THE IMPACT OF ADJUSTMENT FOR COVARIATES ON META-ANALYSIS OF RANDOMISED INTERVENTION STUDIES FOR BINARY OUTCOME

Ly-Mee Yu

Wolfson College

Doctor of Philosophy Thesis

Trinity Term 2015

ABSTRACT

The Impact of Adjustment for Covariates on Meta-Analysis of Randomised Intervention Studies for Binary Outcome

Ly-Mee Yu
Wolfson College
DPhil thesis, Trinity Term, 2015

Background

Covariate adjustment analysis is often used in epidemiological studies but is less common in randomised controlled trials (RCTs) and RCT meta-analyses. There is a lack of consensus on whether the analysis of RCT data should adjust for important baseline covariates. The estimated treatment effect of a binary covariate can differ when logistic regression is carried out, even when the covariate is balanced between treatment groups.

Objectives

The objectives of this study were to examine the factors that affect the impact of adjusted analysis in different RCT scenarios and to explore the impact of adjusted analysis in RCT meta-analysis.

Methods

Simulation and sampling studies were conducted to identify the factors that affect the impact of using an adjusted logistic regression model. Two covariates, one continuous and one binary, were considered simultaneously. The event rate, treatment effect, binary and continuous variable distributions, covariate prognostic strengths, and correlation between the covariates were varied during the simulations. The impact of adjustment on RCT meta-analysis was investigated using individual participant data obtained from the Perinatal Antiplatelet Review of International Studies. Different methods of performing unadjusted and adjusted meta-analysis were compared.

Results

The simulation results suggest that adjustment only has a notable effect in extreme scenarios, such as a very large treatment effect or highly prognostic covariates. The relative difference between the unadjusted and adjusted odds ratios was found to be larger than 50% under these extreme scenarios. Covariate adjustment is likely to have a small effect on meta-analyses with many studies.

Summary

Adjusted analysis should be carried out by design. Performing adjusted analysis in a meta-analysis can be challenging as sufficient information about the covariates is often not available.

"Now to Him who is able to do exceedingly abundantly above all that we ask or think, according to the power that works in us…"

Ephesians 3:20 (NKJV)

"I will praise You, O Lord my God, with all my heart,

And I will glorify Your name forevermore."

Psalms 86:12 (NKJV)

ACKNOWLEDGEMENTS

Undertaking this DPhil has been a long and challenging journey for me. It would not have been possible without the support and encouragement that I received from many people.

I would like to thank my supervisor, Professor Doug Altman, for his patient guidance and advice throughout my studies. It has been a great privilege to work with him for 10 years at the Centre for Statistics in Medicine, University of Oxford. My sincere thanks to Associate Professor Gary Collins, Mrs Merryn Voysey and Associate Professor Jonathan Cook for their chapter reviews and constructive comments. I would also like to thank Professor Jonathan Deeks for his advice during the early stage of this project, and the Nuffield Department of Primary Care Health Sciences for supporting the completion of this thesis.

A very special thank you to Dr Jennifer de Beyer for proofreading in such a short space of time and for hugely improving my thesis. I am also very thankful to the following friends and colleagues for their support of the thesis in various ways: Mrs Nicola Williams, Ms Siu Wah Wong, Ms Candice CH Lam, Ms Barbara Farrell, Ms Patsy Spark, Dr Iveta Simera, Dr Sally Hopewell, and Dr Sue Mallett.

I am indebted to Professor James YW Lau, at the Prince of Wales Hospital in Hong Kong, for his generous support of this DPhil. I am also grateful to members of the Chinese Church in London and the SoBlessed Student Fellowship for their endless prayers and encouragement throughout the past 10 years.

Finally, my heartfelt thanks to my parents and siblings (Esther, Teresa, and Simon) for simply being there for me all along.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

LIST OF TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

LIST OF ABBREVIATIONS

Abbreviation	Full term
ART	Arterial Revascularisation Trial
CLASP	Collaborative Low-dose Aspirin Study in Pregnancy
CI	Confidence interval
CONSORT	Consolidated Standards of Reporting Trials
ECPPA	Estudo Colorativo para Prevencao da Pre-eclampsia com Aspirina
ERASME	Essai Regional Aspirine Mere-Enfant study
IPD	Individual participant data
ISAT	International Subarachnoid Aneurysm Trial
MEADOW	Memantine for Dementia in Adults with Down's Syndrome study
PARIS	Perinatal Antiplatelet Review of International Studies
PSAO	Pregnancy with a serious adverse outcome
RCT	Randomised controlled trial
ROR	Ratio of odds ratios
RPOR	Ratios of pooled odds ratios
SE	Standard error
SST	Spine Stabilisation Trial

CHAPTER 1: INTRODUCTION

1.1 Covariate Adjustment in Randomised Controlled Trials

The randomised controlled trial (RCT) is the established 'gold standard' study design for determining the effect of health care interventions. Randomisation prevents bias in the allocation of patients to treatment groups. It provides unbiased estimates of treatment effects, but does not guarantee comparability between groups, particularly in small trials. Covariate adjustment is less common in RCTs than in epidemiological studies as the latter are prone to substantial bias when unadjusted.

Covariate adjustment should be considered in RCTs (Assmann et al, 2000; Altman, 1985) as it can correct baseline covariate imbalances remaining after randomisation (Senn, 1989), increase statistical power by modelling the variability in the outcome explained by relationships with highly prognostic covariates (Pocock et al, 2002; Altman, 1985), obtain treatment effect estimates that are more relevant for an individual patient (Hauck, Anderson and Marcus, 1998), and account for study design features, such as the use of stratified randomisation for treatment allocation (Kahan and Morris, 2012). Guidelines suggest that adjustment should be pre-specified in the trial protocol (Altman et al, 2001; ICH E9 Expert Working Group, 1999; European Medicines Agency Committee for Proprietary Medicinal Products, 2015).

In practice, however, adjustment is often only performed if baseline imbalance in a covariate is observed. Although a conditional decision to adjust an analysis based on the observation of baseline imbalance is made to remove bias, it can also introduce bias (Pocock et al, 2002). Covariates that are prognostic of the outcome but balanced across randomised groups are rarely routinely adjusted for in RCT analyses. The gain in efficiency obtained by adjusting the analysis depends on the strength of the correlation between the covariates and outcome (Altman, 2005). When procedures such as stratified randomisation or minimisation are performed, analysis without adjustment for stratifying variables may overestimate the standard error (SE) of the treatment effect and distort the P-value (Forsythe and Stitt, 1977; Brennan and Morris, 2012). Covariates that are measured after randomisation in RCTs are considered to be more complicated to analyse because their relationship with the outcome can be confounded by the treatment effect. These types of variables require special attention and need a different analytical approach (Rochon, 1999).

The Consolidated Standards of Reporting Trials (CONSORT) (Altman et al, 2001) statement provides guidance on the reporting of adjusted analyses. It recommends specifying the rationale for any adjusted analysis and the statistical methods used, and clarifying which variables are adjusted. However, information on the extent and quality of such practices in published papers is lacking.

The decision to adjust for baseline covariates in an RCT is a complex one. Unadjusted and adjusted analyses are not directly comparable; unadjusted analyses give population-averaged estimates of the treatment effect, whereas adjusted analyses assess subgroup-specific estimates. The two measures can have the same quantity for continuous outcomes with a normal distribution. Investigators often do not take into account that when an RCT is conducted on a select group of participants, the unadjusted population-average estimates are comparable to estimates that are conditional on specific inclusion and exclusion criteria.

Hauck, Anderson, and Marcus (1998) concluded that adjusted analysis leads to both greater efficiency in hypothesis tests of treatment effect and improved external validity of trials. Although they recommended that the primary analysis should be adjusted for important prognostic covariates to approach the most clinically relevant subject-specific measure of treatment effect, this practice is still uncommon. It is likely that most researchers believe that unadjusted analysis is simpler and more transparent, and that interpreting unadjusted results is more straightforward than interpreting adjusted results.

1.2 Covariate Adjustment for Binary Outcomes

The primary aim of analysing RCT data is to achieve an unbiased, statistically efficient treatment comparison that takes into account baseline factors that predict prognosis, especially those factors that are unbalanced between treatment groups. Adjusting for covariates in continuous data analysis improves

the statistical power of hypothesis tests without altering the magnitude of the treatment effect estimate. However, omitting highly prognostic covariates when analysing binary and time-to-event outcomes can yield an estimated treatment effect that is attenuated towards the null even when these covariates are balanced between the treatment groups (Gail, Wieand and Piantadosi, 1984).

To illustrate this phenomenon, consider an extreme example of an RCT with 400 patients (Tabl1-1), 200 randomised to the treatment and 200 to the placebo, with equal numbers of males (n = 200) and females (n = 200) in each treatment group. Let the overall mortality at 30 days be 40% in the treatment group and 60% in the placebo group, giving a crude odds ratio for the treatment effect of 0.44. Suppose that the mortality rates for males are 10% and 30% in the treatment and placebo groups, respectively, and are 70% and 90% for females. The odds ratio for the treatment effect in males alone is 0.26. The same odds ratio is observed for females alone.

Table 1-1 Hypothetical example of a clinical trial stratified by sex

	Total			Male		Female	
	Dead	Alive	Total	Dead	Alive	Dead	Alive
Treatment	80	120	200	10	90	70	30
Placebo	120	80	200	30	70	90	10
	Odds ratio = 0.44		400	Odds ratio = 0.26		Odds ratio = 0.26	

This phenomenon is referred to as ‘noncollapsibility’ (Greenland, Robins and Pearl, 1999; Guo and Geng, 1995) or ‘mavericks’ (Hauck et al, 1991), and only arises in nonlinear models such as logistic and Cox’s proportional hazard models. Adjusting for these covariates can correct for this ‘heterogeneity bias’ (Steyerberg and Eijkemans, 2004). The conventional adjusted analysis method for a binary outcome is logistic regression. This method induces a gain in statistical power (Hernandez, Steyerberg and Habbema, 2004; Turner et al, 2012a; Kahan et al, 2014), but also reduces the precision of the treatment effect estimate (Chastang, Byar and Piatadosi, 1988; Deeks et al, 2003; Robinson and Jewell, 1991). The loss of precision is greater when the association between the outcome and covariate is strong, or when the outcome event is relatively common. However, the increase in the treatment effect has been shown to dominate the increase in precision, giving an overall increase in statistical power (Steyerberg, Bossuyt and Lee, 2000).

Other simulation studies have shown that adjusting for one baseline prognostic covariate in an RCT with binary outcomes leads to an increase in the power to detect treatment effects. However, the notable effects on the size of the treatment effect have only been recorded when a very strong prognostic covariate was included in the analysis (Chu et al, 2012; Hernandez, Steyerberg and Habbema, 2004; Negassa and Hanley, 2007). It is not known whether the extent of this effect will change when multiple covariates with different degrees of imbalance and different directions in the prognostic relationship to the outcome are included in the adjusted analysis.

1.3 Covariate Adjustment in Meta-analysis of Randomised Controlled Trials

Meta-analyses of estimates from RCTs generally use unadjusted trial estimates of treatment effects even when the original studies present adjusted analyses. This approach is undertaken for practical reasons; for example, adjusted analyses may not be routinely presented, or each study may adjust for different covariates. However, the approach is often supported by the argument that baseline covariate imbalances across all of the trials included in a meta-analysis are unlikely to be important due to the large overall sample size. Most meta-analyses for Cochrane reviews are carried out using software (e.g., RevMan) that requires the input of raw counts of events in each treatment group and calculates unadjusted comparisons only. The covariates available for adjustment may vary even when individual participant data (IPD) are available, so standardised adjustment may not be possible (Simmonds et al, 2005). It is thus customary to assume, if only implicitly, that it is better to use unadjusted RCT results in meta-analyses. This is the recommended approach of the Cochrane Collaboration, for example.

We know that adjustment tends to lead to larger effect sizes in individual trials. However, there is no empirical evidence of the impact of adjustment on meta-analyses, nor how the impact relates to the nature of the covariates and degree of imbalance. Trowman et al (2007) presented an illustrative case study

demonstrating that the pooled treatment effect could be misleading if the baseline imbalance within studies was not accounted for. Their findings show that we need to explore whether ignoring adjusted analyses is a reasonable strategy in meta-analysis.

Adjusted analysis issues are much more complex in meta-analysis due to the diversity of the studies involved. Studies can vary in their inclusion/exclusion criteria, sample size, and randomisation method, all of which may affect the distribution of the covariates between treatments (Steyerberg, Bossuyt and Lee, 2000; Zhang et al, 2005). Nicholl (2007) pointed out the problem of the 'constant risk fallacy' in non-randomised studies: the relationship between a covariate and an outcome is assumed to be similar across the populations, and the possible differences between them are ignored. This fallacy may be extended to the context of meta-analysis, in which the relationship between the outcome and covariates can differ across trials. Adjusting for covariates may exacerbate the difference in the treatment effect estimate.

The statistical properties of covariate adjustment are complex and often poorly understood. There remains confusion as to the most appropriate statistical strategy for selecting which covariates to adjust for. In meta-analysis, the effect of using alternative statistical strategies to perform adjusted analysis has not been explored. The conventional approaches for analysing IPD are multilevel modelling and two-stage methods, in which an adjusted analysis is performed on

each study separately, then the summary statistics are combined using standard summary data meta-analysis techniques.

1.4 Aims

The research project reported in this thesis comprises three research questions to investigate the reporting and role of adjusted analysis in RCTs and meta-analysis:

1. How often are adjusted analyses used in RCTs, and has there been any improvement in the use and reporting of adjusted analysis since the revision of the CONSORT Statement in 2001?
2. What is the impact on the treatment effect and its confidence interval (CI) when one or more prognostic covariates are included in the analysis of an RCT?
3. What is the impact of covariate adjustment, and which approach should be considered when performing RCT meta-analysis?

1.5 Thesis Structure

The rest of this thesis is organised as follows. Chapter 2 reports the results of a literature review examining how frequently adjusted analysis is used in RCTs and how well it is reported. The review was carried out using two cohorts of RCT reports published in December 2000 and December 2006.

Detailed examinations of the factors affecting the impact of adjusted analysis are described in Chapters 3 and 4. Chapter 3 reports on clinical trial scenarios consisting of the treatment, one binary covariate, and one continuous covariate, generated using an adjusted logistic regression model. Chapter 4 extends the investigation in Chapter 3 with additional data sets to assess how the impact of adjustment changes with the degree of correlation between the covariates.

Chapter 5 examines the effect of sample size and imbalance on adjusted analysis using resampling and the data reported in Chapter 3. Chapter 6 investigates the effect of covariate adjustment on RCT meta-analysis using a case study. Chapter 7 summarises the research findings and the issues raised in the previous chapters. It presents conclusions on whether adjusted analysis should be carried out in RCTs and meta-analyses. The limitations of this study and areas of future research are also discussed in Chapter 7.

CHAPTER 2: REVIEW OF COVARIATE ADJUSTMENT IN RANDOMISED CONTROLLED TRIALS

2.1 Introduction

The adjustment of baseline covariates is less common in the analysis of RCTs than of epidemiological studies. However, little is known about the frequency and quality of reporting of adjusted analyses in RCTs. The objective of this review is to compare the frequency of adjusted analyses and the quality of their reporting in two cohorts of published RCTs, before and after the 2001 revision of the CONSORT Statement (Altman et al, 2001).

2.2 Methods

2.2.1 Study selection

Two cohorts of published papers of RCTs were used, (1) a review of a previously identified cohort indexed in PubMed (http://www.ncbi.nlm.nih.gov/pubmed) in December 2000 (Chan and Altman, 2005a; Chan and Altman, 2005b) and (2) a new set of reports of RCTs indexed in PubMed in December 2006, identified using the same search strategy as those identified in 2000 (Appendix A). Both cohorts were identified by searching PubMed using the extended version of Phase 1 of the Cochrane highly sensitive search strategy for trials (Robinson and Dickersin, 2002). The abstracts of the search results for those articles identified in December 2006 were screened. Articles that were obviously not trials were

excluded. The full text of the remaining papers was reviewed to assess their eligibility.

Only papers that met the criteria described above and described RCTs of a parallel group design that were carried out in humans and published in a full publication in English were included. Studies that were published as a letter or brief communication were excluded from the review because they were unlikely to contain all of the information required due to their space constraints. Papers reporting phase I or pilot studies were also excluded. These studies usually have small samples, so their formal analyses are usually limited to descriptive estimates or unadjusted analyses. The main comparison in any RCT should be a direct contrast on the primary outcome between the randomised treatment groups. Studies that did not provide sufficient information about their statistical analysis methods or did not perform a formal comparison between treatment groups were excluded.

2.2.2 Defining adjusted analysis

Papers with trial outcomes that were explicitly reported, in either the methods or results section of the article, to have undergone adjusted analysis for comparisons between randomised groups were defined as including adjusted analyses. Relevant here were references to accounting for covariates in the statistical analysis of the treatment effect or an explicit statement that some

results were adjusted. Analyses that used multiple regression methods to identify prognostic variables or risk factors were not defined as adjusted analyses, as their purpose is not to estimate a direct treatment effect.

2.2.3 Data extraction

2.2.3.1 Study characteristics

Information on the trial characteristics and all outcomes were extracted from the 2006 cohort using the same criteria as those in the 2000 cohort (Chan and Altman, 2005b). An outcome was defined as the primary outcome if this was explicitly specified in the article, it was used in the power calculation or it was explicitly described in the primary study objectives. Multicentre involvement was defined as data being collected from more than one study site. Sample size was defined as the total number of participants randomised in the study. The number of intervention groups, whether the study performed stratified randomisation, whether it was published in a general medicine or specialist journal, the number of outcomes collected, and the type of data for each outcome (binary, continuous or time-to-event) were also collected from each study.

2.2.3.2 Details of the adjusted analysis

One main outcome was selected from each trial to maintain independence between the observations. If more than one outcome underwent adjusted

analysis, then the outcome for use in the review was selected according to the following hierarchy: (1) it was a pre-specified primary outcome, (2) the sample size of the trial was based on the outcome, or (3) the method used to adjust for this outcome was most completely reported. If more than one outcome was equally reported within an article, then the outcome for the review was chosen at random.

2.2.3.2.1 Types of analysis

The methods used to adjust an analysis that were explicitly reported in the methods and results sections were assessed for the papers in both cohorts that described adjusted analysis. Papers were classified as reporting, for the selected outcome, only unadjusted analysis, only adjusted analysis, both, or unspecified/unclear.

2.2.3.2.2 Reasons for adjustment

Each paper's stated reason for using adjusted analysis was recorded. The possible reasons were covariate imbalance across the intervention groups, covariates having prognostic importance to the outcome, or both.

2.2.3.2.3 Covariate details

Each paper was searched for information about the covariates included in the adjusted analysis. Information was often found in the methods section, but was also sometimes embedded in the footnote of a table in the results section. The reason for choosing the covariates used was recorded as explicitly pre-specified in the methods or a published protocol, due to observed baseline imbalance, or based on analysis exploring the association between the covariate and outcome (i.e., suggested by the data). The number of covariates used in the adjusted analysis, whether they were used for stratified randomisation or minimisation, and whether the adjusted analysis included measures collected after randomisation as covariates was also recorded.

2.2.3.2.4 Methods used in the adjusted analysis

The methods used for adjusting for covariates were collected from the papers. If more than one method was mentioned for the same adjusted analysis, then the first method mentioned was included in the review.

2.2.3.2.5 Reporting of adjusted analysis

The methods used to report the unadjusted and adjusted results were recorded, such as the summary statistics, within-group CIs and SEs, between-group treatment effects, CI and SE of the treatment effect, and corresponding P-value.

The degree of adherence of the reporting of the adjusted analyses to the 2001 CONSORT guideline was assessed. For the 2006 cohort, whether each paper was published in a CONSORT-endorsing journal was recorded, based on the journals' 'Instruction to Authors' (assessed June 2008). Trial characteristics, details of outcome measures, and information on adjusted analysis were also extracted.

2.2.4 Data synthesis and analysis

The frequency of adjusted analysis was expressed as the proportion of identified trials that described using adjusted analysis. The two cohorts' trial characteristics and adherence to the 2001 CONSORT Statement were compared using a chi-square test or a Fisher's exact test (if the expected counts were less than five) for categorical data, and a Mann-Whitney test for continuous data. Percentage differences and precision based on 95% CIs were calculated to quantify the change in reporting between 2000 and 2006. Similar analyses were used to compare the trial characteristics of the trials that did and did not report adjusted analysis within each cohort. The data were analysed using Stata version 9 (Stata Corporation, College Station, TX, USA). A P-value of less than 0.05 was considered to indicate statistical significance.

2.3 Results

2.3.1 Trial characteristics

In December 2006, 1,735 citations were identified in PubMed, of which 616 papers were included in the review. The details of the included and excluded articles are shown in Figure 2-1. In 2000, 519 articles were retrieved from 326 journals, yielding 355 parallel group studies for inclusion in the review. In 2006, 616 articles were retrieved from 316 journals, yielding 421 parallel group studies for inclusion (Figure 2-2). Most studies in both cohorts were characterised by two study arms (74% in 2000 vs 78% in 2006), a single study centre (66% in both cohorts) and publication in specialty journals (91% in both cohorts) (Table 2-1). There was a significant increase in the proportion of articles that specified the primary outcome between 2000 and 2006 (51% vs 65%, respectively; $P < 0.0001$). The average sample size and number of trial outcomes were similar in the two cohorts, and most reported outcomes were continuous (about 70%). Fewer studies performed stratified randomisation in 2000 than in 2006 (16% vs 20%, respectively, $P = 0.1$).

Eighty-four articles (24%) in 2000 and 113 articles (27%) in 2006 reported performing adjusted analysis on at least one outcome in the methods, results, or both sections.

2.3.2 Characteristics of trials that did and did not report adjusted analysis

In both cohorts, there was a marked difference between the characteristics of the studies that did and did not report adjusted analysis. A higher proportion of the articles that reported adjusted analysis specified their primary outcomes, involved multiple centres, performed stratified randomisation, and were published in general journals. Trials with adjusted analysis recruited more participants and had fewer outcomes (Figure 2-3).

2.3.3 Consistency of analysis reporting

Among the adjusted analysis papers, 6 (7%) and 4 (4%) papers did not provide details of the statistical methods used for analysis in 2000 and 2006, respectively. The consistency of the type of analysis reported in the methods and results sections was examined for the outcome selected in each trial.

In 2000, 43 out of 79 papers (54%) explicitly specified adjusted analysis in the methods and reported them in the results. Inconsistencies were found between the information reported in the methods and results sections of 36 papers (46%). For example, two papers specified adjusted analysis in the methods, but only reported unadjusted results. Twenty-four (30%) articles did not clearly specify the type of analysis used in the results section.

In 2006, the analysis reporting consistency increased to 69% (74 out of 109 papers, P = 0.06, difference [95% CI] = 13.5% [-0.6% to 27.5%]). Fewer papers did not specify clearly the type of analysis used in the results section for the selected outcome in 2006 (19 out of 109 papers, 17%, P = 0.04, difference [95% CI] = -12.9 [-25.3 to -0.6]). Three articles specified using adjusted analysis in the methods but only reported unadjusted results. The authors of the 19 papers with an inconsistency between the methods and results sections, or unclear results reporting were contacted for clarification, but only three responded.

