]	1.00 [0.84; 1.18]	1.00 [0.86; 1.16]

Table A.10: Simulation II, Scenario A: results under null hypothesis ($n = 1000$); median, 5% and 95% quantiles of empirical distribution of estimated treatment effect based on Mantel-Haenszel method; both odds ratio ψ and risk ratio ϕ as measures of treatment effect.

τ^2	ψ	k	Linear regression test "add selective"	Linear regression test "add all"	Rank correlation test "add selective"	Rank correlation test "add all"	New test
0.00	0.25	5	11.2% / 1.3%	9.1% / 1.8%	5.1% / 1.5%	4.4% / 2.1%	3.0% / 4.3%
0.00	0.25	10	**23.2%** / 0.5%	18.8% / 0.6%	12.0% / 0.5%	10.6% / 0.8%	4.2% / 4.1%
0.00	0.25	20	**39.6%** / 0.0%	**29.7%** / 0.2%	**21.2%** / 0.2%	15.3% / 0.4%	3.1% / 4.5%
0.00	0.25	50	**72.9%** / 0.0%	**59.1%** / 0.0%	**50.4%** / 0.0%	**35.5%** / 0.0%	3.4% / 4.7%
0.00	0.50	5	9.7% / 2.8%	8.0% / 3.5%	7.5% / 2.6%	6.4% / 2.8%	4.3% / 4.1%
0.00	0.50	10	13.4% / 2.1%	10.8% / 3.0%	9.6% / 1.8%	8.1% / 2.3%	6.4% / 5.2%
0.00	0.50	20	17.8% / 1.2%	13.4% / 2.0%	14.3% / 1.2%	10.7% / 1.6%	4.4% / 3.6%
0.00	0.50	50	**30.9%** / 0.3%	**21.6%** / 0.7%	**27.5%** / 0.3%	16.7% / 0.4%	5.0% / 2.8%
0.00	0.67	5	10.4% / 4.7%	9.4% / 5.2%	7.2% / 4.3%	6.7% / 4.9%	4.3% / 4.9%
0.00	0.67	10	10.1% / 3.4%	7.9% / 4.0%	8.7% / 2.4%	7.7% / 3.3%	5.7% / 3.3%
0.00	0.67	20	13.3% / 2.2%	11.0% / 2.9%	11.7% / 1.4%	9.5% / 2.3%	5.3% / 3.1%
0.00	0.67	50	17.6% / 0.3%	10.9% / 1.0%	16.9% / 0.2%	10.9% / 0.8%	5.0% / 2.9%
0.00	1.00	5	6.3% / 5.7%	6.1% / 5.7%	4.5% / 4.2%	4.2% / 4.2%	4.1% / 4.1%
0.00	1.00	10	5.6% / 4.9%	5.3% / 4.6%	4.9% / 4.8%	4.6% / 4.5%	4.8% / 4.7%
0.00	1.00	20	6.3% / 5.9%	6.4% / 5.6%	5.3% / 5.6%	5.0% / 5.8%	4.3% / 4.9%
0.00	1.00	50	4.6% / 7.6%	4.3% / 7.5%	4.4% / 6.8%	4.1% / 6.4%	3.7% / 5.7%
0.05	0.25	5	13.2% / 0.8%	11.0% / 1.3%	5.8% / 1.1%	5.0% / 1.6%	3.9% / 3.4%
0.05	0.25	10	**25.0%** / 0.3%	19.2% / 0.9%	14.0% / 0.5%	11.5% / 1.1%	4.5% / 5.4%
0.05	0.25	20	**41.7%** / 0.2%	**31.4%** / 0.3%	**20.9%** / 0.2%	15.5% / 0.5%	2.8% / 3.5%
0.05	0.25	50	**72.5%** / 0.0%	**56.4%** / 0.0%	**48.2%** / 0.0%	**33.0%** / 0.2%	1.8% / 6.4%
0.05	0.50	5	8.1% / 2.9%	7.3% / 3.4%	5.2% / 1.8%	5.2% / 2.1%	3.0% / 3.5%
0.05	0.50	10	12.9% / 1.8%	10.7% / 2.7%	10.1% / 2.2%	7.5% / 3.1%	4.8% / 4.6%
0.05	0.50	20	17.6% / 1.3%	14.0% / 1.5%	13.8% / 0.8%	11.1% / 1.6%	5.0% / 4.5%
0.05	0.50	50	**31.3%** / 0.2%	**21.6%** / 0.7%	**25.5%** / 0.3%	18.0% / 0.7%	4.0% / 4.8%
0.05	0.67	5	6.9% / 3.3%	6.0% / 3.8%	5.6% / 2.9%	5.2% / 3.3%	3.9% / 4.3%
0.05	0.67	10	10.7% / 3.1%	9.3% / 3.4%	8.8% / 2.6%	7.4% / 3.0%	4.5% / 4.0%
0.05	0.67	20	11.3% / 1.1%	9.3% / 1.4%	10.0% / 1.1%	8.4% / 1.4%	4.6% / 3.2%
0.05	0.67	50	15.9% / 1.2%	11.7% / 1.8%	14.6% / 1.0%	11.3% / 1.4%	3.7% / 4.0%
0.05	1.00	5	5.0% / 6.2%	5.0% / 5.8%	4.6% / 4.4%	4.5% / 4.4%	4.1% / 3.9%
0.05	1.00	10	4.4% / 7.1%	4.2% / 6.7%	5.5% / 6.5%	5.4% / 6.4%	3.6% / 5.6%
0.05	1.00	20	4.7% / 6.3%	4.9% / 6.1%	3.3% / 5.4%	3.1% / 5.7%	2.8% / 4.9%
0.05	1.00	50	5.8% / 5.2%	5.7% / 4.9%	5.8% / 4.0%	5.7% / 3.9%	4.8% / 3.7%
0.25	0.25	5	14.1% / 1.4%	11.8% / 2.4%	6.5% / 1.6%	6.0% / 2.0%	2.5% / 4.4%
0.25	0.25	10	**23.1%** / 0.7%	18.2% / 0.8%	12.8% / 0.8%	10.6% / 0.9%	2.7% / 7.2%
0.25	0.25	20	**39.1%** / 0.4%	**30.9%** / 0.4%	17.1% / 0.4%	12.8% / 0.6%	0.9% / 6.8%
0.25	0.25	50	**72.8%** / 0.0%	**58.9%** / 0.0%	**41.7%** / 0.0%	**30.2%** / 0.0%	0.4% / 12.5%
0.25	0.50	5	8.6% / 2.8%	7.7% / 3.1%	4.8% / 2.1%	4.5% / 2.7%	3.4% / 5.1%
0.25	0.50	10	11.1% / 2.9%	9.8% / 3.4%	9.6% / 2.6%	7.9% / 3.3%	4.0% / 6.1%
0.25	0.50	20	14.1% / 0.6%	11.1% / 1.0%	12.0% / 0.6%	9.3% / 1.0%	2.6% / 4.4%
0.25	0.50	50	**35.8%** / 0.1%	**28.2%** / 0.2%	**29.9%** / 0.1%	**22.6%** / 0.2%	2.4% / 5.4%
0.25	0.67	5	7.7% / 3.2%	7.0% / 3.4%	5.2% / 2.4%	5.1% / 2.7%	3.6% / 2.4%
0.25	0.67	10	8.0% / 2.8%	6.7% / 3.2%	7.6% / 2.8%	6.7% / 3.0%	3.4% / 4.7%
0.25	0.67	20	9.7% / 1.9%	8.3% / 2.0%	7.8% / 1.8%	6.7% / 2.0%	2.3% / 4.2%
0.25	0.67	50	16.9% / 0.5%	13.6% / 0.8%	15.2% / 0.1%	12.5% / 0.2%	3.3% / 4.5%
0.25	1.00	5	5.5% / 4.9%	5.3% / 4.8%	3.7% / 4.2%	3.7% / 4.0%	3.8% / 3.8%
0.25	1.00	10	5.3% / 4.4%	5.3% / 4.6%	4.4% / 3.9%	4.4% / 3.9%	4.5% / 4.1%
0.25	1.00	20	5.7% / 5.7%	5.8% / 5.4%	5.3% / 4.1%	5.3% / 4.7%	4.6% / 3.5%
0.25	1.00	50	4.5% / 2.9%	4.5% / 3.5%	3.9% / 2.9%	3.9% / 2.8%	2.9% / 2.8%

Table A.11: Simulation II, Scenario A: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of one-sided test for bias (lower / upper tail); odds ratio ψ as measure of treatment effect; control event rate equals 0.1; nominal significance level is 5% for each tail; error rates above 20% in bold font.

τ^2	ψ	k	Linear regression test "add selective"	Linear regression test "add all"	Rank correlation test "add selective"	Rank correlation test "add all"	New test
0.00	0.25	5	9.4% / 2.0%	7.4% / 2.7%	6.3% / 2.5%	5.2% / 3.1%	3.9% / 4.4%
0.00	0.25	10	14.9% / 1.3%	11.0% / 2.1%	14.5% / 0.9%	10.7% / 1.6%	5.8% / 4.1%
0.00	0.25	20	**22.7%** / 0.7%	15.7% / 1.3%	19.5% / 0.6%	13.5% / 1.4%	4.3% / 4.1%
0.00	0.25	50	**47.9%** / 0.1%	**29.0%** / 0.4%	**42.7%** / 0.2%	**24.5%** / 0.2%	4.2% / 3.4%
0.00	0.50	5	8.2% / 4.5%	7.3% / 4.7%	6.8% / 2.7%	6.0% / 3.1%	5.0% / 4.0%
0.00	0.50	10	7.8% / 3.8%	6.6% / 4.3%	7.6% / 4.3%	7.0% / 4.6%	5.7% / 5.5%
0.00	0.50	20	9.9% / 3.3%	7.4% / 3.8%	9.7% / 1.6%	7.3% / 2.6%	3.7% / 4.0%
0.00	0.50	50	16.4% / 1.7%	11.6% / 2.9%	14.6% / 1.8%	10.8% / 2.5%	4.4% / 5.4%
0.00	0.67	5	5.9% / 6.4%	5.6% / 6.7%	4.9% / 3.9%	4.9% / 4.3%	4.7% / 4.5%
0.00	0.67	10	6.9% / 3.8%	6.2% / 4.0%	6.3% / 4.1%	5.4% / 4.4%	4.0% / 4.8%
0.00	0.67	20	7.8% / 3.2%	6.0% / 3.3%	6.6% / 2.5%	5.2% / 2.6%	3.9% / 2.9%
0.00	0.67	50	9.6% / 2.7%	8.1% / 3.2%	9.1% / 2.2%	8.0% / 2.6%	4.8% / 3.3%
0.00	1.00	5	4.9% / 7.1%	4.6% / 7.1%	3.6% / 5.5%	3.5% / 5.5%	3.7% / 5.0%
0.00	1.00	10	5.3% / 4.9%	5.2% / 4.9%	5.3% / 4.8%	5.4% / 4.9%	5.0% / 4.6%
0.00	1.00	20	5.6% / 5.7%	5.8% / 5.5%	5.6% / 5.1%	5.4% / 5.1%	5.1% / 4.8%
0.00	1.00	50	6.2% / 4.7%	5.9% / 4.6%	4.7% / 4.9%	4.8% / 5.1%	4.6% / 4.7%
0.05	0.25	5	10.6% / 3.6%	8.5% / 3.8%	6.8% / 2.3%	5.5% / 2.6%	4.0% / 4.9%
0.05	0.25	10	13.2% / 1.2%	10.1% / 1.8%	11.9% / 1.4%	8.9% / 1.7%	4.5% / 4.1%
0.05	0.25	20	**25.6%** / 0.4%	17.6% / 1.0%	18.8% / 0.7%	12.5% / 1.3%	4.2% / 4.3%
0.05	0.25	50	**49.1%** / 0.0%	**30.9%** / 0.4%	**43.8%** / 0.0%	**26.5%** / 0.2%	3.5% / 6.0%
0.05	0.50	5	7.0% / 3.8%	6.1% / 4.1%	4.5% / 3.3%	4.2% / 3.4%	3.2% / 4.1%
0.05	0.50	10	8.7% / 4.0%	7.8% / 4.5%	6.7% / 3.3%	5.9% / 3.6%	3.8% / 5.0%
0.05	0.50	20	8.5% / 1.3%	7.1% / 2.0%	8.1% / 1.8%	6.6% / 2.1%	3.7% / 3.2%
0.05	0.50	50	15.2% / 1.0%	11.4% / 2.0%	13.6% / 1.0%	10.7% / 1.9%	2.4% / 5.9%
0.05	0.67	5	6.7% / 4.0%	6.3% / 4.1%	4.6% / 3.3%	4.6% / 3.3%	4.6% / 4.0%
0.05	0.67	10	7.3% / 4.6%	6.7% / 4.7%	6.2% / 3.5%	5.9% / 4.0%	4.0% / 4.9%
0.05	0.67	20	8.0% / 3.5%	7.2% / 3.7%	6.9% / 2.3%	6.0% / 2.5%	4.0% / 3.4%
0.05	0.67	50	9.6% / 2.6%	8.3% / 3.1%	9.1% / 2.1%	7.8% / 2.6%	3.8% / 4.9%
0.05	1.00	5	4.4% / 4.0%	4.2% / 4.1%	3.4% / 3.7%	3.4% / 3.6%	3.2% / 4.0%
0.05	1.00	10	4.8% / 4.7%	4.7% / 4.6%	4.1% / 4.3%	4.1% / 4.3%	3.7% / 4.1%
0.05	1.00	20	4.9% / 4.3%	4.8% / 4.2%	5.3% / 3.6%	4.9% / 3.7%	5.0% / 3.0%
0.05	1.00	50	4.5% / 5.6%	4.4% / 5.6%	3.6% / 4.2%	3.4% / 4.2%	3.2% / 4.1%
0.25	0.25	5	9.1% / 1.4%	8.3% / 2.0%	6.4% / 1.3%	5.5% / 1.6%	2.6% / 3.7%
0.25	0.25	10	16.7% / 1.2%	13.7% / 1.5%	12.6% / 1.0%	9.8% / 1.1%	2.8% / 7.0%
0.25	0.25	20	**24.8%** / 0.1%	18.6% / 0.2%	18.2% / 0.1%	13.4% / 0.1%	1.5% / 7.3%
0.25	0.25	50	**52.7%** / 0.0%	**38.6%** / 0.0%	**39.4%** / 0.0%	**27.6%** / 0.0%	1.1% / 9.8%
0.25	0.50	5	7.1% / 3.8%	7.0% / 4.1%	5.2% / 3.3%	4.7% / 3.6%	3.2% / 4.2%
0.25	0.50	10	7.7% / 2.9%	6.9% / 3.1%	5.4% / 3.0%	5.2% / 3.2%	3.7% / 5.6%
0.25	0.50	20	12.9% / 1.3%	11.5% / 1.8%	9.0% / 1.3%	7.9% / 1.4%	2.7% / 5.3%
0.25	0.50	50	18.2% / 1.0%	14.5% / 1.3%	14.1% / 1.5%	11.0% / 1.6%	1.5% / 6.2%
0.25	0.67	5	7.2% / 3.9%	6.8% / 4.0%	4.8% / 3.0%	4.6% / 3.3%	4.0% / 3.0%
0.25	0.67	10	6.9% / 3.6%	6.0% / 3.6%	5.8% / 3.3%	5.3% / 3.3%	3.6% / 4.2%
0.25	0.67	20	7.8% / 2.2%	7.2% / 2.6%	5.5% / 1.6%	4.9% / 1.7%	2.6% / 3.3%
0.25	0.67	50	9.3% / 1.4%	8.3% / 1.8%	8.5% / 0.9%	7.3% / 1.1%	2.3% / 3.6%
0.25	1.00	5	5.9% / 5.5%	6.0% / 5.6%	4.3% / 3.2%	4.3% / 3.4%	3.8% / 4.0%
0.25	1.00	10	5.1% / 5.2%	5.1% / 5.2%	5.0% / 4.4%	5.0% / 4.2%	4.8% / 4.1%
0.25	1.00	20	3.9% / 4.8%	4.1% / 4.8%	3.1% / 3.4%	3.1% / 3.4%	2.7% / 3.2%
0.25	1.00	50	5.8% / 5.0%	5.8% / 5.0%	3.9% / 3.1%	3.6% / 2.9%	3.3% / 3.3%

Table A.12: Simulation II, Scenario A: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of one-sided test for bias (lower / upper tail); odds ratio ψ as measure of treatment effect; control event rate equals 0.3; nominal significance level is 5% for each tail; error rates above 20% in bold font.

τ^2	ϕ	k	Linear regression test "add selective"	Linear regression test "add all"	Rank correlation test "add selective"	Rank correlation test "add all"	New test
0.00	0.25	5	16.0% / 1.0%	13.4% / 1.6%	5.7% / 1.3%	4.7% / 1.8%	3.7% / 3.7%
0.00	0.25	10	**26.8%** / 0.3%	**22.4%** / 0.5%	10.9% / 0.8%	9.9% / 1.0%	4.1% / 6.3%
0.00	0.25	20	**44.2%** / 0.0%	**35.5%** / 0.2%	16.1% / 0.3%	13.2% / 0.5%	2.4% / 5.0%
0.00	0.25	50	**80.5%** / 0.0%	**68.3%** / 0.0%	**41.6%** / 0.0%	**31.4%** / 0.0%	2.6% / 5.6%
0.00	0.50	5	11.1% / 2.6%	8.3% / 3.1%	6.1% / 1.9%	6.0% / 2.5%	4.7% / 3.2%
0.00	0.50	10	15.5% / 2.1%	13.0% / 2.9%	11.2% / 1.5%	9.3% / 2.6%	5.4% / 4.5%
0.00	0.50	20	**25.3%** / 0.9%	19.9% / 0.9%	16.3% / 0.5%	12.0% / 0.7%	5.7% / 2.9%
0.00	0.50	50	**41.9%** / 0.0%	**31.7%** / 0.2%	**32.7%** / 0.0%	**21.7%** / 0.3%	5.8% / 4.3%
0.00	0.67	5	8.1% / 3.6%	6.7% / 4.0%	5.5% / 2.3%	5.0% / 2.3%	4.3% / 3.4%
0.00	0.67	10	9.7% / 2.4%	8.9% / 3.0%	8.8% / 3.0%	8.4% / 3.2%	5.4% / 4.6%
0.00	0.67	20	13.8% / 2.9%	10.6% / 3.4%	10.2% / 1.6%	8.2% / 2.2%	3.9% / 3.9%
0.00	0.67	50	**21.6%** / 1.3%	15.7% / 1.9%	17.3% / 1.0%	13.8% / 1.9%	4.8% / 3.6%
0.00	1.00	5	5.9% / 5.8%	5.7% / 5.4%	4.1% / 3.9%	3.9% / 4.0%	4.5% / 4.0%
0.00	1.00	10	5.7% / 5.1%	5.6% / 5.0%	6.5% / 4.1%	6.2% / 4.3%	5.9% / 3.4%
0.00	1.00	20	5.8% / 6.8%	5.6% / 6.4%	5.0% / 5.5%	4.8% / 5.6%	4.8% / 4.8%
0.00	1.00	50	5.7% / 6.5%	5.7% / 6.8%	5.0% / 5.8%	4.8% / 5.6%	4.6% / 5.5%
0.05	0.25	5	17.6% / 0.7%	14.5% / 1.5%	5.5% / 1.3%	6.0% / 2.1%	3.4% / 3.6%
0.05	0.25	10	**26.3%** / 0.5%	19.9% / 0.9%	8.1% / 0.7%	6.1% / 1.2%	2.4% / 5.1%
0.05	0.25	20	**46.2%** / 0.2%	**38.4%** / 0.4%	17.7% / 0.3%	14.3% / 0.4%	2.2% / 4.8%
0.05	0.25	50	**82.4%** / 0.0%	**72.9%** / 0.0%	**44.6%** / 0.0%	**35.1%** / 0.0%	2.1% / 6.8%
0.05	0.50	5	10.8% / 2.2%	9.8% / 2.6%	6.6% / 1.8%	5.8% / 2.3%	4.6% / 2.8%
0.05	0.50	10	12.4% / 2.4%	10.5% / 3.1%	9.1% / 2.6%	7.9% / 3.7%	5.0% / 6.5%
0.05	0.50	20	**21.2%** / 0.6%	16.0% / 1.1%	14.2% / 0.2%	10.4% / 1.2%	3.0% / 4.8%
0.05	0.50	50	**42.5%** / 0.1%	**31.3%** / 0.1%	**31.2%** / 0.0%	**21.5%** / 0.3%	4.0% / 3.5%
0.05	0.67	5	7.0% / 3.7%	6.4% / 4.0%	5.0% / 2.5%	4.8% / 2.8%	4.0% / 3.4%
0.05	0.67	10	11.3% / 3.3%	10.0% / 4.1%	7.8% / 2.3%	7.4% / 3.0%	4.5% / 4.9%
0.05	0.67	20	13.7% / 1.8%	10.7% / 2.1%	9.8% / 1.2%	7.8% / 1.7%	4.4% / 4.5%
0.05	0.67	50	**21.3%** / 1.0%	15.9% / 1.8%	15.0% / 0.9%	10.6% / 1.2%	3.4% / 4.7%
0.05	1.00	5	6.2% / 5.5%	5.9% / 5.4%	4.1% / 3.8%	3.9% / 4.0%	4.2% / 3.6%
0.05	1.00	10	5.1% / 5.1%	5.3% / 5.1%	5.1% / 5.0%	4.8% / 5.0%	5.1% / 4.0%
0.05	1.00	20	7.4% / 5.2%	7.5% / 5.3%	5.2% / 3.9%	5.5% / 4.1%	4.8% / 3.3%
0.05	1.00	50	6.0% / 5.6%	5.9% / 5.7%	5.7% / 4.3%	5.5% / 4.5%	4.7% / 3.9%
0.25	0.25	5	16.9% / 0.7%	13.8% / 1.0%	5.1% / 1.6%	5.2% / 1.5%	2.5% / 5.8%
0.25	0.25	10	**26.7%** / 0.2%	**22.9%** / 0.4%	9.3% / 0.6%	7.3% / 0.8%	1.4% / 7.9%
0.25	0.25	20	**47.8%** / 0.1%	**39.6%** / 0.2%	17.1% / 0.3%	12.9% / 0.6%	1.4% / 9.7%
0.25	0.25	50	**81.2%** / 0.0%	**72.1%** / 0.0%	**39.9%** / 0.0%	**32.3%** / 0.0%	0.2% / 17.6%
0.25	0.50	5	9.5% / 2.8%	8.9% / 3.7%	5.0% / 2.3%	4.8% / 3.0%	3.2% / 4.5%
0.25	0.50	10	13.5% / 1.1%	10.2% / 1.7%	8.2% / 1.0%	6.8% / 1.7%	2.5% / 5.7%
0.25	0.50	20	**20.1%** / 0.8%	16.2% / 1.3%	12.5% / 0.6%	10.2% / 0.9%	2.0% / 5.6%
0.25	0.50	50	**37.9%** / 0.0%	**30.5%** / 0.1%	**25.3%** / 0.0%	19.8% / 0.1%	1.2% / 7.2%
0.25	0.67	5	8.1% / 3.1%	7.6% / 2.9%	4.6% / 2.6%	4.2% / 2.7%	3.3% / 3.1%
0.25	0.67	10	9.9% / 2.9%	8.8% / 3.5%	8.5% / 3.2%	8.4% / 3.9%	4.1% / 5.9%
0.25	0.67	20	13.5% / 0.8%	12.0% / 1.0%	9.6% / 0.4%	8.6% / 0.6%	3.9% / 2.9%
0.25	0.67	50	**20.4%** / 0.3%	16.9% / 0.6%	13.9% / 0.5%	11.2% / 0.7%	1.9% / 5.0%
0.25	1.00	5	6.0% / 5.3%	6.1% / 5.4%	3.6% / 4.6%	3.8% / 4.8%	4.1% / 3.9%
0.25	1.00	10	5.8% / 5.7%	5.5% / 5.7%	4.0% / 5.3%	4.3% / 5.1%	4.3% / 4.1%
0.25	1.00	20	5.3% / 4.9%	5.4% / 5.3%	4.2% / 3.8%	4.4% / 3.8%	3.7% / 3.3%
0.25	1.00	50	4.8% / 4.0%	4.9% / 4.1%	3.8% / 3.2%	3.9% / 3.5%	2.8% / 2.9%

Table A.13: Simulation II, Scenario A: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of one-sided test for bias (lower / upper tail); risk ratio ϕ as measure of treatment effect; control event rate equals 0.1; nominal significance level is 5% for each tail; error rates above 20% in bold font.

τ^2	ϕ	k	Linear regression test "add selective"	Linear regression test "add all"	Rank correlation test "add selective"	Rank correlation test "add all"	New test
0.00	0.25	5	16.1% / 1.5%	12.3% / 2.3%	8.9% / 1.1%	7.4% / 1.5%	4.4% / 3.6%
0.00	0.25	10	**27.1%** / 0.5%	17.6% / 0.7%	19.6% / 0.7%	12.6% / 0.8%	4.9% / 5.1%
0.00	0.25	20	**46.4%** / 0.1%	**31.2%** / 0.3%	**34.8%** / 0.1%	**20.7%** / 0.2%	4.0% / 4.4%
0.00	0.25	50	**84.0%** / 0.0%	**60.3%** / 0.0%	**72.0%** / 0.0%	**47.9%** / 0.0%	4.4% / 3.8%
0.00	0.50	5	10.9% / 3.1%	9.6% / 3.5%	8.1% / 2.3%	6.6% / 2.7%	4.2% / 3.9%
0.00	0.50	10	11.3% / 2.2%	8.5% / 2.4%	9.4% / 2.3%	7.2% / 2.8%	3.8% / 4.3%
0.00	0.50	20	19.5% / 0.8%	15.2% / 0.9%	16.8% / 1.1%	13.4% / 1.6%	5.2% / 4.7%
0.00	0.50	50	**37.0%** / 0.1%	**23.9%** / 0.2%	**31.3%** / 0.0%	**21.2%** / 0.2%	4.3% / 4.1%
0.00	0.67	5	6.9% / 4.1%	6.4% / 4.2%	5.4% / 3.5%	4.9% / 4.1%	4.1% / 4.2%
0.00	0.67	10	8.7% / 3.1%	8.4% / 3.6%	8.3% / 3.4%	7.8% / 3.7%	4.5% / 4.8%
0.00	0.67	20	12.4% / 2.1%	10.3% / 2.3%	9.9% / 1.7%	9.4% / 1.9%	5.5% / 3.8%
0.00	0.67	50	19.8% / 0.7%	15.5% / 1.4%	16.6% / 0.9%	13.1% / 1.4%	4.4% / 4.9%
0.00	1.00	5	5.7% / 4.7%	5.7% / 4.7%	3.9% / 5.3%	4.0% / 5.2%	3.5% / 4.8%
0.00	1.00	10	5.7% / 5.4%	5.5% / 5.5%	5.8% / 5.4%	5.7% / 5.5%	4.5% / 4.3%
0.00	1.00	20	4.6% / 5.9%	4.6% / 5.8%	4.4% / 4.8%	4.4% / 4.9%	4.5% / 4.8%
0.00	1.00	50	5.9% / 6.1%	5.7% / 6.1%	4.8% / 5.1%	4.4% / 5.3%	4.3% / 4.5%
0.05	0.25	5	13.3% / 1.4%	10.9% / 2.2%	7.7% / 1.6%	5.8% / 2.0%	3.1% / 4.5%
0.05	0.25	10	**23.5%** / 0.3%	17.5% / 1.0%	17.7% / 0.5%	12.4% / 0.9%	3.2% / 5.5%
0.05	0.25	20	**44.3%** / 0.1%	**30.7%** / 0.2%	**31.7%** / 0.1%	**21.0%** / 0.3%	2.4% / 5.6%
0.05	0.25	50	**80.5%** / 0.0%	**64.2%** / 0.0%	**70.0%** / 0.0%	**49.1%** / 0.1%	2.2% / 6.9%
0.05	0.50	5	8.5% / 2.3%	7.9% / 2.7%	5.9% / 1.7%	5.7% / 1.9%	3.3% / 3.8%
0.05	0.50	10	11.8% / 1.7%	9.4% / 2.4%	9.2% / 2.0%	7.6% / 2.6%	3.1% / 6.3%
0.05	0.50	20	19.0% / 0.6%	14.7% / 0.9%	13.8% / 0.6%	11.2% / 0.8%	2.5% / 4.5%
0.05	0.50	50	**34.0%** / 0.6%	**25.1%** / 0.6%	**27.1%** / 0.2%	**20.3%** / 0.5%	2.7% / 5.8%
0.05	0.67	5	6.9% / 4.0%	6.5% / 4.5%	5.3% / 3.2%	4.8% / 3.5%	3.8% / 4.4%
0.05	0.67	10	8.5% / 3.3%	7.8% / 3.5%	5.8% / 3.4%	5.6% / 4.2%	3.5% / 5.3%
0.05	0.67	20	12.8% / 1.8%	11.0% / 2.2%	10.0% / 1.7%	8.8% / 2.2%	4.6% / 4.5%
0.05	0.67	50	18.4% / 1.0%	15.7% / 1.3%	14.8% / 0.7%	12.0% / 1.3%	2.4% / 3.7%
0.05	1.00	5	4.7% / 5.2%	4.6% / 5.2%	4.0% / 3.4%	4.0% / 3.5%	3.7% / 3.0%
0.05	1.00	10	3.8% / 5.0%	3.8% / 5.0%	3.8% / 4.1%	3.8% / 4.1%	3.7% / 3.9%
0.05	1.00	20	4.3% / 5.1%	4.5% / 5.1%	3.6% / 4.5%	3.6% / 4.6%	2.4% / 4.3%
0.05	1.00	50	5.0% / 5.0%	4.7% / 5.1%	4.1% / 3.3%	4.0% / 3.3%	3.3% / 3.5%
0.25	0.25	5	13.9% / 0.8%	12.1% / 1.1%	8.5% / 0.7%	7.8% / 0.8%	1.8% / 5.1%
0.25	0.25	10	**24.7%** / 0.3%	19.7% / 0.4%	15.5% / 0.4%	11.9% / 0.4%	1.1% / 9.0%
0.25	0.25	20	**44.2%** / 0.1%	**35.1%** / 0.1%	**28.9%** / 0.0%	**21.9%** / 0.0%	0.9% / 12.8%
0.25	0.25	50	**84.4%** / 0.0%	**70.6%** / 0.0%	**68.0%** / 0.0%	**50.7%** / 0.0%	0.1% / **24.9%**
0.25	0.50	5	8.8% / 2.7%	8.1% / 3.0%	6.4% / 1.8%	6.3% / 1.9%	2.5% / 4.1%
0.25	0.50	10	10.1% / 1.2%	8.9% / 1.7%	9.2% / 1.5%	8.2% / 1.8%	2.0% / 5.5%
0.25	0.50	20	18.0% / 0.7%	15.9% / 0.8%	12.5% / 0.4%	10.6% / 0.5%	1.2% / 6.3%
0.25	0.50	50	**32.9%** / 0.1%	**28.1%** / 0.2%	**24.8%** / 0.0%	**21.0%** / 0.0%	0.7% / 10.3%
0.25	0.67	5	9.5% / 3.2%	9.1% / 3.5%	7.0% / 2.4%	7.0% / 2.5%	4.4% / 3.7%
0.25	0.67	10	9.0% / 2.9%	8.5% / 3.3%	8.4% / 2.3%	7.6% / 2.5%	3.8% / 5.5%
0.25	0.67	20	11.1% / 1.4%	10.3% / 1.4%	8.6% / 1.4%	7.7% / 1.7%	1.7% / 4.6%
0.25	0.67	50	17.0% / 0.9%	15.0% / 1.1%	13.1% / 0.6%	11.1% / 0.8%	1.7% / 6.4%
0.25	1.00	5	3.9% / 5.7%	3.9% / 5.7%	2.1% / 3.3%	2.2% / 3.4%	2.3% / 3.2%
0.25	1.00	10	3.6% / 4.7%	3.7% / 4.6%	4.3% / 3.2%	4.4% / 3.3%	3.8% / 2.7%
0.25	1.00	20	4.4% / 3.8%	4.5% / 3.7%	3.5% / 2.5%	4.0% / 2.6%	3.3% / 2.4%
0.25	1.00	50	3.2% / 4.4%	3.0% / 4.4%	3.1% / 4.0%	3.1% / 4.1%	3.0% / 2.9%

Table A.14: Simulation II, Scenario A: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of one-sided test for bias (lower / upper tail); risk ratio ϕ as measure of treatment effect; control event rate equals 0.3; nominal significance level is 5% for each tail; error rates above 20% in bold font.

τ^2	ψ / ϕ	k	$\widehat{\psi}_{MH}$ ($p_c = 0.1$) 50% [5%; 95%]	$\widehat{\psi}_{MH}$ ($p_c = 0.3$) 50% [5%; 95%]	$\widehat{\phi}_{MH}$ ($p_c = 0.1$) 50% [5%; 95%]	$\widehat{\phi}_{MH}$ ($p_c = 0.3$) 50% [5%; 95%]
0.00	0.25	5	0.25 [0.17; 0.34]	0.25 [0.20; 0.30]	0.25 [0.17; 0.34]	0.25 [0.21; 0.29]
0.00	0.25	10	0.25 [0.19; 0.32]	0.25 [0.22; 0.29]	0.25 [0.19; 0.31]	0.25 [0.22; 0.28]
0.00	0.25	20	0.25 [0.21; 0.29]	0.25 [0.23; 0.27]	0.25 [0.21; 0.29]	0.25 [0.23; 0.27]
0.00	0.25	50	0.25 [0.22; 0.28]	0.25 [0.24; 0.26]	0.25 [0.23; 0.28]	0.25 [0.24; 0.26]
0.00	0.50	5	0.50 [0.38; 0.66]	0.50 [0.43; 0.59]	0.50 [0.38; 0.63]	0.50 [0.43; 0.57]
0.00	0.50	10	0.50 [0.41; 0.60]	0.50 [0.45; 0.56]	0.50 [0.42; 0.59]	0.50 [0.45; 0.55]
0.00	0.50	20	0.50 [0.44; 0.57]	0.50 [0.46; 0.54]	0.50 [0.44; 0.56]	0.50 [0.47; 0.53]
0.00	0.50	50	0.50 [0.46; 0.54]	0.50 [0.48; 0.52]	0.50 [0.47; 0.54]	0.50 [0.48; 0.52]
0.00	0.67	5	0.67 [0.51; 0.85]	0.67 [0.57; 0.79]	0.67 [0.53; 0.83]	0.67 [0.58; 0.76]
0.00	0.67	10	0.67 [0.56; 0.79]	0.67 [0.60; 0.75]	0.67 [0.57; 0.79]	0.67 [0.62; 0.73]
0.00	0.67	20	0.67 [0.60; 0.76]	0.67 [0.62; 0.72]	0.67 [0.59; 0.75]	0.67 [0.63; 0.71]
0.00	0.67	50	0.67 [0.62; 0.72]	0.67 [0.64; 0.70]	0.67 [0.62; 0.71]	0.67 [0.65; 0.69]
0.00	1.00	5	1.01 [0.79; 1.25]	1.00 [0.87; 1.16]	1.00 [0.82; 1.24]	0.99 [0.90; 1.11]
0.00	1.00	10	1.00 [0.86; 1.17]	1.00 [0.90; 1.11]	1.00 [0.86; 1.15]	1.00 [0.93; 1.08]
0.00	1.00	20	1.00 [0.90; 1.11]	1.00 [0.92; 1.07]	1.00 [0.91; 1.11]	1.00 [0.95; 1.05]
0.00	1.00	50	1.00 [0.94; 1.07]	1.00 [0.95; 1.04]	1.00 [0.94; 1.06]	1.00 [0.97; 1.03]
0.05	0.25	5	0.25 [0.17; 0.37]	0.25 [0.19; 0.33]	0.25 [0.16; 0.36]	0.25 [0.19; 0.33]
0.05	0.25	10	0.25 [0.19; 0.32]	0.25 [0.21; 0.30]	0.25 [0.19; 0.33]	0.25 [0.20; 0.30]
0.05	0.25	20	0.25 [0.21; 0.30]	0.25 [0.22; 0.29]	0.25 [0.21; 0.30]	0.25 [0.22; 0.28]
0.05	0.25	50	0.25 [0.22; 0.28]	0.25 [0.23; 0.27]	0.25 [0.22; 0.28]	0.25 [0.23; 0.27]
0.05	0.50	5	0.50 [0.36; 0.69]	0.50 [0.39; 0.66]	0.49 [0.36; 0.68]	0.50 [0.39; 0.63]
0.05	0.50	10	0.50 [0.40; 0.64]	0.50 [0.42; 0.60]	0.50 [0.41; 0.64]	0.50 [0.42; 0.59]
0.05	0.50	20	0.50 [0.43; 0.59]	0.50 [0.44; 0.57]	0.50 [0.43; 0.58]	0.50 [0.44; 0.56]
0.05	0.50	50	0.50 [0.45; 0.55]	0.50 [0.46; 0.54]	0.50 [0.45; 0.55]	0.50 [0.47; 0.54]
0.05	0.67	5	0.67 [0.49; 0.93]	0.67 [0.52; 0.86]	0.66 [0.49; 0.90]	0.67 [0.53; 0.83]
0.05	0.67	10	0.67 [0.53; 0.83]	0.67 [0.56; 0.80]	0.67 [0.54; 0.83]	0.67 [0.57; 0.78]
0.05	0.67	20	0.67 [0.58; 0.78]	0.67 [0.59; 0.76]	0.67 [0.57; 0.77]	0.67 [0.60; 0.75]
0.05	0.67	50	0.67 [0.61; 0.74]	0.67 [0.62; 0.73]	0.67 [0.61; 0.74]	0.67 [0.62; 0.72]
0.05	1.00	5	1.00 [0.75; 1.36]	1.01 [0.79; 1.26]	0.99 [0.77; 1.30]	1.01 [0.82; 1.24]
0.05	1.00	10	1.00 [0.81; 1.24]	1.00 [0.84; 1.19]	1.00 [0.82; 1.22]	1.00 [0.85; 1.17]
0.05	1.00	20	1.00 [0.86; 1.17]	1.00 [0.88; 1.13]	0.99 [0.86; 1.14]	1.00 [0.89; 1.12]
0.05	1.00	50	1.00 [0.90; 1.10]	1.00 [0.93; 1.08]	1.00 [0.92; 1.08]	1.00 [0.93; 1.07]
0.25	0.25	5	0.25 [0.14; 0.42]	0.25 [0.15; 0.41]	0.24 [0.14; 0.42]	0.25 [0.15; 0.40]
0.25	0.25	10	0.25 [0.17; 0.37]	0.26 [0.18; 0.35]	0.26 [0.17; 0.37]	0.25 [0.18; 0.35]
0.25	0.25	20	0.25 [0.19; 0.33]	0.26 [0.20; 0.33]	0.25 [0.19; 0.33]	0.25 [0.20; 0.32]
0.25	0.25	50	0.25 [0.21; 0.30]	0.26 [0.22; 0.30]	0.25 [0.21; 0.30]	0.25 [0.21; 0.30]
0.25	0.50	5	0.50 [0.30; 0.85]	0.50 [0.32; 0.83]	0.50 [0.31; 0.80]	0.50 [0.31; 0.78]
0.25	0.50	10	0.50 [0.35; 0.72]	0.50 [0.36; 0.70]	0.50 [0.35; 0.73]	0.50 [0.35; 0.69]
0.25	0.50	20	0.50 [0.39; 0.64]	0.50 [0.40; 0.64]	0.50 [0.39; 0.66]	0.50 [0.40; 0.63]
0.25	0.50	50	0.50 [0.43; 0.59]	0.51 [0.44; 0.59]	0.50 [0.43; 0.59]	0.50 [0.43; 0.58]
0.25	0.67	5	0.65 [0.41; 1.11]	0.68 [0.43; 1.08]	0.68 [0.41; 1.12]	0.68 [0.44; 1.04]
0.25	0.67	10	0.67 [0.47; 0.96]	0.68 [0.49; 0.93]	0.66 [0.46; 0.94]	0.67 [0.48; 0.94]
0.25	0.67	20	0.67 [0.53; 0.85]	0.68 [0.53; 0.86]	0.66 [0.52; 0.85]	0.67 [0.53; 0.84]
0.25	0.67	50	0.67 [0.57; 0.79]	0.67 [0.59; 0.77]	0.67 [0.57; 0.79]	0.67 [0.58; 0.78]
0.25	1.00	5	1.00 [0.64; 1.64]	0.99 [0.64; 1.57]	0.98 [0.61; 1.63]	1.01 [0.65; 1.59]
0.25	1.00	10	0.99 [0.68; 1.43]	0.98 [0.71; 1.38]	0.99 [0.69; 1.39]	0.98 [0.72; 1.39]
0.25	1.00	20	0.99 [0.76; 1.29]	0.99 [0.78; 1.27]	1.00 [0.78; 1.26]	0.99 [0.79; 1.24]
0.25	1.00	50	1.00 [0.86; 1.17]	0.99 [0.87; 1.16]	1.00 [0.85; 1.16]	1.00 [0.87; 1.16]

Table A.15: Simulation II, Scenario B: results under null hypothesis ($n = 1000$); median, 5% and 95% quantiles of empirical distribution of estimated treatment effect based on Mantel-Haenszel method; both odds ratio ψ and risk ratio ϕ as measures of treatment effect.

τ^2	ψ	k	Linear regression test "add selective"	Linear regression test "add all"	Rank correlation test "add selective"	Rank correlation test "add all"	New test
0.00	0.25	5	11.9% / 2.2%	8.5% / 3.0%	6.1% / 1.8%	5.0% / 2.6%	3.8% / 3.8%
0.00	0.25	10	17.2% / 1.4%	11.7% / 2.2%	12.7% / 1.2%	8.9% / 1.8%	4.6% / 5.3%
0.00	0.25	20	**28.7%** / 0.0%	17.2% / 0.6%	**21.6%** / 0.1%	12.1% / 0.5%	4.8% / 2.0%
0.00	0.25	50	**52.9%** / 0.0%	**28.2%** / 0.5%	**43.1%** / 0.0%	**21.7%** / 0.8%	4.1% / 3.3%
0.00	0.50	5	7.8% / 3.1%	6.1% / 3.5%	5.2% / 2.5%	4.8% / 2.5%	3.9% / 2.6%
0.00	0.50	10	9.4% / 3.8%	6.7% / 4.6%	7.5% / 4.1%	5.7% / 4.9%	3.8% / 6.2%
0.00	0.50	20	12.5% / 2.4%	8.9% / 3.6%	10.6% / 1.9%	7.4% / 3.1%	4.2% / 4.9%
0.00	0.50	50	**20.9%** / 1.0%	11.7% / 2.5%	17.5% / 0.9%	11.2% / 1.6%	4.3% / 3.5%
0.00	0.67	5	5.9% / 4.6%	5.2% / 4.7%	4.7% / 3.2%	4.1% / 3.5%	3.6% / 3.7%
0.00	0.67	10	8.4% / 5.1%	7.0% / 5.7%	7.6% / 4.5%	6.9% / 4.9%	5.1% / 5.8%
0.00	0.67	20	8.9% / 3.9%	7.2% / 5.1%	7.0% / 2.9%	5.9% / 3.2%	4.7% / 4.2%
0.00	0.67	50	13.7% / 2.3%	10.4% / 3.1%	11.4% / 1.3%	8.3% / 2.2%	4.9% / 3.6%
0.00	1.00	5	5.6% / 4.1%	5.4% / 4.1%	4.0% / 2.9%	4.2% / 2.8%	3.7% / 2.8%
0.00	1.00	10	6.3% / 4.9%	6.4% / 4.7%	5.8% / 4.5%	5.7% / 4.3%	5.5% / 4.5%
0.00	1.00	20	6.6% / 5.3%	6.3% / 4.9%	4.7% / 4.3%	4.6% / 4.2%	4.4% / 4.2%
0.00	1.00	50	6.4% / 5.1%	6.2% / 4.9%	5.2% / 4.6%	5.2% / 4.7%	4.9% / 4.1%
0.05	0.25	5	9.2% / 1.6%	6.8% / 2.8%	5.3% / 1.9%	4.3% / 2.3%	2.6% / 4.1%
0.05	0.25	10	18.2% / 0.7%	12.5% / 1.3%	12.1% / 0.8%	8.5% / 1.6%	3.7% / 3.4%
0.05	0.25	20	**26.7%** / 0.6%	16.1% / 1.4%	18.4% / 0.2%	11.3% / 0.7%	3.2% / 4.9%
0.05	0.25	50	**52.2%** / 0.0%	**28.0%** / 0.0%	**41.0%** / 0.0%	**20.6%** / 0.1%	1.5% / 4.6%
0.05	0.50	5	6.8% / 3.1%	5.7% / 3.5%	3.6% / 3.3%	3.3% / 3.5%	2.7% / 3.4%
0.05	0.50	10	10.0% / 2.9%	7.6% / 3.7%	8.5% / 1.8%	7.3% / 2.3%	5.3% / 4.0%
0.05	0.50	20	10.8% / 1.2%	8.8% / 1.4%	7.8% / 1.0%	5.7% / 1.4%	3.1% / 3.0%
0.05	0.50	50	18.9% / 0.7%	12.6% / 1.7%	17.2% / 0.4%	10.5% / 1.0%	2.3% / 5.2%
0.05	0.67	5	6.4% / 5.2%	5.4% / 5.4%	3.8% / 4.3%	3.7% / 4.4%	3.2% / 4.5%
0.05	0.67	10	6.3% / 4.5%	5.7% / 4.5%	5.6% / 3.8%	5.0% / 4.2%	3.8% / 5.4%
0.05	0.67	20	8.2% / 3.1%	6.4% / 3.6%	6.7% / 1.2%	5.6% / 1.7%	3.3% / 2.5%
0.05	0.67	50	10.8% / 2.4%	9.1% / 2.9%	9.1% / 2.3%	7.4% / 2.7%	2.8% / 4.2%
0.05	1.00	5	5.0% / 5.5%	4.9% / 5.6%	3.0% / 4.7%	3.1% / 4.7%	2.8% / 4.6%
0.05	1.00	10	5.0% / 4.1%	5.1% / 4.0%	4.3% / 4.5%	4.2% / 4.5%	4.2% / 3.9%
0.05	1.00	20	5.7% / 4.3%	5.7% / 4.3%	4.2% / 4.0%	4.2% / 3.9%	3.8% / 4.0%
0.05	1.00	50	5.6% / 4.0%	5.5% / 3.9%	4.1% / 3.3%	3.9% / 3.1%	3.8% / 3.1%
0.25	0.25	5	10.8% / 1.7%	9.0% / 2.1%	6.5% / 1.2%	5.4% / 1.5%	3.0% / 4.4%
0.25	0.25	10	15.5% / 0.6%	11.7% / 1.0%	11.2% / 0.7%	7.9% / 1.3%	2.6% / 6.6%
0.25	0.25	20	**25.0%** / 0.2%	17.2% / 0.6%	14.8% / 0.1%	9.8% / 0.2%	1.1% / 8.2%
0.25	0.25	50	**56.8%** / 0.0%	**37.9%** / 0.1%	**36.5%** / 0.0%	**21.2%** / 0.2%	0.2% / 13.4%
0.25	0.50	5	7.0% / 2.9%	6.3% / 3.0%	4.9% / 2.5%	4.7% / 2.7%	2.6% / 3.7%
0.25	0.50	10	9.6% / 1.6%	7.7% / 2.1%	7.9% / 1.5%	7.0% / 1.9%	2.9% / 3.8%
0.25	0.50	20	11.2% / 1.2%	9.3% / 1.7%	8.1% / 1.6%	6.6% / 1.9%	2.0% / 4.8%
0.25	0.50	50	**20.4%** / 0.6%	16.1% / 0.8%	15.4% / 0.5%	11.3% / 0.8%	1.0% / 6.1%
0.25	0.67	5	5.6% / 3.4%	5.4% / 3.9%	2.8% / 2.5%	2.7% / 2.6%	2.6% / 3.7%
0.25	0.67	10	6.3% / 3.7%	5.8% / 4.1%	5.9% / 2.0%	5.4% / 2.3%	3.3% / 3.7%
0.25	0.67	20	7.9% / 3.2%	7.2% / 3.3%	5.6% / 1.5%	5.2% / 1.7%	2.8% / 4.2%
0.25	0.67	50	10.1% / 1.7%	8.1% / 2.1%	7.3% / 1.1%	5.3% / 1.4%	1.3% / 4.6%
0.25	1.00	5	4.9% / 4.7%	5.1% / 4.8%	3.4% / 2.7%	3.5% / 2.6%	3.7% / 2.7%
0.25	1.00	10	4.6% / 5.1%	4.6% / 5.4%	4.9% / 4.3%	5.0% / 4.3%	4.8% / 3.3%
0.25	1.00	20	4.0% / 4.1%	4.0% / 4.4%	2.4% / 2.5%	2.5% / 2.5%	2.7% / 2.5%
0.25	1.00	50	3.4% / 3.3%	3.4% / 3.2%	2.2% / 2.1%	1.9% / 2.0%	1.5% / 1.7%

Table A.16: Simulation II, Scenario B: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of one-sided test for bias (lower / upper tail); odds ratio ψ as measure of treatment effect; control event rate equals 0.1; nominal significance level is 5% for each tail; error rates above 20% in bold font.