Figure 2-1 Flow diagram of the articles eligible for review in December 2006

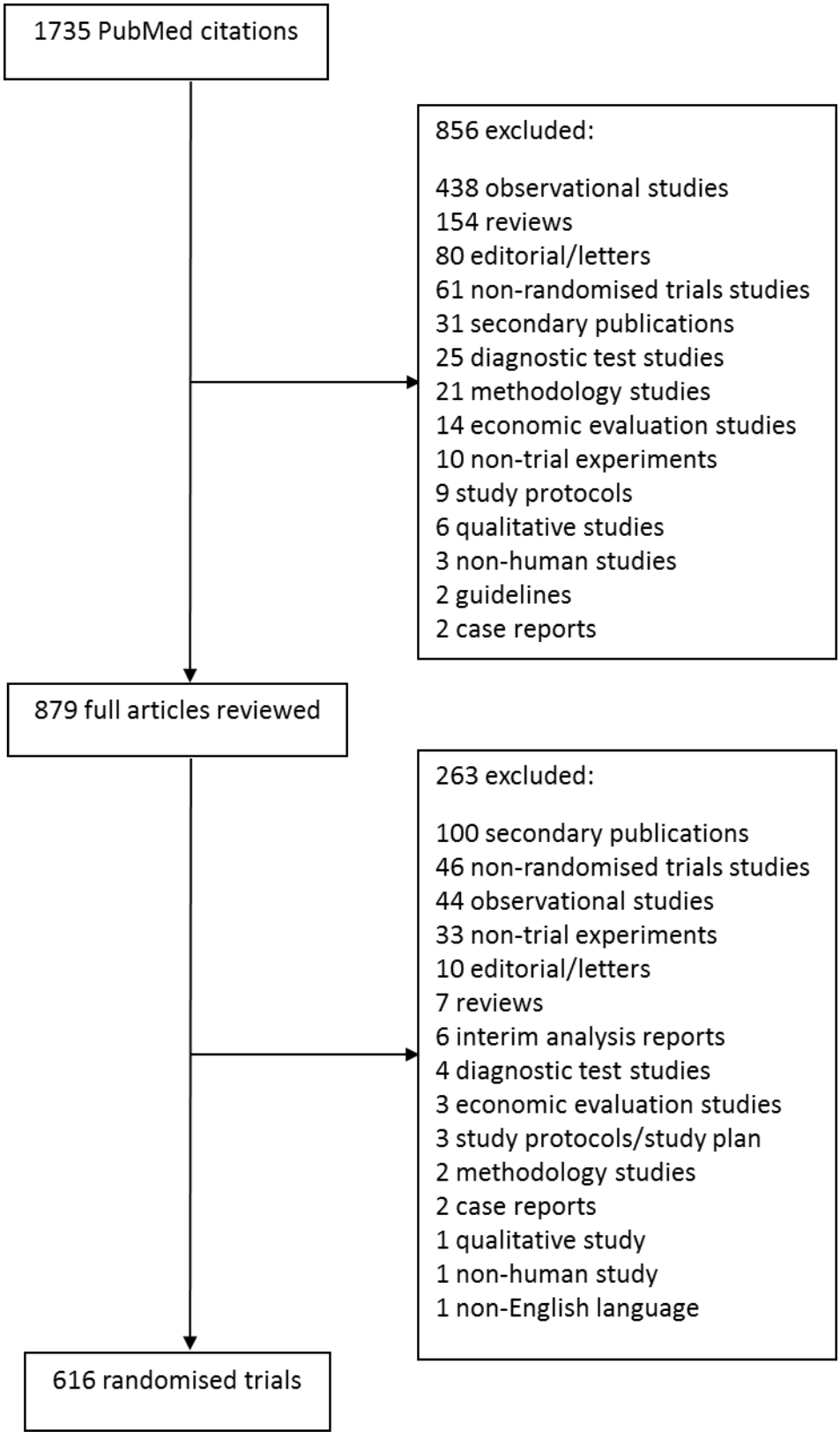
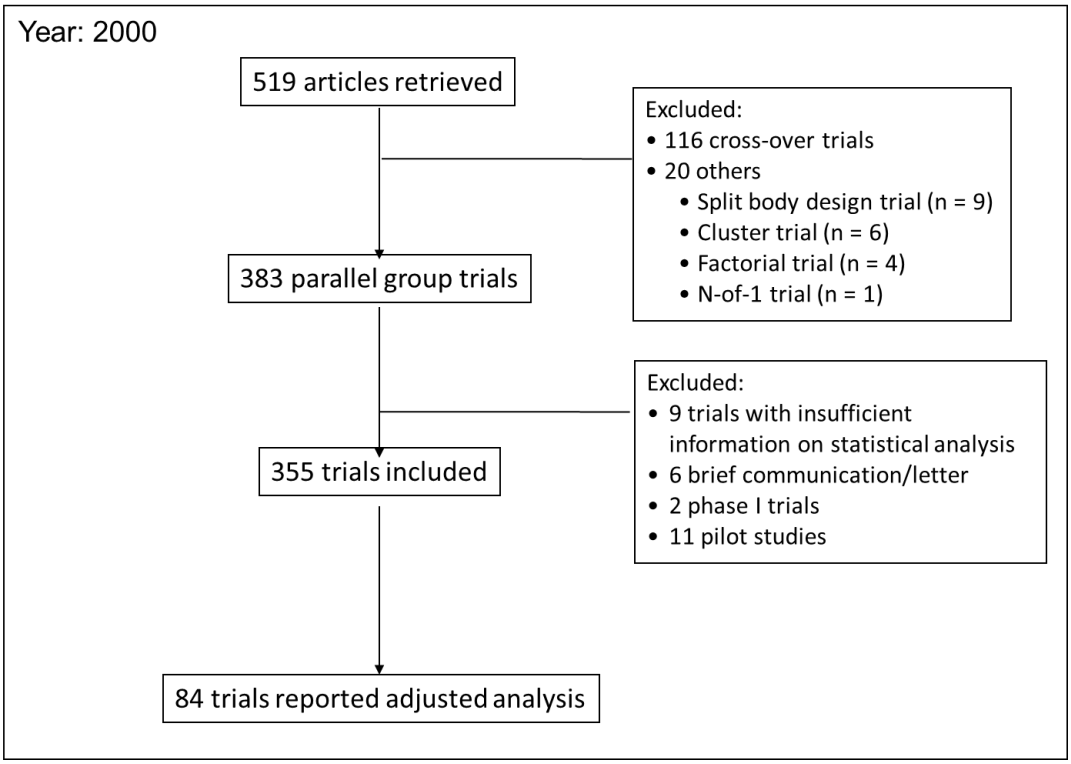


Figure 2-2 Flow diagram of the articles retrieved and included in the review



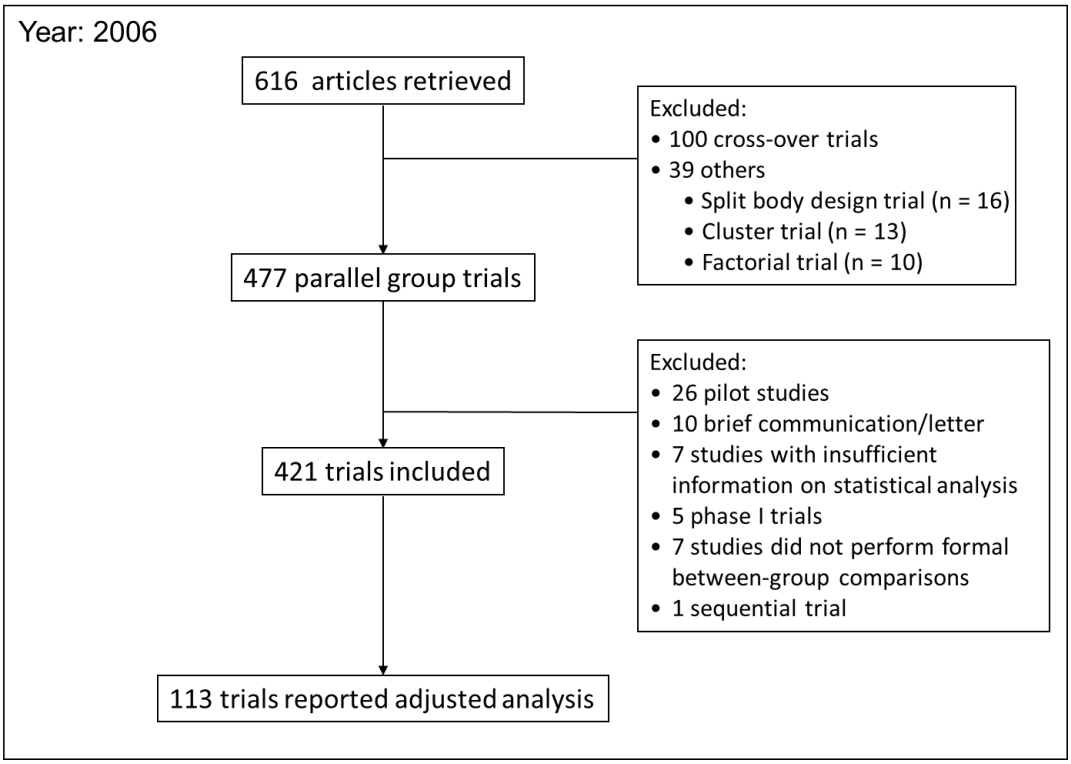


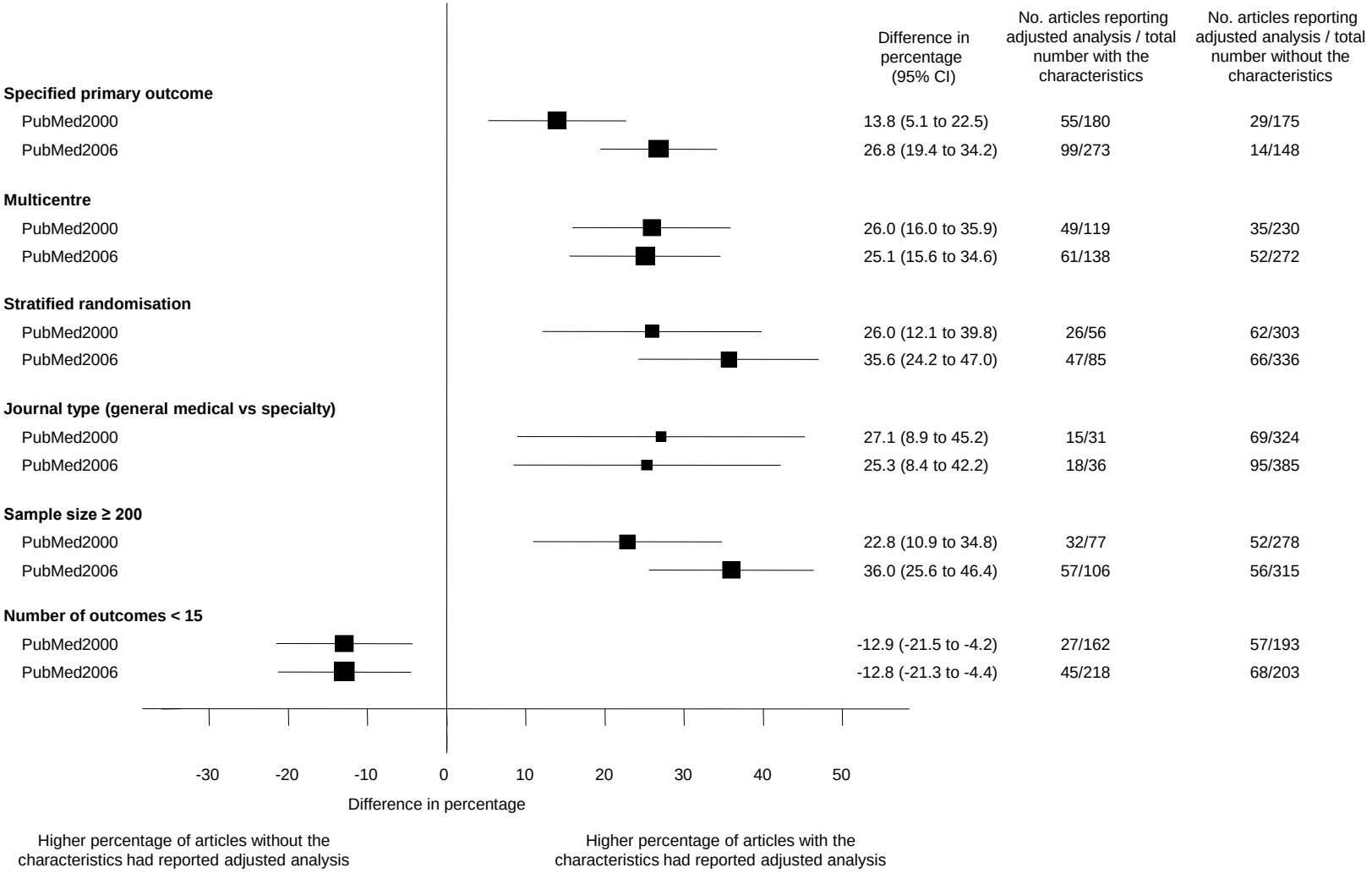
Table 2-1 Characteristics of reports of parallel-group randomised trials by year of publication

Year of Publication	2000 (n = 355)	2006 (n = 421)	% difference† (95% CI)	P-value
Outcome specification				
Primary	180 (50.7%)	273 (64.9%)	14.1 (7.2 to 21.0)	< 0.0001
Unspecified	175 (49.3%)	148 (35.1%)		
Centres involved*				
Multiple centres	119 (34.1%)	138 (33.7%)	-0.4 (-7.2 to 6.3)	0.9
Single centre	230 (65.9%)	272 (66.3%)		
Number of intervention groups	261 (73.5%)	328 (77.9%)		0.2
2	57 (16.1%)	64 (15.2%)		
3	37 (10.4%)	29 (6.9%)		
> 3				
Performed stratified randomisation	56 (15.8%)	85 (20.2%)	4.4 (-1.0 to 9.8)	0.1
Sample size				
< 50	116 (32.7%)	129 (30.6%)		
51 – 150	141 (39.7%)	169 (39.9%)		
151 – 300	49 (13.8%)	52 (12.4%)		
301 – 450	20 (5.6%)	27 (6.6%)		
> 450	29 (8.2%)	44 (10.5%)		
Median (10th to 90th percentile)	91 (27 to 394)	80 (28 to 462)		0.7
Journal type				
General medical	31 (8.7%)	36 (8.6%)	-0.1 (3.8 to -4.2)	0.9
Specialty	324 (91.3%)	385 (91.4%)		
Number of outcomes per trial	15 (1, 131)	14 (1, 372)		0.2
Median (range)				
Type of outcomes	(n = 7132)	(n = 8299)		< 0.0001
Continuous	4984 (69.9%)	5705 (68.7%)		
Binary	1961 (27.5%)	2357 (28.4%)		
Time-to-event	47 (0.6%)	128 (1.5%)		
Ordinal	140 (2.0%)	98 (1.2%)		
Categorical	0	11 (0.1%)		
Adjusted analysis	84 (23.7%)	113 (26.8%)	3.1 (-2.9 to 9.3)	0.3

* Unclear: 6 for year 2000 and 11 for year 2006

† Percentage difference = percentage in 2006 – percentage in 2000

Figure 2-3 Comparison of the characteristics of articles that did and did not report adjusted analysis for trials published in 2000 and 2006



2.3.4 Details of adjusted analysis

Details of the adjusted analyses performed are summarised in Table 2-2. In the 2000 cohort, over 90% of the articles that performed adjusted analysis carried it out on the primary outcome. The majority of the articles with adjusted analyses (80%) did not report why they used adjustment or how they selected the covariates for adjustment. Of the 78 papers that specified the covariates for adjustment, 9 (12%) included covariates that were collected after randomisation. These included change scores from baseline or measures that were assessed during the treatment period, which should have been regarded as outcomes rather than predictors of the primary outcome. Sixteen (20%) papers did not specify the adjustment methods used in the methods section. Fewer than half of the papers (44%) included all of the stratification factors used at randomisation in the adjusted analysis.

Continuing to focus on the 2000 cohort, 83 papers (99%) reported the statistical methods used for adjustment. As their outcomes were predominately continuous, most of the studies used regression methods (ANCOVA, ANOVA, or multiple regressions) for adjustment (Table 2-3). Binary outcomes and time-to-event data were mainly analysed by logistic regression and Cox regression, respectively. Stratified analyses (e.g., Cochrane-Mantel-Haenszel or chi-square analysis) for adjustment were used more often for binary outcomes than other types of outcomes.

Table 2-2 Details of adjusted analysis

Year of publication	2000 (n = 84)*	2006 (n = 113)*	% difference (95% CI)	P-value
Performed adjusted analysis on primary outcome †	50 (91%)	93 (94%)	3.0 (-5.9 to 12.0)	0.5
Reason for adjustment				0.4
Imbalance in covariates	9 (11%)	12 (11%)		
Prognostic covariates	6 (7%)	15 (13%)		
Both	0	3 (3%)		
Other reasons ‡	3 (4%)	4 (4%)		
Not mentioned	66 (78%)	79 (70%)		
Choice of covariates				0.5
All pre-specified	5 (6%)	8 (7%)		
All suggested by data	12 (14%)	20 (18%)		
Combination of pre-specified and post hoc	0	3 (3%)		
Not mentioned	67 (80%)	82 (73%)		
Number of covariates adjusted for §				0.02 ‖
1	39 (46%)	36 (32%)		
2	23 (27%)	33 (29%)		
3-5	14 (17%)	25 (22%)		
6-9	2 (2%)	12 (11%)		
Not mentioned	6 (7%)	7 (6%)		
Covariate used for adjustment				
Outcome assessed at baseline	33/62 (53%)	55/81 (68%)	14.7 (-1.4 to 30.7)	0.07
Centre/country	31/49 (63%)	25/61 (41%)	-22.3 (-40.6 to -39.9)	0.02
Assessed after randomisation	9/78 (12%)	9/107 (8%)	-4.1 (-12.0 to 5.7)	0.5
All stratification factors were adjusted for	11/25 (44%)	20/46 (44%)	-0.5 (-24.7 to 23.6)	1.0
Explicitly specified nature of analysis				
Primary analysis	2	5		
Secondary / sensitivity analysis	5	5		
Type of outcomes				0.8
Binary	13 (16%)	19 (17%)		
Continuous	65 (77%)	81 (72%)		
Ordinal	1 (1%)	3 (3%)		
Time-to-event	5 (6%)	10 (9%)		
Adjusted analysis method used was mentioned for specific outcome in the Method section	62/78¶ (80%)	88/109¶ (81%)	1.2 (-10.4 to 12.9)	0.8

* One adjusted analysis was selected per study

† Number of studies with specified primary outcomes: Year 2000 = 55 and Year 2006 = 99

‡ Year 2000: Clinical relevance (n = 1), significant at 3 weeks after randomisation (n = 1), exploring role of baseline variables (n = 1). Year 2006: Mediated treatment effect on outcome (n = 1), related to compliance/adherence of treatment (n = 2), effect of outcome declines over time (n = 1)

§ Year 2000: 6 studies did not report the number of covariates. Year 2006: 6 studies did not report the number of covariates and 1 stated at least 2 covariates

‖ Mann-Whitney test

¶ Number of studies without a statistical methods section: Year 2000 = 6 and Year 2006 = 5

In the 2006 cohort, there was no evidence of change in the reporting of the reasons for adjustment (30%) or choice of covariates (27%). More trials in 2006 adjusted for covariates that were believed to be correlated with the outcomes (13% vs 7%), but only two articles explicitly stated that the covariates selected for adjustment were pre-specified. More covariates were adjusted for in 2006 than in 2000, especially outcomes collected at baseline, but fewer multicentre studies adjusted for centre effects. The two cohorts had similar statistical method usage patterns (Table 2-3).

Very few articles in either cohort (8% in 2000 and 9% in 2006) explicitly stated whether the adjusted analysis was the primary or secondary analysis. When stratified by type of outcome, the median study sample size was higher for binary and time-to-event outcomes (435 and 318 in the 2000 cohort, and 860 and 306 in the 2006 cohort, respectively) than continuous outcomes (102 in 2000 and 148 in 2006).

2.3.5 Adjusted analysis reporting

Table 2-4 presents the types of results reported in the results sections and abstracts. Fifty-four articles in 2000 reported the results of any adjusted or unadjusted analysis in the results section. Of these, 80% explicitly reported the type of analysis used to derive the P-value, while just under half reported estimated treatment effects (e.g., the odds ratio or difference between means)

and the corresponding CIs. Lack of reporting of results for the selected outcome was more severe in abstracts. Over 80% of the articles did not report either the estimated treatment effect or the corresponding CI in the abstract. Only 31% of the studies reported P-values.

Table 2-3 Methods used in adjusted analysis

Year of publication	2000 (n = 84)	2006 (n = 113)
Continuous data	65 (77.4%)	81 (71.7%)
ANOVA / ANCOVA	50 (76.9%)	56 (69.1%)
Multiple regression method*	7 (10.8%)	19 (23.5%)
Stratified analysis	1 (1.5%)	0
Other†	6 (9.2%)	6 (7.4%)
Not mentioned	1 (1.5%)	0
Binary data	13 (15.5%)	19 (16.8%)
Logistic regression	5 (38.5%)	11 (57.9%)
Stratified analysis (Cochrane-Mantel-Haenszel test)	6 (46.1%)	5 (26.3%)
Other‡	2 (15.4%)	3 (15.8%)
Ordinal data	1 (1.2%)	3 (2.6%)
Stratified analysis (Cochrane-Mantel-Haenszel test)	1 (100%)	1 (33.3%)
Nonlinear mixed effect model	0	1 (33.3%)
Ordinal logistic regression	0	1 (33.3%)
Time to event data	5 (5.9%)	10 (8.9%)
Cox proportional hazard	5 (100%)	9 (90.0%)
Stratified log rank test	0	1 (10.0%)

* Including random effect and mixed effect models
† Including the generalised estimating equation (GEE), generalised linear model, analysis of covariance for rank data, Zellner seemingly unrelated regression, Poisson model and Van Elteren test
‡ Including the non-parametric generalised mixed effect model, GEE, and non-parametric ANCOVA

There was an increase in reporting any adjusted results in abstracts from 2000 to 2006 (Table 2-4). However, many of the papers in both cohorts that used adjustment did not report an adjusted treatment effect. Only 26 of 50 (52%) papers in 2000 and 61 of 93 (66%) papers in 2006 reported the results of any treatment comparison (the treatment effect estimate, CI, or P-value) in their abstract. Of these, 50% and 61% reported an adjusted result in 2000 and 2006, respectively. Only 30% of both cohorts presented the adjusted treatment effect. CIs were rarely provided.

2.3.6 Adherence to the CONSORT guideline

The revised CONSORT Statement indicates how adjusted analysis should be reported. There was a slight improvement in reporting some CONSORT items five years after the 2001 revision, but the overall adherence was still low (Table 2-5). Although fewer articles in 2000 reported that stratified randomisation was performed, the proportion that adjusted for any stratification variables was higher than in the 2006 cohort (46% in 2000 vs 35% in 2006).

Table 2-4 Results reported in the results sections and abstracts of studies that reported adjusted analysis

Year of publication	Results section			Abstract		
	2000 (n = 54)	2006 (n = 89)	P-value	2000 (n = 71)*	2006 (n = 101) †	P-value
Summary statistics for each group			0.7			0.5
Unadjusted only	42 (78%)	70 (80%)		26 (37%)	45 (44%)	
Adjusted only	6 (11%)	12 (14%)		3 (4%)	5 (5%)	
Both	4 (7%)	3 (3%)		0	0	
None/not clear	2 (4%)	3 (3%)		42 (59%)	51 (51%)	
CI/SE within group			0.2			1.0
Unadjusted only	12 (22%)	11 (13%)		2 (3%)	2 (2%)	
Adjusted only	6 (11%)	10 (11%)		2 (3%)	4 (4%)	
Both	1 (2%)	0		0 (%)	0	
None/not clear	35 (65%)	67 (76%)		67 (94%)	95 (94%)	
Treatment effect			0.4			0.1
Unadjusted only	5 (9%)	5 (6%)		3 (4%)	5 (5%)	
Adjusted only	17 (31%)	35 (39%)		5 (7%)	19 (19%)	
Both	4 (7%)	12 (13%)		1 (2%)	1 (1%)	
None/not clear	28 (52%)	37 (42%)		62 (87%)	76 (75%)	
CI/SE of treatment effect			0.6			0.4
Unadjusted only	6 (11%)	5 (6%)		2 (3%)	6 (6%)	
Adjusted only	16 (30%)	24 (27%)		3 (4%)	17 (17%)	
Both	4 (7%)	10 (11%)		1 (2%)	1 (1%)	
None/not clear	28 (52%)	49 (56%)		65 (91%)	77 (76%)	
P-value for treatment effect			0.2			0.2
Unadjusted only	9 (17%)	8 (9%)		9 (13%)	13 (13%)	
Adjusted only	27 (50%)	52 (59%)		13 (18%)	30 (30%)	
Both	7 (13%)	17 (19%)		0	2 (2%)	
None/not clear	11 (20%)	11 (13%)		49 (69%)	56 (55%)	

* Thirteen studies did not report the selected outcome in the abstract

† Two studies did not have an abstract and 10 studies did not report the selected outcome in the abstract

Table 2-5 Adherence to the CONSORT recommendations

Year of publication	2000 (n = 84)	2006 (n = 113)	Relative risk (95% CI)	P-value
Adjusted for any stratification variables*	26 (46%)	30 (35%)	0.76 (0.51, 1.14)	0.2
Specified the rationale for any adjusted analysis	18 (21%)	34 (30%)	1.40 (0.85, 2.31)	0.2
Specified the statistical method used for adjusted analysis	83 (99%)	113 (100%)	1.0 (0.97, 1.03)	1.0
Reported the results from adjusted analysis only†	18 (21%)	29 (26%)	1.20 (0.71, 2.01)	0.5
Reported the results from both adjusted and unadjusted analysis†	4 (5%)	11 (10%)	2.04 (0.67, 6.20)	0.3

* n = 56 for Year 2000 and n = 85 for Year 2006
† Results include the summary in each group, effect size, and CI

Both adjusted and unadjusted results were poorly reported. Only 4 of 84 papers in 2000 and 11 of 113 papers in 2006 reported both results. Twenty-one articles (25%) in 2000 mentioned both adjusted and unadjusted analyses, of which seven reported only the unadjusted results because the results of the two analyses were similar. Similarly, 27 articles in 2006 performed both analyses, of which two reported the adjusted results and five reported the unadjusted results because the results of the two analyses were similar. Four studies in the 2006 cohort reported that the significance of the treatment effect changed after adjusting for covariates.

In 2006, 65 of the 113 (57%) articles that reported adjusted analysis were published in CONSORT-endorsing journals. Among these, 23 (35%) specified their rationale for conducting adjusted analysis, compared with 11 of the 48 (23%)

articles published in journals that did not endorse CONSORT. Slightly more of the articles published in CONSORT-endorsing journals reported both adjusted and unadjusted results than those published in non-endorsing journals (seven vs four articles).

2.4 Summary

The characteristics of the published reports of parallel group randomised trials indexed in PubMed in 2000 and 2006 were found to be similar, although there was a significant improvement in primary outcome specification in 2006. Only a quarter of the reviewed randomised trials reported any covariate adjustment analysis. The prevalence of adjusted analysis in the two cohorts was much lower than the 72% reported in a 1997 review restricted to four high-impact general medical journals (Assmann et al, 2000) and the 64% reported by Austin et al (2010). Mullner, Matthews, and Altman (2002) reviewed 34 scientific medical journals in 1998 with a high impact factor and reported that 31% of their articles specified adjustment for confounding factors. Hernández et al (2005) found that a similar percentage of articles on clinical trials of traumatic brain injury specified adjusted analysis. A more recent review carried out by Saquib, Saquib, and Ioannidis (2013) found that 49% trials, published in 25 biomedical journals with the highest impact factor, had reported adjusted analysis in the Results section. In contrast, the literature review reported here included articles published in journals from all specialties. The frequency of adjusted analysis found in the two

reported cohorts is thus more likely to be representative of the overall randomised trial literature.

The analyses specified in the methods sections of the analysed articles did not necessarily reflect how the reported results were obtained. Often the method was either not clearly specified or the results were obtained from different analyses from those specified in the methods. Readers often trust that the results reported in an article were derived from the analyses specified in the article's methods. The findings of this review show that further clarification for reporting results is needed, especially in studies involving adjusted analysis.

Most of the reviewed articles did not report the rationale for adjustment and the choice of covariates, although there was an improvement in the overall reporting of adjusted analysis in trial reports from 2000 to 2006. This lack of pre-specification echoes the findings of Chan et al (2008). They found that most trials that mentioned adjusted analysis in either the protocol or main article had discrepancies between the two (18 of 28 reviewed trials). Twelve of the 18 trials that published adjusted analyses included covariates that were not pre-specified in the protocol and 10 did not mention the adjusted analysis in the protocol at all. In addition, Saquib, Saquib, and Ioannidis (2013) also concluded in their review that information regarding which covariates were included in the adjusted model was lacking.