τ^2	ψ	k	Linear regression test "add selective"	Linear regression test "add all"	Rank correlation test "add selective"	Rank correlation test "add all"	New test
0.00	0.25	5	7.7% / 4.1%	6.4% / 4.4%	6.1% / 3.3%	4.7% / 3.6%	3.5% / 4.7%
0.00	0.25	10	8.5% / 1.7%	6.3% / 2.5%	7.4% / 2.1%	5.8% / 2.7%	4.1% / 4.8%
0.00	0.25	20	14.2% / 1.1%	9.5% / 2.2%	13.1% / 0.7%	9.9% / 1.1%	4.1% / 2.5%
0.00	0.25	50	**21.6%** / 0.4%	13.3% / 2.0%	**21.1%** / 0.5%	12.3% / 1.3%	4.7% / 3.7%
0.00	0.50	5	6.3% / 3.8%	5.7% / 4.0%	4.4% / 3.5%	4.2% / 3.7%	3.7% / 4.0%
0.00	0.50	10	7.5% / 4.0%	6.2% / 4.4%	6.2% / 3.8%	5.5% / 4.1%	4.4% / 5.1%
0.00	0.50	20	6.7% / 3.4%	6.0% / 3.9%	6.3% / 2.4%	5.2% / 3.4%	4.1% / 4.6%
0.00	0.50	50	9.5% / 2.3%	7.3% / 3.3%	8.0% / 1.6%	6.3% / 2.5%	4.0% / 3.9%
0.00	0.67	5	5.6% / 4.3%	5.4% / 4.3%	4.8% / 3.1%	4.7% / 3.1%	4.1% / 2.9%
0.00	0.67	10	4.9% / 4.0%	4.8% / 4.3%	5.4% / 4.0%	5.2% / 4.2%	5.1% / 4.2%
0.00	0.67	20	6.0% / 4.6%	5.1% / 5.0%	3.7% / 3.6%	3.1% / 3.8%	2.7% / 4.0%
0.00	0.67	50	6.9% / 3.8%	5.7% / 4.5%	5.9% / 3.1%	5.0% / 4.1%	3.3% / 4.8%
0.00	1.00	5	4.4% / 4.4%	4.4% / 4.4%	4.5% / 3.8%	4.4% / 3.6%	4.3% / 3.6%
0.00	1.00	10	6.2% / 5.8%	6.1% / 5.8%	5.3% / 4.9%	5.3% / 4.9%	5.0% / 4.6%
0.00	1.00	20	5.0% / 5.4%	5.0% / 5.4%	4.2% / 4.4%	4.2% / 4.4%	3.9% / 4.4%
0.00	1.00	50	4.4% / 5.0%	4.5% / 5.0%	4.3% / 4.1%	4.3% / 4.0%	4.1% / 4.0%
0.05	0.25	5	8.8% / 2.7%	7.0% / 3.0%	5.6% / 2.0%	4.6% / 2.6%	3.5% / 2.9%
0.05	0.25	10	9.7% / 2.0%	7.6% / 2.4%	8.8% / 2.1%	8.1% / 2.7%	4.0% / 4.4%
0.05	0.25	20	13.6% / 1.4%	10.8% / 2.1%	9.6% / 0.9%	7.0% / 1.1%	2.2% / 4.6%
0.05	0.25	50	**22.8%** / 0.5%	16.9% / 0.9%	18.3% / 0.3%	11.5% / 0.7%	1.5% / 5.8%
0.05	0.50	5	5.8% / 3.3%	5.6% / 3.6%	3.3% / 2.1%	3.0% / 2.2%	2.2% / 2.7%
0.05	0.50	10	6.9% / 3.8%	6.1% / 4.4%	6.0% / 2.8%	5.4% / 3.1%	3.8% / 4.4%
0.05	0.50	20	6.9% / 3.1%	5.6% / 3.7%	4.7% / 2.6%	4.2% / 2.8%	2.6% / 3.9%
0.05	0.50	50	10.3% / 1.7%	9.0% / 2.1%	8.6% / 1.4%	7.5% / 1.8%	3.5% / 4.5%
0.05	0.67	5	4.4% / 3.9%	4.2% / 3.9%	2.6% / 3.1%	2.5% / 3.1%	2.0% / 3.5%
0.05	0.67	10	4.6% / 3.2%	4.6% / 3.3%	3.6% / 3.2%	3.5% / 3.6%	2.5% / 3.9%
0.05	0.67	20	4.6% / 3.6%	4.2% / 3.8%	3.6% / 1.9%	2.9% / 2.0%	2.1% / 3.3%
0.05	0.67	50	8.2% / 3.2%	7.8% / 3.4%	5.5% / 1.9%	4.9% / 2.1%	3.5% / 3.0%
0.05	1.00	5	4.5% / 4.2%	4.5% / 4.2%	3.5% / 2.9%	3.5% / 2.9%	3.2% / 2.6%
0.05	1.00	10	3.9% / 4.6%	3.8% / 4.6%	3.1% / 4.2%	3.0% / 4.2%	3.2% / 4.5%
0.05	1.00	20	4.4% / 4.3%	4.4% / 4.3%	3.3% / 3.0%	3.3% / 3.0%	3.3% / 2.6%
0.05	1.00	50	5.3% / 4.9%	5.4% / 4.8%	4.2% / 3.4%	4.1% / 3.3%	4.0% / 3.3%
0.25	0.25	5	9.4% / 2.0%	8.6% / 2.4%	5.0% / 1.3%	4.5% / 1.4%	2.3% / 3.7%
0.25	0.25	10	10.6% / 1.8%	9.6% / 2.1%	8.0% / 1.0%	6.5% / 1.5%	1.4% / 5.0%
0.25	0.25	20	15.3% / 0.5%	12.5% / 0.9%	8.4% / 0.7%	6.7% / 0.9%	0.6% / 5.1%
0.25	0.25	50	**28.1%** / 0.1%	**22.9%** / 0.5%	18.4% / 0.1%	13.4% / 0.5%	0.1% / 10.6%
0.25	0.50	5	6.0% / 2.8%	6.0% / 2.9%	3.3% / 1.8%	3.1% / 1.9%	3.0% / 3.1%
0.25	0.50	10	6.8% / 3.5%	6.4% / 3.9%	5.2% / 3.0%	5.0% / 3.0%	3.3% / 5.2%
0.25	0.50	20	7.4% / 2.3%	6.9% / 2.6%	4.5% / 1.4%	4.1% / 1.4%	1.7% / 3.7%
0.25	0.50	50	11.4% / 1.5%	10.0% / 1.9%	5.8% / 0.8%	5.1% / 0.9%	1.0% / 4.2%
0.25	0.67	5	5.7% / 4.5%	5.7% / 4.7%	2.8% / 2.5%	2.8% / 2.7%	2.5% / 3.3%
0.25	0.67	10	5.1% / 3.3%	4.8% / 3.5%	2.3% / 2.1%	2.0% / 2.1%	1.7% / 2.4%
0.25	0.67	20	7.3% / 3.1%	7.2% / 3.5%	3.5% / 2.0%	3.5% / 2.1%	2.6% / 3.5%
0.25	0.67	50	8.1% / 2.1%	7.8% / 2.4%	4.7% / 1.1%	4.4% / 1.2%	1.9% / 2.9%
0.25	1.00	5	5.8% / 5.4%	5.7% / 5.4%	3.3% / 3.4%	3.3% / 3.4%	3.5% / 3.2%
0.25	1.00	10	3.8% / 4.8%	3.9% / 4.9%	2.7% / 3.8%	2.7% / 3.7%	2.8% / 3.6%
0.25	1.00	20	5.2% / 4.3%	5.2% / 4.3%	2.2% / 1.8%	2.1% / 1.8%	2.2% / 1.9%
0.25	1.00	50	3.7% / 4.0%	3.7% / 4.1%	2.6% / 3.2%	2.6% / 3.2%	1.9% / 3.2%

Table A.17: Simulation II, Scenario B: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of one-sided test for bias (lower / upper tail); odds ratio ψ as measure of treatment effect; control event rate equals 0.3; nominal significance level is 5% for each tail; error rates above 20% in bold font.

τ^2	ϕ	k	Linear regression test "add selective"	Linear regression test "add all"	Rank correlation test "add selective"	Rank correlation test "add all"	New test
0.00	0.25	5	13.3% / 1.2%	10.2% / 2.2%	7.9% / 1.6%	7.0% / 2.2%	5.4% / 3.5%
0.00	0.25	10	18.9% / 0.8%	12.8% / 1.2%	13.0% / 1.3%	8.7% / 2.1%	4.8% / 4.7%
0.00	0.25	20	**32.3%** / 0.4%	19.9% / 0.9%	**21.6%** / 0.6%	12.9% / 1.0%	4.4% / 3.8%
0.00	0.25	50	**62.0%** / 0.0%	**34.3%** / 0.2%	**46.7%** / 0.0%	**20.8%** / 0.2%	4.0% / 3.5%
0.00	0.50	5	8.5% / 3.4%	6.9% / 4.1%	6.0% / 2.7%	5.0% / 2.9%	4.0% / 3.3%
0.00	0.50	10	10.9% / 2.6%	8.8% / 3.2%	8.4% / 2.5%	6.6% / 3.1%	4.8% / 4.3%
0.00	0.50	20	15.3% / 1.3%	9.9% / 2.3%	11.5% / 1.3%	7.6% / 1.9%	4.2% / 3.5%
0.00	0.50	50	**23.9%** / 0.5%	13.6% / 0.9%	**20.8%** / 0.3%	11.5% / 0.5%	3.3% / 3.9%
0.00	0.67	5	6.8% / 3.7%	6.4% / 3.9%	5.0% / 2.3%	4.6% / 2.7%	3.0% / 3.2%
0.00	0.67	10	8.7% / 4.0%	7.5% / 4.9%	6.7% / 3.6%	5.2% / 4.2%	4.5% / 4.6%
0.00	0.67	20	9.8% / 3.2%	7.9% / 3.9%	7.7% / 2.8%	5.5% / 3.3%	4.9% / 4.4%
0.00	0.67	50	14.2% / 1.6%	11.2% / 1.9%	12.0% / 1.0%	9.4% / 1.7%	4.9% / 3.0%
0.00	1.00	5	6.3% / 4.7%	6.3% / 4.6%	4.6% / 4.0%	4.6% / 4.1%	4.1% / 3.7%
0.00	1.00	10	4.9% / 6.4%	4.5% / 6.2%	4.7% / 4.5%	4.3% / 4.5%	4.4% / 4.5%
0.00	1.00	20	5.5% / 4.9%	5.3% / 4.9%	3.5% / 3.9%	3.5% / 3.8%	4.2% / 3.7%
0.00	1.00	50	4.8% / 4.8%	4.8% / 4.7%	4.5% / 4.2%	4.4% / 3.8%	3.7% / 3.9%
0.05	0.25	5	12.0% / 1.4%	8.7% / 1.9%	6.5% / 0.8%	5.5% / 1.2%	3.2% / 2.7%
0.05	0.25	10	17.0% / 0.8%	11.9% / 1.5%	10.1% / 1.0%	8.1% / 2.3%	3.0% / 5.3%
0.05	0.25	20	**29.8%** / 0.2%	19.7% / 0.6%	18.7% / 0.3%	11.0% / 0.7%	2.7% / 4.4%
0.05	0.25	50	**60.5%** / 0.0%	**35.4%** / 0.1%	**42.9%** / 0.0%	**21.6%** / 0.0%	2.1% / 5.7%
0.05	0.50	5	8.7% / 2.6%	7.4% / 3.0%	5.0% / 2.1%	4.7% / 2.4%	2.9% / 3.1%
0.05	0.50	10	10.9% / 1.6%	8.4% / 2.1%	8.8% / 1.4%	7.3% / 2.2%	4.7% / 3.7%
0.05	0.50	20	16.0% / 1.4%	12.3% / 2.5%	11.2% / 0.9%	7.9% / 1.4%	2.7% / 3.9%
0.05	0.50	50	**25.0%** / 0.2%	15.8% / 1.3%	19.1% / 0.6%	12.2% / 1.3%	3.1% / 4.2%
0.05	0.67	5	6.5% / 3.0%	5.7% / 3.2%	3.5% / 2.5%	3.1% / 2.7%	2.9% / 2.5%
0.05	0.67	10	7.3% / 3.0%	6.3% / 3.6%	6.1% / 3.6%	5.3% / 3.8%	4.0% / 4.5%
0.05	0.67	20	10.9% / 2.7%	8.8% / 3.7%	8.0% / 2.1%	6.2% / 3.1%	3.3% / 3.7%
0.05	0.67	50	13.1% / 0.8%	9.5% / 1.5%	8.8% / 0.5%	7.4% / 0.8%	3.2% / 2.9%
0.05	1.00	5	5.1% / 4.3%	5.1% / 4.3%	3.1% / 3.5%	3.0% / 3.6%	3.0% / 3.0%
0.05	1.00	10	4.3% / 5.2%	4.3% / 5.4%	3.5% / 3.9%	3.4% / 3.9%	4.1% / 4.0%
0.05	1.00	20	4.4% / 5.1%	4.4% / 5.1%	2.8% / 3.9%	2.5% / 4.0%	2.9% / 3.5%
0.05	1.00	50	3.8% / 5.8%	3.8% / 5.6%	3.0% / 4.1%	3.1% / 4.1%	2.5% / 3.6%
0.25	0.25	5	11.0% / 0.7%	8.8% / 1.7%	5.9% / 0.8%	5.1% / 1.4%	2.0% / 5.0%
0.25	0.25	10	18.4% / 0.6%	12.8% / 0.8%	10.4% / 0.5%	8.2% / 0.6%	1.6% / 5.7%
0.25	0.25	20	**32.8%** / 0.1%	**25.0%** / 0.2%	16.9% / 0.3%	9.4% / 0.3%	0.4% / 9.3%
0.25	0.25	50	**66.3%** / 0.0%	**48.3%** / 0.0%	**39.8%** / 0.0%	**24.5%** / 0.0%	0.0% / 19.0%
0.25	0.50	5	6.9% / 2.8%	6.3% / 3.5%	5.3% / 1.8%	4.9% / 2.1%	2.9% / 3.7%
0.25	0.50	10	10.4% / 2.0%	9.1% / 2.7%	7.4% / 2.0%	6.3% / 2.4%	2.8% / 4.6%
0.25	0.50	20	14.7% / 0.8%	12.1% / 1.1%	9.0% / 0.5%	7.6% / 0.6%	1.0% / 4.0%
0.25	0.50	50	**24.2%** / 0.4%	17.9% / 0.7%	15.7% / 0.4%	11.1% / 0.9%	0.4% / 8.7%
0.25	0.67	5	6.8% / 3.9%	6.5% / 3.9%	3.1% / 2.1%	3.1% / 2.3%	2.8% / 3.3%
0.25	0.67	10	6.1% / 2.6%	5.6% / 2.9%	5.3% / 2.3%	4.6% / 2.4%	2.2% / 4.2%
0.25	0.67	20	7.8% / 2.0%	6.7% / 2.1%	5.5% / 1.1%	5.2% / 1.3%	0.8% / 3.2%
0.25	0.67	50	12.8% / 1.1%	11.3% / 1.4%	8.4% / 1.2%	6.8% / 1.5%	0.9% / 5.8%
0.25	1.00	5	4.7% / 4.1%	4.5% / 3.8%	2.8% / 3.1%	2.8% / 3.1%	3.0% / 2.7%
0.25	1.00	10	4.7% / 5.3%	4.5% / 5.5%	2.8% / 3.2%	2.9% / 3.3%	3.7% / 3.6%
0.25	1.00	20	4.0% / 3.7%	4.0% / 3.9%	2.0% / 2.6%	2.1% / 2.8%	2.3% / 2.2%
0.25	1.00	50	3.5% / 4.4%	3.7% / 4.5%	2.6% / 3.3%	2.7% / 3.4%	2.7% / 3.1%

Table A.18: Simulation II, Scenario B: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of one-sided test for bias (lower / upper tail); risk ratio ϕ as measure of treatment effect; control event rate equals 0.1; nominal significance level is 5% for each tail; error rates above 20% in bold font.

τ^2	ϕ	k	Linear regression test "add selective"	Linear regression test "add all"	Rank correlation test "add selective"	Rank correlation test "add all"	New test
0.00	0.25	5	9.5% / 1.3%	7.2% / 1.8%	7.3% / 1.1%	6.0% / 1.7%	3.2% / 3.1%
0.00	0.25	10	12.9% / 1.1%	7.5% / 2.2%	10.7% / 1.0%	7.8% / 1.6%	3.4% / 3.4%
0.00	0.25	20	**20.6%** / 0.6%	13.4% / 1.3%	17.0% / 0.5%	10.5% / 0.7%	4.5% / 5.7%
0.00	0.25	50	**41.6%** / 0.1%	**24.2%** / 0.2%	**37.0%** / 0.1%	**21.9%** / 0.4%	5.3% / 4.0%
0.00	0.50	5	7.8% / 4.5%	6.6% / 4.6%	5.3% / 3.0%	5.1% / 3.4%	4.0% / 3.9%
0.00	0.50	10	8.1% / 2.2%	6.9% / 2.9%	8.8% / 2.7%	7.1% / 2.9%	4.5% / 4.3%
0.00	0.50	20	12.2% / 1.8%	10.0% / 2.7%	10.3% / 1.6%	8.1% / 2.2%	4.7% / 2.7%
0.00	0.50	50	18.6% / 1.0%	12.5% / 1.7%	16.9% / 0.7%	10.3% / 1.5%	4.6% / 4.6%
0.00	0.67	5	6.6% / 3.8%	5.9% / 4.0%	4.6% / 2.0%	4.5% / 2.2%	3.7% / 2.9%
0.00	0.67	10	6.4% / 3.1%	5.7% / 3.7%	6.0% / 2.8%	5.4% / 2.9%	4.5% / 3.3%
0.00	0.67	20	8.3% / 3.4%	7.4% / 3.8%	7.9% / 2.5%	7.0% / 2.7%	5.2% / 3.9%
0.00	0.67	50	10.7% / 1.9%	8.1% / 2.3%	10.3% / 1.7%	8.4% / 2.2%	3.8% / 3.6%
0.00	1.00	5	4.4% / 4.4%	4.4% / 4.4%	3.6% / 3.2%	3.6% / 3.2%	3.5% / 3.4%
0.00	1.00	10	4.0% / 6.4%	4.0% / 6.4%	4.9% / 5.0%	5.0% / 4.9%	4.6% / 5.0%
0.00	1.00	20	5.5% / 4.7%	5.5% / 4.8%	4.3% / 4.4%	4.4% / 4.3%	3.7% / 4.4%
0.00	1.00	50	5.0% / 4.3%	4.9% / 4.2%	4.8% / 4.3%	4.7% / 4.3%	4.0% / 4.2%
0.05	0.25	5	10.2% / 1.3%	8.6% / 1.6%	7.4% / 1.2%	6.2% / 1.6%	3.3% / 4.1%
0.05	0.25	10	15.0% / 1.1%	11.3% / 1.6%	11.7% / 0.9%	8.8% / 1.3%	3.7% / 5.1%
0.05	0.25	20	**22.2%** / 0.2%	15.0% / 0.6%	14.9% / 0.3%	9.6% / 0.6%	1.4% / 4.4%
0.05	0.25	50	**42.4%** / 0.1%	**27.9%** / 0.4%	**33.0%** / 0.2%	**20.8%** / 0.2%	0.7% / 7.7%
0.05	0.50	5	7.4% / 2.9%	6.9% / 3.1%	4.1% / 1.8%	4.0% / 1.8%	3.2% / 3.0%
0.05	0.50	10	8.5% / 2.4%	7.2% / 3.0%	6.0% / 2.8%	5.4% / 3.0%	2.8% / 3.8%
0.05	0.50	20	12.1% / 1.6%	10.1% / 2.0%	7.5% / 1.4%	6.9% / 1.5%	2.3% / 4.2%
0.05	0.50	50	17.4% / 0.9%	13.9% / 1.1%	13.6% / 0.8%	10.2% / 0.9%	1.3% / 4.1%
0.05	0.67	5	6.0% / 3.7%	6.0% / 3.9%	4.4% / 2.3%	3.9% / 2.7%	3.4% / 2.9%
0.05	0.67	10	5.6% / 2.9%	5.1% / 3.0%	4.8% / 2.3%	4.5% / 2.4%	3.0% / 4.1%
0.05	0.67	20	7.7% / 2.6%	7.0% / 2.9%	5.1% / 1.5%	4.6% / 1.7%	2.3% / 3.3%
0.05	0.67	50	11.9% / 1.7%	10.7% / 2.0%	7.5% / 1.3%	6.6% / 1.7%	2.7% / 3.6%
0.05	1.00	5	5.1% / 5.4%	5.1% / 5.4%	3.7% / 3.7%	3.6% / 3.8%	3.6% / 3.7%
0.05	1.00	10	5.2% / 5.7%	5.3% / 5.8%	3.3% / 4.9%	3.3% / 4.9%	3.7% / 4.5%
0.05	1.00	20	3.4% / 6.4%	3.4% / 6.4%	2.6% / 3.8%	2.6% / 3.8%	2.4% / 3.0%
0.05	1.00	50	3.7% / 4.1%	3.9% / 4.1%	2.6% / 2.5%	2.7% / 2.5%	2.3% / 2.1%
0.25	0.25	5	8.7% / 1.3%	7.6% / 1.5%	4.9% / 1.2%	4.3% / 1.4%	1.3% / 4.8%
0.25	0.25	10	12.7% / 0.9%	11.1% / 1.2%	9.0% / 1.3%	6.9% / 1.6%	0.9% / 8.4%
0.25	0.25	20	**25.7%** / 0.5%	**21.4%** / 0.6%	15.1% / 0.1%	12.8% / 0.2%	0.2% / 12.8%
0.25	0.25	50	**54.2%** / 0.0%	**46.3%** / 0.0%	**34.8%** / 0.0%	**26.8%** / 0.0%	0.0% / **22.4%**
0.25	0.50	5	7.0% / 3.1%	6.8% / 3.1%	4.7% / 2.4%	4.3% / 2.6%	2.7% / 4.4%
0.25	0.50	10	9.5% / 1.9%	9.0% / 2.2%	5.7% / 1.3%	5.6% / 1.4%	1.8% / 4.6%
0.25	0.50	20	11.4% / 1.2%	10.5% / 1.3%	7.0% / 0.7%	6.5% / 0.9%	0.9% / 4.7%
0.25	0.50	50	**23.7%** / 0.4%	**21.5%** / 0.5%	15.5% / 0.1%	14.2% / 0.5%	0.5% / 8.9%
0.25	0.67	5	6.1% / 3.5%	6.1% / 3.8%	4.7% / 1.9%	4.5% / 2.0%	3.8% / 3.2%
0.25	0.67	10	6.3% / 3.4%	6.3% / 3.5%	4.5% / 2.5%	4.4% / 2.6%	2.3% / 5.0%
0.25	0.67	20	8.0% / 2.3%	7.5% / 2.6%	4.6% / 1.5%	4.3% / 1.7%	1.1% / 4.3%
0.25	0.67	50	12.2% / 0.4%	12.0% / 0.6%	7.7% / 0.6%	7.3% / 0.6%	0.9% / 4.0%
0.25	1.00	5	4.1% / 4.3%	4.0% / 4.4%	2.8% / 2.9%	2.7% / 3.0%	2.6% / 3.0%
0.25	1.00	10	4.2% / 3.7%	4.2% / 3.8%	2.9% / 2.3%	2.9% / 2.3%	2.5% / 2.3%
0.25	1.00	20	3.6% / 5.0%	3.7% / 5.0%	2.6% / 3.7%	2.6% / 3.8%	2.3% / 3.7%
0.25	1.00	50	3.3% / 2.2%	3.4% / 2.2%	1.8% / 1.6%	1.7% / 1.6%	1.3% / 1.8%

Table A.19: Simulation II, Scenario B: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of one-sided test for bias (lower / upper tail); risk ratio ϕ as measure of treatment effect; control event rate equals 0.3; nominal significance level is 5% for each tail; error rates above 20% in bold font.

τ^2	ψ / ϕ	k	$\widehat{\psi}_{MH}$ ($p_c = 0.1$) 50% [5%; 95%]	$\widehat{\psi}_{MH}$ ($p_c = 0.3$) 50% [5%; 95%]	$\widehat{\phi}_{MH}$ ($p_c = 0.1$) 50% [5%; 95%]	$\widehat{\phi}_{MH}$ ($p_c = 0.3$) 50% [5%; 95%]
0.00	0.25	5	0.25 [0.21; 0.30]	0.25 [0.22; 0.28]	0.25 [0.21; 0.29]	0.25 [0.22; 0.28]
0.00	0.25	10	0.25 [0.22; 0.28]	0.25 [0.23; 0.27]	0.25 [0.22; 0.28]	0.25 [0.23; 0.27]
0.00	0.25	20	0.25 [0.23; 0.27]	0.25 [0.24; 0.26]	0.25 [0.23; 0.27]	0.25 [0.24; 0.26]
0.00	0.25	50	0.25 [0.24; 0.26]	0.25 [0.24; 0.26]	0.25 [0.24; 0.26]	0.25 [0.24; 0.26]
0.00	0.50	5	0.50 [0.43; 0.58]	0.50 [0.46; 0.54]	0.50 [0.43; 0.58]	0.50 [0.46; 0.54]
0.00	0.50	10	0.50 [0.45; 0.55]	0.50 [0.47; 0.53]	0.50 [0.46; 0.54]	0.50 [0.48; 0.52]
0.00	0.50	20	0.50 [0.47; 0.54]	0.50 [0.48; 0.52]	0.50 [0.47; 0.53]	0.50 [0.48; 0.52]
0.00	0.50	50	0.50 [0.48; 0.52]	0.50 [0.49; 0.51]	0.50 [0.48; 0.52]	0.50 [0.49; 0.51]
0.00	0.67	5	0.67 [0.58; 0.77]	0.67 [0.62; 0.72]	0.67 [0.59; 0.76]	0.67 [0.63; 0.71]
0.00	0.67	10	0.67 [0.62; 0.73]	0.67 [0.63; 0.71]	0.67 [0.62; 0.73]	0.67 [0.64; 0.70]
0.00	0.67	20	0.67 [0.63; 0.71]	0.67 [0.64; 0.69]	0.67 [0.64; 0.71]	0.67 [0.65; 0.69]
0.00	0.67	50	0.67 [0.65; 0.69]	0.67 [0.66; 0.69]	0.67 [0.65; 0.69]	0.67 [0.66; 0.68]
0.00	1.00	5	1.00 [0.88; 1.15]	1.00 [0.93; 1.08]	1.01 [0.88; 1.12]	1.00 [0.94; 1.06]
0.00	1.00	10	1.00 [0.93; 1.08]	1.00 [0.95; 1.05]	1.00 [0.93; 1.07]	1.00 [0.96; 1.04]
0.00	1.00	20	1.00 [0.95; 1.06]	1.00 [0.97; 1.03]	1.00 [0.96; 1.05]	1.00 [0.97; 1.03]
0.00	1.00	50	1.00 [0.97; 1.04]	1.00 [0.98; 1.02]	1.00 [0.97; 1.03]	1.00 [0.99; 1.02]
0.05	0.25	5	0.25 [0.18; 0.33]	0.25 [0.19; 0.32]	0.25 [0.18; 0.33]	0.25 [0.19; 0.32]
0.05	0.25	10	0.25 [0.20; 0.30]	0.25 [0.21; 0.30]	0.25 [0.20; 0.31]	0.25 [0.21; 0.30]
0.05	0.25	20	0.25 [0.22; 0.29]	0.25 [0.22; 0.29]	0.25 [0.22; 0.29]	0.25 [0.22; 0.29]
0.05	0.25	50	0.25 [0.23; 0.28]	0.25 [0.23; 0.27]	0.25 [0.23; 0.28]	0.25 [0.23; 0.27]
0.05	0.50	5	0.50 [0.38; 0.67]	0.50 [0.39; 0.63]	0.50 [0.38; 0.67]	0.50 [0.39; 0.64]
0.05	0.50	10	0.50 [0.41; 0.61]	0.50 [0.42; 0.60]	0.50 [0.41; 0.61]	0.50 [0.42; 0.59]
0.05	0.50	20	0.50 [0.44; 0.57]	0.50 [0.44; 0.57]	0.50 [0.44; 0.58]	0.50 [0.44; 0.57]
0.05	0.50	50	0.50 [0.46; 0.55]	0.50 [0.46; 0.54]	0.50 [0.46; 0.55]	0.50 [0.46; 0.54]
0.05	0.67	5	0.66 [0.51; 0.87]	0.66 [0.53; 0.85]	0.67 [0.51; 0.86]	0.67 [0.52; 0.83]
0.05	0.67	10	0.67 [0.56; 0.81]	0.67 [0.57; 0.80]	0.67 [0.56; 0.81]	0.67 [0.56; 0.80]
0.05	0.67	20	0.67 [0.58; 0.77]	0.67 [0.59; 0.77]	0.67 [0.58; 0.77]	0.67 [0.59; 0.76]
0.05	0.67	50	0.67 [0.62; 0.73]	0.67 [0.62; 0.73]	0.67 [0.62; 0.73]	0.67 [0.62; 0.73]
0.05	1.00	5	0.99 [0.77; 1.30]	1.00 [0.78; 1.30]	0.99 [0.77; 1.28]	0.99 [0.79; 1.27]
0.05	1.00	10	1.00 [0.82; 1.20]	1.00 [0.84; 1.20]	1.00 [0.83; 1.21]	1.00 [0.84; 1.20]
0.05	1.00	20	1.00 [0.88; 1.15]	1.00 [0.88; 1.15]	1.00 [0.88; 1.14]	1.00 [0.88; 1.12]
0.05	1.00	50	1.00 [0.92; 1.08]	1.01 [0.93; 1.09]	1.00 [0.92; 1.09]	1.00 [0.92; 1.08]
0.25	0.25	5	0.25 [0.15; 0.42]	0.26 [0.15; 0.43]	0.25 [0.15; 0.42]	0.25 [0.14; 0.41]
0.25	0.25	10	0.25 [0.16; 0.37]	0.25 [0.18; 0.38]	0.25 [0.17; 0.37]	0.25 [0.17; 0.36]
0.25	0.25	20	0.25 [0.19; 0.34]	0.26 [0.20; 0.34]	0.25 [0.19; 0.33]	0.25 [0.19; 0.33]
0.25	0.25	50	0.25 [0.21; 0.31]	0.26 [0.22; 0.30]	0.25 [0.21; 0.30]	0.25 [0.21; 0.30]
0.25	0.50	5	0.50 [0.29; 0.85]	0.51 [0.30; 0.82]	0.50 [0.29; 0.84]	0.51 [0.30; 0.85]
0.25	0.50	10	0.49 [0.33; 0.76]	0.50 [0.34; 0.73]	0.50 [0.34; 0.75]	0.50 [0.34; 0.73]
0.25	0.50	20	0.51 [0.38; 0.68]	0.50 [0.39; 0.66]	0.49 [0.37; 0.67]	0.51 [0.38; 0.67]
0.25	0.50	50	0.51 [0.42; 0.60]	0.51 [0.43; 0.61]	0.50 [0.41; 0.60]	0.50 [0.42; 0.60]
0.25	0.67	5	0.67 [0.39; 1.11]	0.68 [0.41; 1.12]	0.67 [0.40; 1.14]	0.67 [0.40; 1.09]
0.25	0.67	10	0.67 [0.47; 1.00]	0.68 [0.46; 0.99]	0.68 [0.44; 1.00]	0.67 [0.46; 0.98]
0.25	0.67	20	0.67 [0.50; 0.89]	0.68 [0.51; 0.91]	0.67 [0.51; 0.88]	0.67 [0.50; 0.87]
0.25	0.67	50	0.67 [0.56; 0.81]	0.68 [0.57; 0.80]	0.67 [0.56; 0.80]	0.67 [0.55; 0.80]
0.25	1.00	5	0.99 [0.60; 1.65]	1.01 [0.60; 1.68]	0.99 [0.59; 1.67]	1.02 [0.60; 1.67]
0.25	1.00	10	1.00 [0.69; 1.50]	1.00 [0.68; 1.49]	1.01 [0.67; 1.47]	1.00 [0.67; 1.44]
0.25	1.00	20	1.00 [0.75; 1.29]	0.99 [0.76; 1.31]	1.00 [0.74; 1.32]	1.00 [0.75; 1.34]
0.25	1.00	50	1.00 [0.83; 1.21]	1.00 [0.83; 1.19]	0.99 [0.83; 1.20]	1.00 [0.83; 1.18]

Table A.20: Simulation II, Scenario C: results under null hypothesis ($n = 1000$); median, 5% and 95% quantiles of empirical distribution of estimated treatment effect based on Mantel-Haenszel method; both odds ratio ψ and risk ratio ϕ as measures of treatment effect.

τ^2	ψ	k	Linear regression test "add selective"	Linear regression test "add all"	Rank correlation test "add selective"	Rank correlation test "add all"	New test
0.00	0.25	5	9.0% / 2.9%	6.3% / 3.7%	5.2% / 1.8%	4.6% / 2.4%	3.5% / 2.3%
0.00	0.25	10	8.1% / 2.6%	4.6% / 4.8%	5.5% / 2.3%	3.4% / 3.5%	4.1% / 4.7%
0.00	0.25	20	10.9% / 2.0%	5.1% / 5.3%	6.4% / 1.3%	3.3% / 3.2%	4.8% / 2.9%
0.00	0.25	50	12.9% / 1.8%	3.2% / 5.8%	11.2% / 0.4%	3.2% / 2.2%	3.5% / 2.8%
0.00	0.50	5	6.7% / 4.0%	5.7% / 4.5%	4.7% / 3.1%	4.0% / 3.3%	4.4% / 3.2%
0.00	0.50	10	6.0% / 4.3%	5.0% / 5.0%	4.8% / 2.9%	3.7% / 3.5%	3.8% / 3.2%
0.00	0.50	20	6.5% / 3.5%	4.6% / 5.0%	3.9% / 1.7%	2.4% / 2.2%	2.6% / 2.6%
0.00	0.50	50	8.2% / 3.0%	4.2% / 4.6%	6.8% / 1.0%	3.4% / 2.3%	3.6% / 2.2%
0.00	0.67	5	4.8% / 3.2%	4.2% / 3.8%	3.7% / 2.5%	3.4% / 2.5%	3.2% / 2.2%
0.00	0.67	10	5.5% / 3.8%	4.9% / 4.5%	3.8% / 3.1%	3.4% / 3.6%	2.9% / 3.4%
0.00	0.67	20	7.0% / 4.8%	5.2% / 5.8%	5.6% / 3.2%	4.5% / 3.9%	4.8% / 3.7%
0.00	0.67	50	6.9% / 4.5%	4.5% / 5.9%	4.8% / 2.6%	3.7% / 3.4%	3.3% / 3.6%
0.00	1.00	5	5.3% / 5.0%	5.2% / 5.0%	4.2% / 2.7%	4.2% / 2.6%	3.8% / 2.3%
0.00	1.00	10	5.8% / 5.0%	5.7% / 4.6%	4.3% / 3.7%	4.4% / 3.6%	4.4% / 4.2%
0.00	1.00	20	5.3% / 5.7%	5.3% / 5.5%	2.9% / 4.2%	2.8% / 3.9%	2.5% / 3.8%
0.00	1.00	50	5.2% / 5.8%	5.2% / 5.5%	2.7% / 2.7%	2.6% / 2.5%	2.7% / 2.5%
0.05	0.25	5	7.4% / 3.6%	5.3% / 5.2%	3.0% / 2.6%	2.3% / 3.3%	2.2% / 3.6%
0.05	0.25	10	6.4% / 2.5%	4.2% / 3.7%	2.9% / 0.8%	2.3% / 1.9%	2.1% / 2.4%
0.05	0.25	20	8.1% / 1.2%	4.2% / 2.8%	3.4% / 0.4%	1.9% / 0.6%	1.3% / 1.3%
0.05	0.25	50	12.5% / 0.6%	5.0% / 2.7%	6.6% / 0.1%	2.6% / 0.5%	0.6% / 2.6%
0.05	0.50	5	6.1% / 3.8%	5.2% / 4.3%	4.3% / 2.0%	3.9% / 2.3%	3.3% / 2.5%
0.05	0.50	10	4.8% / 3.8%	4.1% / 4.8%	3.1% / 1.8%	2.4% / 2.1%	2.1% / 2.7%
0.05	0.50	20	7.1% / 2.1%	5.4% / 3.3%	2.7% / 0.7%	2.3% / 0.8%	1.5% / 1.3%
0.05	0.50	50	6.4% / 1.8%	4.3% / 3.4%	2.7% / 0.6%	1.7% / 1.2%	0.8% / 1.6%
0.05	0.67	5	4.9% / 3.9%	4.5% / 4.3%	2.4% / 2.5%	2.4% / 2.5%	1.6% / 2.8%
0.05	0.67	10	5.5% / 3.5%	5.1% / 3.9%	2.7% / 1.7%	2.7% / 1.9%	1.7% / 2.2%
0.05	0.67	20	5.1% / 1.4%	4.1% / 1.8%	2.2% / 1.2%	2.0% / 1.5%	1.3% / 1.8%
0.05	0.67	50	5.2% / 2.1%	4.1% / 2.7%	2.3% / 0.6%	1.8% / 0.7%	1.3% / 1.0%
0.05	1.00	5	4.0% / 5.0%	4.0% / 5.1%	1.9% / 3.7%	2.0% / 3.8%	2.1% / 3.4%
0.05	1.00	10	3.5% / 3.9%	3.6% / 3.9%	1.4% / 1.9%	1.4% / 1.7%	1.2% / 1.4%
0.05	1.00	20	3.1% / 4.3%	3.1% / 4.4%	1.0% / 1.4%	1.0% / 1.5%	1.1% / 1.3%
0.05	1.00	50	2.9% / 4.0%	2.9% / 4.0%	0.8% / 1.1%	0.8% / 1.1%	0.7% / 1.2%
0.25	0.25	5	6.6% / 3.7%	6.1% / 4.1%	3.1% / 1.0%	2.9% / 1.4%	1.8% / 3.9%
0.25	0.25	10	7.1% / 1.9%	6.1% / 2.5%	2.8% / 1.2%	2.3% / 1.6%	1.1% / 2.8%
0.25	0.25	20	9.7% / 1.0%	7.5% / 2.3%	3.2% / 0.4%	1.7% / 0.6%	0.5% / 3.5%
0.25	0.25	50	16.0% / 0.6%	9.7% / 1.4%	2.8% / 0.2%	1.2% / 0.5%	0.1% / 5.5%
0.25	0.50	5	4.9% / 3.4%	4.6% / 4.0%	2.1% / 1.4%	2.1% / 1.5%	1.4% / 2.3%
0.25	0.50	10	4.8% / 2.8%	4.5% / 3.5%	2.7% / 1.8%	2.6% / 1.9%	1.2% / 2.4%
0.25	0.50	20	4.7% / 1.6%	4.0% / 2.1%	1.0% / 0.9%	0.9% / 1.1%	0.2% / 1.8%
0.25	0.50	50	8.1% / 1.3%	6.8% / 2.0%	2.5% / 0.2%	1.8% / 0.5%	0.1% / 3.0%
0.25	0.67	5	4.1% / 3.4%	3.7% / 3.9%	1.4% / 2.3%	1.3% / 2.4%	0.9% / 2.7%
0.25	0.67	10	4.8% / 3.7%	4.7% / 3.9%	1.7% / 1.3%	1.5% / 1.4%	0.9% / 2.3%
0.25	0.67	20	5.1% / 2.9%	4.9% / 3.4%	1.3% / 1.2%	0.9% / 1.3%	0.7% / 1.8%
0.25	0.67	50	5.4% / 1.9%	4.9% / 2.2%	1.5% / 0.7%	1.2% / 0.9%	0.2% / 2.2%
0.25	1.00	5	3.9% / 4.4%	3.9% / 4.5%	1.6% / 2.1%	1.6% / 2.1%	1.7% / 2.0%
0.25	1.00	10	3.9% / 3.0%	4.0% / 3.0%	1.9% / 1.1%	1.9% / 1.1%	1.7% / 1.1%
0.25	1.00	20	3.2% / 4.2%	3.3% / 4.2%	1.6% / 0.9%	1.5% / 0.8%	1.6% / 0.6%
0.25	1.00	50	3.2% / 2.6%	3.3% / 2.8%	0.6% / 1.2%	0.6% / 1.3%	0.6% / 1.0%

Table A.21: Simulation II, Scenario C: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of one-sided test for bias (lower / upper tail); odds ratio ψ as measure of treatment effect; control event rate equals 0.1; nominal significance level is 5% for each tail; error rates above 20% in bold font.