Most of the reviewed articles that gave their reason for adjustment or choice of covariates did not follow the CONSORT recommendations (ICH E9 Expert Working Group, 1999; European Medicines Agency Committee for Proprietary Medicinal Products, 2003). Few of the reviewed studies performed and reported adjusted analysis adequately. For example, if procedures such as stratified randomisation or minimisation methods are used, stratifying variables should be adjusted to prevent overestimatation of the SE of the treatment effect and O-value distortion (Forsythe and Stitt, 1977). However, the reviewed trials that performed these procedures often did not adjust for stratification or minimisation factors. Furthermore, covariates assessed after randomisation require a different analytical approach because their relationship with the study outcome can be confounded by treatment (ICH E9 Expert Working Group, 1999; European Medicines Agency Committee for Proprietary Medicinal Products, 2003; Rochon, 1999; Yusuf et al, 1991). However, some of the reviewed trials included these kinds of covariates in their analyses, as has been documented by others (van Walraven et al, 2004; Hirji and Fagerland, 2009).

The reporting of adjusted analysis was generally comparable between the cohorts of trials published before and after the 2001 revision of the CONSORT Statement. The CONSORT Statement recommends reporting main results using treatment group summary statistics, the treatment effect, and the CI. However,

the reviewed trials often either did not report these measures for their main results or reported them unclearly in both the results and abstract. Much more attention has been given to other issues in systematic reviews, such as the adequacy and transparency of sample size calculation, blinding, and randomisation methods (Moher, Jones and Lepage, 2001; Kane, Wang and Garrard, 2007), which may explain the lack of improvement in reporting adjusted analysis.

The strength of this literature review is the inclusion of articles published in journals from all specialties. The findings are therefore likely to be more representative of the quality of reporting of adjusted analysis than findings that focus on top-tier journals. The major limitation of this review is the length of time since it was carried out. Although no similar detailed review has been carried out since 2006, recent reviews have shown that reporting is generally improving slowly (To et al, 2013; Turner et al, 2012b; Saquib, Saquib, and Ioannidis, 2013). Another limitation of this study is that, apart from the trial characteristics for the 2000 cohort, the data were extracted by a single reviewer. However, the reviewer revisited the data extraction a few months after the first extraction in a quality assurance procedure.

CHAPTER 3: FACTORS AFFECTING THE IMPACT OF BASELINE COVARIATES

3.1 Introduction

The difference between the treatment effects in RCTs before and after covariate adjustment has been explored mathematically (Gail, Wieand and Piantadosi, 1984; Neuhaus, Kalbfleisch and Hauck, 1991) and with simulations (Hernández, Steyerberg and Habbema, 2004; Negassa and Hanley, 2007; Steyerberg, Bossuyt and Lee, 2000; Chu et al, 2012). However, mathematical exploration is complex and not necessarily reproducible, and previously conducted simulations have adjusted for only one covariate. It is possible that the impact of single-covariate adjustment diminishes when more covariates are added, as the direction of the effects may cancel each other out. However, simulations have not yet been carried out to investigate the effect of multicovariate adjustment.

The aim of this simulation study is to explore the factors that can affect the extent of the empirical difference between unadjusted and covariate-adjusted treatment effects. The following questions will be investigated in this chapter:

1. Does the difference between the unadjusted and adjusted treatment effects vary with the overall event rate?
2. Does the prognostic strength and distribution of the covariates affect this difference?

3. Is the difference between the unadjusted and adjusted treatment effects larger when adjusting for two highly prognostic factors with the same direction of association (e.g., both prognostic factors have a negative or positive effect on the outcome) than when adjusting for two variables with the same prognostic strength but the opposite direction of association (i.e., one factor is negatively associated with the outcome and the other is positively associated)?
4. Does the impact of adjustment depend on the strength of the treatment effect?

3.2 Setting and Notation

Simulations were conducted based on a two-group parallel RCT setting comparing a new treatment (T = 1) with a standard or control treatment (T = 0). A practical example is a clinical trial to assess the effect of beta-blockers for preventing heart failure in patients with hypertension. This trial randomises participants to either a beta-blocker agent or placebo. At the end of follow-up, information on whether the patients have developed heart failure or not is collected. Possible risk factors for heart failure include sex, age, and hypertension status at baseline.

Let N be the total sample size; Y_i be a binary outcome for subject i, where 1 is defined as having an event (e.g., developed heart failure) and 0 as no event (e.g., no heart failure); X_{1i} and X_{2i} be the pre-specified baseline covariates, where X_{1i} is

a binary covariate (e.g., X_1 = 0 represents female and X_1 = 1 represents male) with percentage P_{X_1} for $X_1 = 1$ (e.g., the prevalence of males in the study), and X_{2i} is a continuous covariate that is normally distributed with mean 0 and variance σ^2 (e.g., this can be the standardised age of the participants at randomisation); and R be the overall event rate, defined as the percentage of patients for whom an outcome occurs during the study.

The unadjusted treatment effect is obtained with the following logistic regression model:

$$Logit(Y|\ T) = \alpha + \theta_u T \qquad \text{(Eq. 3-1)}$$

where θ_u is the unadjusted log odds ratio of T. The adjusted treatment effect is obtained with the following logistic regression model:

$$Logit(Y|\ T, X_1, X_2) = \alpha + \theta_a T + \gamma_{1\_a} X_1 + \gamma_{2\_a} X_2 \qquad \text{(Eq. 3-2)}$$

where θ_a, γ_{1_a}, and γ_{2_a}, are the adjusted log odds ratios of T, X_1, and X_2, respectively.

3.3 Simulation Procedure

The simulation procedure specified by Burton et al (2006) was followed. Independent data sets were generated to represent different scenarios by varying

the overall event rate, adjusted treatment and covariate effects, and covariate distributions (Table 3-1). The overall event rate (R) ranged from 10% to 90%, representing a range from rare to common events. The adjusted odds ratio of the binary covariate, X_1, was set to $OR_{X_1\_a}$ = 0.2 for a very large negative prognostic effect, $OR_{X_1\_a}$ = 0.67 for a moderately negative effect, $OR_{X_1\_a}$ = 1.09 for a small positive effect, $OR_{X_1\_a}$ = 1.5 for a moderately positive effect, or $OR_{X_1\_a}$ = 5 for a very large positive effect. The overall prevalence of this covariate (P_{X_1}) was set to 20%, 30%, 40%, 50%, 60%, 70%, or 80%. The adjusted odds ratio of the continuous covariate, X_2, was set to $OR_{X_2\_a}$ = 1.2 or 2, for a moderate or strong effect, respectively. Its variance, σ^2, was set to 1 to indicate low variability or 10 to indicate high variability, to explore how the distribution of the continuous covariate affected the adjusted treatment effect. The variable with large variance can be transformed to the standard Normal form with zero mean and unit variance (i.e., $\mathrm{N}(0, 1)$), which the adjusted odds ratios of 1.2 and 2 become 1.78 and 8.95, respectively.

The prognostic strength values were based on published studies. For example, the adjusted odds ratio of the effect of systolic blood pressure on pregnancy with a serious adverse outcome (PSAO) measured by the Collaborative Low-dose Aspirin Study in Pregnancy (CLASP) trial (CLASP, 1994) was 1.013, with a standard deviation of the baseline systolic blood pressure of 17 mmHg. The corresponding standardised adjusted odds ratio was 1.25. Christensen et al (1985) assessed the effect of azathioprine versus placebo on the survival of

patients with primary biliary cirrhosis and found that the adjusted hazard ratio of baseline log serum bilirubin was 12.3. Although they studied a time-to-event outcome and the standard deviation of the covariate was not reported, their results do demonstrate that extreme scenarios can occur, if rarely. Moderate (OR_{T_a} = 0.7), large (OR_{T_a} = 2), and very large (OR_{T_a} = 5) adjusted treatment effect values were tested.

The values for each parameter (listed in Table 3-1) were combined to give 3,780 scenarios for simulation. The simulations were based on the assumption that the distributions of the covariates were balanced between the treatment and control groups. The covariates were assumed to be independent of one another (i.e., no correlation between them). It was assumed that there were no interactions between the treatment effect and either covariate.

Table 3-1 Values of the parameters in the simulated population data sets

Parameters		**Values**
Overall event rate, R		10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%
Binary covariate, X_1	Adjusted odds ratio (prognostic strength), $OR_{X_1\_a} = e^{\gamma_{1\_a}}$	5 (strongly positive) 1.5 (moderately positive) 1.09 (weakly positive) 0.67 (moderately negative) 0.2 (strongly negative)
	% with X_1 = 1, P_{X_1}	20%, 30%, 40%, 50%, 60%, 70%, 80%
Continuous covariate, X_2	Adjusted odds ratio (prognostic strength), $OR_{X_2\_a} = e^{\gamma_{2\_a}}$	2 (strongly positive) 1.2 (moderately positive)
	$X_2 \sim N(0, \sigma^2)$	$N(0, 1)$, $N(0, 10)$
Adjusted odds ratio for T, $OR_{T\_a} = e^{\theta_a}$		0.7, 2, 5

3.3.1 Generating the population data sets

A population data set of 50,000 cases was created for each of the 3,780 scenarios. These population data sets can be regarded as the truth of their respective scenarios.

The relationship between the covariates and the binary outcome in each population data set was obtained using logistic regression (Eq. 3-2):

$$Logit(Y|\ T, X_1, X_2) = \alpha + \theta_a T + \gamma_{1\_a} X_1 + \gamma_{2\_a} X_2 \qquad \text{(Eq. 3-3)}$$

Due to the number of data sets involved, the intercept value for each model was first identified to reduce computer processing time. Box 3-1 lists the procedure for deriving the intercept values. The intercepts were used as the initial values for simulating the population data sets. For each population data set, 50,000 observations for T, X_1, and X_2 were generated. A fully adjusted logistic regression model was fitted to the data, and the coefficients obtained were compared with the parameter values (θ_a, γ_{1_a}, and γ_{2_a}). This process was iterated until all of the specified parameter values were within ± 0.01 of their true values. Full details of the program are stated in Appendix B.

Box 3-1: Procedure for identifying the intercept of each model before generating population data

1. Simulate 1,000 observations for T, X_1, and X_2
2. Set a range of intercept values from -6 to 6

Fit a fully adjusted model using coefficients based on one set of values from

3. Table 3-1
4. Calculate the overall event rate
5. Repeat the process 500 times
6. Calculate the mean event rate
7. Store the intercept that corresponds to the mean event rate

3.3.2 Outcome of interest

Unadjusted and adjusted logistic regression analysis was performed on each population data set and the corresponding odds ratios were calculated. The difference between the unadjusted and adjusted results was expressed as a ratio of the two odds ratios (ROR) (Schulz et al, 1995):

$$ROR = \frac{OR_{T\_u}}{OR_{T\_a}} \qquad \text{(Eq. 3-4)}$$

where OR_{T_u} is the unadjusted odds ratio of the treatment and OR_{T_a} is the adjusted odds ratio of treatment. The difference can also be expressed as a percentage difference between the unadjusted and adjusted odds ratios, calculated as (ROR−1) x 100%.

ROR = 1 when the unadjusted and adjusted odds ratios are equal. ROR < 1 indicates that the unadjusted odds ratio is smaller than the adjusted odds ratio. ROR > 1 occurs when the unadjusted odds ratio is larger than the adjusted odds ratio.

Taking into account the direction of the adjusted odds ratio of the treatment effect, ROR < 1 and OR_{T_a} < 1 indicates that the unadjusted odds ratio is further away from no treatment effect (i.e., an odds ratio of 1) than the adjusted odds ratio (Table 3-2). Conversely, ROR < 1 and OR_{T_a} > 1 indicates that the unadjusted odds ratio is closer to no treatment effect than the adjusted odds ratio.

Table 3-2 Interpretation of the ratio of odds ratios according to the direction of the adjusted odds ratio

Direction of the adjusted odds ratio (OR_{T_a})	**Magnitude of ROR**	
	< 1	**> 1**
< 1	OR_{T_u} is further away from no treatment effect than OR_{T_a}	OR_{T_u} is closer to no treatment effect than OR_{T_a}
> 1	OR_{T_u} is closer to no treatment effect than OR_{T_a}	OR_{T_u} is further away from no treatment effect than OR_{T_a}

3.3.3 Presentation of results

The results obtained from the simulation are mostly presented in graphical form. Each plot consists of ROR plotted against event rate (R) by the distribution of the binary covariate (P_{X_1}). A line was fitted through these points using a locally weighted regression method to provide a visual impression of the relationship

between ROR and event rate, called the ROR curve. Each colour on the graphs represents a different value of P_{X_1}. Sixty-three scenarios (9 levels of $R \times 7$ levels of P_{X_1}) are depicted on each plot. Sixty plots were generated to present all 3,780 scenarios.

R version 3.1.1 was used to generate and analyse the data sets and create the graphs.

3.4 Simulation Results

The interpretation of the results will begin by looking at the relative difference in the overall event rate produced using unadjusted and adjusted analysis, using a scenario more commonly seen in clinical trials. How the relative difference changes as the distribution and prognostic strength of each covariate varies (both while the other covariate is held constant and with both varying) will then be discussed. Finally, whether the results are affected by the magnitude of the treatment effect will be reported.

3.4.1 Varying the overall event rate

To investigate the effect of varying the overall event rate, R, of the outcome of interest, simulations were performed with the overall event rate set from 10% to 90%.

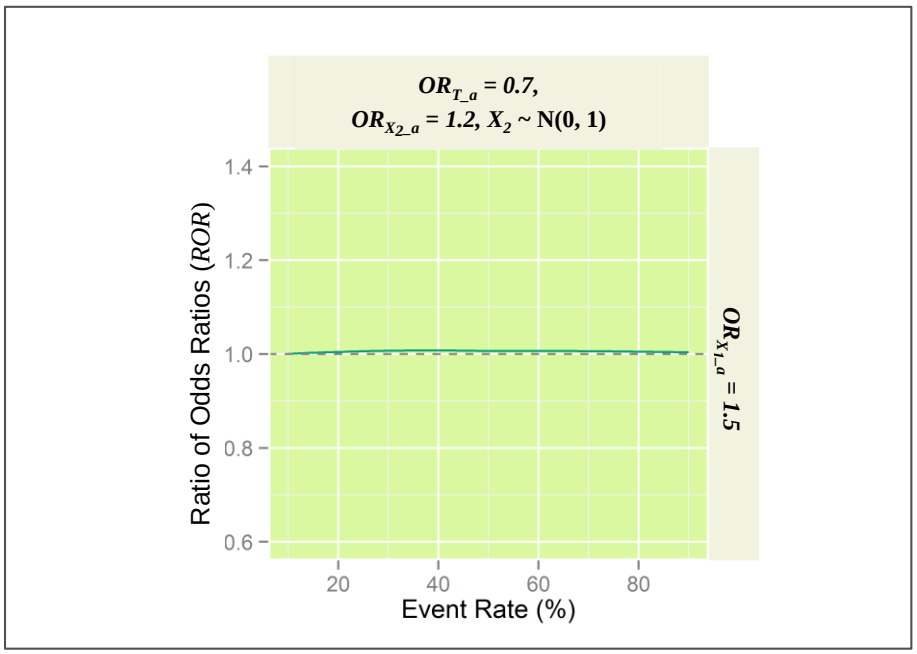
Table 3-3 shows ROR for T in a typical clinical trial scenario with a moderate adjusted treatment effect (OR_{T_a} = 0.7), in which both of the covariates had a positive prognostic impact on the outcome ($OR_{X_1\_a}$ = 1.5 and $OR_{X_2\_a}$ = 1.2), both groups had an even distribution of X_1 (P_{X_1} = 50%), and X_2 had a small variance (σ^2 = 1). The unadjusted and adjusted results were similar to one another under this scenario. The odds ratios derived from the adjusted analysis were expected to be near to 0.7 because the simulation was based on this value. Without adjusting for both covariates, the unadjusted odds ratio for treatment was slightly higher than the corresponding adjusted odds ratio at each event rate. ROR ranged between 1.001 and 1.008, and was slightly lower at extreme event rates (ROR = 1.001 at R = 10%, 1.003 at R = 20%, and 1.004 at R = 90%) than at event rates between 30% and 70%.

Table 3-3 Ratio of the unadjusted and adjusted odds ratios at different overall event rate values, calculated in a typical clinical trial scenario (OR_{T_a} = 0.7, $OR_{X_1\_a}$ = 1.5, P_{X_1}= 0.5, $OR_{X_2\_a}$ = 1.2, and σ^2 = 1)

Overall event rate (R)	Unadjusted odds ratio (OR_{T_u})	Adjusted odds ratio (OR_{T_a})	Ratio of odds ratios (ROR)
10%	0.702	0.701	1.001
20%	0.707	0.705	1.003
30%	0.699	0.694	1.008
40%	0.712	0.707	1.008
50%	0.707	0.703	1.006
60%	0.709	0.704	1.007
70%	0.698	0.694	1.006
80%	0.703	0.699	1.005
90%	0.707	0.705	1.004

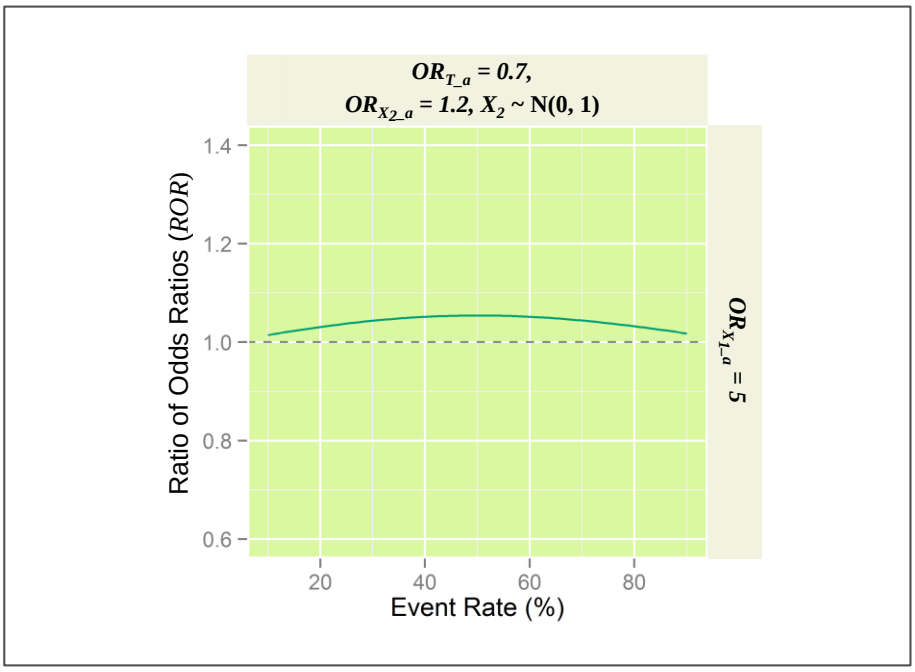
Figure 3-1 graphically presents the change in ROR shown in Table 3-3. The solid green line represents the ROR curve that explains the relationship between ROR and event rate. The dotted line is the 'line of no difference', which marks where the unadjusted and adjusted odds ratios were equal (ROR = 1 or $OR_{T\_u} = OR_{T\_a}$). The ROR curve lies just above the equity line, indicating that the unadjusted odds ratio was greater than the adjusted odds ratio. However, the unadjusted odds ratio was closer to no treatment effect because the adjusted odds ratio was less than 1.

Figure 3-1 Unadjusted and adjusted simulation results using the values from Table 3-3



When the prognostic impact of X_1 was increased to $OR_{X_1\_a}$ = 5 (Figure 3-2), a similar pattern of variation in ROR was seen for varying event rates. Figure 3-2 shows a clear convex relationship between ROR and R, with ROR maximised (1.053, i.e., OR_{T_a} = 0.701 vs OR_{T_u} = 0.739) at R = 50%. This can be interpreted as the unadjusted odds ratio being 5.3% higher, or closer to null, than the adjusted odds ratio. When R = 10%, however, ROR = 1.015, indicating that the unadjusted odds ratio was only 1.5% higher than the adjusted odds ratio.

Figure 3-2 Ratio of the unadjusted and adjusted odds ratios as the overall event rate changes, calculated in a typical clinical trial scenario (OR_{T_a} = 0.7, $OR_{X_1\_a}$ = 5, P_{X_1}= 0.5, $OR_{X_2\_a}$ = 1.2, and σ^2 = 1)



3.4.2 Changes in prognostic strength and covariate distribution

3.4.2.1 Varying the distribution and prognostic strength of the binary covariate, X_1

ROR varied with the prevalence of the binary covariate, X_1 (Figure 3-3). The percentage of X_1 that maximised *ROR* depended on event rate. When X_1 had a high prevalence (e.g., P_{X_1} = 80%) the difference between the unadjusted and adjusted treatment effects was maximised at a high event rate of 70% (highlighted in yellow in Table 3-4). Similarly, when the prevalence of X_1 was low (i.e., 20%), then the largest difference between the unadjusted and adjusted treatment effects was observed at a low event rate (R = 20%). The largest values for *ROR* were observed when X_1 prevalence was 60% and event rate was 60% (Table 3-4). Overall, *ROR* shifted further away from 1 when the prevalence of the binary variable was varied than when the overall event rate was varied (section 3.4.1), although the values were still small.

Figure 3-3 Ratio of the unadjusted and adjusted odds ratios as the overall event rate changes, calculated in a typical clinical trial scenario with varying distribution of P_{X_1}

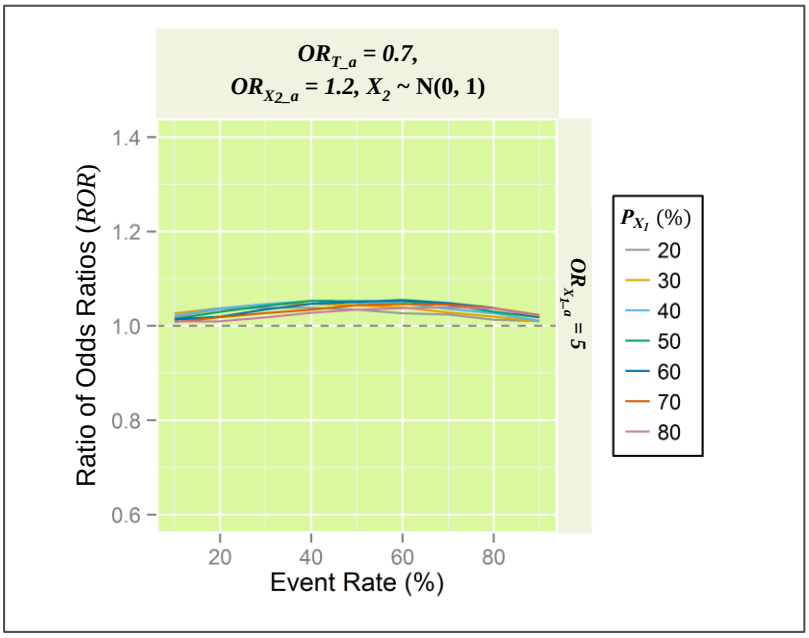


Table 3-4 Ratio of the unadjusted and adjusted odds ratios at different event rate values, as plotted in Figure 3-3

Overall event rate (R)	Distribution of X_1 (P_{X_1})						
	20%	30%	40%	50%	60%	70%	80%
10%	1.019	1.027	1.021	1.015	1.014	1.009	1.009
20%	1.039	1.038	1.038	1.030	1.020	1.019	1.010
30%	1.038	1.046	1.046	1.043	1.035	1.027	1.018
40%	1.036	1.047	1.053	1.052	1.047	1.035	1.028
50%	1.034	1.044	1.050	1.053	1.051	1.044	1.034
60%	1.027	1.039	1.048	1.052	1.055	1.047	1.038
70%	1.024	1.028	1.036	1.045	1.048	1.045	1.040
80%	1.013	1.020	1.027	1.030	1.038	1.038	1.037
90%	1.010	1.009	1.011	1.018	1.018	1.023	1.021

The values highlighted in yellow are the highest ROR at each value of P_{X_1}

Figure 3-4 shows the ROR curve at varying prognostic strengths of the binary covariate, X_1, when (a) P_{X_1}= 50% and (b) 20% < P_{X_1} < 80%. In all of the plots, the further away the ROR curve is from the line of no difference (ROR = 1), the larger the difference between the unadjusted and adjusted odds ratios. The results show that when all of the other parameters were held constant, ROR increased as the prognostic effect of X_1 on the outcome increased. The largest difference between the unadjusted and adjusted treatment effects occurred when X_1 was strongly prognostic (i.e., when $OR_{X_1\_a}$ = 0.2 or 5). ROR varied from 0.997 ($OR_{X_1\_a}$ = 1.09, P_{x1} = 30%, and R = 40%) to 1.055 ($OR_{X_1\_a}$ = 5, P_{X_1}= 60%, and R = 60%), depending on both the overall event rate and the percentage distribution of X_1, but it followed the pattern observed in Figure 3-3. However, the difference recorded between the adjusted and unadjusted analyses was small, with a maximum difference of 6%.

3.4.2.2 Varying the distribution and prognostic strength of the continuous covariate, X_2

When the adjusted treatment effect was moderate (OR_{T_a} = 0.7), the presence of a highly prognostic continuous covariate led to larger unadjusted treatment effect estimates than the presence of a moderately prognostic continuous covariate. Figure 3-5 presents situations where X_2 is moderately prognostic ($OR_{X_2\_a}$ = 1.2, panels 1 and 2) or more strongly prognostic ($OR_{X_2\_a}$ = 2.0, panels 3 and 4), and where the prognostic continuous variable has small variance (panels 1 and 3) or large variance (panels 2 and 4).

Unlike the binary covariate, the variability of X_2 greatly affected the difference between the unadjusted and adjusted treatment effects. When the standardised odds ratio was considered, the prognostic effect was larger in a covariate with a large variance than in a covariate with the same mean but a small variance before standardisation. The ROR distribution for a moderately prognostic continuous covariate ($OR_{X_2\_a}$ = 1.2) with large variance ($\mathrm{N}(0, 10)$) was similar to the ROR distribution when X_2 was more strongly prognostic ($OR_{X_2\_a}$ = 2) but had a small variance ($\mathrm{N}(0, 1)$), because the standardised $OR_{X_2\_a}$ of the former was 1.82. The curves in Figure 3-5 have roughly similar shapes, with larger ROR values at outcome event rates closest to 50%. This shape does not appear to have been influenced by the prevalence of X_1. When X_2 had both large variance and a large prognostic effect on the outcome ($OR_{X_2\_a}$ = 2 and $\mathrm{N}(0, 10)$, standardised odds ratio = 9.3), the difference between the unadjusted and adjusted odds ratios ranged from 6.9% to 18.3%.

Figure 3-4 **Ratio of the unadjusted and adjusted odds ratios as the overall event rate changes, calculated in a typical clinical trial scenario with varying distributions of P_{X_1} and $OR_{X_1\_a}$**

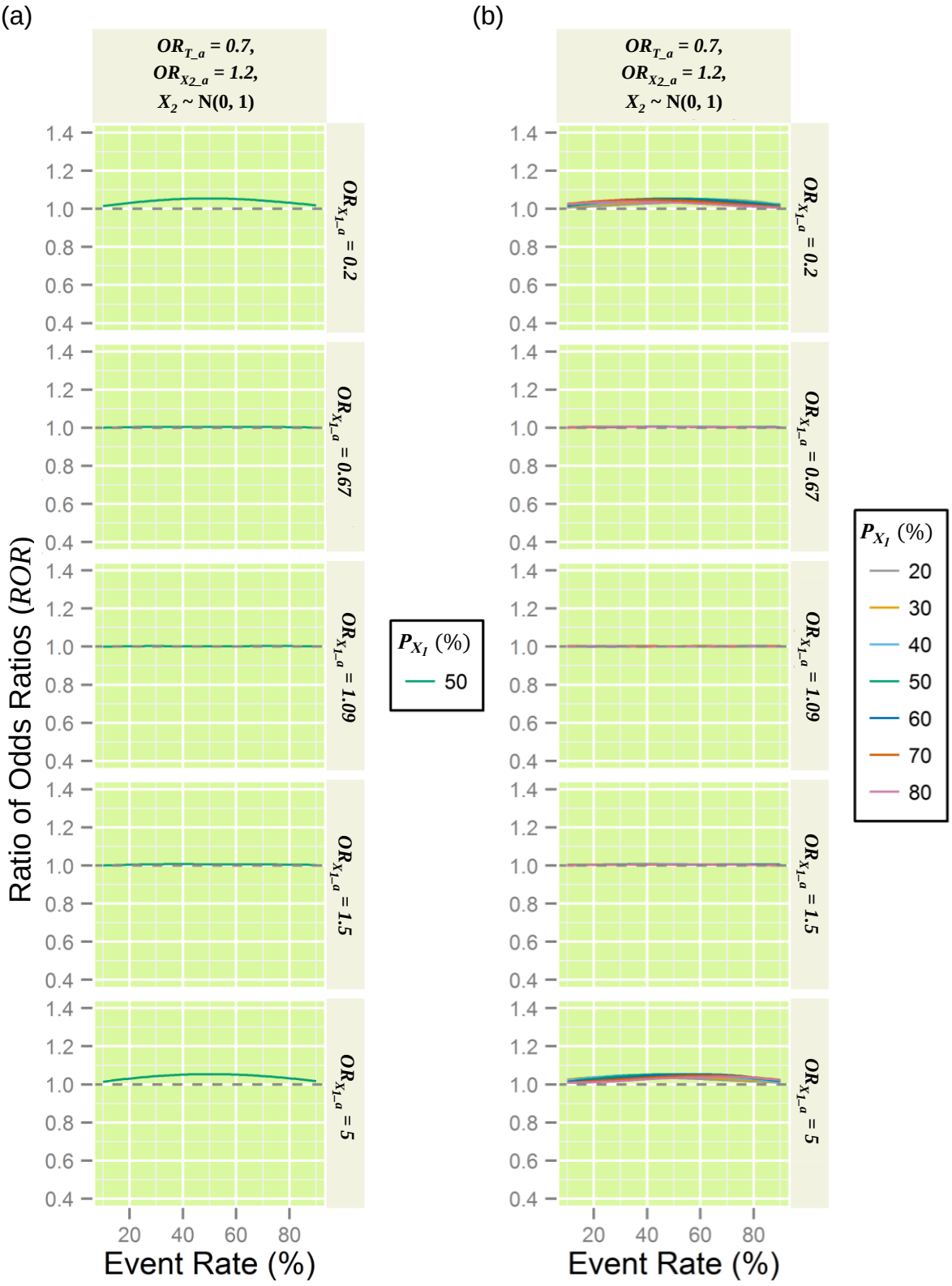
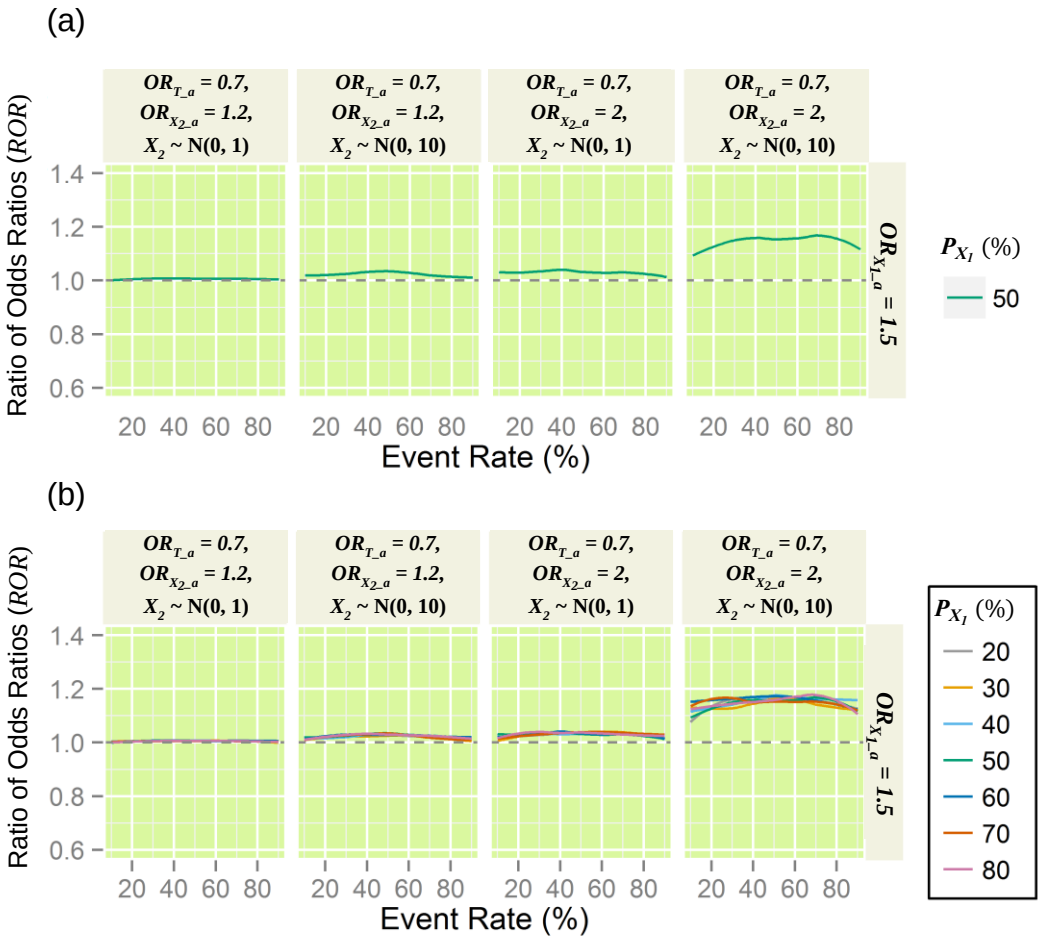


Figure 3-5 **Ratio of the unadjusted and adjusted odds ratios as the overall event rate changes, calculated in a typical clinical trial scenario, varying the distribution and prognostic strength of the continuous covariate**



3.4.3 Effect of the direction of the association between the covariates and treatment effect

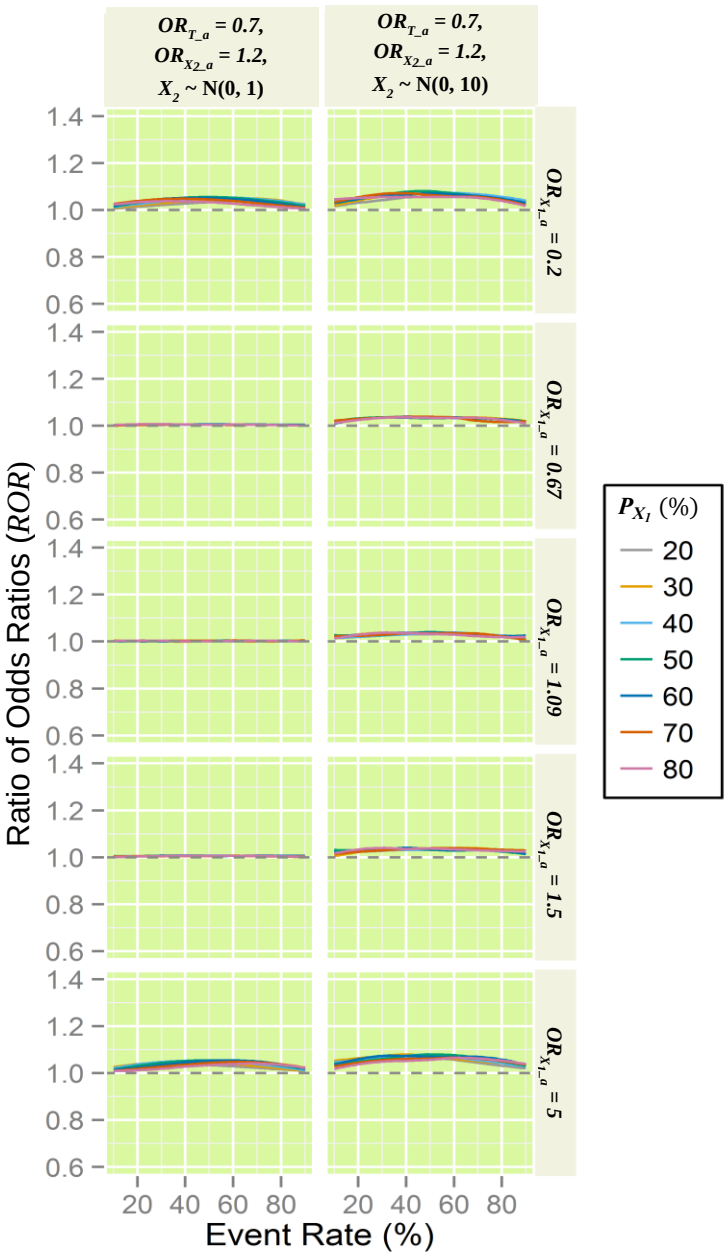
Figure 3-6 shows the effect of the direction of the association between the covariates and outcome on the shape of the *ROR* curve. In the figure, the prognostic strength of the binary covariate varies from a strongly negative

association ($OR_{X_1\_a}$ = 0.2, top row) to moderate associations (0.67 ≤ $OR_{X_1\_a}$ ≤ 1.5, middle three rows) to a strongly positive association ($OR_{X_1\_a}$ = 5, bottom row). The effect of the continuous covariate (X_2) on the outcome is held constant at $OR_{X_2\_a}$ = 1.2, with variance 1 (left column) and 10 (right column), to test the situation of adjusting for two covariates that are both positively associated with the outcome or have opposite associations with the outcome.

The ROR curve remains above the line of no difference for all of the tested combinations. When both covariates were adjusted for, a strongly positive covariate did not cancel out a strongly negative covariate. Instead, their effect was additive. This result implies that ROR had the same value whether both covariates were positively or negatively associated with the outcome, or whether one was positively and the other negatively associated.

So, when the two covariates were uncorrelated, the difference between the unadjusted treatment effect and the treatment effect adjusted for both covariates was the difference between the unadjusted treatment effect and the treatment effect adjusted for the binary covariate, plus the difference between the unadjusted treatment effect and the treatment effect adjusted for the continuous covariate. This accumulation of differences did not depend on the direction of the covariates' effects.

Figure 3-6 **Ratio of the unadjusted and adjusted odds ratios as the overall event rate changes, calculated in a typical clinical trial scenario, varying the distribution and prognostic strength of the binary covariate**



3.4.4 Varying the magnitude of the treatment effect

The ROR values reported in Table 3-5 show that when the odds ratio of the continuous covariate was fixed at 1.2, the relative difference between the unadjusted and adjusted treatment effects depended on the size and direction of the treatment effect. The unadjusted odds ratio was higher than the adjusted odds ratio (ROR > 1) when there was a small negative treatment effect (OR_{T_a} = 0.7), but was lower than the adjusted odds ratio (ROR < 1) when the treatment effect was large and positive (OR_{T_a} = 2 or 5).

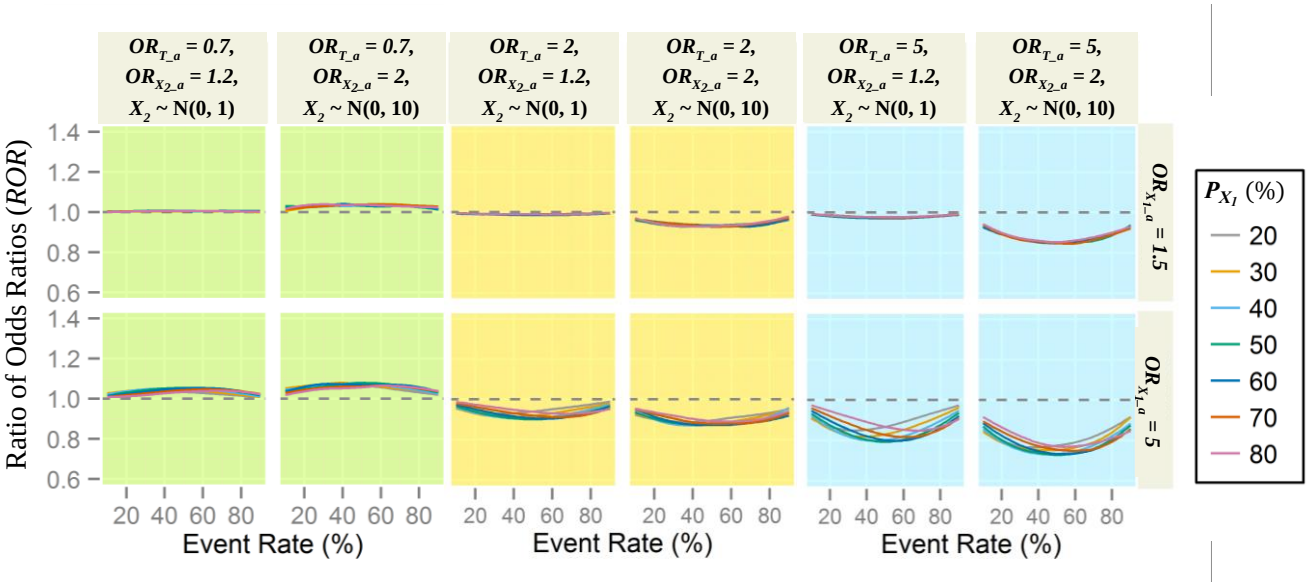
In both cases, the unadjusted results were closer to the no treatment effect than the adjusted results. Adjustment had a more pronounced effect when the binary outcome was highly prognostic. The highly prognostic binary covariate had a greater effect on the observed difference between the adjusted and unadjusted treatment effects when the treatment was highly effective. For example, when the overall event rate was 50%, there was only a small difference in ROR between $OR_{X_1\_a}$ = 1.5 and 5 for OR_{T_a} = 0.7 (i.e., 1.006 vs 1.007, respectively), but a large difference when OR_{T_a} = 5 (i.e., 0.972 vs 0.795, respectively).

Table 3-5 Ratio of the unadjusted and adjusted odds ratios as the overall event rate and treatment effect change, calculated in a typical clinical trial scenario (P_{X_1}= 0.5, $OR_{X_{2\_a}}$ = 1.2, and σ^2 = 1) with a moderately strongly prognostic X_1

Overall event rate (R)	$OR_{X_{1\_a}}$ = 1.5			$OR_{X_{1\_a}}$ = 5		
	OR_{T_a} = 0.7	2	5	OR_{T_a} = 0.7	2	5
10%	1.001	0.996	0.993	1.003	0.955	0.934
20%	1.003	0.994	0.981	1.005	0.937	0.877
30%	1.008	0.988	0.977	1.005	0.920	0.834
40%	1.008	0.989	0.973	1.004	0.905	0.803
50%	1.006	0.986	0.972	1.007	0.892	0.795
60%	1.007	0.985	0.975	1.005	0.892	0.804
70%	1.006	0.988	0.978	1.006	0.897	0.834
80%	1.005	0.991	0.985	1.002	0.908	0.877
90%	1.004	0.995	0.991	1.004	0.944	0.938

To illustrate the influence of the treatment effect further, Figure 3-7 shows scenarios in three clinical trial settings, a fairly typical treatment effect of OR_{T_a} = 0.7 (green), a strong treatment effect of OR_{T_a} = 2.0 (yellow), and a very strong treatment effect of OR_{T_a} = 5.0 (blue), with two prognostic strengths for each covariate. The ROR curve only changes slightly as P_{X_1} changes when $OR_{X_{1\_a}}$ is moderate, but varies more when $OR_{X_{1\_a}}$ is large. There is a further increase in the relative difference between the unadjusted and adjusted treatment effects when the prognostic strengths of both covariates and the treatment effect are increased. The largest relative difference found between the two estimates was 23%, at OR_{T_a} = 5, $OR_{X_{1\_a}}$ = 5, $OR_{X_{2\_a}}$ = 2, P_{X_1}= 50%, and R = 50% (the lower right blue panel in Figure 3-7).

Figure 3-7 Ratio of the unadjusted and adjusted odds ratios as the overall event rate changes, at different treatment effect strengths



3.5 Summary of the Simulation Results

Figure 3-8 summarises the *ROR* results for all of the tested scenarios, comparing moderate (OR_{T_a} = 0.7, green panels), strong (OR_{T_a} = 2, yellow panels), and very strong (OR_{T_a} = 5, blue panels) treatment effects. The binary covariate, X_1, varies from very strongly negatively prognostic (top row) to very strongly positively prognostic (bottom row). The prognostic strength and distribution of the continuous covariate, X_2, change in each column. Each cell shows a sub-plot of the relationship between *ROR* and overall event rate at a particular X_1 distribution.

ROR ranged between 0.474 and 1.197, depending on the size of the adjusted treatment effect. All of the scenarios yielded *ROR* > 1 for negative treatment

effects and < 1 for positive treatment effects. The unadjusted treatment effect was therefore consistently lower (closer to the no treatment effect) than the adjusted treatment effect. The magnitude of ROR depended on the combination of the overall event rate (R), the prognostic impact of the covariates (γ_{1_a} and γ_{2_a}), and the distribution of the covariates (σ^2 and P_{X_1}). The maximum differences between the unadjusted and adjusted treatment effects observed at OR_{T_a} = 0.7, 2, and 5 were 19.7% (ROR = 1.197), 28.6% (ROR = 0.714), and 52.6% (ROR = 0.474), respectively. These differences occurred when both covariates were highly prognostic and variable. In the most extreme scenario, the unadjusted odds ratio was less than half the adjusted odds ratio (2.37 vs 5).

Table 3-6 summarises how ROR was affected by each parameter. P_{X_1} only affected ROR greatly when the adjusted treatment effect was large. The direction of the treatment effect influenced the ROR direction.

Figure 3-8 **Ratio of the unadjusted and adjusted odds ratios as the overall event rate changes for all of the scenarios**

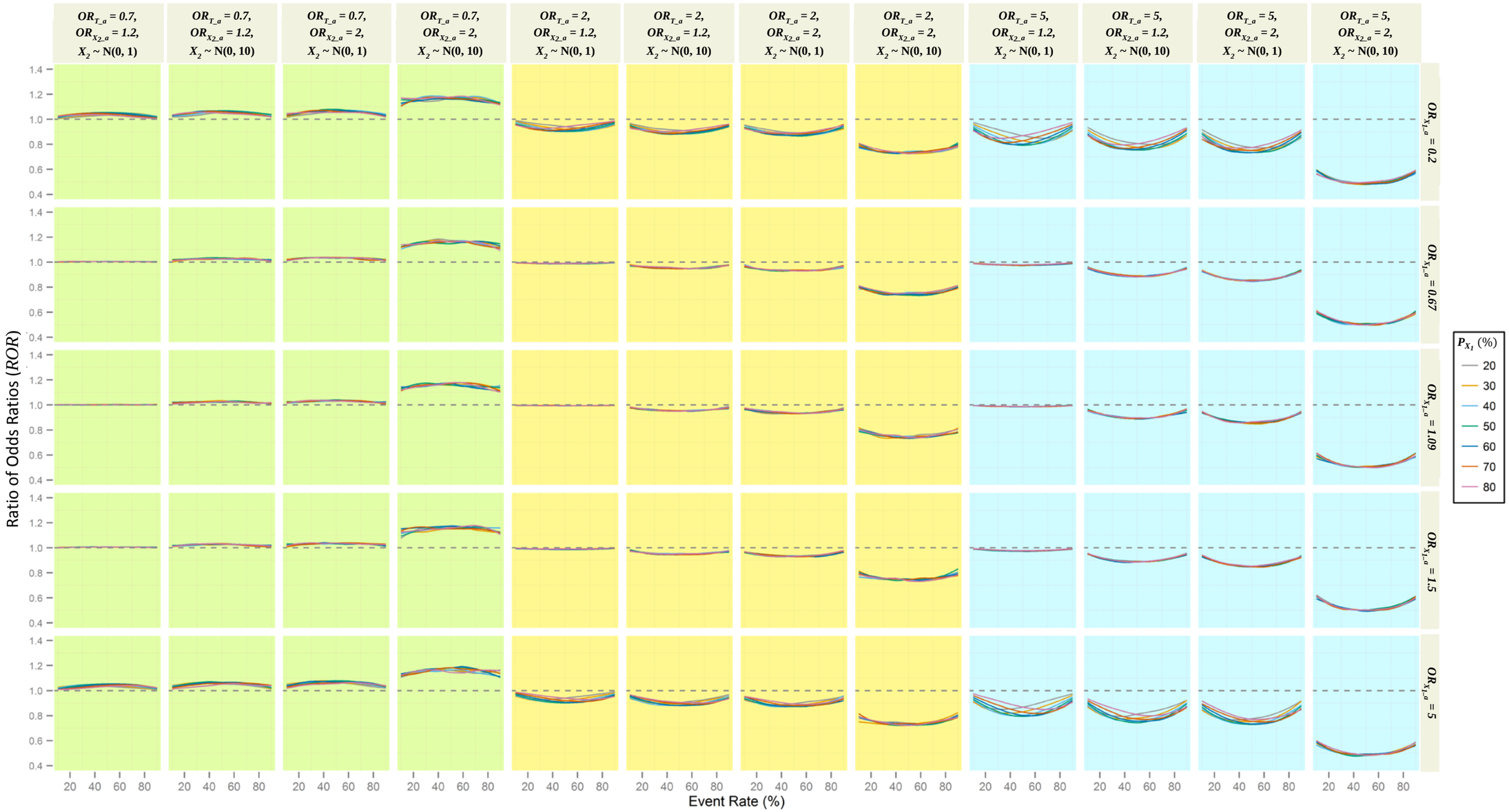


Table 3-7 summarises the percentage difference ((ROR−1) x 100%) between the unadjusted and adjusted treatment effects. The reported mean is the average percentage difference and the corresponding reported standard deviation represents its variability as the event rate and prevalence of the binary covariate change. The results were pooled for OR_{X1_a} = 0.67 and 1.5, and OR_{X1_a} = 0.2 and 5, as they have the same prognostic strength to the outcome. The green cell shows a typical clinical trial, which has an average percentage difference of only 0.45% and a maximum percentage difference of no more than 1%. The orange cell shows a trial with a very large adjusted treatment effect, which has a maximum percentage difference of no more than 3%. The blue cell shows the most extreme trial scenario (OR_{T_a} = 5), which, as discussed, has an unadjusted odds ratio that is less than half the adjusted odds ratio.

Table 3-6 Summary of how the event rate and covariate distribution affect the impact of adjustment when the covariates are uncorrelated and perfectly balanced across the treatment groups

	ROR*
$OR_{T\_a} < 1$	↑ as OR_{T_a} ↑
R	↑ as R approaches 50%; ↓ as R moves away from 50%
P_{X_1}	Unaffected by P_{X_1} unless OR_{T_a} is very large
$OR_{X_2\_a}$	↑ as $OR_{X_2\_a}$↑
$OR_{X_1\_a}$	↑ as $OR_{X_1\_a}$ moves away from 1 ; ↓ as $OR_{X_1\_a}$ moves towards 1
σ^2	↑ as σ^2↑
$OR_{T\_a} > 1$	↓ as OR_{T_a} ↓
R	↑ as R approaches 50%; ↓ as R moves away from 50%
P_{X_1}	Unaffected by P_{X_1} unless OR_{T_a} is very large
$OR_{X_2\_a}$	↓ as $OR_{X_2\_a}$ ↑
$OR_{X_1\_a}$	↓ as $OR_{X_1\_a}$ moves away from 1 ; ↑ as $OR_{X_1\_a}$ moves towards 1
σ^2	↓ as σ^2↑

* $ROR > 1$: $OR_{T\_u} > OR_{T\_a}$; $ROR = 1$: $OR_{T\_u} = OR_{T\_a}$; $ROR > 1$: $OR_{T\_u} > OR_{T\_a}$

Table 3-7 Mean (standard deviation) [minimum to maximum] percentage difference between the unadjusted and adjusted treatment effect at different combinations of event rate, X_1 prevalence, adjusted treatment effect, X_1 and X_2 prognostic strength, and X_2 variance values

OR_{X1_a}	σ^2	OR_{X2_a}	OR_{T_a}		
			0.7	2	5
1.09	1	1.2	0.24% (0.16%) [-0.3% to 5.9%]	-0.41% (0.20%) [-0.8% to -0.1%]	-0.94% (0.35%) [-1.7% to -0.2%]
		2	2.7% (0.90%) [0.6% to 4.5%]	-5.16% (1.50%) [-7.2% to -2.1%]	-10.90% (3.26%) [-15.7% to -4.6%]
	10	1.2	2.0% (0.70%) [0.5% to 3.7%]	-3.65% (1.08%) [-5.4% to -1.3%[	-8.02% (2.46%) -11.3% to -2.9%
		2	15.03% (2.0%) [9.7% to 19.1%]	-23.49% (2.41%) [-27.0% to -17.9%]	-45.96% (3.58%) [-50.1% to -38.2%]
0.67 or 1.5	1	1.2	0.45% (0.20%) [-0.1% to 0.9%]	-0.83% (0.32%) [-1.5% to -0.2%]	-1.77% (0.63%) [-2.8% to -0.3%]
		2	2.92% (0.89%) [0.4% to 4.6%]	-5.44% (1.43%) [-7.9% to -1.7%]	-11.61% (3.04%) [-15.4% to -5.4%]
	10	1.2	2.18% (0.77%) [-0.2% to 3.7%]	-4.03% (1.10%) [-5.7% to -1.4%]	-8.64% (2.48%) [-12.2% to -3.5%]
		2	14.77% (2.21%) [6.9% to 18.6%]	-23.48% (2.33%) [-27.2% to -16.7%]	-45.98% (3.93%) [-50.5% to -37.6%]
0.2 or 5	1	1.2	3.28 (1.36%) [0.6% to 5.5%]	-5.99 (2.39%) [-9.8% to -1.0%]	-12.63% (5.04%) [-20.9% to -2.4%]
		2	5.32% (1.58%) [1.7% to 8.2%]	-9.48% (2.54%) [-13.5% to -3.9%]	-19.78% (5.25%) [-27.1% to -8.0%]
	10	1.2	4.69% (1.48%) [1.3% to 7.6%]	-8.41% (2.43%) [-12.2% to -3.1%]	-17.60% (5.10%) [-25.3% to -6.4%]
		2	15.5% (2.23%) [10.0% to 19.7%]	-24.57% (2.64%) [-28.6% to -16.9%]	-47.72% (3.73%) [-52.6% to -39.3%]

3.6 Summary

The simulations reported in this chapter showed that when both covariates were balanced across both treatment groups and the two covariates were uncorrelated, the treatment effect tended to be closer to no treatment effect if the covariates were not adjusted for in the analysis. This is consistent with simulation results published by others using hypothetical data and assuming that covariates are balanced across the treatment groups (Hernández, Steyerberg and Habbema, 2004; Negassa and Hanley, 2007; Pang, Kaufman and Platt, 2013).