τ^2	ψ	k	Linear regression test "add selective"	Linear regression test "add all"	Rank correlation test "add selective"	Rank correlation test "add all"	New test
0.00	0.25	5	7.0% / 4.5%	5.7% / 5.3%	4.1% / 2.4%	3.7% / 3.1%	3.6% / 3.0%
0.00	0.25	10	6.3% / 3.5%	5.1% / 5.1%	4.3% / 2.3%	3.5% / 3.0%	3.4% / 3.2%
0.00	0.25	20	8.7% / 3.4%	5.1% / 4.8%	5.6% / 1.8%	3.6% / 2.5%	3.4% / 3.1%
0.00	0.25	50	9.7% / 3.0%	5.0% / 4.9%	8.1% / 1.0%	4.0% / 2.5%	3.0% / 3.0%
0.00	0.50	5	6.7% / 4.2%	6.4% / 4.2%	5.2% / 3.2%	4.6% / 3.5%	4.6% / 2.8%
0.00	0.50	10	5.8% / 4.1%	5.6% / 4.7%	4.2% / 3.4%	3.7% / 3.6%	3.8% / 3.5%
0.00	0.50	20	5.2% / 3.5%	3.9% / 4.2%	3.5% / 2.3%	2.8% / 2.6%	3.0% / 2.7%
0.00	0.50	50	6.8% / 3.4%	5.3% / 4.1%	5.1% / 2.1%	3.8% / 2.5%	3.3% / 2.7%
0.00	0.67	5	3.6% / 5.2%	3.3% / 5.4%	2.3% / 3.3%	2.2% / 3.4%	2.1% / 3.1%
0.00	0.67	10	5.0% / 4.7%	4.5% / 4.9%	3.6% / 3.9%	3.4% / 3.9%	3.5% / 4.1%
0.00	0.67	20	5.4% / 4.8%	4.9% / 5.3%	3.9% / 2.1%	3.2% / 2.3%	3.5% / 2.3%
0.00	0.67	50	7.0% / 5.2%	5.5% / 5.6%	4.9% / 3.4%	3.5% / 3.8%	3.6% / 4.1%
0.00	1.00	5	5.2% / 5.5%	5.2% / 5.5%	2.9% / 4.3%	2.8% / 4.2%	3.0% / 4.3%
0.00	1.00	10	4.7% / 5.0%	4.7% / 4.8%	2.8% / 3.6%	2.8% / 3.6%	2.9% / 3.3%
0.00	1.00	20	4.1% / 5.0%	4.1% / 5.0%	2.4% / 2.8%	2.4% / 2.8%	2.5% / 2.8%
0.00	1.00	50	5.9% / 4.4%	5.9% / 4.3%	3.5% / 2.5%	3.3% / 2.5%	3.5% / 2.5%
0.05	0.25	5	4.5% / 4.0%	3.8% / 4.6%	2.5% / 1.9%	2.2% / 2.2%	2.1% / 2.9%
0.05	0.25	10	6.1% / 2.9%	5.2% / 3.9%	2.7% / 1.2%	2.6% / 1.5%	1.4% / 1.9%
0.05	0.25	20	7.5% / 2.2%	6.0% / 3.1%	2.7% / 0.8%	2.0% / 1.3%	1.1% / 2.2%
0.05	0.25	50	8.0% / 1.3%	5.5% / 2.4%	2.7% / 0.8%	1.2% / 1.1%	0.5% / 2.2%
0.05	0.50	5	4.5% / 4.2%	4.4% / 4.5%	2.7% / 1.4%	2.5% / 1.4%	2.0% / 1.5%
0.05	0.50	10	5.0% / 3.7%	4.6% / 3.9%	2.4% / 1.4%	2.2% / 1.8%	1.9% / 1.9%
0.05	0.50	20	5.1% / 3.3%	4.8% / 3.5%	1.4% / 0.6%	1.4% / 0.6%	0.9% / 0.6%
0.05	0.50	50	5.9% / 2.0%	4.8% / 2.4%	1.6% / 0.9%	1.5% / 1.0%	0.7% / 1.6%
0.05	0.67	5	4.6% / 4.8%	4.6% / 4.8%	3.0% / 1.6%	2.9% / 1.6%	2.7% / 1.8%
0.05	0.67	10	3.8% / 4.2%	3.7% / 4.3%	1.6% / 1.7%	1.6% / 1.8%	1.4% / 1.7%
0.05	0.67	20	4.7% / 4.1%	4.6% / 4.3%	0.9% / 0.8%	0.8% / 0.9%	1.2% / 1.0%
0.05	0.67	50	3.9% / 2.9%	3.6% / 3.2%	1.3% / 0.6%	1.2% / 0.8%	0.8% / 1.2%
0.05	1.00	5	3.9% / 4.8%	3.8% / 4.8%	2.1% / 1.5%	2.1% / 1.5%	1.9% / 1.5%
0.05	1.00	10	4.7% / 4.3%	4.7% / 4.3%	1.5% / 2.2%	1.5% / 2.1%	1.5% / 2.3%
0.05	1.00	20	3.8% / 3.7%	3.8% / 3.8%	1.3% / 1.5%	1.3% / 1.5%	1.1% / 1.5%
0.05	1.00	50	3.3% / 3.0%	3.3% / 3.0%	1.3% / 0.9%	1.3% / 0.9%	1.3% / 0.9%
0.25	0.25	5	5.1% / 3.9%	4.6% / 4.2%	2.1% / 1.4%	2.0% / 1.4%	1.3% / 2.2%
0.25	0.25	10	7.2% / 2.7%	6.5% / 3.1%	2.1% / 0.4%	1.9% / 0.4%	1.2% / 2.0%
0.25	0.25	20	7.7% / 1.2%	6.9% / 1.8%	2.0% / 0.1%	1.6% / 0.2%	0.3% / 1.8%
0.25	0.25	50	11.1% / 0.4%	9.5% / 0.7%	1.1% / 0.4%	0.8% / 0.5%	0.0% / 4.9%
0.25	0.50	5	5.9% / 4.2%	5.7% / 4.5%	1.8% / 1.3%	1.6% / 1.3%	1.7% / 1.4%
0.25	0.50	10	4.8% / 4.4%	4.8% / 4.5%	1.6% / 1.8%	1.6% / 2.1%	1.0% / 2.4%
0.25	0.50	20	5.5% / 2.9%	5.4% / 3.0%	1.0% / 0.6%	1.0% / 0.8%	0.2% / 1.4%
0.25	0.50	50	6.1% / 2.2%	5.6% / 2.3%	0.4% / 0.5%	0.3% / 0.8%	0.1% / 2.0%
0.25	0.67	5	5.2% / 3.4%	5.2% / 3.5%	2.3% / 1.4%	2.3% / 1.4%	1.6% / 1.8%
0.25	0.67	10	3.7% / 3.1%	3.8% / 3.1%	1.5% / 1.5%	1.4% / 1.6%	1.0% / 1.7%
0.25	0.67	20	5.3% / 2.3%	5.3% / 2.3%	1.7% / 0.7%	1.7% / 0.8%	1.3% / 0.9%
0.25	0.67	50	4.5% / 2.2%	4.4% / 2.3%	1.0% / 0.5%	1.0% / 0.5%	0.8% / 1.2%
0.25	1.00	5	4.1% / 4.6%	4.2% / 4.5%	1.7% / 1.6%	1.7% / 1.6%	1.3% / 1.3%
0.25	1.00	10	4.0% / 5.2%	4.1% / 5.2%	1.4% / 1.5%	1.4% / 1.5%	1.2% / 1.3%
0.25	1.00	20	2.9% / 3.3%	3.0% / 3.4%	0.7% / 0.5%	0.7% / 0.4%	0.8% / 0.3%
0.25	1.00	50	3.8% / 3.4%	3.8% / 3.4%	0.5% / 0.2%	0.6% / 0.2%	0.4% / 0.2%

Table A.22: Simulation II, Scenario C: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of one-sided test for bias (lower / upper tail); odds ratio ψ as measure of treatment effect; control event rate equals 0.3; nominal significance level is 5% for each tail; error rates above 20% in bold font.

τ^2	ϕ	k	Linear regression test "add selective"	"add all"	Rank correlation test "add selective"	"add all"	New test
0.00	0.25	5	7.1% / 3.4%	5.6% / 4.8%	4.7% / 2.0%	3.3% / 3.0%	3.9% / 3.4%
0.00	0.25	10	7.6% / 2.0%	4.0% / 4.4%	5.4% / 1.6%	2.7% / 3.9%	3.1% / 3.7%
0.00	0.25	20	11.2% / 1.9%	4.2% / 5.0%	6.8% / 0.8%	2.9% / 3.3%	3.7% / 3.0%
0.00	0.25	50	13.8% / 1.7%	3.6% / 7.8%	11.1% / 0.4%	2.3% / 2.9%	3.9% / 3.8%
0.00	0.50	5	6.6% / 4.0%	5.1% / 4.8%	4.8% / 2.9%	3.6% / 3.4%	4.0% / 3.5%
0.00	0.50	10	7.9% / 4.6%	5.7% / 5.5%	6.2% / 3.1%	4.7% / 4.0%	4.9% / 3.8%
0.00	0.50	20	6.5% / 2.4%	4.4% / 4.6%	5.6% / 1.3%	3.2% / 2.3%	3.3% / 3.0%
0.00	0.50	50	8.8% / 3.0%	4.3% / 5.0%	7.4% / 1.1%	3.7% / 2.4%	3.8% / 2.7%
0.00	0.67	5	5.0% / 5.2%	4.4% / 5.3%	3.1% / 2.9%	2.9% / 3.1%	2.8% / 3.0%
0.00	0.67	10	7.2% / 4.4%	5.8% / 5.0%	5.6% / 3.6%	4.3% / 4.1%	4.4% / 3.7%
0.00	0.67	20	7.1% / 4.2%	5.5% / 5.4%	4.3% / 2.3%	3.1% / 2.9%	3.4% / 2.5%
0.00	0.67	50	7.3% / 4.4%	5.7% / 5.7%	6.6% / 2.7%	3.9% / 3.7%	4.1% / 3.7%
0.00	1.00	5	5.8% / 5.3%	5.7% / 5.1%	3.6% / 2.7%	3.6% / 2.8%	3.3% / 3.0%
0.00	1.00	10	4.8% / 5.6%	4.7% / 5.7%	3.5% / 4.3%	3.5% / 4.6%	3.7% / 4.0%
0.00	1.00	20	4.3% / 6.2%	4.2% / 6.1%	2.9% / 3.8%	2.9% / 3.6%	3.0% / 3.9%
0.00	1.00	50	3.8% / 4.9%	3.8% / 4.8%	3.2% / 3.4%	3.0% / 3.3%	2.7% / 2.9%
0.05	0.25	5	6.0% / 2.9%	4.7% / 3.3%	3.2% / 2.4%	2.4% / 2.9%	2.4% / 3.6%
0.05	0.25	10	7.4% / 1.2%	4.9% / 3.0%	3.1% / 1.1%	2.6% / 1.9%	1.7% / 3.4%
0.05	0.25	20	8.9% / 1.9%	4.1% / 3.4%	3.3% / 0.2%	1.2% / 0.5%	1.3% / 1.7%
0.05	0.25	50	13.4% / 0.4%	4.8% / 2.8%	6.6% / 0.0%	1.5% / 0.8%	0.7% / 2.5%
0.05	0.50	5	4.9% / 3.6%	4.4% / 4.2%	1.4% / 2.9%	1.2% / 3.0%	1.2% / 3.2%
0.05	0.50	10	6.2% / 3.1%	4.7% / 3.6%	3.7% / 1.6%	3.2% / 1.9%	2.5% / 2.7%
0.05	0.50	20	6.6% / 2.2%	4.9% / 2.7%	3.5% / 0.7%	2.5% / 0.9%	2.0% / 1.3%
0.05	0.50	50	7.1% / 1.5%	5.0% / 2.7%	3.3% / 0.5%	1.7% / 0.9%	1.2% / 1.6%
0.05	0.67	5	3.3% / 4.0%	3.2% / 4.4%	2.3% / 2.2%	2.0% / 2.4%	1.9% / 2.3%
0.05	0.67	10	5.0% / 3.4%	4.6% / 3.8%	2.7% / 1.9%	2.3% / 2.0%	2.1% / 2.4%
0.05	0.67	20	5.0% / 2.3%	4.4% / 2.8%	2.1% / 1.1%	2.0% / 1.3%	1.6% / 1.2%
0.05	0.67	50	5.9% / 2.5%	4.5% / 3.3%	1.7% / 1.3%	1.2% / 2.0%	0.7% / 2.6%
0.05	1.00	5	3.7% / 5.7%	3.8% / 5.7%	1.7% / 3.5%	1.7% / 3.6%	1.7% / 3.2%
0.05	1.00	10	3.8% / 2.9%	3.9% / 2.8%	1.9% / 1.5%	2.0% / 1.4%	1.9% / 1.6%
0.05	1.00	20	4.1% / 4.2%	4.2% / 4.2%	1.1% / 1.8%	1.1% / 1.7%	0.7% / 1.5%
0.05	1.00	50	3.6% / 3.3%	3.7% / 3.3%	1.7% / 0.7%	1.7% / 0.7%	1.4% / 0.7%
0.25	0.25	5	5.9% / 2.5%	4.7% / 3.3%	2.4% / 1.0%	1.7% / 1.4%	1.2% / 1.9%
0.25	0.25	10	7.1% / 1.6%	6.4% / 3.1%	1.9% / 0.7%	1.6% / 1.0%	0.7% / 2.8%
0.25	0.25	20	10.7% / 1.4%	7.6% / 2.2%	2.2% / 0.3%	1.4% / 1.2%	0.2% / 4.0%
0.25	0.25	50	18.7% / 0.2%	11.3% / 1.1%	2.7% / 0.1%	1.2% / 0.4%	0.0% / 6.5%
0.25	0.50	5	5.3% / 3.0%	5.0% / 3.6%	2.4% / 1.9%	2.5% / 2.2%	1.4% / 2.7%
0.25	0.50	10	5.3% / 2.5%	4.8% / 3.0%	2.0% / 0.8%	1.8% / 1.2%	1.3% / 2.4%
0.25	0.50	20	5.5% / 1.8%	4.7% / 2.0%	1.6% / 0.6%	1.5% / 0.6%	0.6% / 1.5%
0.25	0.50	50	7.8% / 1.1%	6.2% / 1.7%	2.1% / 0.6%	1.3% / 1.0%	0.3% / 2.8%
0.25	0.67	5	4.6% / 3.3%	4.6% / 3.4%	1.5% / 1.6%	1.5% / 1.9%	1.2% / 2.0%
0.25	0.67	10	5.1% / 2.5%	5.2% / 3.2%	1.6% / 1.0%	1.6% / 1.3%	0.9% / 1.5%
0.25	0.67	20	3.8% / 1.9%	3.6% / 2.3%	1.4% / 0.8%	1.0% / 1.0%	0.9% / 1.1%
0.25	0.67	50	5.4% / 1.9%	4.9% / 1.9%	2.0% / 0.8%	1.7% / 1.0%	0.3% / 1.5%
0.25	1.00	5	3.7% / 4.1%	3.5% / 4.2%	1.5% / 1.8%	1.5% / 1.9%	1.6% / 1.9%
0.25	1.00	10	4.5% / 3.6%	4.7% / 3.9%	1.1% / 1.7%	1.2% / 1.6%	1.2% / 1.7%
0.25	1.00	20	2.5% / 3.0%	2.6% / 3.1%	1.3% / 0.9%	1.4% / 1.0%	1.4% / 0.9%
0.25	1.00	50	2.3% / 2.9%	2.3% / 3.2%	0.9% / 0.8%	0.9% / 0.8%	0.7% / 0.8%

Table A.23: Simulation II, Scenario C: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of one-sided test for bias (lower / upper tail); risk ratio ϕ as measure of treatment effect; control event rate equals 0.1; nominal significance level is 5% for each tail; error rates above 20% in bold font.

τ^2	ϕ	k	Linear regression test "add selective"	Linear regression test "add all"	Rank correlation test "add selective"	Rank correlation test "add all"	New test
0.00	0.25	5	5.8% / 4.3%	4.1% / 4.8%	4.4% / 2.7%	3.3% / 3.5%	3.0% / 3.2%
0.00	0.25	10	6.8% / 3.2%	5.0% / 5.0%	5.7% / 2.0%	3.9% / 3.4%	3.7% / 4.1%
0.00	0.25	20	8.4% / 2.8%	4.5% / 4.7%	5.1% / 1.1%	2.4% / 2.6%	2.7% / 3.2%
0.00	0.25	50	12.6% / 0.3%	3.7% / 4.7%	11.4% / 0.3%	4.0% / 1.3%	3.1% / 1.7%
0.00	0.50	5	6.6% / 3.1%	5.9% / 3.6%	4.6% / 2.4%	4.0% / 2.9%	3.4% / 3.3%
0.00	0.50	10	6.5% / 4.2%	5.3% / 5.3%	5.1% / 3.3%	3.6% / 3.6%	4.1% / 3.3%
0.00	0.50	20	6.3% / 2.8%	4.8% / 4.0%	5.2% / 1.7%	4.2% / 2.1%	2.9% / 1.9%
0.00	0.50	50	7.4% / 4.0%	5.1% / 5.9%	6.1% / 1.9%	4.2% / 2.8%	3.9% / 2.8%
0.00	0.67	5	4.8% / 3.8%	4.4% / 4.1%	3.1% / 2.7%	2.9% / 2.7%	3.1% / 2.4%
0.00	0.67	10	5.9% / 5.3%	4.6% / 5.6%	3.9% / 4.1%	3.6% / 4.6%	4.0% / 4.6%
0.00	0.67	20	5.6% / 5.3%	4.8% / 6.0%	3.6% / 2.7%	3.1% / 3.1%	3.3% / 3.4%
0.00	0.67	50	6.5% / 3.9%	4.7% / 4.8%	5.7% / 1.7%	4.5% / 2.5%	4.4% / 2.4%
0.00	1.00	5	4.3% / 3.5%	4.3% / 3.5%	3.4% / 2.6%	3.3% / 2.5%	2.9% / 3.0%
0.00	1.00	10	3.7% / 5.4%	3.8% / 5.3%	2.8% / 4.0%	2.8% / 4.0%	2.5% / 3.9%
0.00	1.00	20	5.2% / 6.0%	5.2% / 6.0%	3.1% / 3.3%	3.1% / 3.3%	3.5% / 3.3%
0.00	1.00	50	4.7% / 4.2%	4.7% / 4.2%	2.9% / 2.4%	2.8% / 2.3%	3.1% / 2.5%
0.05	0.25	5	6.9% / 3.2%	5.8% / 3.8%	3.3% / 2.0%	3.1% / 2.2%	1.8% / 2.3%
0.05	0.25	10	5.9% / 2.7%	4.1% / 3.5%	2.4% / 1.0%	1.8% / 1.2%	1.1% / 2.4%
0.05	0.25	20	7.5% / 2.3%	5.5% / 3.3%	3.0% / 0.4%	1.8% / 0.7%	0.5% / 1.6%
0.05	0.25	50	11.9% / 0.9%	7.5% / 1.5%	3.9% / 0.2%	2.2% / 0.5%	0.3% / 2.5%
0.05	0.50	5	4.7% / 4.4%	4.6% / 4.6%	2.1% / 1.9%	2.1% / 2.1%	1.5% / 2.2%
0.05	0.50	10	6.0% / 2.8%	5.6% / 3.1%	1.9% / 1.6%	1.8% / 1.7%	1.7% / 1.9%
0.05	0.50	20	5.6% / 2.8%	5.0% / 3.3%	1.7% / 0.4%	1.6% / 0.4%	0.3% / 1.1%
0.05	0.50	50	6.0% / 1.9%	5.2% / 2.9%	1.7% / 0.5%	1.2% / 0.7%	0.3% / 1.4%
0.05	0.67	5	3.9% / 3.8%	3.8% / 4.0%	1.4% / 1.9%	1.4% / 2.0%	1.5% / 2.0%
0.05	0.67	10	4.4% / 3.1%	4.4% / 3.2%	1.6% / 0.9%	1.4% / 0.9%	1.3% / 1.3%
0.05	0.67	20	4.0% / 2.7%	3.7% / 2.8%	1.6% / 0.7%	1.3% / 0.7%	1.2% / 1.0%
0.05	0.67	50	5.4% / 2.2%	4.7% / 2.3%	1.7% / 1.1%	1.6% / 1.2%	0.7% / 1.6%
0.05	1.00	5	4.3% / 3.4%	4.3% / 3.4%	1.5% / 1.9%	1.6% / 1.9%	1.9% / 2.1%
0.05	1.00	10	4.8% / 4.2%	4.9% / 4.2%	1.0% / 1.0%	1.0% / 1.0%	1.0% / 0.9%
0.05	1.00	20	3.5% / 3.7%	3.5% / 3.9%	1.3% / 1.3%	1.3% / 1.3%	0.9% / 1.1%
0.05	1.00	50	4.3% / 3.0%	4.5% / 3.0%	0.9% / 0.4%	1.0% / 0.4%	1.1% / 0.3%
0.25	0.25	5	5.4% / 3.2%	5.0% / 3.3%	1.8% / 1.3%	1.8% / 1.4%	0.8% / 3.0%
0.25	0.25	10	6.9% / 1.9%	6.0% / 2.1%	2.7% / 0.7%	2.2% / 0.8%	0.8% / 2.6%
0.25	0.25	20	9.1% / 1.3%	8.0% / 1.6%	1.5% / 0.2%	1.4% / 0.3%	0.1% / 3.0%
0.25	0.25	50	18.0% / 0.2%	15.1% / 0.3%	3.6% / 0.4%	2.5% / 0.5%	0.1% / 8.7%
0.25	0.50	5	5.2% / 2.7%	5.1% / 2.7%	2.3% / 2.1%	2.3% / 2.1%	1.8% / 3.0%
0.25	0.50	10	6.2% / 3.3%	6.0% / 3.4%	2.6% / 1.5%	2.6% / 1.5%	1.2% / 2.1%
0.25	0.50	20	5.6% / 1.9%	5.2% / 2.1%	0.9% / 0.2%	0.9% / 0.3%	0.6% / 1.5%
0.25	0.50	50	7.8% / 1.2%	7.0% / 1.3%	1.2% / 0.4%	1.1% / 0.5%	0.0% / 2.2%
0.25	0.67	5	4.5% / 4.3%	4.4% / 4.3%	2.4% / 1.4%	2.4% / 1.4%	1.6% / 1.5%
0.25	0.67	10	5.3% / 2.7%	5.2% / 2.8%	1.9% / 1.0%	1.8% / 1.0%	1.1% / 1.2%
0.25	0.67	20	4.4% / 3.4%	4.2% / 3.6%	1.6% / 0.1%	1.3% / 0.1%	0.8% / 0.6%
0.25	0.67	50	6.5% / 1.0%	6.4% / 1.0%	0.9% / 0.4%	0.9% / 0.4%	0.5% / 1.4%
0.25	1.00	5	4.6% / 4.5%	4.7% / 4.5%	2.7% / 2.2%	2.7% / 2.3%	2.3% / 2.0%
0.25	1.00	10	2.2% / 3.7%	2.2% / 3.7%	0.8% / 1.5%	0.8% / 1.5%	1.1% / 1.2%
0.25	1.00	20	2.2% / 3.1%	2.2% / 3.1%	0.7% / 1.3%	0.7% / 1.3%	0.8% / 0.9%
0.25	1.00	50	1.6% / 2.5%	1.6% / 2.5%	0.9% / 0.4%	0.9% / 0.4%	1.5% / 0.4%

Table A.24: Simulation II, Scenario C: results under null hypothesis ($n = 1000$); proportion of simulated meta-analyses with significant result of one-sided test for bias (lower / upper tail); risk ratio ϕ as measure of treatment effect; control event rate equals 0.3; nominal significance level is 5% for each tail; error rates above 20% in bold font.

A.2 Publication bias present

The following abbreviations are used in tables listed in this section.