Negassa and Henley (2007) carried out a simulation to investigate the effect of an omitted covariate on the odds ratio. They compared scenarios in which the adjusted odds ratio of the exposed was 2 and the odds ratio of the covariate ranged from 2 to 10. A binary covariate with a prevalence of 50% and a continuous covariate with variances of 1 and 5 were considered separately in the simulation. The event rate in the unexposed group was set at 40%. They found that the percentage ‘bias’, which was defined as the percentage difference between the empirical median unadjusted and adjusted log odds ratios, increased with the odds ratio of the covariate and was more apparent when the continuous covariate was highly variable. In the most extreme case, they found a percentage bias of over 80%.

Hernández, Steyerberg and Habbema (2004) reported similar results. They found differences of 7.7% and 9.5% between the unadjusted and adjusted odds

ratios when the unadjusted odds ratio was set to 1.4 and 1.7, respectively, with a strong unadjusted covariate effect (odds ratio = 5). In the most extreme case, when the unadjusted covariate effect was very strong (odds ratio = 30), the percentage difference was 42.9%. They also examined how the impact was affected by a changing event rate and covariate prevalence, but the focus was primarily on sample size reduction from adjusted analysis.

Pang, Kaufman and Platt (2013) carried out a more detailed investigation of the impact of covariate adjustment. They quantified the noncollapsibility effect as the difference between the marginal log odds ratio estimated from a marginal structural model and the unadjusted log odds ratio, using analytical and graphical approaches. They found that the noncollapsibility effect ranged from -0.3 to 0.3, depending on the event rate, prevalence, and prognostic strength of the binary covariate.

The results from these prior studies may not be directly comparable with the results obtained from the simulation study in this chapter as the difference between the unadjusted and adjusted results was calculated differently and different numbers of parameters and their values were used. However, their conclusions were broadly similar to what was found in this chapter. This simulation study took the exploration of the factors that affect the impact that adjustment has further than the published simulation studies by simultaneously considering two covariates.

In conclusion, this chapter has shown that the extent of the difference between unadjusted and adjusted results depends on a combination of factors. The relative difference between the unadjusted and adjusted odds ratios was shown to reach more than 50% in extreme scenarios only, such as very large treatment effects and highly prognostic covariates. In most common clinical trial scenarios, the differences due to noncollapsibility were small. The direction of association between the covariates and outcome did not affect the difference between the unadjusted and adjusted results.

CHAPTER 4: IMPACT OF CORRELATIONS BETWEEN COVARIATES ON COVARIATE ADJUSTMENT ANALYSIS

4.1 Introduction

The simulation study performed in Chapter 3 was based on the assumption that the two covariates were independent of one another (i.e., that there was no correlation between them). This chapter explores the role of the correlation between the covariates in adjusted analysis. The specific aims of this investigation are as follows:

1. Determine whether a correlation between the covariates affects adjusted analysis.
2. If there is an effect, determine whether and how it is varied by the magnitude or direction of the correlation.

4.2 Method

New population data sets were generated by incorporating different correlation values into the data simulation programme introduced in Chapter 3.

4.2.1 Determining the correlation coefficients

The magnitude of the correlation between the covariates used to generate the data sets was determined using an assessment of empirical data from four RCTs, the International Subarachnoid Aneurysm Trial (ISAT) (Molyneux et al, 2005), the Spine Stabilisation Trial (SST) (Fairbank et al, 2005), the Arterial Revascularisation Trial (ART) (Taggart et al, 2010), and the Memantine for Dementia in Adults with Down's Syndrome Study (MEADOWS) (Hanney et al, 2012). These trials were selected because they give a general impression of the level of correlation between covariates across a diverse range of disease specialties (neurosurgery, lower back pain, dementia, and cardiothoracic surgery). Table 4-1 shows the correlation matrix of the baseline covariates used in the randomisation process from the ISAT trial as an example. The correlation matrices for the other three studies are presented in Appendix C. The correlation coefficients from these studies lay between -0.37 to 0.3. Correlation coefficients of -0.3, -0.1, 0.1 and 0.3 were therefore chosen for generating data sets for the simulation study.

Table 4-1: Correlation matrix of the baseline covariates in ISAT

	Age at randomisation	Male	WFNS*	Lumen size
Age at randomisation	1.0000			
Male	-0.1221	1.0000		
WFNS*	0.0841	-0.0268	1.0000	
Lumen size	0.0125	0.0054	0.0629	1.0000

* WFNS = World Federation of Neurosurgical Societies grade scale

4.2.2 Data generation

A similar simulation procedure to that described in Chapter 3 was used. Intercept values were identified to serve as initial values for generating the population data sets. Population data sets of 50,000 cases with a treatment group, one binary covariate and one continuous covariate were generated. Each data set was generated from a multivariate normal distribution with variance-covariance matrices that gave the same data structure as in Chapter 3 (i.e., 1:1 treatment allocation and balanced distribution of covariates across the treatment groups).

The correlation between the two covariates was also incorporated into the simulation programme. The treatment group and binary covariate were generated as continuous variables, then dichotomised. The variance-covariance structure between the continuous and binary covariates was obtained from a separate simulation, as the correct correlation differed from the variance-covariance structure of two continuous covariates. Only two levels of binary covariate prevalence were incorporated, P_{X_1} = 20% and 50%. The values for the other parameters were set as in Chapter 3 and are displayed in Table 4-2. Details of the procedure are reported in Box 4-1. Appendix D describes the simulation programme.

Table 4-2 Values of the parameters used in the simulation

Parameters		**Values**
Correlation coefficient, ρ		-0.3, -0.1, 0, 0.1, 0.3
Overall event rate, R		10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%
Binary covariate, X_1	Adjusted odds ratio (prognostic strength), $OR_{X_1\_a} = e^{\gamma_{1\_a}}$	5 (strongly positive) 1.5 (moderately positive) 1.09 (weakly positive) 0.67 (moderately negative) 0.2 (strongly negative)
	% with $X_1 = 1$, P_{X_1}	20%, 50%
Continuous covariate, X_2	Adjusted odds ratio (prognostic strength), $OR_{X_2\_a} = e^{\gamma_{2\_a}}$	2 (strongly positive) 1.2 (moderately positive)
	$X_2 \sim N(0, \sigma^2)$	N(0, 1), N(0, 10)
Adjusted odds ratio for T, $OR_{T\_a} = e^{\theta_a}$		0.7, 2, 5

Box 4-1: Procedure for generating the population data, incorporating a correlation between the two covariates

Set the covariance between X_1 and X_2. Set $cov(T, X_1)$ and $cov(T, X_2)$ to zero.

1. Generate multivariate normal distribution data for T, X_1 and X_2
2. Dichotomise T and X_1
3. Calculate the correlation between X_1 and X_2
4. Store the covariance and correlation
5. Generate population data using the derived variance-covariance matrix
6. Fit the adjusted and unadjusted regression
7. Calculate and store the results

4.3 Results

4.3.1 Typical clinical trial scenario

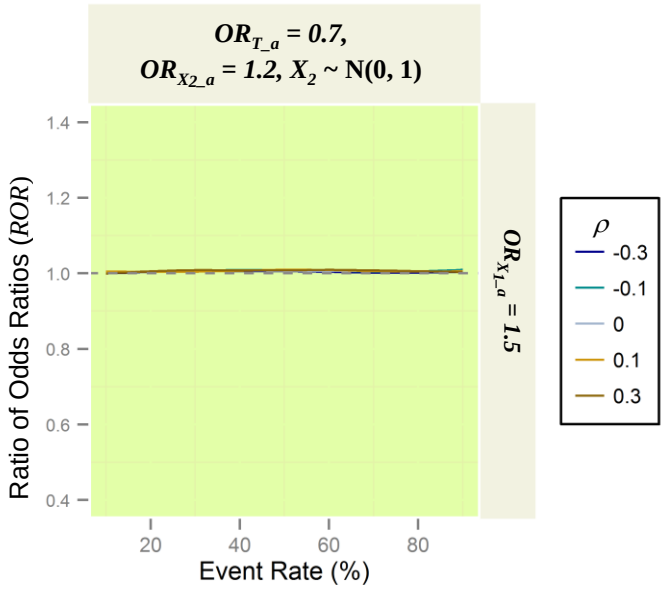
Table 4-3 shows the results from analysing the simulated population data for a typical clinical trial scenario with a moderate treatment effect, moderately prognostic binary covariate, and moderately prognostic continuous covariate. The unadjusted treatment effects were higher than the adjusted results. However, the differences were small, leading to slight changes in ROR as the level of covariate correlation changed. ROR ranged from 1.0059 to 1.0101.

The results presented in Figure 4-1 show the relationship between ROR and event rate. Similar to the graphical presentation in Chapter 3, the grey dotted line represents the line of no difference, where the unadjusted and adjusted treatment effects are the same and ROR = 1.0. The coloured curves above the dotted line are ROR curves and describe the relationship between ROR and event rate at different levels of correlation between the binary and continuous covariates. The blue and dark blue curves represent negative correlations between the two covariates of -0.1 and -0.3, respectively. The brown and dark brown curves represent positive correlations of 0.1 and 0.3, respectively. The light grey line shows the uncorrelated covariate results and corresponds to the results reported in Chapter 3. These results suggest that adjustment had a minor effect under this RCT scenario, regardless of the overall event rate.

Table 4-3 Odds ratios at different covariate correlations (OR_{T_a} = 0.7, $OR_{X_2\_a}$ = 1.2, σ^2 = 1, $OR_{X_1\_a}$ = 1.5, P_{X_1}= 50%, and R = 50%)

Odds ratio/ Ratio of odds ratios	Correlation, (ρ)				
	-0.3	-0.1	0	0.1	0.3
Unadjusted odds ratio (OR_{T_u})	0.7058	0.7103	0.7074	0.7023	0.7053
Adjusted odd ratio (OR_{T_u})	0.7010	0.7046	0.7033	0.6953	0.6995
Ratio of odds ratios (ROR)	1.0068	1.0080	1.0059	1.0101	1.0082

Figure 4-1 Ratio of the unadjusted and adjusted odds ratios as the overall event rate changes, calculated in a typical clinical trial scenario (P_{X_1}= 50%) and with varying correlation between the two covariates



4.3.1.1 Changes in prognostic strength and covariate distribution.

4.3.1.1.1 Varying the distribution and prognostic strength of the binary covariate, X_1

Figure 4-2 compares the relationship between ROR and event rate at two prevalence levels of a binary covariate of moderate prognostic strength ($OR_{X_1\_a}$ = 1.5), with varying levels of correlation between the binary and continuous covariates. The prevalence of the binary covariate was set to 20% (light green background) or 50% (dark green background). The relationship between ROR and event rate was unchanged by the tested levels of covariate correlation and X_1 prevalence tested.

Similar results were found when the ROR curves at different correlation coefficients were compared with one another at different levels of X_1 prognostic strength (Figure 4-3). Under each set of scenarios, all of the ROR curves overlapped. The correlation between the covariates therefore did not affect ROR in scenarios with a modest treatment effect on the outcome.

Figure 4-2 **Ratio of the unadjusted and adjusted odds ratios as the overall event rate changes, calculated in a typical clinical trial scenario (OR_{T_a} = 0.7, $OR_{X_1\_a}$ = 1.5, $OR_{X_2\_a}$= 1.2, and σ^2 = 1) with varying correlation strength between the two covariates, at two P_{X_1} values**

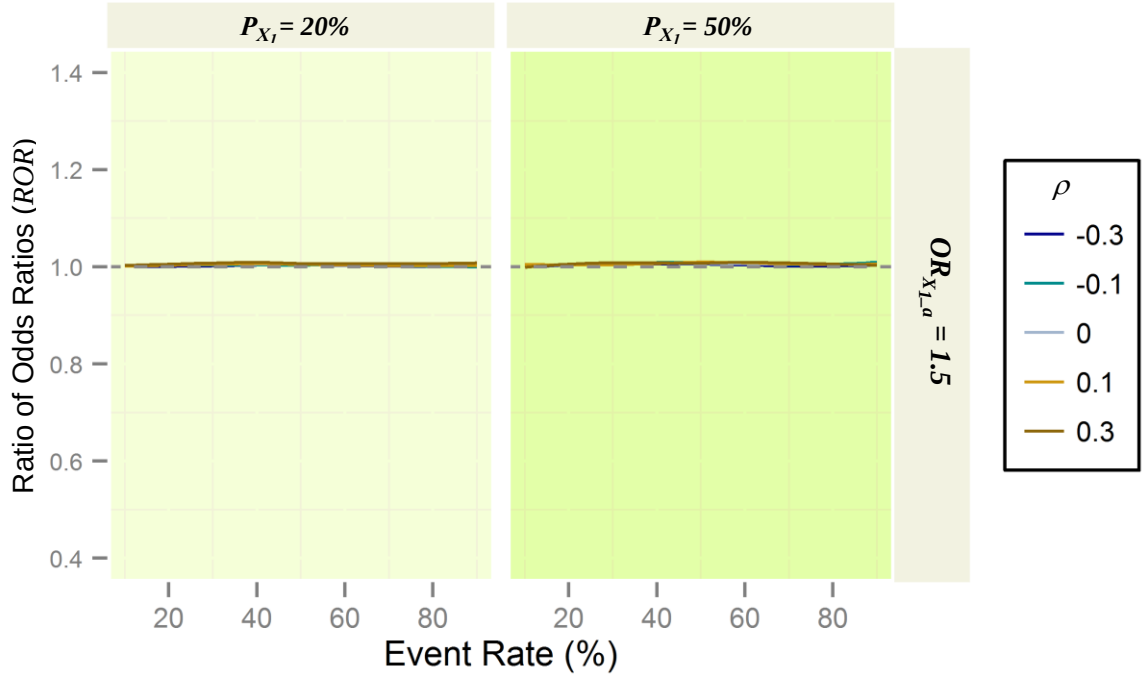
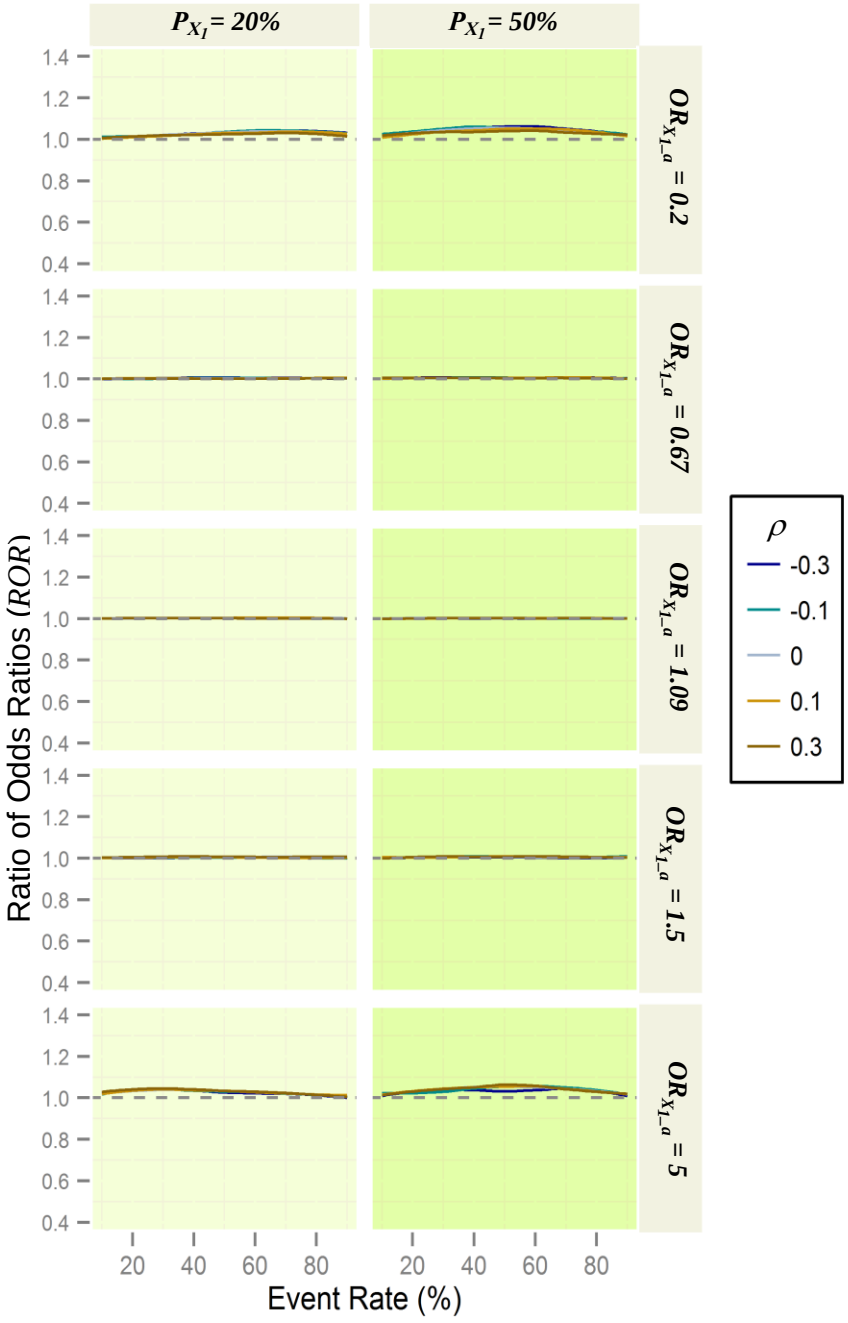


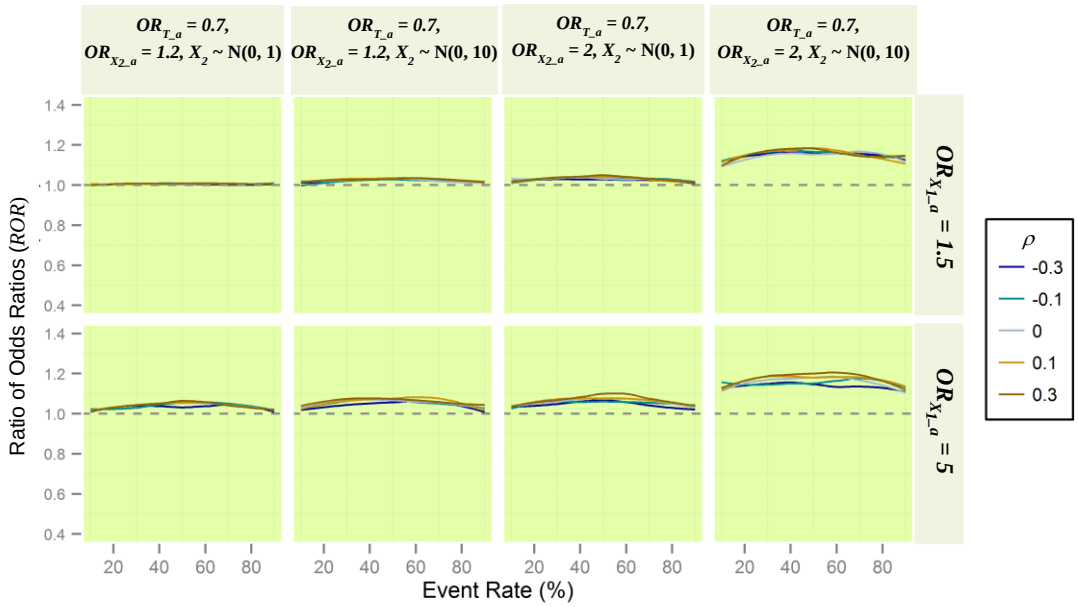
Figure 4-3 **Ratio of the unadjusted and adjusted odds ratios as the overall event rate changes, calculated in a typical trial setting (OR_{T_a} = 0.7, $OR_{X_2\_a}$ = 1.2, and σ^2 = 1) with varying correlation between the two covariates, distribution of P_{X_1}, and $OR_{X_1\_a}$**



4.3.1.1.2 Varying the distribution and prognostic strength of the continuous covariate, X_2

The analysis was repeated for the continuous covariate (Figure 4-4). Again, ROR was largely unaffected by the changes in correlation when the prognostic strength was moderate to strong. However, some variation between the ROR curves across the correlation coefficients did occur when both covariates were considered simultaneously and both had a very strong prognostic effect on the outcome. Some differences were observed when both covariates were very highly prognostic to the outcome ($OR_{X_1\_a}$ = 5, $OR_{X_2\_a}$ = 2, and σ^2 = 10, which is equivalent to a standardised odds ratio of 9.35). The ROR curves moved further away from 1 when there was a positive correlation between the two covariates (ρ = 0.1 or 0.3), and moved closer to 1 when there was an inverse relationship between the covariates (ρ = -0.1 or -0.3).

Figure 4-4 **Ratio of the unadjusted and adjusted odds ratios as the overall event rate changes, calculated in a typical trial scenario (OR_{T_a} = 0.7, and P_{X_1} = 50%) with varying correlation between the two covariates, $OR_{X_1\_a}$, $OR_{X_2\_a}$, and the distribution of X_2**



4.3.2 Large treatment effect scenario

4.3.2.1 Moderate prognostic effect of the covariates

When the treatment effect was large (OR_{T_a} = 5) and the prognostic strength for both covariates was moderate ($OR_{X_2\_a}$ = 1.5 or 1.2), the impact of covariate adjustment was marginally affected by the magnitude of the correlation between the covariates (Figure 4-5). The dark brown *ROR* curve in Figure 4-5, which corresponds to a correlation coefficient of 0.3, is very slightly separated from the dark blue curve, which corresponds to correlation coefficient of -0.3. This is also illustrated by the results reported in Table 4-4: at P_{X_1} = 20% and event rate = 50%, *ROR* decreased from 0.9860 to 0.9684 when the correlation coefficient between

the covariates increased from -0.3 to 0.3. However, the difference in ROR was small. Table 4-4 shows that the same trend also existed for P_{X_1}= 50%.

Figure 4-5 Ratio of the unadjusted and adjusted odds ratios as the overall event rate changes, calculated in a clinical trial with a very large treatment effect (OR_{T_a} = 5) and covariates with a moderate prognostic effect ($OR_{X_1\_a}$ = 1.2, P_{X_1} = 50%, $OR_{X_2\_a}$ = 1.5, and σ^2 = 1), with varying correlation between the two covariates

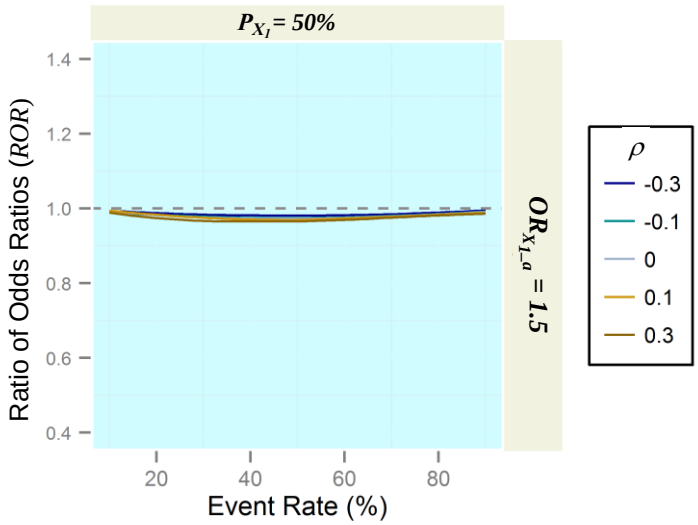


Table 4-4 Change in the odds ratios and ratios of odds ratios as the covariate correlation changes (OR_{T_a} = 5, $OR_{X_2\_a}$ = 1.2, σ^2 = 1, $OR_{X_1\_a}$ = 1.5, and R = 50%)

Odds ratio/ ratio of odds ratios	Correlation, (ρ)				
	-0.3	-0.1	0	0.1	0.3
P_{X_1} = 20%					
Unadjusted odds ratio (OR_{T_u})	4.9047	4.9461	4.9131	4.8564	4.8419
Adjusted odds ratio (OR_{T_a})	4.9744	5.0425	5.0204	4.9913	4.9998
Ratio of odds ratio (ROR)	0.9860	0.9809	0.9786	0.9730	0.9684
P_{X_1} = 50%					
Unadjusted odds ratio (OR_{T_u})	4.9339	4.8530	4.8998	4.8974	4.8598
Adjusted odds ratio (OR_{T_a})	5.0188	4.9898	5.0385	5.0154	5.0264
Ratio of odds ratio (ROR)	0.9830	0.9726	0.9725	0.9718	0.9669

The brightness of the cell colour indicates how ROR decreases as the correlation coefficient increases

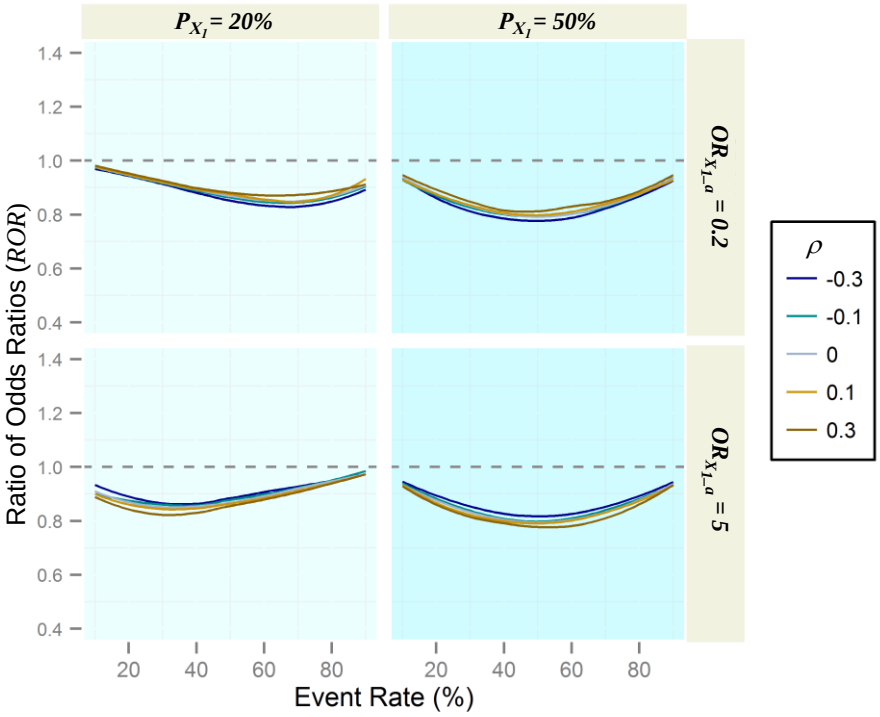
4.3.2.2 Effect of a strongly prognostic binary covariate

When the binary covariate was highly prognostic, some differences in the ROR curves were observed as the level of correlation changed. Figure 4-6 shows the ROR curves at different correlation coefficients and either P_{X_1} = 20% (light blue background) or P_{X_1} = 50% (dark blue background). In all four plots, the shape of the ROR curve at every correlation strength is similar to the curve for no correlation between the covariates. All of the curves show a convex relationship between ROR and event rate, and lie below 1. The depth of the curve depended on factors such as the direction and magnitude of the correlation, the event rate, and the prevalence and direction of the prognostic strength of the binary covariate.

At both P_{X_1} = 20% and 50%, the ROR curves were deepest for the negative correlation (ρ = -0.3), as the binary covariate was negatively correlated with both the outcome and the continuous covariate (blue curves in the top two graphs of Figure 4-6). The curves for a negative association with the outcome but no, or a positive, correlation between the covariates were slightly shallower. The order of the ROR curves was reversed in the presence of a highly positive prognostic binary covariate (bottom two panels of Figure 4-6): the ROR curve with the largest positive correlation coefficient (brown curves) is the furthest from the line at ROR = 1.

When P_{X_1} = 50%, the variation in ROR was maximised when the event rate approached 50% and was minimised at extreme event rate values (< 20% and > 80%). In contrast, when P_{X_1} = 20%, the variation in ROR was greater at one end of the event rate range (e.g., 70% for $OR_{X_1\_a}$ = 0.2) for both adjusted odds ratios of X_1 than at the opposite end of the event range (e.g. 20%).

Figure 4-6 **Ratio of the unadjusted and adjusted odds ratios as the overall event rate changes, calculated in a clinical trial with a very large treatment effect (OR_{T_a} = 5) and a continuous covariate with a moderate prognostic effect ($OR_{X_2\_a}$ = 1.2 and σ^2 = 1) with varying correlation between the two covariates, distribution of P_{X_1}, and a binary covariate that either has a very strongly negative or very strongly positive prognostic effect**

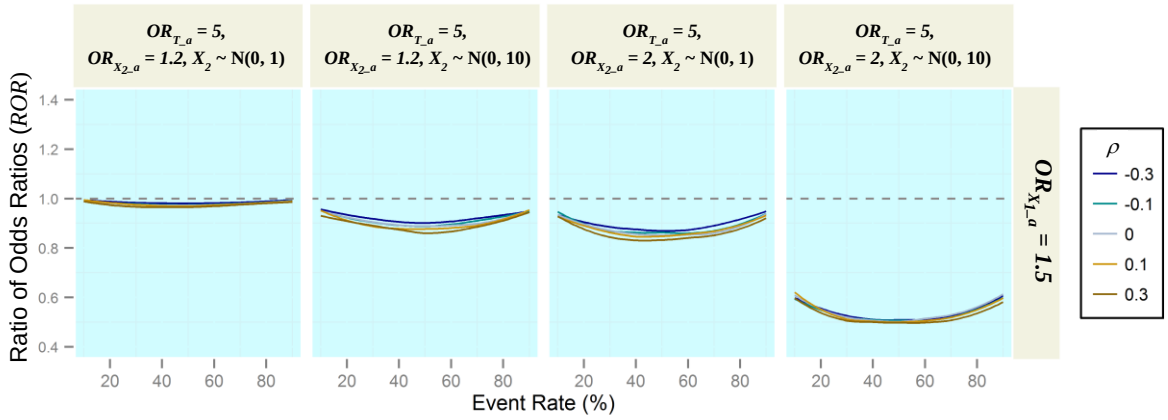


4.3.2.3 Effect of a strongly prognostic continuous covariate

Figure 4-7 presents the ROR curves at different correlation coefficients and varying prognostic strength and variance of the continuous covariate. The middle two plots show the ROR curve changing as the correlation coefficient changes when the continuous covariate is moderately ($OR_{X_2\_a}$ = 1.2) or strongly ($OR_{X_2\_a}$ = 2) associated with the outcome. As the second plot ($OR_{X_2\_a}$ = 1.2 and σ^2 = 10)

has a standardised odds ratio of 1.82, it is not surprising that the ROR curves in this plot are similar to the curves for scenarios with $OR_{X_2\_a}$ = 2 and σ^2 = 1. At an extreme prognostic effect (i.e., $OR_{X_2\_a}$ = 2 and σ^2 = 10), the difference in the ROR curves at each correlation coefficient value becomes small again. The impact of covariate adjustment may have become dominated by the extreme prognostic effect of the continuous covariate.

Figure 4-7 Ratio of the unadjusted and adjusted odds ratios as the overall event rate changes, calculated in a clinical trial with a very large treatment effect (OR_{T_a} = 5) and the binary covariate with a moderately prognostic effect ($OR_{X_1\_a}$ = 1.5, P_{X_1} = 50%, and σ^2 = 1), with varying correlation between the two covariates and a very strongly positively prognostic continuous covariate



4.3.2.4 Effect of strongly prognostic continuous and binary covariates

The ROR curve changed more obviously as the correlation coefficient changed when both covariates were strongly prognostic of the outcome. Figure 4-8 shows

overlapping ROR curves when $OR_{X_2\_a}$ = 1.2 and σ^2 = 1, but clearly separate curves when $OR_{X_2\_a}$ = 2 and σ^2 = 10.

When $OR_{X_2\_a}$ = 1.2 and σ^2 = 10, the ROR curves for each correlation coefficient were further from one another at event rates close to 50% and were closer together at both extreme ends of the event rate range (second panel of Figure 4-8). A similar pattern was seen when $OR_{X_2\_a}$ = 2 and σ^2 = 1. However, when $OR_{X_2\_a}$ = 2 and σ^2 = 10 (far right panel of Figure 4-8), the ROR curves for each correlation coefficient lay a similar distance from one another for all values of event rate. When the directions of the relationships between the two covariates and the outcome were opposite to one another, the ROR curves with negative correlation coefficients lay further away from the line of ROR = 1 and the ROR curve with no correlation. In contrast, the ROR curves with positive coefficients lay closer to the line of ROR = 1, but did not overlap with the ROR curve with no correlation in the opposite direction.

These patterns are further illustrated in Table 4-5, which shows results for a subset of the scenarios (OR_{T_a} = 5, $OR_{X_2\_a}$ = 2, σ^2 = 1, and P_{X_1} = 50%). The orange shading in the cells, from dark to light, represents ROR values from high to low. ROR increased towards 1 when $OR_{X_1\_a}$ = 0.2, indicated that the difference between the unadjusted and adjusted odds ratios decreased as the correlation

coefficients moved from -0.3 to 0.3. In contrast, ROR decreased when $OR_{X_1\_a}$ = 5, indicating that the difference widened as the correlation coefficient moved from -0.3 to 0.3. The largest difference between the unadjusted and adjusted odds ratios occurred at an event rate of 50% (highlighted in green in Table 4-5) for all correlation coefficients. The results for other scenarios are reported in Appendix E.

Figure 4-8 Ratio of the unadjusted and adjusted odds ratios as the overall event rate changes, calculated in a clinical trial with a very large treatment effect (OR_{T_a} = 5) and covariates with a moderate prognostic effect ($OR_{X_1\_a}$ = 0.2 and 5, P_{X_1} = 50%, and σ^2 = 1), with varying correlation between the two covariates and a binary covariate with either a very strongly negative or positive prognostic effect

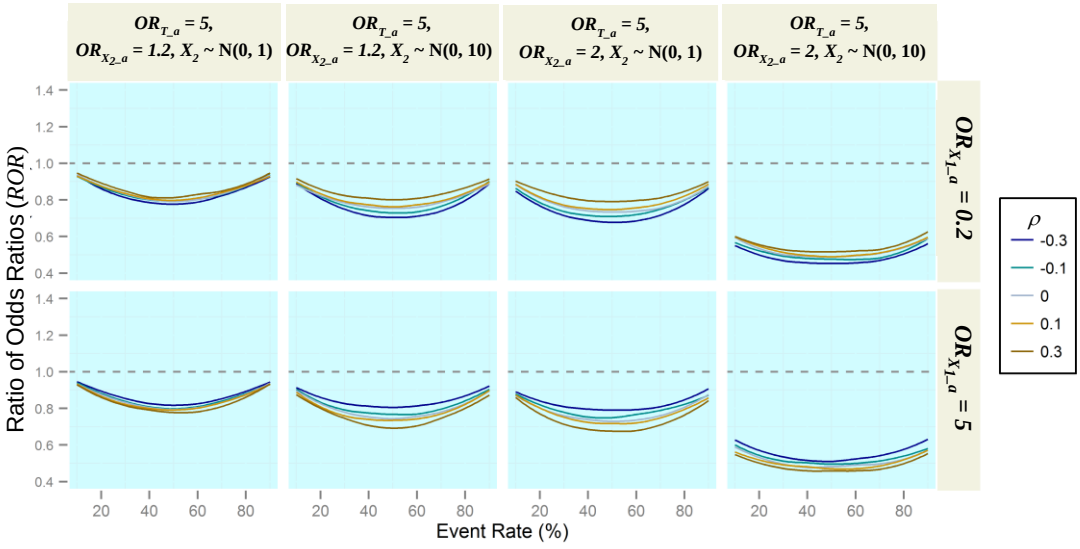


Table 4-5 Odds ratios and ratios of odds ratios at different levels of covariate correlation (OR_{T_a} = 5, $OR_{X_2\_a}$ = 2, σ^2 = 1, $OR_{X_1\_a}$ = 0.2 and 5, and P_{X_1} = 50%)

Event rate (R)	OR / ROR	$OR_{X_1\_a}$ = 0.2					$OR_{X_1\_a}$ = 5				
		Correlation between X_1 and X_2 (ρ)					Correlation between X_1 and X_2 (ρ)				
		-0.3	-0.1	0	0.1	0.3	-0.3	-0.1	0	0.1	0.3
10%	OR_{T_u}	4.2525	4.2926	4.4039	4.4692	4.483	4.4772	4.3706	4.3877	4.3675	4.3129
	OR_{T_a}	5.0063	4.9714	4.9769	5.0244	4.9679	5.0204	4.9610	5.0304	5.0138	5.0108
	ROR	**0.8494**	**0.8635**	**0.8849**	**0.8895**	**0.9024**	**0.8918**	**0.8810**	**0.8722**	**0.8711**	**0.8607**
30%	OR_{T_u}	3.5463	3.6686	3.8275	3.9263	4.107	4.0743	3.9056	3.7981	3.6888	3.5744
	OR_{T_a}	4.9600	4.9968	4.9923	5.0480	5.0058	5.0314	4.9828	4.9908	4.9530	5.0224
	ROR	**0.7150**	**0.7342**	**0.7667**	**0.7778**	**0.8205**	**0.8098**	**0.7838**	**0.7610**	**0.7448**	**0.7117**
50%	OR_{T_u}	3.3656	3.5268	3.6419	3.7051	3.9428	3.9428	3.7333	3.6234	3.5658	3.3713
	OR_{T_a}	4.9639	4.9734	4.9565	4.9515	4.9550	4.9803	5.0415	4.9729	4.9774	5.0193
	ROR	**0.6780**	**0.7091**	**0.7348**	**0.7483**	**0.7957**	**0.7917**	**0.7405**	**0.7286**	**0.7164**	**0.6717**
70%	OR_{T_u}	3.5495	3.7833	3.7894	3.9075	4.0471	4.0572	3.8783	3.8244	3.7460	3.4959
	OR_{T_a}	4.9496	5.0369	5.0380	5.0309	5.0053	5.0128	4.9501	5.0043	5.0400	4.9789
	ROR	**0.7171**	**0.7511**	**0.7522**	**0.7767**	**0.8086**	**0.8094**	**0.7835**	**0.7642**	**0.7433**	**0.7022**
90%	OR_{T_u}	4.3262	4.3754	4.3606	4.3872	4.4875	4.5385	4.4004	4.3553	4.3103	4.1787
	OR_{T_a}	5.0103	5.0264	4.9828	4.9486	4.9908	4.9988	5.0113	4.9838	5.0385	4.9749
	ROR	**0.8635**	**0.8705**	**0.8751**	**0.8866**	**0.8992**	**0.9079**	**0.8781**	**0.8739**	**0.8555**	**0.8400**

4.4 Summary of the Simulation Results

Figures 4-9 and 4-10 summarise the findings from the simulations at P_{X_1} = 20% and 50%, respectively. Both figures suggest that the correlation between the covariates made little difference to the overall impact on covariate adjustment, except when both covariates and the treatment effect had high odds ratios. The impact of the correlation between the covariates on adjusted analysis depended on the size of the treatment effect and the prognostic strength of the covariates on the outcome. The covariate distributions and event rate also affected the impact, but their effect was dependent on the treatment effect and covariate prognostic strength. For example, when the treatment effect was small or both covariates were moderately prognostic, the correlation between the covariates did not affect the analysis at any covariate distribution or event rate. When the treatment effect was large and both covariates were strongly prognostic, the difference between the unadjusted and adjusted results varied more with the correlation coefficients of the covariates at lower event rates when the prevalence of the binary covariate was low (Figure 4-9). However, there was more variation in the difference as the event rate approached 50% when the prevalence of the binary covariate was 50% (Figure 4-10).

In extreme scenarios, the direction of the correlation between the two covariates affected ROR, depending on the directional association of the covariates with the

outcome. This effect also increased as the magnitude of association or correlation increased.

Figure 4-9 Ratios of the unadjusted and adjusted odds ratios as the overall event rate changes at different levels of covariate correlation for all of the scenarios investigated (P_{X_1} = 20%)

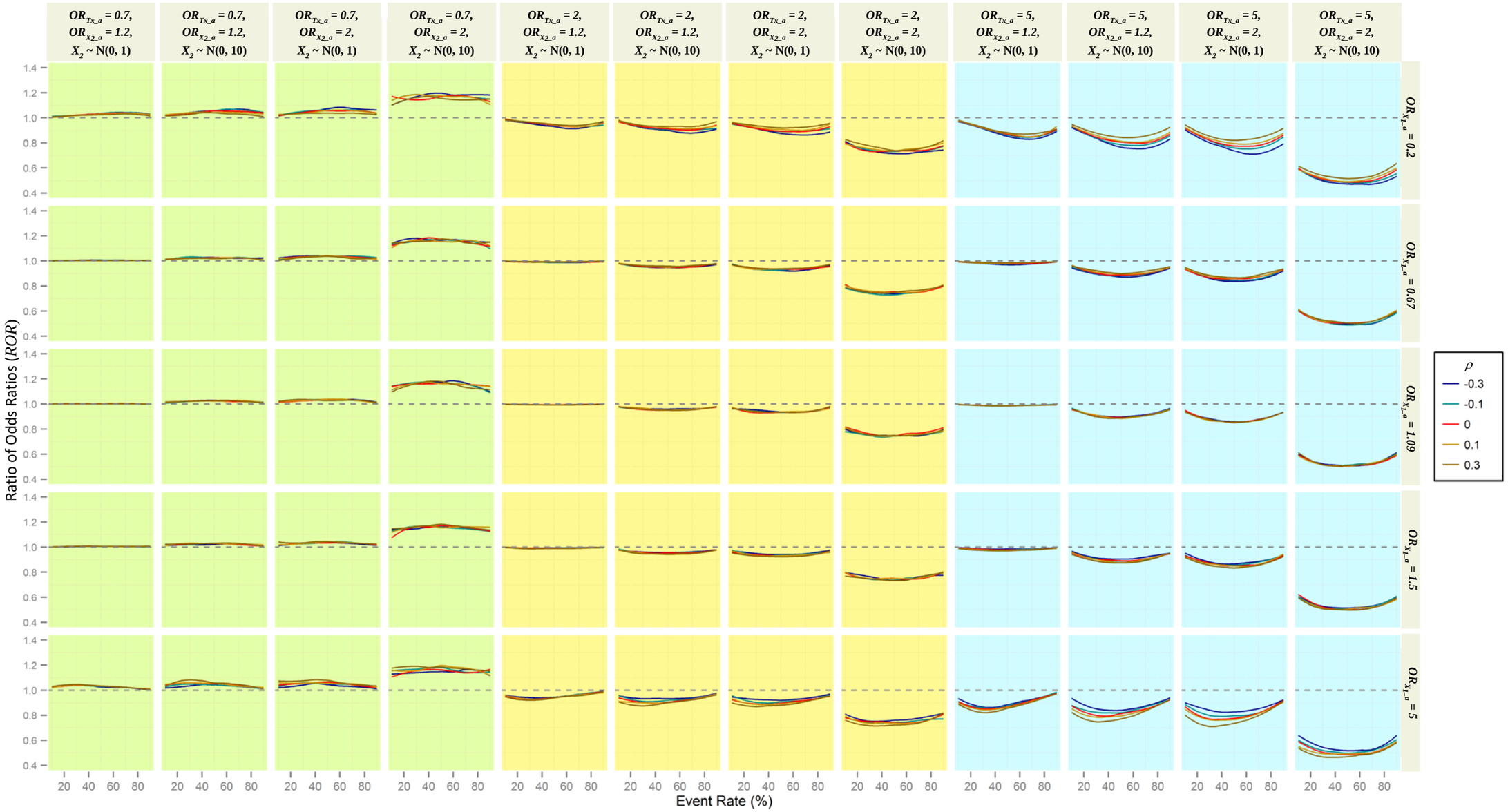
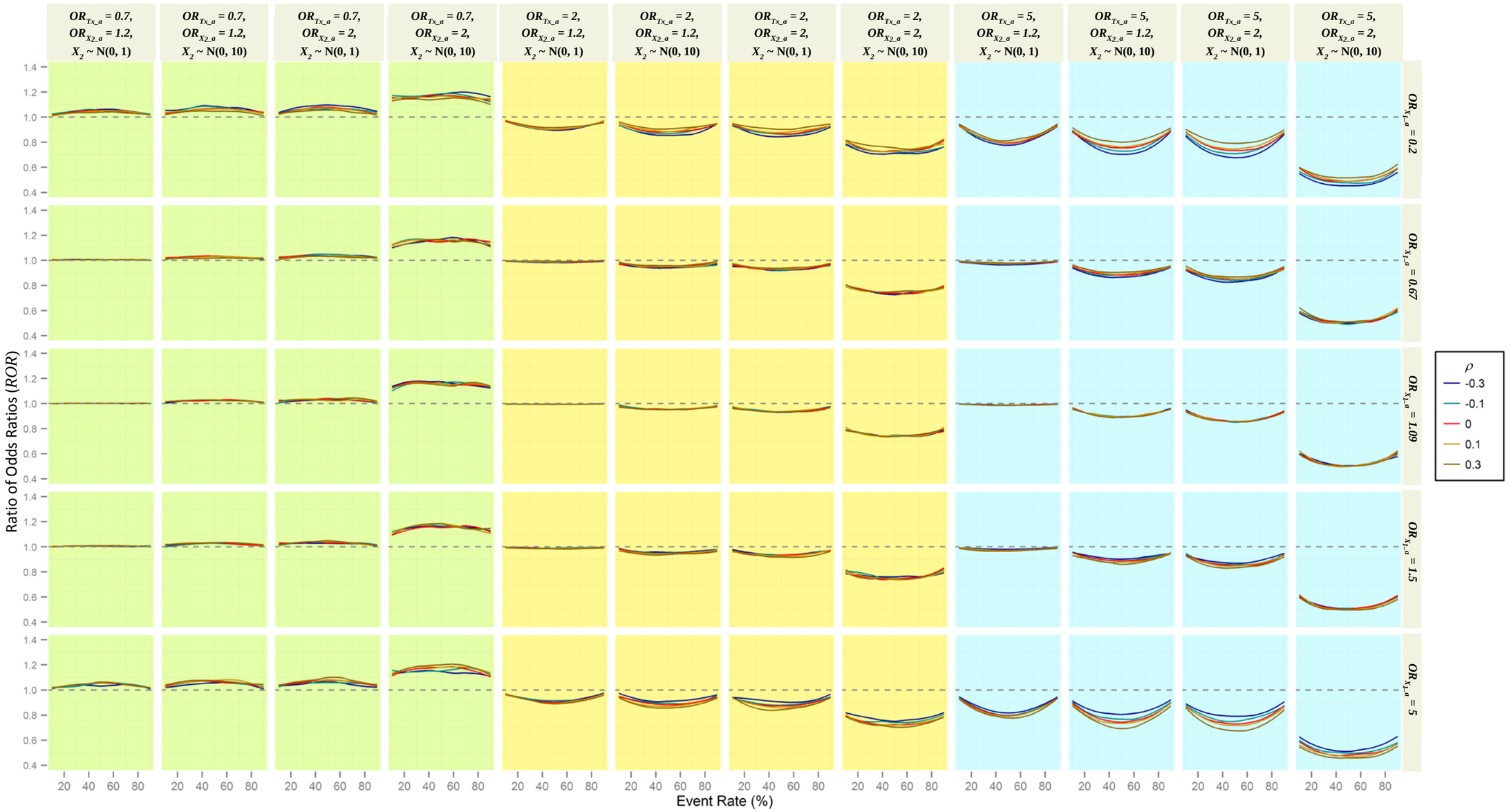


Figure 4-10 Ratios of the unadjusted and adjusted odds ratios as the overall event rate changes at different levels of covariate correlation for all of the scenarios investigated (P_{X_1} = 50%)



4.5 Summary

Published studies based on the simulation of hypothetical data have focused on one covariate (Hernández, Steyerberg and Habbema, 2004; Negassa and Hanley, 2007; Pang, Kaufman and Platt, 2013), so have not investigated whether the correlation between covariates can affect adjusted analysis. Including two covariates and the correlation between them in the adjusted model produces a theoretical scenario that is closer to real practice.

Prior studies that have used data from other clinical trials or observational studies are likely to have encountered correlations between the covariates (e.g., O'Conner, Mentz and Whellan, 2011; Kahan et al, 2014; Steyerberg, Bossuyt and Lee, 2000; Roozenbeek et al, 2009; Thompson et al, 2014), but have not fully examined the correlation structure of the covariates and its effects. The main aim of most of these studies was to assess the gain in power after adjustment. A change in an adjustment due to correlations between the covariates is also likely to affect power.

The results of the simulations in this chapter have shown that, in extreme scenarios, the difference between the unadjusted and adjusted treatment effects can be increased or reduced when the correlation between the covariates is considered with the parameters assessed in Chapter 3. The maximum percentage difference obtained between scenarios that did and did not take into

account correlation between the covariates was 7.6%. Larger differences are likely if higher correlation coefficients are considered.

CHAPTER 5: IMPACT OF SAMPLE SIZE AND COVARIATE IMBALANCE ON ADJUSTED ANALYSIS

5.1 Introduction

Thus far, the impact of covariate adjustment has been investigated using population data generated to represent different scenarios in randomised trials. Each generated data set contained a large number of observations and the distributions of the two covariates were perfectly balanced between the two treatment groups. The observed differences between the unadjusted and adjusted results therefore related solely to noncollapsibility.

The likelihood of covariate imbalance is closely linked to sample size. Covariate imbalance is likely to occur in smaller studies as they suffer from more sampling errors (Roberts and Torgerson, 1999). Treatment effect estimates are thus affected if any of a small study's covariates are highly associated with the outcome and are not handled properly (Pocock et al, 2002).

In this chapter, the change in the impact of covariate adjustment in the presence of unbalanced covariates and varying sample sizes is investigated. This chapter aims to answer the following specific questions:

1. What is the impact of sample size on covariate adjustment when both covariates are balanced?
2. What is the extent of covariate imbalance in relation to sample size?
3. How is the impact of covariate adjustment affected by the degree of covariate imbalance?

5.2 Methods

5.2.1 Resampling procedures

The population data sets generated and reported in Chapter 3 were used as the basis for the resampling study in this chapter. Table 5-1 lists the scenarios included in this investigation and their parameter combinations, which are similar to those described in Chapter 3. Scenarios with a moderate treatment effect (OR_{T_a} = 0.7) were excluded because the results reported in prior chapters showed very little difference between the adjusted and unadjusted results in these scenarios. The prevalence of the binary covariate was fixed at 50% (P_{X_1} = 50%) for the same reason. The investigation included 360 scenarios.

The population data set for each scenario was randomly resampled with replacement, stratified by treatment group. Samples of sizes 100 (50 per group), 200 (100 per group), 500 (250 per group), 1,000 (500 per group), and 5,000 (2500

per group) were drawn for each scenario. The sampling method may have introduced covariate imbalance between the two treatment groups because no constraints were placed on the covariates to maintain a balanced structure across the two groups. This approach is similar to using a randomisation process for treatment allocation in a clinical trial without covariate stratification. It provides information on the likelihood that balanced covariates can be achieved when this method of randomisation is used.