Given values:

τ^2	Between-trial variance
ψ	Underlying odds ratio
k	Number of trials in meta-analysis
orig.	original critical values
adjusted	rejection limits based on empirical 5% and 95% quantiles of test statistics derived from simulations under null hypothesis

Estimated values:

% selected	Proportion of published trials
$\widehat{\psi}_{MH}$	Mantel-Haenszel odds ratio

τ^2	ψ	k (% selected)	$\widehat{\psi}_{MH}$ 50% [5%; 95%]	Linear regression test ("add all") orig. / adjusted	Rank correlation test ("add all") orig. / adjusted	New test
0.00	0.25	5 (100.0%)	0.24 [0.13; 0.42]	19.0% / 10.1%	6.4% / 6.4%	7.0%
0.00	0.25	10 (90.9%)	0.24 [0.16; 0.36]	28.7% / 13.2%	9.5% / 6.5%	7.4%
0.00	0.25	20 (90.9%)	0.24 [0.18; 0.32]	41.7% / 13.6%	13.3% / 10.0%	7.4%
0.00	0.25	50 (90.9%)	0.25 [0.20; 0.29]	66.7% / 22.5%	37.1% / 19.1%	7.4%
0.00	0.50	5 (83.3%)	0.43 [0.24; 0.65]	18.0% / 9.8%	9.6% / 9.6%	9.3%
0.00	0.50	10 (83.3%)	0.42 [0.29; 0.58]	28.9% / 12.9%	20.4% / 9.3%	15.0%
0.00	0.50	20 (83.3%)	0.42 [0.33; 0.53]	46.1% / 29.4%	35.4% / 26.5%	20.9%
0.00	0.50	50 (80.6%)	0.43 [0.37; 0.50]	74.0% / 46.8%	73.6% / 40.5%	51.2%
0.00	0.67	5 (71.4%)	0.53 [0.31; 0.78]	19.5% / 14.1%	11.2% / 11.2%	10.9%
0.00	0.67	10 (76.9%)	0.54 [0.37; 0.71]	29.3% / 17.0%	23.6% / 13.2%	18.6%
0.00	0.67	20 (74.1%)	0.54 [0.42; 0.66]	46.1% / 32.7%	37.5% / 27.1%	29.6%
0.00	0.67	50 (73.5%)	0.54 [0.47; 0.62]	77.6% / 62.9%	74.5% / 58.2%	59.7%
0.00	1.00	5 (62.5%)	0.68 [0.43; 0.97]	16.8% / 15.1%	13.4% / 13.4%	10.4%
0.00	1.00	10 (58.8%)	0.70 [0.50; 0.90]	29.3% / 23.0%	23.3% / 13.1%	19.0%
0.00	1.00	20 (57.1%)	0.70 [0.57; 0.85]	43.1% / 39.6%	38.7% / 38.7%	31.6%
0.00	1.00	50 (56.8%)	0.70 [0.62; 0.79]	79.6% / 81.4%	79.8% / 83.1%	68.2%
0.05	0.25	5 (100.0%)	0.24 [0.13; 0.43]	19.0% / 10.5%	6.3% / 6.3%	5.4%
0.05	0.25	10 (90.9%)	0.25 [0.16; 0.37]	27.4% / 11.3%	9.6% / 6.5%	6.7%
0.05	0.25	20 (90.9%)	0.25 [0.18; 0.33]	42.3% / 15.9%	13.6% / 11.6%	6.4%
0.05	0.25	50 (90.9%)	0.25 [0.20; 0.30]	70.7% / 22.1%	38.5% / 16.9%	7.9%
0.05	0.50	5 (83.3%)	0.41 [0.23; 0.66]	16.2% / 10.3%	8.6% / 8.6%	9.3%
0.05	0.50	10 (83.3%)	0.42 [0.28; 0.59]	26.9% / 17.7%	18.5% / 10.3%	14.5%
0.05	0.50	20 (83.3%)	0.42 [0.32; 0.53]	42.2% / 25.8%	31.1% / 20.6%	20.3%
0.05	0.50	50 (80.6%)	0.42 [0.36; 0.50]	70.9% / 47.4%	71.4% / 46.2%	45.6%
0.05	0.67	5 (71.4%)	0.53 [0.29; 0.80]	17.8% / 11.0%	10.1% / 10.1%	9.1%
0.05	0.67	10 (71.4%)	0.52 [0.37; 0.71]	28.5% / 21.4%	23.5% / 16.3%	14.2%
0.05	0.67	20 (71.4%)	0.53 [0.41; 0.66]	43.1% / 27.1%	34.8% / 24.8%	26.3%
0.05	0.67	50 (72.5%)	0.53 [0.45; 0.61]	72.6% / 49.8%	69.9% / 49.2%	50.5%
0.05	1.00	5 (55.6%)	0.66 [0.40; 1.00]	16.4% / 13.0%	11.8% / 11.8%	10.7%
0.05	1.00	10 (58.8%)	0.67 [0.49; 0.88]	27.8% / 20.0%	22.9% / 18.2%	19.9%
0.05	1.00	20 (57.1%)	0.68 [0.54; 0.81]	42.3% / 36.2%	37.5% / 37.2%	28.3%
0.05	1.00	50 (56.2%)	0.68 [0.58; 0.77]	72.1% / 66.0%	72.5% / 75.6%	59.2%
0.25	0.25	5 (100.0%)	0.23 [0.12; 0.44]	20.8% / 12.8%	7.0% / 7.0%	7.8%
0.25	0.25	10 (90.9%)	0.24 [0.14; 0.37]	30.8% / 12.3%	9.6% / 8.4%	7.5%
0.25	0.25	20 (90.9%)	0.24 [0.17; 0.34]	45.4% / 17.1%	15.9% / 12.2%	6.6%
0.25	0.25	50 (89.3%)	0.24 [0.19; 0.30]	68.1% / 23.2%	37.6% / 15.7%	8.3%
0.25	0.50	5 (83.3%)	0.39 [0.20; 0.68]	18.7% / 13.7%	10.0% / 3.6%	8.3%
0.25	0.50	10 (76.9%)	0.39 [0.24; 0.60]	26.8% / 12.2%	17.4% / 8.3%	9.1%
0.25	0.50	20 (80.0%)	0.39 [0.29; 0.53]	44.5% / 19.6%	31.3% / 18.4%	14.3%
0.25	0.50	50 (79.4%)	0.39 [0.32; 0.47]	67.2% / 38.6%	68.0% / 39.5%	30.1%
0.25	0.67	5 (71.4%)	0.47 [0.25; 0.79]	14.3% / 9.8%	7.5% / 7.5%	7.9%
0.25	0.67	10 (71.4%)	0.47 [0.30; 0.67]	26.3% / 20.0%	20.5% / 15.6%	13.4%
0.25	0.67	20 (69.0%)	0.47 [0.35; 0.61]	43.0% / 28.7%	33.5% / 23.8%	17.2%
0.25	0.67	50 (70.4%)	0.48 [0.39; 0.57]	63.0% / 37.9%	66.0% / 43.3%	32.8%
0.25	1.00	5 (55.6%)	0.60 [0.33; 0.93]	15.5% / 12.2%	9.9% / 9.9%	9.5%
0.25	1.00	10 (55.6%)	0.59 [0.41; 0.82]	25.8% / 21.5%	22.5% / 22.6%	14.9%
0.25	1.00	20 (55.6%)	0.59 [0.46; 0.75]	36.3% / 24.5%	30.9% / 28.0%	20.2%
0.25	1.00	50 (55.6%)	0.60 [0.51; 0.70]	58.5% / 41.7%	62.1% / 61.8%	40.0%

Table A.25: Simulation I: results under alternative hypothesis of moderate selection bias ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.1; nominal significance level is 10%.

τ^2	ψ	k (% selected)	$\widehat{\psi}_{MH}$ 50% [5%; 95%]	Linear regression test ("add all") orig. / adjusted	Rank correlation test ("add all") orig. / adjusted	New test
0.00	0.25	5 (100.0%)	0.24 [0.15; 0.37]	14.1% / 10.3%	9.9% / 9.9%	6.7%
0.00	0.25	10 (100.0%)	0.24 [0.18; 0.32]	15.3% / 9.8%	15.0% / 6.2%	8.6%
0.00	0.25	20 (100.0%)	0.25 [0.20; 0.30]	23.1% / 11.4%	23.3% / 10.2%	9.0%
0.00	0.25	50 (98.0%)	0.25 [0.22; 0.28]	36.4% / 9.7%	45.8% / 10.2%	10.3%
0.00	0.50	5 (100.0%)	0.46 [0.32; 0.63]	11.2% / 8.2%	8.1% / 3.2%	7.8%
0.00	0.50	10 (90.9%)	0.47 [0.37; 0.58]	17.4% / 10.1%	15.9% / 9.0%	11.0%
0.00	0.50	20 (90.9%)	0.47 [0.40; 0.55]	26.5% / 18.6%	23.1% / 17.2%	15.3%
0.00	0.50	50 (90.9%)	0.47 [0.42; 0.52]	42.5% / 27.7%	44.5% / 20.3%	22.7%
0.00	0.67	5 (83.3%)	0.61 [0.43; 0.79]	12.1% / 8.7%	9.4% / 9.4%	8.3%
0.00	0.67	10 (83.3%)	0.60 [0.49; 0.73]	19.8% / 14.9%	17.4% / 12.3%	13.5%
0.00	0.67	20 (80.0%)	0.60 [0.52; 0.69]	31.0% / 27.4%	27.8% / 18.8%	19.5%
0.00	0.67	50 (80.6%)	0.60 [0.55; 0.66]	56.7% / 42.0%	55.6% / 39.7%	39.8%
0.00	1.00	5 (62.5%)	0.78 [0.58; 1.00]	12.2% / 13.4%	9.5% / 8.9%	7.8%
0.00	1.00	10 (58.8%)	0.78 [0.64; 0.93]	21.6% / 20.6%	19.5% / 14.8%	16.8%
0.00	1.00	20 (57.1%)	0.79 [0.68; 0.89]	39.2% / 41.0%	32.8% / 36.1%	28.4%
0.00	1.00	50 (56.2%)	0.79 [0.73; 0.85]	72.6% / 70.1%	65.1% / 62.2%	58.5%
0.05	0.25	5 (100.0%)	0.24 [0.15; 0.38]	12.9% / 9.5%	8.9% / 3.3%	6.6%
0.05	0.25	10 (100.0%)	0.25 [0.18; 0.34]	18.2% / 9.4%	15.3% / 9.7%	7.7%
0.05	0.25	20 (100.0%)	0.25 [0.20; 0.31]	26.1% / 10.2%	24.0% / 8.4%	7.6%
0.05	0.25	50 (98.0%)	0.25 [0.21; 0.28]	38.9% / 13.1%	48.0% / 9.3%	7.6%
0.05	0.50	5 (100.0%)	0.46 [0.30; 0.63]	14.0% / 10.8%	10.4% / 4.3%	8.6%
0.05	0.50	10 (90.9%)	0.46 [0.34; 0.60]	18.5% / 12.1%	15.1% / 8.2%	9.7%
0.05	0.50	20 (90.9%)	0.47 [0.38; 0.56]	29.8% / 17.9%	23.5% / 13.5%	12.6%
0.05	0.50	50 (89.3%)	0.46 [0.41; 0.52]	45.1% / 25.9%	43.4% / 20.9%	19.6%
0.05	0.67	5 (83.3%)	0.57 [0.40; 0.77]	11.7% / 8.5%	8.3% / 3.3%	7.5%
0.05	0.67	10 (76.9%)	0.58 [0.45; 0.72]	20.5% / 16.5%	16.6% / 11.7%	12.3%
0.05	0.67	20 (80.0%)	0.58 [0.49; 0.68]	29.2% / 19.3%	23.0% / 19.1%	15.6%
0.05	0.67	50 (79.4%)	0.59 [0.53; 0.65]	53.8% / 34.3%	48.0% / 32.8%	31.1%
0.05	1.00	5 (62.5%)	0.75 [0.53; 0.97]	11.2% / 7.5%	9.2% / 8.2%	7.9%
0.05	1.00	10 (58.8%)	0.75 [0.60; 0.90]	19.3% / 16.6%	16.6% / 16.6%	14.0%
0.05	1.00	20 (55.6%)	0.75 [0.64; 0.85]	29.4% / 23.8%	25.9% / 23.4%	21.2%
0.05	1.00	50 (56.2%)	0.75 [0.68; 0.83]	55.8% / 50.2%	53.1% / 53.4%	43.8%
0.25	0.25	5 (100.0%)	0.25 [0.13; 0.42]	17.4% / 9.5%	10.4% / 4.1%	6.9%
0.25	0.25	10 (100.0%)	0.24 [0.15; 0.37]	23.3% / 9.2%	15.8% / 8.5%	8.6%
0.25	0.25	20 (95.2%)	0.24 [0.18; 0.32]	29.7% / 10.0%	23.9% / 7.9%	6.2%
0.25	0.25	50 (96.2%)	0.24 [0.20; 0.30]	50.6% / 11.0%	55.1% / 8.8%	6.5%
0.25	0.50	5 (83.3%)	0.42 [0.26; 0.66]	14.1% / 11.5%	9.1% / 3.2%	8.1%
0.25	0.50	10 (83.3%)	0.42 [0.30; 0.58]	20.9% / 11.6%	15.6% / 9.0%	8.7%
0.25	0.50	20 (83.3%)	0.43 [0.33; 0.55]	28.0% / 15.3%	24.1% / 13.3%	8.3%
0.25	0.50	50 (84.7%)	0.43 [0.36; 0.51]	45.7% / 21.4%	45.7% / 21.7%	12.3%
0.25	0.67	5 (83.3%)	0.52 [0.33; 0.79]	14.0% / 10.0%	10.5% / 10.5%	8.2%
0.25	0.67	10 (76.9%)	0.52 [0.37; 0.70]	20.8% / 14.1%	17.1% / 11.6%	11.3%
0.25	0.67	20 (74.1%)	0.52 [0.41; 0.65]	28.7% / 17.1%	23.0% / 17.7%	10.6%
0.25	0.67	50 (74.6%)	0.52 [0.45; 0.61]	38.6% / 18.6%	39.3% / 29.4%	14.3%
0.25	1.00	5 (55.6%)	0.65 [0.43; 0.91]	11.9% / 10.9%	10.3% / 10.3%	7.0%
0.25	1.00	10 (55.6%)	0.64 [0.47; 0.84]	18.4% / 10.5%	13.3% / 13.1%	10.1%
0.25	1.00	20 (55.6%)	0.65 [0.52; 0.79]	22.8% / 15.2%	16.8% / 17.9%	10.7%
0.25	1.00	50 (55.6%)	0.65 [0.56; 0.73]	34.2% / 19.3%	32.6% / 34.2%	18.6%

Table A.26: Simulation I: results under alternative hypothesis of moderate selection bias ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.3; nominal significance level is 10%.

τ^2	ψ	k (% selected)	$\widehat{\psi}_{MH}$ 50% [5%; 95%]	Linear regression test ("add all") orig. / adjusted	Rank correlation test ("add all") orig. / adjusted	New test
0.00	0.25	5 (83.3%)	0.22 [0.12; 0.37]	25.2% / 11.8%	6.4% / 6.4%	7.3%
0.00	0.25	10 (76.9%)	0.23 [0.15; 0.32]	35.0% / 14.9%	11.6% / 9.3%	8.4%
0.00	0.25	20 (76.9%)	0.22 [0.17; 0.29]	55.3% / 20.4%	19.5% / 13.6%	8.3%
0.00	0.25	50 (76.9%)	0.22 [0.19; 0.27]	86.9% / 40.6%	54.0% / 31.0%	11.6%
0.00	0.50	5 (62.5%)	0.37 [0.21; 0.57]	23.8% / 14.8%	11.6% / 11.6%	11.6%
0.00	0.50	10 (62.5%)	0.38 [0.26; 0.52]	39.0% / 23.1%	28.1% / 15.1%	20.8%
0.00	0.50	20 (62.5%)	0.39 [0.30; 0.48]	61.7% / 46.6%	48.6% / 39.6%	38.4%
0.00	0.50	50 (61.7%)	0.39 [0.33; 0.45]	93.6% / 78.7%	90.6% / 65.2%	78.9%
0.00	0.67	5 (55.6%)	0.46 [0.27; 0.69]	21.7% / 15.9%	12.4% / 12.4%	11.6%
0.00	0.67	10 (52.6%)	0.47 [0.32; 0.64]	39.9% / 24.9%	29.1% / 16.1%	26.9%
0.00	0.67	20 (52.6%)	0.48 [0.38; 0.59]	64.9% / 52.5%	53.8% / 40.2%	43.7%
0.00	0.67	50 (51.8%)	0.48 [0.41; 0.55]	94.0% / 86.0%	91.5% / 78.4%	81.3%
0.00	1.00	5 (35.7%)	0.60 [0.35; 0.87]	22.8% / 19.2%	14.8% / 14.8%	13.7%
0.00	1.00	10 (35.7%)	0.61 [0.44; 0.81]	37.4% / 32.6%	31.8% / 16.0%	26.5%
0.00	1.00	20 (35.1%)	0.62 [0.48; 0.76]	60.2% / 56.0%	50.5% / 50.5%	44.5%
0.00	1.00	50 (35.7%)	0.62 [0.53; 0.72]	93.7% / 94.9%	89.9% / 92.1%	81.2%
0.05	0.25	5 (83.3%)	0.21 [0.12; 0.37]	23.2% / 12.0%	6.1% / 6.1%	6.7%
0.05	0.25	10 (76.9%)	0.22 [0.14; 0.33]	34.0% / 11.8%	9.4% / 6.5%	7.7%
0.05	0.25	20 (76.9%)	0.22 [0.16; 0.30]	56.2% / 23.9%	19.7% / 15.2%	8.3%
0.05	0.25	50 (75.8%)	0.22 [0.18; 0.27]	83.8% / 35.5%	51.2% / 25.4%	11.2%
0.05	0.50	5 (62.5%)	0.37 [0.20; 0.58]	22.1% / 13.5%	10.7% / 10.7%	9.2%
0.05	0.50	10 (62.5%)	0.37 [0.25; 0.54]	37.7% / 24.7%	24.8% / 13.8%	19.8%
0.05	0.50	20 (62.5%)	0.38 [0.28; 0.49]	65.5% / 47.7%	47.5% / 34.9%	36.3%
0.05	0.50	50 (61.0%)	0.38 [0.32; 0.45]	89.0% / 73.0%	86.7% / 65.5%	69.0%
0.05	0.67	5 (55.6%)	0.46 [0.26; 0.67]	23.8% / 15.4%	12.1% / 12.1%	10.7%
0.05	0.67	10 (52.6%)	0.46 [0.32; 0.61]	37.9% / 29.3%	27.9% / 20.9%	21.7%
0.05	0.67	20 (52.6%)	0.46 [0.35; 0.57]	62.3% / 46.8%	50.0% / 39.3%	38.8%
0.05	0.67	50 (52.1%)	0.46 [0.39; 0.54]	89.6% / 77.4%	87.1% / 69.0%	74.8%
0.05	1.00	5 (38.5%)	0.58 [0.35; 0.87]	20.3% / 13.8%	10.9% / 10.9%	10.0%
0.05	1.00	10 (37.0%)	0.59 [0.42; 0.78]	35.3% / 27.4%	26.6% / 20.3%	22.5%
0.05	1.00	20 (36.4%)	0.59 [0.46; 0.72]	55.1% / 49.4%	49.3% / 49.4%	40.6%
0.05	1.00	50 (36.0%)	0.60 [0.51; 0.67]	87.4% / 84.2%	85.3% / 86.8%	75.3%
0.25	0.25	5 (83.3%)	0.21 [0.11; 0.38]	21.5% / 14.0%	6.5% / 6.5%	6.6%
0.25	0.25	10 (76.9%)	0.21 [0.13; 0.33]	36.9% / 14.6%	10.7% / 7.9%	7.2%
0.25	0.25	20 (74.1%)	0.21 [0.15; 0.30]	53.0% / 22.9%	17.0% / 11.4%	6.3%
0.25	0.25	50 (74.6%)	0.21 [0.17; 0.27]	82.5% / 35.7%	42.7% / 19.7%	6.2%
0.25	0.50	5 (62.5%)	0.33 [0.17; 0.57]	23.3% / 15.2%	10.7% / 3.3%	9.2%
0.25	0.50	10 (62.5%)	0.34 [0.21; 0.50]	34.5% / 14.0%	19.3% / 8.8%	14.4%
0.25	0.50	20 (60.6%)	0.35 [0.25; 0.45]	54.3% / 25.4%	37.1% / 23.5%	21.1%
0.25	0.50	50 (60.2%)	0.34 [0.28; 0.41]	80.9% / 58.4%	76.4% / 51.5%	44.0%
0.25	0.67	5 (55.6%)	0.40 [0.21; 0.66]	19.7% / 12.9%	10.3% / 10.3%	10.0%
0.25	0.67	10 (52.6%)	0.41 [0.27; 0.59]	33.3% / 26.1%	23.1% / 16.4%	15.1%
0.25	0.67	20 (51.3%)	0.41 [0.30; 0.54]	51.5% / 39.0%	39.7% / 30.5%	27.7%
0.25	0.67	50 (51.5%)	0.41 [0.34; 0.49]	80.8% / 60.3%	79.2% / 58.8%	53.6%
0.25	1.00	5 (38.5%)	0.50 [0.28; 0.82]	16.7% / 11.9%	9.9% / 9.9%	9.2%
0.25	1.00	10 (38.5%)	0.52 [0.35; 0.73]	28.9% / 24.2%	20.5% / 20.4%	16.0%
0.25	1.00	20 (37.7%)	0.52 [0.39; 0.66]	44.0% / 32.5%	38.0% / 34.7%	25.8%
0.25	1.00	50 (37.9%)	0.52 [0.44; 0.61]	72.2% / 59.1%	73.9% / 73.5%	50.3%

Table A.27: Simulation I: results under alternative hypothesis of strong selection bias ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.1; nominal significance level is 10%.

τ^2	ψ	k (% selected)	$\widehat{\psi}_{MH}$ 50% [5%; 95%]	Linear regression test ("add all") orig. / adjusted	Rank correlation test ("add all") orig. / adjusted	New test
0.00	0.25	5 (100.0%)	0.23 [0.15; 0.34]	14.2% / 9.7%	9.7% / 9.7%	8.1%
0.00	0.25	10 (90.9%)	0.24 [0.18; 0.31]	21.7% / 11.8%	18.5% / 8.2%	11.5%
0.00	0.25	20 (90.9%)	0.24 [0.20; 0.29]	30.6% / 11.7%	30.2% / 12.4%	11.4%
0.00	0.25	50 (92.6%)	0.24 [0.21; 0.27]	51.9% / 18.8%	62.8% / 15.8%	21.0%
0.00	0.50	5 (83.3%)	0.44 [0.31; 0.57]	16.5% / 10.9%	12.7% / 4.6%	10.4%
0.00	0.50	10 (76.9%)	0.45 [0.35; 0.55]	27.6% / 18.1%	25.6% / 12.9%	17.4%
0.00	0.50	20 (76.9%)	0.45 [0.38; 0.52]	42.1% / 33.4%	38.3% / 27.9%	26.5%
0.00	0.50	50 (75.8%)	0.45 [0.40; 0.49]	73.1% / 57.4%	74.1% / 46.0%	54.0%
0.00	0.67	5 (62.5%)	0.56 [0.39; 0.73]	18.5% / 13.9%	12.7% / 12.7%	10.9%
0.00	0.67	10 (62.5%)	0.56 [0.45; 0.68]	30.4% / 22.8%	24.6% / 19.0%	20.4%
0.00	0.67	20 (62.5%)	0.57 [0.49; 0.65]	52.1% / 44.0%	42.0% / 33.4%	34.8%
0.00	0.67	50 (61.7%)	0.57 [0.51; 0.62]	85.9% / 75.6%	80.5% / 67.6%	70.3%
0.00	1.00	5 (38.5%)	0.71 [0.52; 0.93]	18.4% / 18.9%	12.5% / 12.0%	11.2%
0.00	1.00	10 (37.0%)	0.73 [0.57; 0.86]	30.3% / 29.4%	24.8% / 18.9%	21.5%
0.00	1.00	20 (35.7%)	0.72 [0.62; 0.83]	52.7% / 53.4%	44.2% / 47.2%	38.3%
0.00	1.00	50 (35.8%)	0.72 [0.66; 0.79]	90.0% / 89.1%	83.1% / 81.0%	76.8%
0.05	0.25	5 (100.0%)	0.24 [0.15; 0.36]	14.6% / 10.2%	9.2% / 4.0%	5.8%
0.05	0.25	10 (90.9%)	0.24 [0.17; 0.32]	22.6% / 12.9%	18.4% / 12.2%	9.7%
0.05	0.25	20 (90.9%)	0.24 [0.19; 0.29]	32.6% / 13.3%	28.6% / 10.2%	10.6%
0.05	0.25	50 (90.9%)	0.24 [0.21; 0.27]	55.9% / 20.9%	66.6% / 17.3%	15.9%
0.05	0.50	5 (71.4%)	0.43 [0.29; 0.59]	16.4% / 11.0%	10.4% / 3.5%	8.7%
0.05	0.50	10 (76.9%)	0.44 [0.33; 0.55]	26.1% / 17.5%	19.3% / 11.0%	13.2%
0.05	0.50	20 (74.1%)	0.43 [0.36; 0.52]	38.9% / 25.8%	34.3% / 23.6%	21.5%
0.05	0.50	50 (74.6%)	0.44 [0.39; 0.49]	70.3% / 48.7%	69.5% / 42.9%	44.4%
0.05	0.67	5 (62.5%)	0.53 [0.36; 0.73]	14.9% / 10.9%	11.8% / 3.6%	8.7%
0.05	0.67	10 (62.5%)	0.54 [0.41; 0.66]	26.7% / 20.7%	22.5% / 16.1%	16.8%
0.05	0.67	20 (62.5%)	0.54 [0.45; 0.63]	46.8% / 35.6%	37.1% / 32.1%	28.0%
0.05	0.67	50 (61.0%)	0.54 [0.48; 0.60]	73.5% / 55.5%	70.0% / 55.9%	54.4%
0.05	1.00	5 (38.5%)	0.67 [0.49; 0.89]	17.5% / 14.5%	11.9% / 11.0%	10.6%
0.05	1.00	10 (38.5%)	0.68 [0.53; 0.84]	25.6% / 23.1%	21.8% / 21.6%	17.6%
0.05	1.00	20 (37.0%)	0.68 [0.58; 0.79]	43.9% / 38.2%	37.0% / 34.9%	30.6%
0.05	1.00	50 (37.0%)	0.68 [0.62; 0.75]	74.0% / 68.9%	71.0% / 71.3%	62.2%
0.25	0.25	5 (100.0%)	0.22 [0.12; 0.38]	15.4% / 9.3%	9.7% / 2.9%	6.0%
0.25	0.25	10 (90.9%)	0.23 [0.15; 0.34]	27.1% / 10.8%	18.7% / 7.9%	7.5%
0.25	0.25	20 (90.9%)	0.23 [0.17; 0.31]	37.9% / 12.8%	32.3% / 9.7%	7.8%
0.25	0.25	50 (87.7%)	0.23 [0.19; 0.28]	63.2% / 21.2%	69.2% / 17.3%	11.0%
0.25	0.50	5 (71.4%)	0.39 [0.23; 0.59]	17.0% / 12.9%	11.6% / 4.7%	7.5%
0.25	0.50	10 (71.4%)	0.39 [0.28; 0.54]	25.0% / 14.1%	20.8% / 11.3%	12.1%
0.25	0.50	20 (71.4%)	0.39 [0.31; 0.49]	39.8% / 21.7%	30.7% / 19.0%	13.0%
0.25	0.50	50 (71.4%)	0.39 [0.34; 0.46]	61.4% / 33.9%	64.5% / 39.1%	28.0%
0.25	0.67	5 (62.5%)	0.46 [0.29; 0.68]	15.0% / 11.6%	9.6% / 9.6%	8.2%
0.25	0.67	10 (58.8%)	0.47 [0.34; 0.63]	24.4% / 14.8%	18.7% / 14.1%	12.5%
0.25	0.67	20 (58.8%)	0.47 [0.37; 0.60]	36.5% / 22.3%	27.8% / 20.3%	15.7%
0.25	0.67	50 (58.8%)	0.47 [0.41; 0.55]	57.9% / 34.9%	58.7% / 49.1%	32.6%
0.25	1.00	5 (41.7%)	0.58 [0.38; 0.81]	13.8% / 13.1%	8.2% / 8.2%	7.9%
0.25	1.00	10 (40.0%)	0.57 [0.43; 0.77]	22.3% / 13.7%	19.2% / 19.0%	14.1%
0.25	1.00	20 (40.0%)	0.58 [0.47; 0.70]	30.3% / 21.0%	23.2% / 26.3%	15.0%
0.25	1.00	50 (39.7%)	0.58 [0.51; 0.66]	48.1% / 32.4%	48.0% / 49.7%	32.3%

Table A.28: Simulation I: results under alternative hypothesis of strong selection bias ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.3; nominal significance level is 10%.