Logistic regression models with and without covariate adjustment were fitted for each sample. The unadjusted and adjusted treatment effects and SEs were calculated and stored. The process was repeated 20,000 times for sample sizes less than 5,000 and 5,000 times for sample sizes of 5,000, using different seed numbers for each resampling process. Randomly sampling small sample sizes from extreme scenarios generated some extreme X_1 distributions, such as less than 10% prevalence. The fitted values at 0 or 1 could then produce extreme results at low event rates, leading to the problems of separation and lack of convergence (Heinze and Schemper, 2002). Instances of nonconvergence were discarded and new samples were generated to replace them. Samples that produced extreme results were retained. The resampling procedure is presented in Box 5-1 and the corresponding programme is reported in Appendix F.

Table 5-1 Parameter values for this resampling study

Parameters		**Values**
Sample size		100, 200, 500, 1,000, 5,000
Overall event rate, R		10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%
Binary covariate, X_1	Adjusted odds ratio (prognostic strength), $OR_{X_1\_a} = e^{\gamma_{1\_a}}$	5 (strongly positive) 1.5 (moderately positive) 1.09 (weakly positive) 0.67 (moderately negative) 0.2 (strongly negative)
	% with $X_1 = 1$, P_{X_1}	50%
Continuous covariate, X_2	Adjusted odds ratio (prognostic strength), $OR_{X_2\_a} = e^{\gamma_{2\_a}}$	2 (strongly positive) 1.2 (moderately positive)
	$X_2 \sim N(0, \sigma^2)$	N(0, 1), N(0,10)
Adjusted odds ratio for T, $OR_{T\_a} = e^{\theta_a}$		2, 5

Box 5-1 Procedure to collect samples of various sizes from the population data sets

1. Select a random sample of size n, sample with replacement and stratify by treatment group
2. Fit an unadjusted logistic regression model and store the log odds ratio of treatment, τ_u, and the corresponding SE, se(τ_u)
3. Fit a logistic regression adjusted for X_1 and X_2, and store the log odds ratio, τ_a, and the corresponding SE, se(τ_a)
4. Compute the sample level results (Table 5-2)
5. Repeat the process 20,000 times for n = 100, 200, 500, and 1,000, and 5,000 times for n = 5,000
6. Exclude observations below the 2.5 percentile or above the 97.5 percentile in the ROR distribution
7. Calculate the scenario-level summary results (Table 5-2)

5.2.2 Outcomes of interest

The covariate imbalance (see section 5.1.2.2 below) and ROR (Eq. 3-3 in Chapter 3) were computed for each sample. ROR values greater than 2, 5, and 10 were considered to be extreme results. The regression coefficients of the unadjusted and adjusted treatment effects, associated SEs (se(τ_u) and se(τ_a), respectively), and ROR results were averaged over the samples, using the 2.5% trimmed distribution (2.5% on each side) of ROR. Studies have shown that maximum likelihood logistic regression may not be appropriate for studies with a small

sample size (Cox and Snell, 1989), as this method can produce biased estimates. However, it is probably the most commonly used practice, and ROR is unlikely to be affected when the estimated treatment effects are biased in both the unadjusted and adjusted analyses.

5.2.2.1 Trimmed values

The trimmed mean is a family of central tendency measures (Bickel and Lehmann, 1975) that involves calculating the mean after discarding a proportion of observations at the two extreme ends of the distribution. The trimmed mean is a more robust measure and is less sensitive to extreme deviations and heavy-tailed distributions than the ordinary sample mean, which includes all observations (Stigler, 1973).

Using regular logistic regression for small samples, especially when R and P_{X_1} are small, can produce extreme effect sizes that may affect the results. Therefore, the observations included in the calculations for each scenario were based on a 2.5% trimmed distribution of the 20,000 (or 5,000) calculated ROR values (excluding the largest 2.5% and smallest 2.5% of the ROR values). Trimming 2.5% from each side of a distribution has been shown to maintain the estimator's efficiency (Sawilowsky, 2002).

5.2.2.2 Measures of imbalance

The magnitude of the imbalance in each sample was computed. Imbalance in the binary covariate (D_{X_1}) was measured as the percentage with X_1 = 1 in the treatment group minus the percentage in the control group. Negative values of D_{X_1} indicate that a higher proportion were allocated to the control group than the treatment group. Positive values indicate that the treatment group has a higher percentage than the control group. For example, in a randomised trial scenario, the percentage of females allocated to each treatment group should be 50%. However, if randomisation does not include stratification by sex, the treatment group could end up with 42% females and the placebo group with 52% females, for example. The magnitude of this imbalance is -10%. The two groups were considered to be balanced when D_{X_1} was in the range -0.5% to 0.5%.

Imbalance in the continuous covariate (D_{X_2}) was measured by the standardised mean (the mean divided by its standard deviation) in the treatment group minus the standardised mean in the control group. The two groups were considered to be balanced when D_{X_2} was in the range -0.05 to 0.05.

5.2.3 Data analysis

The extreme ROR values before trimming and the summarised ROR values before and after trimming were summarised at the global level (over all of the scenarios

and samples) by sample size. The covariate balance and imbalance after trimming was also summarised over all of the scenarios (Table 5-2).

The impact of the sample size on the adjusted analysis in each scenario was explored by comparing the mean of ROR ($\overline{ROR}$) by sample size. ($\overline{ROR}$) was calculated by averaging ROR over the number of trimmed samples (n_{trm}). The standard deviation (SD_{ROR}) and minimum and maximum values of ROR were calculated. Additional summary statistics and adjusted and unadjusted odds ratios were also computed. The outcome measures and their calculations are presented in Table 5-2.

Those samples that fulfilled the balanced criteria for both covariates (-0.5% ≤ D_{X_1} < 0.5% and -0.05 ≤ D_{X_2} < 0.05) were selected to form the balanced covariates cohort, which was used to explore the impact of sample size on adjusted analysis when both covariates were balanced.

The direction and magnitude of the relationship between ROR and covariate imbalance were estimated using a regression model:

$$ROR = \phi_0 + \phi_1 D_{X_1} + \phi_2 D_{X_2} + \varepsilon \qquad \text{(Eq. 5-1)}$$

where ϕ_0 is the intercept, representing the ROR value when both covariates are balanced (D_{X_1} = 0 and D_{X_2} ≈ 0); ϕ_1 and ϕ_2 are the slopes for D_{X_1} and D_{X_2} that correspond to the change in ROR for every unit of imbalance in X_1 and X_2, respectively; and ε is the residual error term.

The percentage difference between the unadjusted and adjusted analyses due to noncollapsibility was calculated as $(\phi_0 - 1) \times 100\%$. The percentage difference due to imbalance was calculated as $(\phi_1 D_{X_1} + \phi_2 D_{X_2}) \times 100\%$. The total percentage difference was calculated as the sum of the two, $(\phi_0 - 1) \times 100\% + \phi_1 D_{X_1} + \phi_2 D_{X_2}) \times 100\%$.

Table 5-2 Summary of the outcome measures calculated from the samples

Outcome measures	Description	Equation	Notation
At the sample level			
ROR	Ratio of odds ratios	$\frac{OR_{T\_u}}{OR_{T\_a}}$	OR_{T_u} = unadjusted odds ratio of T (Treatment) OR_{T_a} = adjusted odds ratio of T
D_{X_1}	Difference in X_1 between groups	$P_{X_1\mid T=1} - P_{X_1\mid T=0}$	$P_{X_1\mid T=1}$= % of X1 = 1 in T $P_{X_1\mid T=0}$ = % of X_1 = 1 in C (Control)
D_{X_2}	Standardised difference in X_2 between groups	$\frac{\mu_{X_2\mid T=1}}{SD_{X_2\mid T=1}} - \frac{\mu_{X_2\mid T=0}}{SD_{X_2\mid T=0}}$	$\mu_{X_2\mid T=1}$ = mean of X_2 in T $SD_{X_2\mid T=1}$= SD of X_2 in T $\mu_{X_2\mid T=0}$ = mean of X_2 in C $SD_{X_2\mid T=0}$ = SD of X_2 in C
At the scenario level			
$\overline{ROR}$	Mean ROR	$\frac{\sum_{i=1}^{n_{trm}} ROR_i}{n_{trm}}$	n_{trm} = number of trimmed samples ROR_i = ROR for sample i
$SDROR$	Standard deviation of ROR	$\sqrt{\frac{\sum_{i=1}^{n_{trm}} (ROR_i - \overline{ROR})^2}{n_{trm} - 1}}$	
$MinROR$	Minimum of ROR		
$MaxROR$	Maximum of ROR		
$\overline{OR}_{T\_u}$	Mean unadjusted odds ratio of T	$e^{\frac{\sum_i^{ntrm} \theta_{u_i}}{n_{trm}}}$	θ_{u_i} = unadjusted log odds ratio for sample i
$\overline{OR}_{T\_a}$	Mean adjusted odds ratio of T	$e^{\frac{\sum_i^{ntrm} \theta_{a_i}}{n_{trm}}}$	θ_{a_i}= adjusted log odds ratio for sample i
$\bar{D}_{X_1}$	Mean D_{X_1}	$e^{\frac{\sum_i^{ntrm} D_{X_{1i}}}{n_{trm}}}$	$D_{X_{1i}}$ = difference in X_1 for sample i
$\bar{D}_{X_2}$	Mean D_{X_2}	$e^{\frac{\sum_i^{ntrm} D_{X_{2i}}}{n_{trm}}}$	$D_{X_{1i}}$ = difference in X_2 for sample i
ϕ_1	Slope for D_{X_1}	Eq. 5-1	
ϕ_2	Slope for D_{X_2}	Eq. 5-1	

5.3 Results

5.3.1 Extreme values and trimming

There were no convergence issues during resampling, although there were warnings about precision issues. Warnings occurred in extreme cases, especially when the sample size was small.

Table 5-3 shows the number of scenarios that yielded ROR values larger than 2, 5, and 10, by sample size. Extreme values occurred more often with small sample sizes than large sample sizes. In samples of size 100, 250 of the 360 scenarios (69%) included at least one sample with ROR > 2. The extreme scenarios, such as those with extreme event rates or strongly prognostic covariates, were more likely to produce an unadjusted odds ratio more than twice as large as the adjusted odds ratio (ROR > 2). For example, when samples of size 100 were selected, 32% of the 360 scenarios had extreme event rates (R = 10% or 90%), 46% contained a strongly prognostic binary covariate ($OR_{X_1\_a}$ = 0.2 or 5), and 62% had a strongly prognostic continuous outcome ($OR_{X_2\_a}$ = 2). Twenty-nine percent and 19% of the scenarios with a sample size of 100 had ROR > 5 and 10, respectively. The number of large ROR values decreased as the sample size increased. No extreme ROR values were observed in scenarios with a sample size of 1,000 or more.

Although extreme values of ROR occurred in many scenarios with small sample sizes, very few of the samples in these scenarios had extreme ROR values. For example, a median of 13 of the 20,000 samples of size 100 from each scenario had ROR > 2 across the 250 scenarios out of the total scenarios (Table 5-3). In the most extreme scenario, at most 1,311 (6.6%) of the samples had ROR > 2. The number of samples with extreme ROR values also decreased as the sample size increased.

Table 5-3 Mean number (%) of scenarios and median (range) number of samples in these scenarios with large and extreme ROR values

Sample size	ROR > 2	ROR > 5	ROR > 10
100			
No. of scenarios, n (%)*	250 (69.4%)	106 (29.4%)	69 (19.2%)
Median no. of samples (range), n†	13 (1 to 1311)	3 (1 to 236)	3 (1 to 122)
200			
No. of scenarios, n (%)*	75 (20.8%)	11 (3.1%)	1 (0.3%)
Median no. of samples (range), n†	5 (1 to 230)	2 (1 to 4)	1 (1 to 1)
500			
No. of scenarios, n (%)*	7 (1.9%)	0	0
Median no. of samples (range), n†	2 (1 to 5)	-	-
1,000			
No. of scenarios, n (%)*	0	0	0
5,000			
No. of scenarios, n (%)*	0	0	0

* Total number of scenarios = 360

† Total number of samples = 20,000

Although few extreme ROR values were observed, these extreme results produced undefined or extreme ROR values that affected the mean ROR. As

shown in Table 5-4, the mean ROR for all of the samples of size 100 from all of the scenarios after trimming was much larger than the mean ROR when all of the samples of size 100 were considered before trimming were considered. Trimming was conducted by excluding the observations with ROR values in the largest and smallest 2.5% of the distribution. The mean after trimming was more robust and less affected by extreme ROR values than the untrimmed mean. The mean was largely unaffected by trimming once the sample size increased to 200.

Table 5-4 Empirical mean (minimum and maximum) ratio of odds ratios over all of the scenarios by sample size

Sample size	Before trimming	After trimming
100	7.2e+256 (0.4785 to 2.6e+259)	0.8381 (0.4704 to 0.9891)
200	0.8440 (0.4742 to 0.9920)	0.8426 (0.4710 to 0.9917)
500	0.8456 (0.4744 to 0.9944)	0.8454 (0.4733 to 0.9946)
1,000	0.8464 (0.4738 to 0.9958)	0.8463 (0.8929 to 0.9959)
5,000	0.8472 (0.4738 to 0.9970)	0.8471 (0.4737 to 0.9971)

5.3.2 Summary of covariate imbalance

The degree of imbalance varied with sample size, but was consistent across the scenarios (Appendix G). Table 5-5 summarises the global mean difference between the two treatment groups, standard deviation, and minimum and maximum imbalance for both covariates by sample size. As expected, on average both covariates were balanced across both treatment groups ($\overline{D}_{X_1}$= 0 and $\overline{D}_{X_2}$= 0). However, there was considerably more variation in the magnitude of the imbalance in scenarios with small sample sizes than larger sample sizes.

The average maximum imbalance difference between the treatment and control groups was 38 percentage points for the binary covariate. The maximum standardised difference for the continuous covariate was 0.82 units. At a sample size of 5,000, the mean range of the imbalance decreased to $\pm$ 5% for the binary covariate and was negligible for the continuous covariate.

Table 5-5 Global mean, standard deviation, and minimum and maximum differences in the binary and continuous covariates for samples from all of the scenarios, by sample size

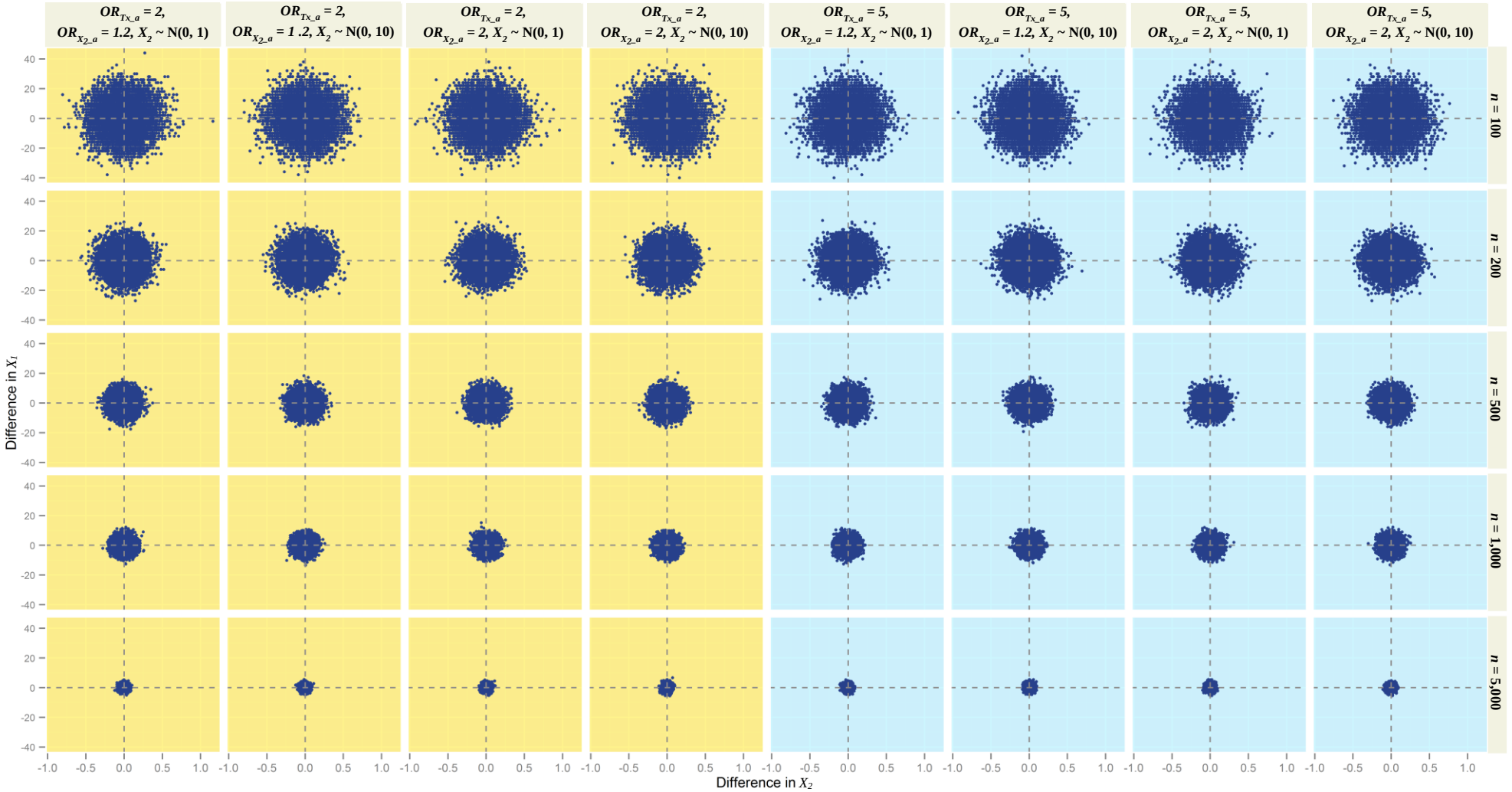
	Sample size				
	N = 100	**N = 200**	**N = 500**	**N = 1,000**	**N = 5,000**
$\bar{D}_{X_1}$					
*n**	360	360	360	360	360
Mean	-0.0003	-0.0023	0.0019	-0.0009	-0.001
SD	9.780	6.894	4.354	3.077	1.375
Minimum†	-38.400	-27.100	-17.068	-12.111	-4.934
Maximum‡	38.450	27.083	17.068	12.009	4.931
$\bar{D}_{X_2}$					
*n**	360	360	360	360	360
Mean	-0.0026	-0.0014	-0.0006	-0.0003	0
SD	0.197	0.137	0.086	0.060	0.027
Minimum†	-0.825	-0.539	-0.345	-0.238	-0.096
Maximum‡	0.791	0.533	0.324	0.225	0.093

* Total number of scenarios
† Negative values show that the mean in the control group is greater than the mean in the treatment group
‡ Positive values show that the mean in the treatment group is greater than the mean in the control group

The relationship between the variation in the difference and sample size is depicted in Figure 5-1, which shows a set of scatter plots of D_{X_2} against D_{X_1} for

samples from scenarios with $OR_{X_1\_a}$ = 1.5, R = 50%, and P_{X_1}= 50%. The variation in the imbalance declined with sample size in all of the scenarios. Both covariates had a symmetrical imbalance distribution, suggesting that imbalance had an equal chance of favouring either group.

Figure 5-1 Scatter plots of the difference in X_1 against the difference in X_2 measured at each sample size in scenarios with $OR_{X_1\_a}$ = 1.5, P_{X_1} = 50%, and R = 50%



5.3.3 Effect of sample size on balanced covariates

Table 5-6 shows that there was a low likelihood of balancing all of the covariates when the distribution of the covariates between the groups was not controlled during randomisation. Only 1.6% of the samples of size 100 and 200 from all of the scenarios had both covariates within the balance thresholds across the treatment groups. This percentage increased with sample size. At a sample size of 5,000, a quarter of the samples contained balanced covariates.

Table 5-6 Summary of the number (%†) of samples over the 360 scenarios with balanced covariates

	Sample size				
	N = 100	**N = 200**	**N = 500**	**N = 1,000**	**N = 5,000**
n_{trm}†	19,000	19,000	19,000	19,000	4,750
Mean	307 (1.6%)	305 (1.6%)	893 (4.7%)	1,409 (7.4%)	1,235 (26%)
SD	17 (0.1%)	17 (0.1%)	31 (0.2%)	42 (0.2%)	43 (0.9%)
Minimum	260 (1.4%)	262 (1.4%)	810 (4.3%)	1,268 (6.7%)	1,089 (23%)
Maximum	368 (1.9%)	354 (1.9%)	1,002 (5.3%)	1,530 (8.1%)	1,341 (28%)

† Number of samples after trimming

All of the samples with balanced covariates were used to form the balanced covariates cohort. The average unadjusted and adjusted odds ratios and ROR were calculated for this cohort. Table 5-7 presents these results for scenarios in which both covariates moderately affected the outcome.

As trimming was based on the ROR distribution, the unadjusted and adjusted mean odds ratios were likely to be affected by extreme values. This was particularly apparent in scenarios with a low or high event rate and a small sample size. For example, when OR_{T_a} = 5 and sample size = 100, scenarios with event rates of 10% and 90% gave a mean adjusted odds ratio of approximately 50 and 35, respectively (highlighted in yellow in Table 5-7). As the sample size increases, the influence of the extreme values decreased dramatically. The median values better represented the 'average' odds ratios of the samples. The results indicate that both the median unadjusted and adjusted odds ratios were biased when the sample size was small. A larger magnitude of bias was observed for an adjusted odds ratio of 5.

The mean ROR was less affected by sample size, although there was a slight downwards bias in ROR as the sample size decreased. For example, when OR_{T_a} = 5 and R = 10%, the mean ROR for sample sizes 100, 200, 500, 1,000, and 5,000 were 0.9621, 0.9750, 0.9852, 0.9880, and 0.9919, respectively (Table 5-7). The corresponding ROR standard deviation was generally small and decreased as the sample size increased.

Figure 5-2 shows the mean ROR curves. The curves for each sample size almost overlap with the total population ROR curve (black) for all of the scenarios.

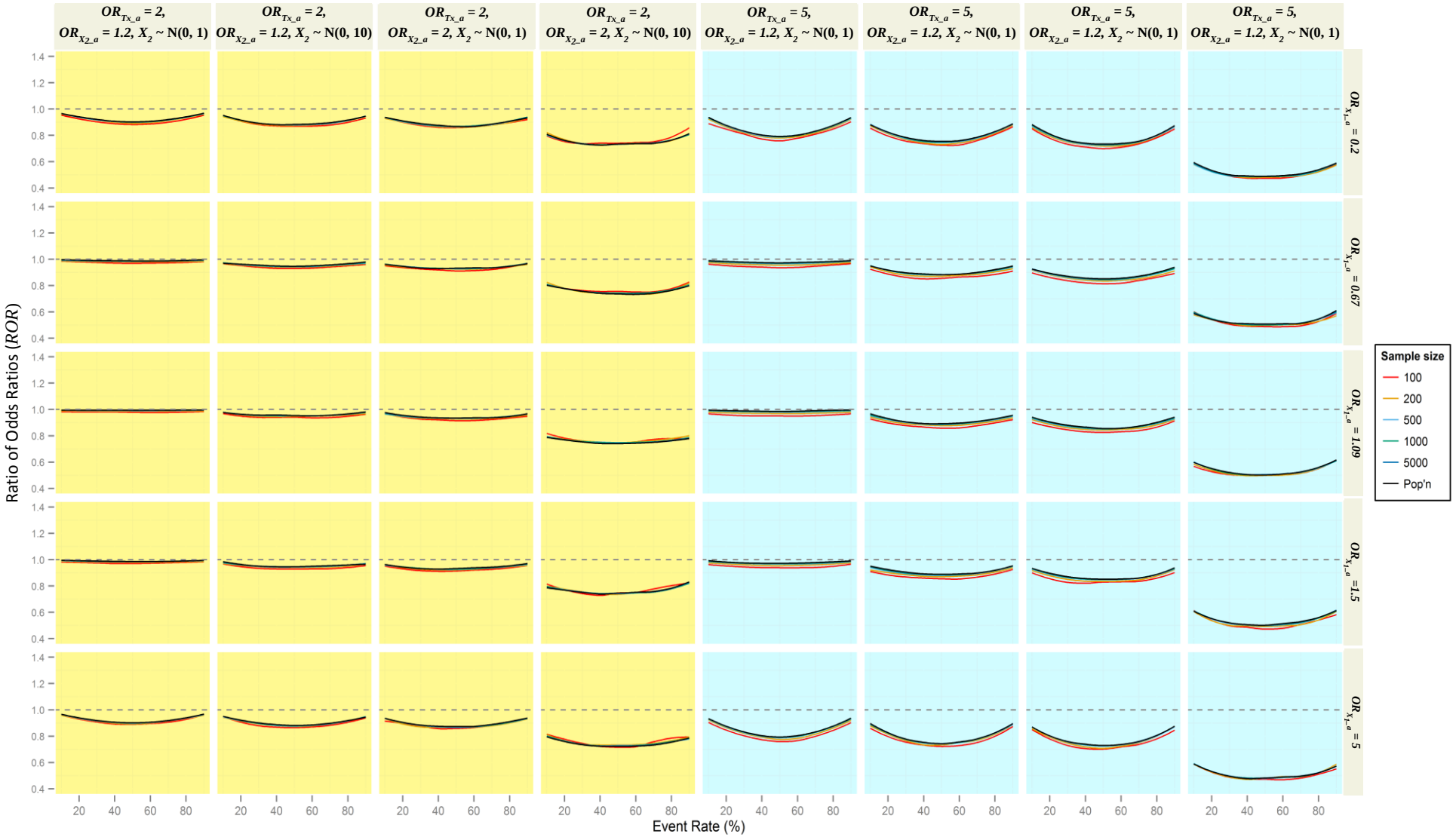
Therefore, when both covariates were balanced across the treatment group, the sample size did not greatly affect ROR.