τ^2	ψ	k (% selected)	$\widehat{\psi}_{MH}$ 50% [5%; 95%]	Linear regression test ("add all") orig. / adjusted	Rank correlation test ("add all") orig. / adjusted	New test
0.00	0.25	5 (83.3%)	0.23 [0.14; 0.36]	18.3% / 13.9%	6.9% / 9.1%	6.7%
0.00	0.25	10 (90.9%)	0.23 [0.16; 0.32]	34.6% / 13.4%	19.5% / 8.6%	10.2%
0.00	0.25	20 (87.0%)	0.23 [0.18; 0.29]	59.1% / 17.8%	33.8% / 14.2%	12.3%
0.00	0.25	50 (86.2%)	0.24 [0.20; 0.27]	94.4% / 35.0%	76.6% / 28.7%	23.5%
0.00	0.50	5 (71.4%)	0.43 [0.27; 0.58]	18.2% / 12.7%	12.0% / 3.7%	10.0%
0.00	0.50	10 (71.4%)	0.43 [0.32; 0.53]	33.3% / 18.1%	26.3% / 14.0%	20.1%
0.00	0.50	20 (71.4%)	0.42 [0.35; 0.50]	56.0% / 33.5%	44.9% / 28.0%	31.9%
0.00	0.50	50 (71.4%)	0.42 [0.37; 0.47]	90.8% / 64.8%	85.8% / 59.7%	60.9%
0.00	0.67	5 (62.5%)	0.52 [0.36; 0.70]	17.2% / 10.6%	11.6% / 2.5%	10.0%
0.00	0.67	10 (58.8%)	0.53 [0.41; 0.65]	32.4% / 24.5%	26.6% / 14.6%	24.3%
0.00	0.67	20 (58.8%)	0.52 [0.44; 0.62]	51.8% / 40.7%	44.5% / 35.7%	34.4%
0.00	0.67	50 (58.1%)	0.53 [0.47; 0.58]	87.7% / 78.7%	84.0% / 72.9%	72.2%
0.00	1.00	5 (35.7%)	0.68 [0.49; 0.89]	16.1% / 14.2%	11.5% / 11.5%	11.4%
0.00	1.00	10 (35.7%)	0.69 [0.55; 0.84]	29.5% / 28.8%	24.2% / 24.2%	21.7%
0.00	1.00	20 (36.4%)	0.69 [0.59; 0.79]	51.0% / 45.7%	44.5% / 44.5%	39.1%
0.00	1.00	50 (36.0%)	0.69 [0.63; 0.75]	83.7% / 84.8%	81.7% / 84.4%	74.3%
0.05	0.25	5 (83.3%)	0.23 [0.14; 0.37]	20.0% / 13.2%	9.4% / 9.4%	7.0%
0.05	0.25	10 (90.9%)	0.23 [0.16; 0.33]	31.2% / 8.5%	17.4% / 8.8%	10.2%
0.05	0.25	20 (87.0%)	0.23 [0.18; 0.29]	61.6% / 18.9%	35.9% / 14.4%	11.6%
0.05	0.25	50 (86.2%)	0.23 [0.20; 0.27]	94.2% / 36.7%	75.8% / 30.3%	17.3%
0.05	0.50	5 (71.4%)	0.41 [0.26; 0.59]	16.8% / 13.6%	10.7% / 2.8%	8.6%
0.05	0.50	10 (71.4%)	0.41 [0.30; 0.54]	31.3% / 17.5%	23.8% / 11.1%	16.4%
0.05	0.50	20 (71.4%)	0.41 [0.34; 0.49]	54.7% / 26.3%	41.4% / 25.3%	24.7%
0.05	0.50	50 (70.4%)	0.41 [0.37; 0.47]	88.3% / 62.7%	82.7% / 54.1%	53.9%
0.05	0.67	5 (62.5%)	0.51 [0.34; 0.71]	17.5% / 16.1%	12.8% / 3.1%	10.2%
0.05	0.67	10 (58.8%)	0.51 [0.39; 0.63]	28.2% / 17.3%	24.1% / 17.7%	17.8%
0.05	0.67	20 (58.8%)	0.50 [0.42; 0.60]	49.8% / 34.5%	41.9% / 29.2%	29.9%
0.05	0.67	50 (58.1%)	0.51 [0.45; 0.56]	85.9% / 72.7%	82.9% / 69.3%	63.8%
0.05	1.00	5 (38.5%)	0.64 [0.44; 0.87]	14.6% / 14.6%	10.3% / 10.3%	9.7%
0.05	1.00	10 (38.5%)	0.65 [0.51; 0.81]	25.9% / 28.3%	21.1% / 16.3%	19.9%
0.05	1.00	20 (37.0%)	0.65 [0.56; 0.75]	44.4% / 45.8%	38.0% / 43.9%	31.0%
0.05	1.00	50 (36.8%)	0.65 [0.59; 0.72]	76.8% / 75.6%	74.5% / 73.2%	62.8%
0.25	0.25	5 (83.3%)	0.22 [0.12; 0.38]	19.3% / 11.5%	9.2% / 2.5%	6.8%
0.25	0.25	10 (83.3%)	0.22 [0.15; 0.33]	32.5% / 13.9%	15.5% / 8.6%	7.8%
0.25	0.25	20 (83.3%)	0.22 [0.16; 0.29]	62.3% / 18.2%	32.1% / 18.0%	6.5%
0.25	0.25	50 (83.3%)	0.22 [0.18; 0.27]	94.6% / 35.8%	76.8% / 28.6%	6.9%
0.25	0.50	5 (71.4%)	0.37 [0.23; 0.56]	16.8% / 12.3%	10.7% / 10.7%	9.7%
0.25	0.50	10 (66.7%)	0.37 [0.26; 0.50]	27.0% / 14.8%	20.0% / 10.7%	12.6%
0.25	0.50	20 (69.0%)	0.37 [0.29; 0.47]	50.1% / 30.4%	37.7% / 25.8%	15.1%
0.25	0.50	50 (67.6%)	0.37 [0.32; 0.43]	85.4% / 48.7%	76.0% / 44.6%	28.2%
0.25	0.67	5 (62.5%)	0.45 [0.28; 0.67]	14.8% / 11.8%	11.7% / 3.2%	8.6%
0.25	0.67	10 (55.6%)	0.44 [0.32; 0.60]	25.0% / 21.1%	19.9% / 14.9%	12.9%
0.25	0.67	20 (57.1%)	0.45 [0.35; 0.55]	42.8% / 34.2%	35.6% / 27.5%	17.8%
0.25	0.67	50 (56.8%)	0.45 [0.38; 0.51]	77.5% / 52.0%	72.4% / 48.7%	34.6%
0.25	1.00	5 (41.7%)	0.56 [0.35; 0.83]	12.8% / 10.9%	8.9% / 8.9%	7.4%
0.25	1.00	10 (40.0%)	0.56 [0.42; 0.74]	22.4% / 21.0%	19.0% / 19.0%	13.8%
0.25	1.00	20 (39.2%)	0.56 [0.45; 0.68]	31.3% / 28.0%	28.5% / 25.5%	18.2%
0.25	1.00	50 (39.1%)	0.56 [0.50; 0.64]	59.5% / 61.9%	59.7% / 63.7%	35.0%

Table A.29: Simulation II, Scenario A: results under alternative hypothesis of strong selection bias ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.1; nominal significance level is 10%.

τ^2	ψ	k (% selected)	$\widehat{\psi}_{MH}$ 50% [5%; 95%]	Linear regression test ("add all") orig. / adjusted	Rank correlation test ("add all") orig. / adjusted	New test
0.00	0.25	5 (100.0%)	0.25 [0.18; 0.33]	12.4% / 11.5%	9.5% / 3.6%	8.4%
0.00	0.25	10 (100.0%)	0.25 [0.20; 0.30]	17.6% / 10.4%	14.3% / 9.4%	10.5%
0.00	0.25	20 (100.0%)	0.25 [0.21; 0.28]	21.9% / 10.3%	18.0% / 12.8%	10.1%
0.00	0.25	50 (98.0%)	0.25 [0.22; 0.27]	43.7% / 14.2%	37.0% / 12.0%	9.6%
0.00	0.50	5 (83.3%)	0.48 [0.37; 0.60]	12.3% / 8.9%	10.2% / 3.7%	9.6%
0.00	0.50	10 (90.9%)	0.48 [0.40; 0.55]	19.2% / 15.3%	17.6% / 12.5%	14.5%
0.00	0.50	20 (87.0%)	0.48 [0.42; 0.53]	28.3% / 23.6%	24.1% / 21.4%	17.7%
0.00	0.50	50 (86.2%)	0.47 [0.44; 0.51]	54.0% / 35.6%	47.5% / 34.5%	33.6%
0.00	0.67	5 (71.4%)	0.61 [0.49; 0.72]	14.3% / 12.1%	12.5% / 12.5%	11.7%
0.00	0.67	10 (71.4%)	0.61 [0.52; 0.69]	21.8% / 20.3%	18.9% / 13.9%	16.7%
0.00	0.67	20 (71.4%)	0.61 [0.54; 0.67]	37.5% / 35.7%	31.5% / 29.2%	26.6%
0.00	0.67	50 (70.4%)	0.60 [0.57; 0.64]	68.6% / 58.8%	63.1% / 55.0%	57.1%
0.00	1.00	5 (38.5%)	0.78 [0.63; 0.94]	15.4% / 15.4%	10.7% / 9.5%	10.0%
0.00	1.00	10 (37.0%)	0.78 [0.67; 0.89]	25.5% / 25.2%	21.0% / 14.4%	20.4%
0.00	1.00	20 (35.7%)	0.78 [0.70; 0.86]	43.7% / 40.7%	37.0% / 35.3%	35.5%
0.00	1.00	50 (35.7%)	0.78 [0.74; 0.83]	77.2% / 74.3%	76.6% / 77.8%	73.7%
0.05	0.25	5 (100.0%)	0.25 [0.17; 0.35]	10.1% / 7.5%	7.1% / 1.7%	6.1%
0.05	0.25	10 (100.0%)	0.25 [0.19; 0.32]	15.2% / 10.9%	11.7% / 8.1%	7.9%
0.05	0.25	20 (100.0%)	0.25 [0.20; 0.29]	23.2% / 9.1%	17.9% / 8.2%	6.9%
0.05	0.25	50 (96.2%)	0.25 [0.22; 0.27]	49.8% / 12.3%	41.5% / 11.3%	6.5%
0.05	0.50	5 (83.3%)	0.46 [0.34; 0.60]	12.2% / 10.0%	9.7% / 9.7%	8.7%
0.05	0.50	10 (83.3%)	0.46 [0.38; 0.56]	18.6% / 15.2%	17.4% / 12.2%	13.1%
0.05	0.50	20 (83.3%)	0.46 [0.40; 0.53]	26.7% / 19.7%	21.2% / 17.4%	13.7%
0.05	0.50	50 (83.3%)	0.46 [0.42; 0.51]	53.2% / 34.1%	47.4% / 36.2%	25.9%
0.05	0.67	5 (71.4%)	0.58 [0.43; 0.73]	15.9% / 13.8%	9.5% / 9.5%	8.6%
0.05	0.67	10 (71.4%)	0.58 [0.48; 0.69]	19.5% / 15.9%	18.0% / 13.3%	16.5%
0.05	0.67	20 (69.0%)	0.58 [0.51; 0.65]	29.6% / 24.2%	26.7% / 22.5%	21.0%
0.05	0.67	50 (68.5%)	0.57 [0.53; 0.62]	56.9% / 49.6%	54.7% / 45.2%	42.5%
0.05	1.00	5 (41.7%)	0.72 [0.57; 0.89]	14.2% / 15.0%	11.2% / 11.2%	10.2%
0.05	1.00	10 (38.5%)	0.72 [0.61; 0.85]	17.1% / 17.1%	15.8% / 15.8%	14.7%
0.05	1.00	20 (38.5%)	0.73 [0.65; 0.81]	31.3% / 32.6%	28.4% / 28.5%	24.7%
0.05	1.00	50 (37.6%)	0.73 [0.67; 0.78]	53.6% / 57.1%	54.0% / 61.1%	47.2%
0.25	0.25	5 (100.0%)	0.25 [0.14; 0.38]	12.5% / 11.0%	8.4% / 5.6%	6.7%
0.25	0.25	10 (100.0%)	0.24 [0.17; 0.33]	17.8% / 8.7%	14.4% / 8.1%	8.6%
0.25	0.25	20 (95.2%)	0.24 [0.19; 0.31]	31.6% / 13.6%	22.1% / 9.9%	6.1%
0.25	0.25	50 (94.3%)	0.24 [0.21; 0.29]	65.1% / 15.7%	52.6% / 14.5%	6.5%
0.25	0.50	5 (83.3%)	0.42 [0.28; 0.60]	11.7% / 9.4%	8.6% / 8.6%	6.1%
0.25	0.50	10 (76.9%)	0.41 [0.31; 0.54]	16.6% / 12.4%	13.3% / 9.7%	9.2%
0.25	0.50	20 (76.9%)	0.42 [0.35; 0.51]	26.2% / 17.2%	21.4% / 14.8%	10.2%
0.25	0.50	50 (76.9%)	0.42 [0.37; 0.47]	49.7% / 27.1%	47.2% / 30.2%	16.3%
0.25	0.67	5 (62.5%)	0.50 [0.34; 0.69]	10.8% / 10.4%	8.4% / 8.4%	7.0%
0.25	0.67	10 (66.7%)	0.50 [0.38; 0.64]	15.1% / 14.6%	13.6% / 9.4%	9.8%
0.25	0.67	20 (64.5%)	0.50 [0.41; 0.60]	21.0% / 17.2%	19.6% / 19.7%	11.4%
0.25	0.67	50 (63.3%)	0.50 [0.44; 0.56]	42.8% / 30.7%	43.9% / 37.2%	20.3%
0.25	1.00	5 (41.7%)	0.61 [0.44; 0.81]	11.4% / 10.3%	8.4% / 8.4%	7.8%
0.25	1.00	10 (41.7%)	0.61 [0.48; 0.76]	13.7% / 13.5%	11.3% / 11.3%	8.9%
0.25	1.00	20 (42.6%)	0.61 [0.52; 0.72]	16.3% / 17.8%	14.9% / 23.3%	10.1%
0.25	1.00	50 (41.3%)	0.61 [0.54; 0.67]	33.2% / 29.3%	36.4% / 41.7%	21.9%

Table A.30: Simulation II, Scenario A: results under alternative hypothesis of strong selection bias ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.3; nominal significance level is 10%.

τ^2	ψ	k (% selected)	$\widehat{\psi}_{MH}$ 50% [5%; 95%]	Linear regression test ("add all") orig. / adjusted	Rank correlation test ("add all") orig. / adjusted	New test
0.00	0.25	5 (100.0%)	0.24 [0.17; 0.32]	13.5% / 12.5%	8.3% / 8.3%	7.6%
0.00	0.25	10 (100.0%)	0.25 [0.20; 0.31]	15.7% / 9.2%	12.8% / 8.1%	9.4%
0.00	0.25	20 (95.2%)	0.25 [0.21; 0.29]	26.5% / 11.3%	16.8% / 12.2%	9.6%
0.00	0.25	50 (96.2%)	0.25 [0.22; 0.27]	49.8% / 14.0%	37.3% / 12.9%	13.0%
0.00	0.50	5 (83.3%)	0.48 [0.36; 0.59]	15.9% / 14.7%	9.5% / 9.5%	9.2%
0.00	0.50	10 (83.3%)	0.47 [0.40; 0.56]	22.1% / 17.6%	18.6% / 13.4%	15.9%
0.00	0.50	20 (83.3%)	0.48 [0.42; 0.54]	33.3% / 23.9%	29.1% / 22.2%	21.0%
0.00	0.50	50 (84.7%)	0.48 [0.44; 0.51]	64.6% / 49.5%	57.6% / 39.7%	41.4%
0.00	0.67	5 (71.4%)	0.61 [0.49; 0.73]	15.2% / 15.2%	11.1% / 11.1%	9.3%
0.00	0.67	10 (71.4%)	0.61 [0.51; 0.70]	26.9% / 21.5%	23.1% / 17.8%	20.7%
0.00	0.67	20 (71.4%)	0.61 [0.55; 0.67]	46.5% / 40.2%	39.3% / 37.2%	35.1%
0.00	0.67	50 (70.4%)	0.61 [0.57; 0.65]	78.9% / 66.7%	74.3% / 62.6%	66.5%
0.00	1.00	5 (38.5%)	0.79 [0.64; 0.94]	20.1% / 19.6%	14.1% / 14.1%	12.9%
0.00	1.00	10 (36.4%)	0.79 [0.68; 0.90]	30.4% / 27.8%	25.6% / 18.6%	23.8%
0.00	1.00	20 (35.7%)	0.80 [0.72; 0.87]	53.5% / 49.1%	47.9% / 48.0%	45.5%
0.00	1.00	50 (35.7%)	0.80 [0.75; 0.84]	87.8% / 85.6%	84.1% / 82.0%	81.4%
0.05	0.25	5 (100.0%)	0.24 [0.17; 0.34]	14.0% / 12.8%	9.7% / 9.7%	7.4%
0.05	0.25	10 (100.0%)	0.25 [0.19; 0.32]	16.2% / 9.0%	11.2% / 7.2%	7.1%
0.05	0.25	20 (95.2%)	0.24 [0.20; 0.29]	28.9% / 11.5%	18.1% / 8.3%	6.0%
0.05	0.25	50 (94.3%)	0.25 [0.22; 0.28]	55.9% / 18.0%	41.7% / 19.7%	10.4%
0.05	0.50	5 (83.3%)	0.47 [0.34; 0.61]	11.0% / 10.8%	7.8% / 7.8%	6.6%
0.05	0.50	10 (83.3%)	0.46 [0.37; 0.56]	19.1% / 11.4%	16.1% / 8.5%	11.5%
0.05	0.50	20 (83.3%)	0.47 [0.40; 0.54]	33.0% / 23.6%	26.6% / 24.5%	16.3%
0.05	0.50	50 (82.0%)	0.46 [0.42; 0.51]	61.7% / 42.9%	55.3% / 42.1%	29.5%
0.05	0.67	5 (71.4%)	0.58 [0.44; 0.73]	15.1% / 14.4%	11.4% / 11.4%	11.4%
0.05	0.67	10 (71.4%)	0.58 [0.47; 0.69]	22.2% / 20.9%	18.5% / 18.5%	14.7%
0.05	0.67	20 (69.0%)	0.58 [0.50; 0.65]	34.9% / 29.3%	30.6% / 28.3%	23.0%
0.05	0.67	50 (67.6%)	0.58 [0.53; 0.62]	68.1% / 55.3%	65.6% / 59.8%	49.5%
0.05	1.00	5 (38.5%)	0.73 [0.57; 0.91]	12.9% / 13.0%	9.0% / 9.0%	8.4%
0.05	1.00	10 (40.0%)	0.74 [0.62; 0.85]	22.3% / 22.6%	20.2% / 20.3%	18.3%
0.05	1.00	20 (37.7%)	0.74 [0.66; 0.82]	35.3% / 34.4%	30.3% / 32.5%	26.9%
0.05	1.00	50 (37.9%)	0.74 [0.69; 0.79]	64.5% / 64.1%	62.5% / 66.1%	55.7%
0.25	0.25	5 (100.0%)	0.24 [0.13; 0.38]	12.8% / 10.7%	6.6% / 5.0%	5.2%
0.25	0.25	10 (90.9%)	0.24 [0.16; 0.34]	17.3% / 10.3%	11.1% / 9.1%	6.0%
0.25	0.25	20 (95.2%)	0.24 [0.18; 0.30]	34.3% / 12.3%	20.9% / 13.2%	5.6%
0.25	0.25	50 (92.6%)	0.24 [0.20; 0.28]	68.7% / 20.9%	49.2% / 17.7%	6.4%
0.25	0.50	5 (83.3%)	0.42 [0.26; 0.61]	13.3% / 11.8%	7.3% / 7.3%	6.1%
0.25	0.50	10 (76.9%)	0.41 [0.30; 0.54]	15.4% / 11.0%	12.9% / 9.0%	7.3%
0.25	0.50	20 (76.9%)	0.42 [0.33; 0.50]	29.0% / 18.7%	22.6% / 18.9%	8.0%
0.25	0.50	50 (76.9%)	0.41 [0.36; 0.47]	58.2% / 30.4%	52.0% / 34.5%	13.7%
0.25	0.67	5 (62.5%)	0.51 [0.33; 0.70]	12.3% / 12.4%	7.9% / 7.9%	6.6%
0.25	0.67	10 (62.5%)	0.50 [0.38; 0.64]	14.6% / 13.7%	12.1% / 9.9%	8.0%
0.25	0.67	20 (64.5%)	0.50 [0.42; 0.60]	26.2% / 20.7%	22.1% / 20.7%	10.2%
0.25	0.67	50 (63.3%)	0.50 [0.44; 0.56]	48.0% / 38.8%	47.8% / 47.1%	20.9%
0.25	1.00	5 (41.7%)	0.62 [0.43; 0.82]	10.0% / 10.1%	7.9% / 7.9%	6.7%
0.25	1.00	10 (41.7%)	0.62 [0.49; 0.78]	15.0% / 16.3%	12.9% / 12.9%	10.6%
0.25	1.00	20 (41.7%)	0.62 [0.52; 0.73]	21.2% / 24.7%	18.6% / 27.1%	11.8%
0.25	1.00	50 (41.3%)	0.61 [0.55; 0.68]	34.6% / 41.6%	39.4% / 51.1%	22.5%

Table A.31: Simulation II, Scenario B: results under alternative hypothesis of strong selection bias ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.1; nominal significance level is 10%.

τ^2	ψ	k (% selected)	$\widehat{\psi}_{MH}$ 50% [5%; 95%]	Linear regression test ("add all") orig. / adjusted	Rank correlation test ("add all") orig. / adjusted	New test
0.00	0.25	5 (100.0%)	0.25 [0.20; 0.30]	8.1% / 8.0%	7.2% / 7.2%	7.2%
0.00	0.25	10 (100.0%)	0.25 [0.22; 0.29]	12.2% / 12.5%	11.7% / 9.2%	10.9%
0.00	0.25	20 (100.0%)	0.25 [0.23; 0.27]	11.3% / 8.9%	9.8% / 8.3%	7.3%
0.00	0.25	50 (100.0%)	0.25 [0.24; 0.26]	17.3% / 12.1%	15.1% / 10.1%	8.4%
0.00	0.50	5 (100.0%)	0.50 [0.43; 0.58]	10.6% / 9.6%	8.4% / 8.4%	8.3%
0.00	0.50	10 (100.0%)	0.49 [0.44; 0.55]	11.4% / 9.9%	9.6% / 7.3%	9.2%
0.00	0.50	20 (95.2%)	0.49 [0.46; 0.53]	13.1% / 12.3%	11.0% / 10.6%	10.1%
0.00	0.50	50 (94.3%)	0.50 [0.47; 0.52]	21.0% / 14.5%	19.1% / 16.3%	13.9%
0.00	0.67	5 (83.3%)	0.65 [0.56; 0.73]	12.4% / 12.4%	9.8% / 9.8%	8.7%
0.00	0.67	10 (83.3%)	0.65 [0.59; 0.71]	16.7% / 17.6%	15.3% / 10.3%	13.9%
0.00	0.67	20 (83.3%)	0.65 [0.61; 0.69]	27.8% / 27.8%	21.8% / 27.3%	20.9%
0.00	0.67	50 (83.3%)	0.65 [0.62; 0.68]	50.6% / 47.3%	44.7% / 44.7%	40.4%
0.00	1.00	5 (38.5%)	0.86 [0.74; 0.95]	16.6% / 18.2%	13.6% / 13.6%	13.0%
0.00	1.00	10 (37.0%)	0.86 [0.78; 0.93]	26.3% / 23.0%	22.4% / 16.1%	22.5%
0.00	1.00	20 (36.4%)	0.86 [0.81; 0.91]	52.4% / 53.2%	43.8% / 45.9%	42.9%
0.00	1.00	50 (36.0%)	0.86 [0.83; 0.90]	81.4% / 83.2%	78.6% / 80.8%	77.3%
0.05	0.25	5 (100.0%)	0.25 [0.19; 0.33]	11.1% / 9.8%	7.3% / 7.3%	7.1%
0.05	0.25	10 (100.0%)	0.25 [0.20; 0.30]	11.0% / 9.6%	8.5% / 6.2%	7.7%
0.05	0.25	20 (100.0%)	0.25 [0.22; 0.29]	11.6% / 8.7%	8.9% / 10.0%	7.0%
0.05	0.25	50 (100.0%)	0.25 [0.23; 0.27]	18.9% / 11.1%	13.3% / 9.8%	6.3%
0.05	0.50	5 (100.0%)	0.49 [0.38; 0.61]	9.8% / 11.6%	6.5% / 6.5%	6.1%
0.05	0.50	10 (90.9%)	0.49 [0.41; 0.58]	11.5% / 10.9%	10.2% / 7.3%	9.1%
0.05	0.50	20 (95.2%)	0.49 [0.43; 0.55]	13.1% / 12.4%	9.8% / 10.0%	6.2%
0.05	0.50	50 (92.6%)	0.49 [0.45; 0.53]	24.2% / 13.9%	20.3% / 16.3%	10.8%
0.05	0.67	5 (83.3%)	0.62 [0.49; 0.75]	11.1% / 13.1%	6.7% / 6.7%	6.4%
0.05	0.67	10 (83.3%)	0.62 [0.54; 0.71]	14.4% / 17.1%	12.0% / 12.6%	10.7%
0.05	0.67	20 (80.0%)	0.62 [0.55; 0.69]	21.1% / 23.4%	18.2% / 21.7%	13.9%
0.05	0.67	50 (78.1%)	0.62 [0.58; 0.66]	38.5% / 28.1%	37.6% / 38.3%	29.8%
0.05	1.00	5 (41.7%)	0.77 [0.64; 0.91]	11.8% / 12.5%	7.5% / 7.5%	7.0%
0.05	1.00	10 (40.0%)	0.78 [0.69; 0.88]	14.2% / 16.9%	12.1% / 12.1%	11.7%
0.05	1.00	20 (39.2%)	0.78 [0.72; 0.85]	23.5% / 26.0%	18.9% / 26.0%	16.7%
0.05	1.00	50 (39.7%)	0.78 [0.74; 0.83]	41.3% / 39.7%	42.2% / 45.7%	39.1%
0.25	0.25	5 (100.0%)	0.25 [0.16; 0.40]	10.4% / 9.2%	5.8% / 5.8%	5.4%
0.25	0.25	10 (100.0%)	0.25 [0.18; 0.34]	13.0% / 10.4%	8.0% / 8.2%	7.3%
0.25	0.25	20 (100.0%)	0.25 [0.20; 0.31]	17.2% / 12.2%	11.0% / 10.0%	4.5%
0.25	0.25	50 (98.0%)	0.25 [0.22; 0.29]	32.9% / 13.2%	20.4% / 11.7%	6.9%
0.25	0.50	5 (83.3%)	0.44 [0.29; 0.64]	9.0% / 9.1%	6.3% / 6.3%	5.7%
0.25	0.50	10 (83.3%)	0.45 [0.34; 0.57]	12.3% / 10.6%	9.7% / 10.5%	6.6%
0.25	0.50	20 (83.3%)	0.44 [0.36; 0.53]	15.5% / 11.6%	10.1% / 12.5%	6.6%
0.25	0.50	50 (83.3%)	0.44 [0.39; 0.50]	27.8% / 15.9%	27.6% / 27.2%	8.3%
0.25	0.67	5 (71.4%)	0.52 [0.37; 0.74]	10.7% / 9.9%	7.1% / 7.1%	6.3%
0.25	0.67	10 (71.4%)	0.53 [0.41; 0.67]	11.8% / 13.1%	9.8% / 14.5%	8.0%
0.25	0.67	20 (69.0%)	0.53 [0.44; 0.62]	16.5% / 12.6%	13.5% / 15.6%	7.6%
0.25	0.67	50 (69.4%)	0.53 [0.48; 0.59]	25.8% / 19.7%	29.1% / 32.2%	15.6%
0.25	1.00	5 (45.5%)	0.65 [0.46; 0.83]	10.5% / 9.8%	7.3% / 7.3%	6.9%
0.25	1.00	10 (45.5%)	0.65 [0.50; 0.79]	11.3% / 13.1%	9.8% / 14.1%	8.3%
0.25	1.00	20 (43.5%)	0.64 [0.55; 0.74]	14.0% / 14.5%	14.3% / 20.9%	11.6%
0.25	1.00	50 (43.5%)	0.64 [0.58; 0.71]	20.5% / 25.9%	24.8% / 35.7%	17.1%

Table A.32: Simulation II, Scenario B: results under alternative hypothesis of strong selection bias ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.3; nominal significance level is 10%.

τ^2	ψ	k (% selected)	$\widehat{\psi}_{MH}$ 50% [5%; 95%]	Linear regression test ("add all") orig. / adjusted	Rank correlation test ("add all") orig. / adjusted	New test
0.00	0.25	5 (100.0%)	0.25 [0.20; 0.30]	9.5% / 10.5%	7.2% / 7.2%	7.1%
0.00	0.25	10 (100.0%)	0.25 [0.22; 0.28]	10.1% / 10.7%	6.7% / 6.7%	6.9%
0.00	0.25	20 (100.0%)	0.25 [0.23; 0.27]	8.7% / 8.4%	5.9% / 9.4%	6.1%
0.00	0.25	50 (98.0%)	0.25 [0.24; 0.26]	7.3% / 8.7%	5.4% / 8.2%	6.9%
0.00	0.50	5 (100.0%)	0.50 [0.43; 0.56]	10.1% / 9.9%	6.3% / 6.3%	6.0%
0.00	0.50	10 (90.9%)	0.50 [0.46; 0.54]	12.4% / 12.7%	8.9% / 9.5%	9.8%
0.00	0.50	20 (90.9%)	0.50 [0.47; 0.53]	16.5% / 18.6%	11.3% / 17.7%	12.5%
0.00	0.50	50 (90.9%)	0.50 [0.48; 0.52]	27.7% / 34.5%	23.9% / 28.0%	24.7%
0.00	0.67	5 (83.3%)	0.66 [0.57; 0.73]	12.6% / 14.5%	8.8% / 8.8%	8.3%
0.00	0.67	10 (83.3%)	0.66 [0.61; 0.71]	18.8% / 19.4%	13.6% / 13.6%	15.2%
0.00	0.67	20 (83.3%)	0.66 [0.62; 0.70]	32.6% / 31.0%	22.9% / 22.9%	23.7%
0.00	0.67	50 (82.0%)	0.66 [0.64; 0.68]	60.7% / 61.7%	52.9% / 61.7%	54.5%
0.00	1.00	5 (35.7%)	0.90 [0.77; 0.98]	23.7% / 23.1%	14.2% / 14.2%	13.5%
0.00	1.00	10 (37.0%)	0.90 [0.83; 0.96]	47.0% / 42.8%	32.0% / 32.0%	32.4%
0.00	1.00	20 (36.4%)	0.91 [0.86; 0.95]	78.6% / 77.8%	61.1% / 69.6%	61.3%
0.00	1.00	50 (35.8%)	0.90 [0.88; 0.93]	99.0% / 99.0%	95.3% / 96.8%	95.1%
0.05	0.25	5 (100.0%)	0.25 [0.19; 0.33]	8.2% / 7.9%	5.2% / 5.2%	4.7%
0.05	0.25	10 (100.0%)	0.25 [0.20; 0.30]	9.4% / 12.3%	5.1% / 7.9%	5.2%
0.05	0.25	20 (100.0%)	0.25 [0.22; 0.29]	8.8% / 10.6%	2.9% / 9.3%	2.9%
0.05	0.25	50 (98.0%)	0.25 [0.23; 0.27]	8.6% / 9.4%	5.2% / 11.5%	2.5%
0.05	0.50	5 (100.0%)	0.49 [0.37; 0.63]	7.6% / 8.4%	4.6% / 4.6%	5.0%
0.05	0.50	10 (90.9%)	0.49 [0.41; 0.59]	10.8% / 12.1%	5.6% / 9.6%	5.1%
0.05	0.50	20 (90.9%)	0.49 [0.43; 0.56]	13.2% / 11.9%	7.1% / 14.9%	5.4%
0.05	0.50	50 (89.3%)	0.49 [0.45; 0.54]	22.2% / 25.5%	14.3% / 32.0%	10.0%
0.05	0.67	5 (83.3%)	0.63 [0.50; 0.79]	11.6% / 12.6%	6.6% / 6.6%	6.7%
0.05	0.67	10 (76.9%)	0.64 [0.54; 0.74]	14.5% / 14.7%	7.3% / 12.3%	8.3%
0.05	0.67	20 (76.9%)	0.64 [0.56; 0.71]	19.5% / 21.7%	12.2% / 22.3%	12.4%
0.05	0.67	50 (76.9%)	0.64 [0.59; 0.69]	39.8% / 43.0%	36.8% / 55.1%	32.7%
0.05	1.00	5 (41.7%)	0.80 [0.65; 0.93]	12.0% / 14.6%	6.6% / 6.6%	5.7%
0.05	1.00	10 (40.0%)	0.80 [0.70; 0.90]	19.3% / 23.3%	12.6% / 24.5%	12.1%
0.05	1.00	20 (40.8%)	0.80 [0.73; 0.87]	26.6% / 35.3%	20.9% / 42.9%	19.9%
0.05	1.00	50 (40.0%)	0.80 [0.76; 0.85]	49.1% / 56.5%	54.7% / 74.5%	54.1%
0.25	0.25	5 (100.0%)	0.25 [0.15; 0.42]	8.1% / 8.0%	2.6% / 2.6%	3.0%
0.25	0.25	10 (100.0%)	0.25 [0.17; 0.36]	9.9% / 10.5%	2.8% / 6.1%	3.3%
0.25	0.25	20 (95.2%)	0.25 [0.19; 0.32]	11.5% / 7.8%	2.7% / 8.4%	2.2%
0.25	0.25	50 (96.2%)	0.25 [0.20; 0.30]	18.4% / 14.2%	6.6% / 14.7%	1.7%
0.25	0.50	5 (83.3%)	0.45 [0.26; 0.68]	11.5% / 12.7%	6.2% / 6.2%	5.5%
0.25	0.50	10 (83.3%)	0.44 [0.32; 0.60]	10.2% / 11.9%	4.9% / 9.4%	3.8%
0.25	0.50	20 (83.3%)	0.45 [0.35; 0.55]	13.7% / 16.1%	6.1% / 21.0%	3.4%
0.25	0.50	50 (82.0%)	0.45 [0.39; 0.51]	19.9% / 16.3%	15.3% / 27.1%	4.5%
0.25	0.67	5 (71.4%)	0.54 [0.35; 0.76]	10.3% / 13.7%	5.4% / 14.7%	4.8%
0.25	0.67	10 (71.4%)	0.55 [0.40; 0.71]	12.5% / 14.5%	6.6% / 17.8%	5.3%
0.25	0.67	20 (69.0%)	0.54 [0.44; 0.66]	14.7% / 15.5%	7.9% / 21.0%	5.1%
0.25	0.67	50 (68.5%)	0.54 [0.47; 0.61]	20.5% / 21.1%	22.3% / 49.4%	12.9%
0.25	1.00	5 (45.5%)	0.66 [0.44; 0.85]	9.0% / 10.8%	5.4% / 5.4%	5.1%
0.25	1.00	10 (45.5%)	0.67 [0.50; 0.81]	11.9% / 14.9%	6.3% / 17.1%	4.8%
0.25	1.00	20 (44.4%)	0.66 [0.54; 0.76]	13.0% / 17.0%	10.1% / 29.1%	8.4%
0.25	1.00	50 (43.9%)	0.66 [0.59; 0.73]	18.3% / 24.0%	29.3% / 57.6%	22.4%

Table A.33: Simulation II, Scenario C: results under alternative hypothesis of strong selection bias ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.1; nominal significance level is 10%.

τ^2	ψ	k (% selected)	$\widehat{\psi}_{MH}$ 50% [5%; 95%]	Linear regression test ("add all") orig. / adjusted	Rank correlation test ("add all") orig. / adjusted	New test
0.00	0.25	5 (100.0%)	0.25 [0.23; 0.27]	11.0% / 10.3%	7.1% / 7.1%	7.4%
0.00	0.25	10 (100.0%)	0.25 [0.23; 0.27]	7.3% / 7.2%	7.2% / 9.0%	6.6%
0.00	0.25	20 (100.0%)	0.25 [0.24; 0.26]	10.4% / 10.5%	6.1% / 9.8%	6.0%
0.00	0.25	50 (100.0%)	0.25 [0.24; 0.26]	8.6% / 9.2%	6.0% / 9.2%	6.3%
0.00	0.50	5 (100.0%)	0.50 [0.46; 0.54]	9.1% / 9.0%	5.8% / 5.8%	5.6%
0.00	0.50	10 (100.0%)	0.50 [0.47; 0.53]	9.9% / 9.4%	6.9% / 6.9%	7.3%
0.00	0.50	20 (95.2%)	0.50 [0.48; 0.52]	12.1% / 13.2%	7.0% / 12.0%	7.0%
0.00	0.50	50 (98.0%)	0.50 [0.49; 0.51]	13.4% / 13.2%	9.3% / 13.2%	9.4%
0.00	0.67	5 (100.0%)	0.67 [0.61; 0.72]	11.4% / 13.0%	8.3% / 8.3%	7.7%
0.00	0.67	10 (90.9%)	0.67 [0.63; 0.70]	11.1% / 11.5%	7.8% / 10.9%	7.6%
0.00	0.67	20 (90.9%)	0.67 [0.64; 0.69]	17.3% / 18.2%	11.7% / 15.8%	11.5%
0.00	0.67	50 (90.9%)	0.67 [0.65; 0.68]	30.8% / 29.9%	25.5% / 28.3%	25.2%
0.00	1.00	5 (38.5%)	0.93 [0.85; 0.99]	25.1% / 24.7%	14.3% / 14.3%	13.6%
0.00	1.00	10 (37.0%)	0.93 [0.88; 0.97]	49.1% / 50.0%	33.8% / 40.3%	34.3%
0.00	1.00	20 (35.7%)	0.94 [0.91; 0.96]	79.3% / 81.7%	60.2% / 69.3%	61.2%
0.00	1.00	50 (35.7%)	0.94 [0.92; 0.95]	98.4% / 98.0%	95.4% / 96.8%	95.3%
0.05	0.25	5 (100.0%)	0.25 [0.19; 0.33]	8.3% / 10.2%	3.9% / 3.9%	3.9%
0.05	0.25	10 (100.0%)	0.25 [0.21; 0.30]	8.6% / 9.1%	3.4% / 6.6%	3.4%
0.05	0.25	20 (100.0%)	0.25 [0.22; 0.29]	9.5% / 11.0%	3.1% / 7.6%	3.5%
0.05	0.25	50 (100.0%)	0.25 [0.23; 0.27]	8.7% / 10.7%	3.8% / 12.4%	2.5%
0.05	0.50	5 (100.0%)	0.49 [0.39; 0.62]	9.5% / 10.4%	4.0% / 4.0%	4.0%
0.05	0.50	10 (100.0%)	0.50 [0.42; 0.59]	9.1% / 10.8%	3.8% / 8.2%	3.4%
0.05	0.50	20 (95.2%)	0.50 [0.44; 0.56]	7.7% / 8.8%	2.7% / 13.0%	1.9%
0.05	0.50	50 (96.2%)	0.50 [0.46; 0.54]	9.8% / 11.5%	4.6% / 13.7%	3.4%
0.05	0.67	5 (83.3%)	0.64 [0.51; 0.79]	9.4% / 11.0%	5.0% / 5.0%	4.6%
0.05	0.67	10 (83.3%)	0.64 [0.55; 0.75]	10.1% / 11.2%	4.7% / 8.8%	4.2%
0.05	0.67	20 (87.0%)	0.64 [0.58; 0.72]	12.1% / 14.0%	5.3% / 16.9%	4.8%
0.05	0.67	50 (84.7%)	0.65 [0.60; 0.69]	18.9% / 23.5%	14.4% / 30.0%	12.9%
0.05	1.00	5 (45.5%)	0.82 [0.69; 0.93]	11.5% / 12.5%	6.3% / 6.3%	5.9%
0.05	1.00	10 (41.7%)	0.82 [0.73; 0.91]	14.2% / 16.2%	8.2% / 12.7%	8.5%
0.05	1.00	20 (41.7%)	0.82 [0.75; 0.89]	19.3% / 23.0%	13.7% / 28.4%	13.4%
0.05	1.00	50 (42.0%)	0.82 [0.77; 0.86]	29.4% / 35.9%	36.3% / 63.9%	35.2%
0.25	0.25	5 (100.0%)	0.25 [0.15; 0.41]	8.3% / 10.6%	3.4% / 3.4%	4.0%
0.25	0.25	10 (100.0%)	0.25 [0.17; 0.37]	9.2% / 7.8%	2.1% / 5.2%	2.5%
0.25	0.25	20 (100.0%)	0.26 [0.20; 0.33]	9.3% / 8.1%	2.0% / 10.8%	2.6%
0.25	0.25	50 (98.0%)	0.26 [0.22; 0.30]	12.2% / 9.3%	1.8% / 9.8%	2.2%
0.25	0.50	5 (83.3%)	0.46 [0.29; 0.68]	8.5% / 7.7%	3.6% / 3.6%	3.6%
0.25	0.50	10 (90.9%)	0.46 [0.33; 0.62]	11.1% / 11.1%	4.4% / 11.1%	4.1%
0.25	0.50	20 (87.0%)	0.46 [0.36; 0.57]	10.7% / 10.9%	5.4% / 17.4%	3.0%
0.25	0.50	50 (86.2%)	0.46 [0.39; 0.53]	11.4% / 11.9%	9.5% / 29.4%	4.0%
0.25	0.67	5 (71.4%)	0.56 [0.36; 0.77]	10.0% / 10.9%	4.3% / 4.3%	3.9%
0.25	0.67	10 (71.4%)	0.56 [0.41; 0.73]	9.7% / 14.3%	5.2% / 14.1%	4.0%
0.25	0.67	20 (74.1%)	0.55 [0.45; 0.67]	12.7% / 13.2%	5.6% / 14.8%	3.9%
0.25	0.67	50 (72.5%)	0.56 [0.48; 0.63]	12.1% / 15.5%	13.9% / 32.3%	8.3%
0.25	1.00	5 (50.0%)	0.67 [0.46; 0.87]	7.9% / 9.5%	4.5% / 11.2%	4.3%
0.25	1.00	10 (45.5%)	0.67 [0.51; 0.81]	8.6% / 9.8%	5.5% / 9.7%	4.2%
0.25	1.00	20 (46.5%)	0.67 [0.56; 0.78]	10.5% / 15.3%	7.2% / 23.0%	6.1%
0.25	1.00	50 (45.9%)	0.67 [0.60; 0.74]	12.1% / 13.9%	22.8% / 54.0%	16.8%

Table A.34: Simulation II, Scenario C: results under alternative hypothesis of strong selection bias ($n = 1000$); proportion of simulated meta-analyses with significant result of two-sided test for bias; odds ratio as measure of treatment effect; control event rate of 0.3; nominal significance level is 10%.