Table 5-7 Mean {median} odds ratios for P_{X_1}= 50%, $OR_{X_1\_a}$ = 1.5, $OR_{X_2\_a}$ = 1.2, and σ^2 = 1 when both covariates are balanced

	OR_{T_a} = 2					OR_{T_a} = 5				
Sample size	100	200	500	1,000	5,000	100	200	500	1,000	5,000
R = 10%										
n (%)*	317 (1.7%)	311 (1.6%)	900 (4.7%)	1,448 (7.6%)	1,291 (27.2%)	298 (1.6%)	329 (1.7%)	866 (4.6%)	1,412 (7.4%)	1,239 (26.1%)
Mean (OR_{T_u})	3.19 {1.87}	2.02 {1.94}	1.99 {2.00}	2.04 {2.05}	1.99 {1.99}	47.3 {5.44}	6.59 {4.83}	5.06 {4.85}	5.04 {4.99}	4.99 {4.99}
Mean (OR_{T_a})	3.27 {1.98}	2.04 {1.99}	2.00 {2.01}	2.05 {2.05}	2.00 {2.00}	49.3 {5.67}	6.76 {5.04}	5.13 {4.92}	5.10 (5.06)	5.03 {5.03}
$\overline{ROR}$	0.9826	0.9890	0.9931	0.9944	0.9955	0.9621	0.9750	0.9852	0.9880	0.9919
SD_{ROR}	0.0531	0.0252	0.0113	0.0081	0.0049	0.0591	0.0339	0.0157	0.0103	0.0053
R = 30%										
n (%)	315 (1.7%)	288 (1.5%)	922 (4.9%)	1,461 (7.7%)	1,253 (26.4%)	322 (1.7%)	270 (1.4%)	876 (4.6%)	1,432 (7.5%)	1296 (27.3%)
Mean (OR_{T_u})	2.05 {2.02}	2.03 {2.03}	2.00 {1.99}	2.00 {2.00}	1.99 {1.99}	5.04 {4.89}	4.92 {4.85}	4.98 {4.94}	4.94 {4.91}	4.89 {4.90}
Mean (OR_{T_a})	2.10 {2.08}	2.07 {2.07}	2.03 {2.03}	2.03 {2.03}	2.02 {2.01}	5.34 {5.30}	5.13 {5.04}	5.13 {5.09}	5.08 {5.06}	5.01 {5.02}
$\overline{ROR}$	0.9761	0.9816	0.9865	0.9874	0.9883	0.9459	0.9604	0.9708	0.9742	0.9765
SD_{ROR}	0.0332	0.0204	0.0108	0.0088	0.0054	0.0553	0.0340	0.0193	0.0138	0.0067
R = 50%										
n (%)	296(1.6%)	291 (1.5%)	898 (4.7%)	1,437 (7.6%)	1,265 (26.6%)	304 (1.6%)	306 (1.6%)	930 (4.9%)	1,416 (7.5%)	1,231 (25.9%)
Mean (OR_{T_u})	2.02 {2.07}	2.02 {2.00}	1.99 {2.00}	1.99 {1.97}	1.96 {1.96}	5.09 {5.06}	5.10 {5.01}	4.93 {4.95}	4.88 {4.91}	4.90 {4.90}
Mean (OR_{T_a})	2.09 {2.12}	2.07 {2.07}	2.03 {2.02}	2.00 {2.00}	1.99 {1.99}	5.42 {5.36}	5.34 {5.32}	5.10 {5.11}	5.03 {5.05}	5.04 {5.04}
$\overline{ROR}$	0.9690	0.9785	0.9842	0.9856	0.9865	0.9405	0.9570	0.9670	0.9703	0.9723
SD_{ROR}	0.0317	0.0216	0.0114	0.0090	0.0051	0.0588	0.0342	0.0199	0.0148	0.0072
R = 70%										
n (%)	319 (1.7%)	334 (1.8%)	940 (4.9%)	1,377 (7.2%)	1,291 (27.2%)	326 (1.7%)	311 (1.6%)	910 (4.8%)	1,461 (7.7%)	1,294 (27.2%)
Mean (OR_{T_u})	1.97 {1.95}	2.06 {2.03}	2.00 {1.99}	2.00 {1.99}	1.99 {199}	5.51 {5.23}	5.09 {4.89}	4.91 {4.90}	4.85 {4.84}	4.86 {4.85}
Mean (OR_{T_a})	2.02 {2.03}	2.10 {2.05}	2.02 {2.02}	2.02 {2.02}	2.01 {2.01}	5.87 {5.41}	5.29 {5.11}	5.06 {5.03}	4.98 {4.98}	4.97 {4.95}
$\overline{ROR}$	0.9742	0.9815	0.9870	0.9880	0.9884	0.9398	0.9636	0.9705	0.9741	0.9773
SD_{ROR}	0.0366	0.0197	0.0107	0.0081	0.0050	0.0555	0.0330	0.0195	0.0135	0.0068
R = 90%										
n (%)	323 (1.7%)	325 (1.7%)	904 (4.8%)	1,478 (7.8%)	1,341 (28.2%)	314 (1.7%)	307 (1.6%)	925 (4.9%)	1,468 (7.7%)	1,274 (26.8%)
Mean (OR_{T_u})	3.06 {1.74}	2.20 {2.11}	2.01 {1.98}	2.05 {2.03}	2.00 {2.00}	34.2 {4.57}	9.01 {4.92}	5.11 {4.95}	5.03 {4.96}	4.95 {4.94}
Mean (OR_{T_a})	3.12 {1.75}	2.23 {2.12}	2.02 {2.00}	2.06 {2.05}	2.01 {2.02}	35.4 {4.94}	9.23 {5.11}	5.18 {5.02}	5.10 {5.03}	5.00 {4.00}
$\overline{ROR}$	0.9843	0.9869	0.9929	0.9941	0.9953	0.9664	0.9772	0.9848	0.9879	0.9902
SD_{ROR}	0.0507	0.0251	0.0105	0.0075	0.0048	0.0628	0.0287	0.0153	0.0104	0.0056

* Number (%) of samples with balanced covariates. Yellow indicates a very high odds ratio.

Figure 5-2 Mean ratio of the unadjusted and adjusted odds ratios as the overall event rate changes, calculated when both covariates are balanced, P_{X_1} = 50%, and with varying treatment effects and sample sizes



5.3.4 Impact of sample size with an imbalance of covariates

The analysis in Table 5-7 was repeated for all of the covariates, balanced and unbalanced. The results, reported in Table 5-8, were similar to those for the balanced covariate cohort. However, the standard deviations differed. For example, the ROR standard deviation for the scenario OR_{T_a} = 2, sample size = 100, and R = 10% was 0.053 for the complete sample cohort (Table 5-8) and 0.102 for the balanced cohort (Table 5-7). Figure 5-3 presents a series of dot plots that show the spread of the ROR distribution as the sample size and other scenario parameters varied, with an event rate of 50%. It shows a clear decrease in variability as the sample size increased. ROR varied the least in scenarios where both covariates had a moderate effect on the outcome ($OR_{X_1\_a}$ = 1.5, $OR_{X_2\_a}$ = 1.2, and σ^2 = 1). As the prognostic strength of these covariates increased, the ROR variability also increased.

Figure 5-4 shows the ROR curve for each scenario, grouped by sample size, when including samples both with and without covariate balance. When OR_{T_a} = 2, X_2 = 2, and σ^2 = 10, it shows a slight increase in ROR when sample size = 100, compared with other sample size values, especially at extreme event rate values. This was likely due to an increase in the variability of ROR with an imbalance of covariates, which drove the mean ROR away from the population value (Figure 5-3).

Overall, the results in Figure 5-4 are similar to those in Figure 5-2 for the balanced cohort, suggesting that sample size had no impact on ROR, regardless of whether the baseline covariates were balanced or not.

Table 5-8 Mean {median} odds ratios for P_{X_1}= 50%, $OR_{X_{1\_a}}$ = 1.5, $OR_{X_{2\_a}}$ = 1.2, and σ^2 = 1 when both balance and imbalance of covariates are included*

	OR_{T_a} = 2					OR_{T_a} = 5				
Sample size	100	200	500	1,000	5,000	100	200	500	1,000	5,000
R = 10%										
Mean (OR_{T_u})	3.11 {2.09}	2.11 {2.03}	2.01 {1.99}	2.00 {2.00}	1.99 {1.99}	44.21 {5.27}	7.58 {5.26}	5.21 {5.03}	5.07 {5.01}	4.99 {4.98}
Mean (OR_{T_a})	3.17 {2.09}	2.13 {2.04}	2.03 {2.01}	2.01 {2.01}	2.00 {2.00}	46.10 {5.57}	7.75 {5.29}	5.27 {5.11}	5.12 {5.06}	5.03 {5.01}
$\overline{ROR}$	0.9865	0.9902	0.9935	0.9945	0.9956	0.9655	0.9801	0.9878	0.9904	0.9928
SD_{ROR}	0.1019	0.0527	0.0259	0.0167	0.0068	0.1086	0.0537	0.0269	0.0176	0.0072
R = 30%										
Mean (OR_{T_u})	2.03 {2.02}	2.01 {2.00}	2.00 {1.99}	2.00 {1.99}	2.00 {2.00}	5.22 {4.94}	5.03 {4.99}	4.95 {4.91}	4.92 {4.91}	4.90 {4.90}
Mean (OR_{T_a})	2.09 {2.08}	2.05 {2.04}	2.03 {2.02}	2.02 {2.02}	2.02 {2.02}	5.53 {5.33}	5.23 {5.15}	5.09 {5.06}	5.05 {5.04}	5.02 {5.02}
$\overline{ROR}$	0.9767	0.9822	0.9855	0.9872	0.9877	0.9470	0.9623	0.9712	0.9741	0.9762
SD_{ROR}	0.0728	0.0431	0.0241	0.0166	0.0072	0.0810	0.0489	0.0273	0.0188	0.0081
R = 50%										
Mean(OR_{T_u})	1.99 {1.94}	1.99 {1.99}	1.97 {1.97}	1.97 {1.97}	1.97 {1.97}	5.07 {4.99}	5.01 {4.97}	4.93 {4.95}	4.91 {4.91}	4.90 {4.90}
Mean(OR_{T_a})	2.05 {2.05}	2.03 {2.02}	2.00 {2.00}	2.00 {2.00}	2.00 {2.00}	5.41 {5.33}	5.24 {5.20}	5.11 {5.10}	5.07 {5.06}	5.04 {5.04}
$\overline{ROR}$	0.9741	0.9809	0.9843	0.9857	0.9864	0.9403	0.9568	0.9664	0.9698	0.9717
SD_{ROR}	0.0703	0.0419	0.0239	0.0164	0.0073	0.0799	0.0491	0.0279	0.0193	0.0085
R = 70%										
Mean(OR_{T_u})	2.02 {1.99}	2.01 {1.99}	2.00 {1.99}	1.99 {1.99}	1.98 {1.99}	5.21 {4.94}	4.97 {4.88}	4.90 {4.89}	4.88 {4.87}	4.86 {4.86}
Mean(OR_{T_a})	2.08 {2.06}	2.05 {2.03}	2.03 {2.03}	2.02 {2.01}	2.01 {2.01}	5.52 {5.30}	5.16 {5.10}	5.04 {5.02}	5.00 {4.99}	4.97 {4.97}
$\overline{ROR}$	0.9771	0.9823	0.9859	0.9871	0.9879	0.9464	0.9639	0.9729	0.9760	0.9780
SD_{ROR}	0.0717	0.0420	0.0236	0.0161	0.0070	0.0824	0.0484	0.0275	0.0186	0.0081
R = 90%										
Mean(OR_{T_u})	2.02 {1.99}	2.01 {1.99}	2.00 {1.99}	1.99 {1.99}	1.98 {1.99}	5.21 {4.94}	4.97 {4.88}	4.90 {4.89}	4.88 {4.87}	4.86 {4.86}
Mean(OR_{T_a})	2.08 {2.06}	2.05 {2.03}	2.03 {2.03}	2.02 {2.01}	2.01 {2.01}	5.52 {5.30}	5.16 {5.10}	5.04 {5.02}	5.00 {4.99}	4.97 {4.97}
$\overline{ROR}$	0.9851	0.9892	0.9930	0.9945	0.9953	0.9624	0.9776	0.9854	0.9880	0.9903
SD_{ROR}	0.1003	0.0514	0.0252	0.0165	0.0066	0.1103	0.0539	0.0274	0.0174	0.0072

* Total trimmed samples = 19,000 for sample size ≤ 1,000 and 4,750 for sample size = 5,000. Cell highlighted in yellow indicate very high odds ratios.

Figure 5-3 Dot plots of the ratios of odds ratios calculated using all of the samples from selected scenarios, grouped by sample size

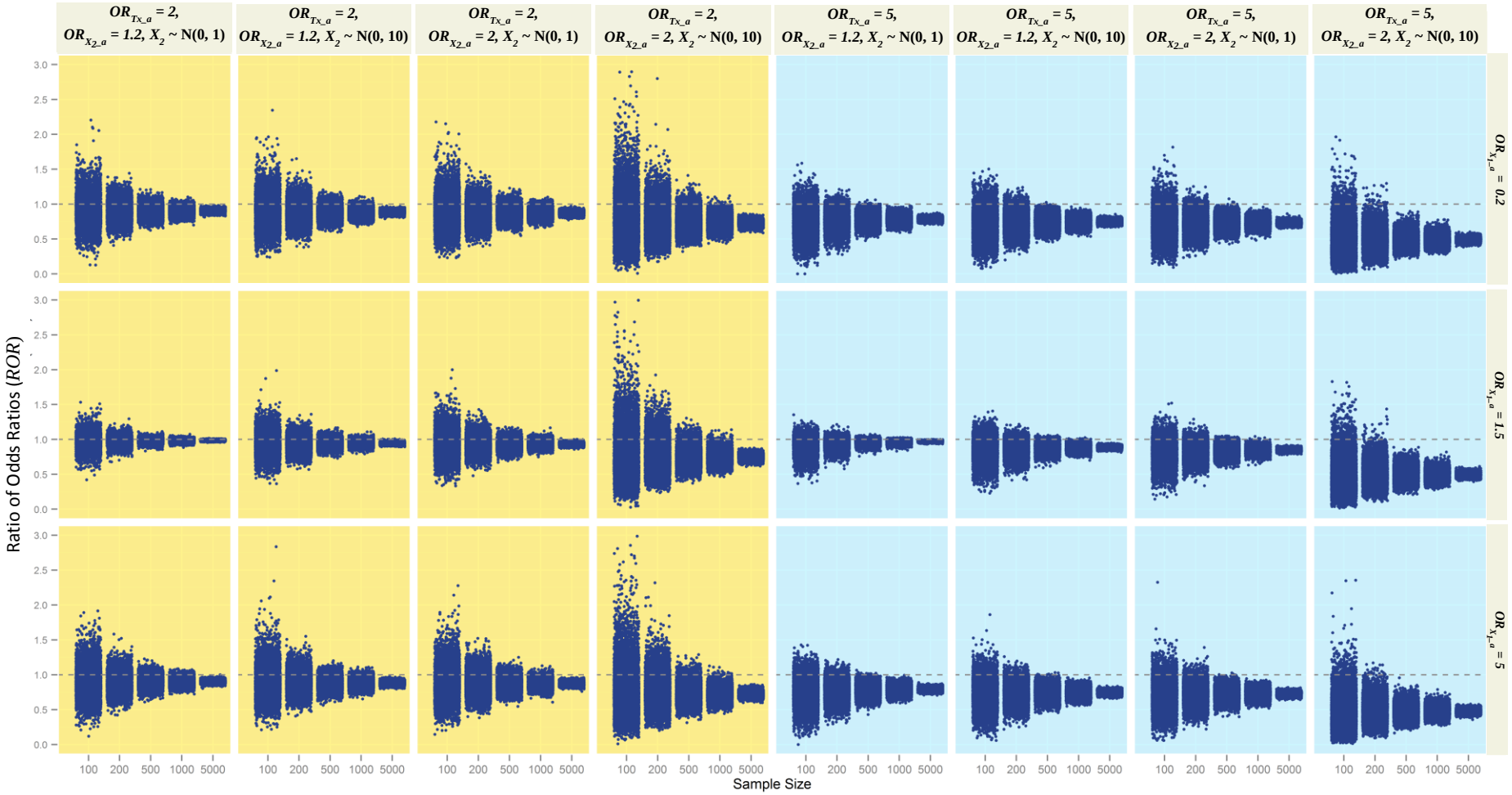
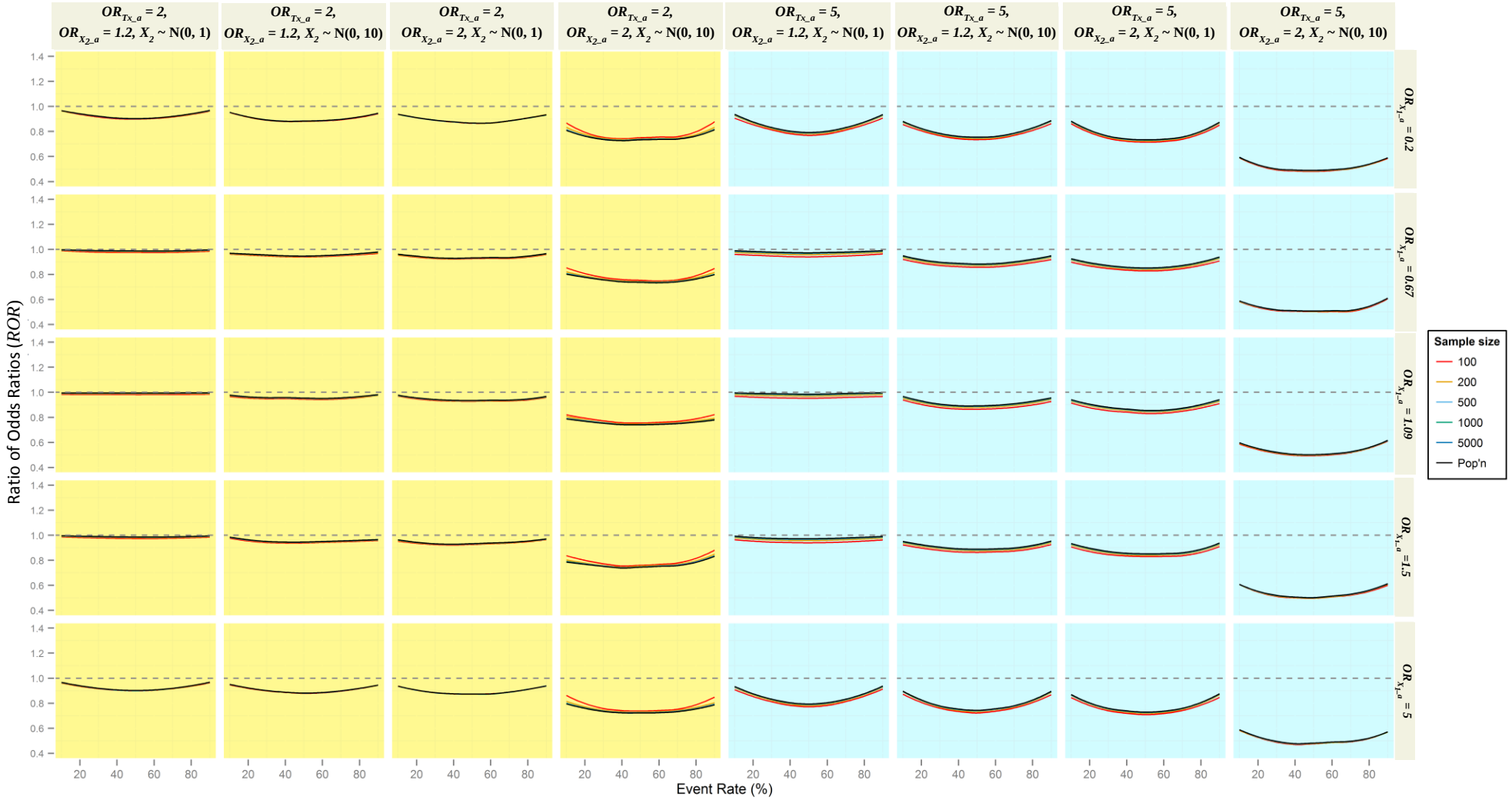


Figure 5-4 **Mean ratio of the unadjusted and adjusted odds ratios as the overall event rate changes, calculated using samples with both balanced and imbalanced covariates from trial scenarios with P_{X_1} = 50% and varying treatment effect and sample size**



5.3.5 Impact of an imbalance of covariates

5.3.5.1 Both covariates have moderate prognostic effects

Figure 5-5 shows scatter plots of ROR obtained from 20,000 samples of size 100 against the level of imbalance in the (a) binary and (b) continuous covariates in a scenario with a strong treatment effect and covariates with moderate prognostic effects on the outcome. The vertical dotted line in each plot shows where the covariates were balanced across the groups. The horizontal dotted line represents no difference between the unadjusted and adjusted odds ratios. In Figure 5-5(a), there is a positive linear relationship between ROR and the difference in X_1 (D_{X_1}). When both covariates were moderately positively associated with the outcome, ROR increased as the imbalance moved from a negative to a positive value. Figure 5-5(b) shows the same relationship between ROR and the difference in X_2 (D_{X_2}).

To illustrate the relationship in further detail, summaries of ROR categorised by the level of imbalance in X_1 and X_2 from one scenario are presented in Table 5-9. The results show that the mean ROR varied with the degree of imbalance in both covariates. The mean ROR ranged from 0.9086 to 1.0502, representing the extreme ends of negative and positive imbalance in both covariates, respectively. When both covariates were balanced, the mean ROR was 0.9690 (highlighted in yellow in Table 5-9). The mean ROR dropped to 0.9086 (highlighted in light orange) when the control group had greater imbalance between the covariates

than the treatment group. However, when the treatment group had greater imbalance, the mean ROR increased to 1.0502 (highlighted in orange).

The results can be interpreted in the context of odds ratios. If the adjusted odds ratio was kept constant at 2, the unadjusted odds ratio of the treatment was 1.938 when both covariates were balanced and was biased downwards to 1.82 and upwards to 2.1 for large negative and large positive imbalances, respectively. The mean ROR tended to increase as the difference in either covariate moved from a large negative imbalance to a large positive imbalance.

Figure 5-5 Scatter plots of the ratio of the unadjusted and adjusted odds ratios against the difference in (a) X_1 and (b) X_2 between the treatment and control group when OR_{T_a} = 2, $OR_{X_1\_a}$ = 1.5, $OR_{X_2\_a}$ = 1.2, σ^2 = 1, R = 50%, P_{X_1} = 50%, and sample size = 100

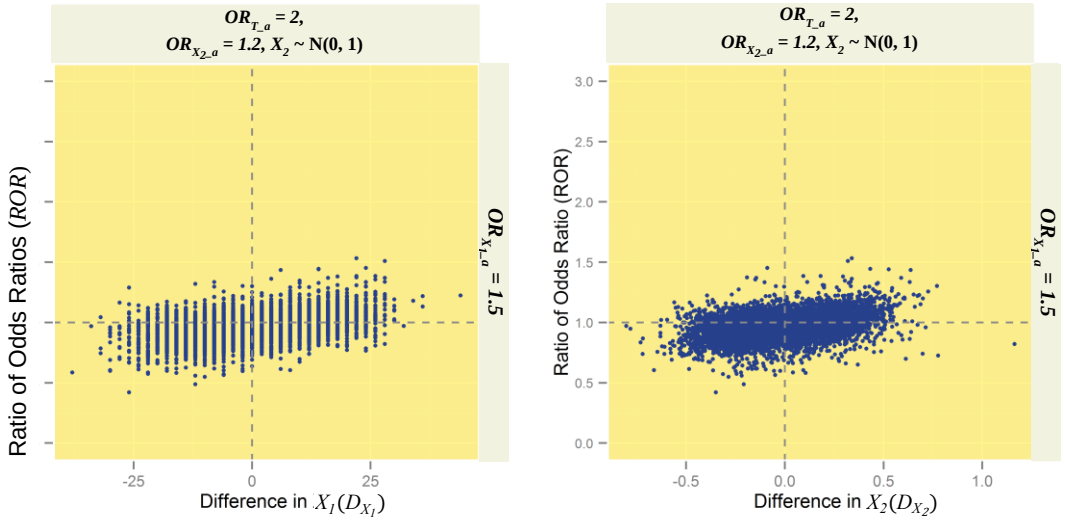


Table 5-9 Ratio of the unadjusted and adjusted odds ratios as the overall event rate changes, when OR_{T_a} = 2, $OR_{X_1\_a}$ = 1.5, $OR_{X_2\_a}$ = 1.2, σ^2 = 1, R = 50%, P_{X_1} = 50%, and sample size = 100

	Imbalance in X_1					
	D_{X_1} < -5	-5 ≤ D_{X_1} < -0.5	-0.05 ≤ D_{X_1} < -0.05	0.5 ≤ D_{X_1} < 5	D_{X_1} ≥ 5	Total
Imbalance in X_2						
D_{X_2} < -0.2						
n (%)	849 (4.5%)	465 (2.5%)	270 (1.4%)	488 (2.6%)	991 (5.2%)	3,063 (16.1%)
$\overline{ROR}$	0.9086	0.9260	0.9252	0.9420	0.9692	0.9376
$SDROR$	0.074	0.066	0.067	0.068	0.080	0.0773
-0.2 ≤ D_{X_2} < -0.05						
n (%)	1479 (7.8%)	751 (3.9%)	370 (1.9%)	751 (3.9%)	1,512 (8.0%)	4,863 (25.6%)
$\overline{ROR}$	0.9206	0.9399	0.9527	0.9639	0.9977	0.9567
$SDROR$	0.060	0.050	0.049	0.045	0.062	0.0643
-0.05 ≤ D_{X_2} < 0.05						
n (%)	1153 (6.1%)	563 (3.0%)	296 (1.6%)	588 (3.1%)	1,107 (5.8%)	3,707 (19.5%)
$\overline{ROR}$	0.9337	0.9591	0.9690	0.9828	1.0119	0.9715
$SDROR$	0.060	0.039	0.032	0.031	0.049	0.0572
0.05 ≤ D_{X_2} < 2						
n (%)	1405 (7.4%)	709 (3.7%)	376 (2.0%)	677 (3.6%)	1,368 (7.2%)	4,535 (23.8%)
$\overline{ROR}$	0.9555	0.9784	0.9952	1.0034	1.0350	0.9935
$SDROR$	0.065	0.044	0.035	0.035	0.045	0.0593
D_{X_2} ≥ 0.2						
n (%)	947 (5.0%)	440 (2.3%)	214 (1.1%)	422 (2.2%)	809 (4.3%)	2,832 (14.9%)
$\overline{ROR}$	0.9822	1.0092	1.0134	1.0346	1.0502	1.0160
$SDROR$	0.080	0.064	0.061	0.058	0.059	0.0728
Total						
n (%)	5,833 (30.7%)	2,928 (15.4%)	1,526 (8.0%)	2,926 (15.4%)	5,788 (30.5%)	1,9000
$\overline{ROR}$	0.9398	0.9611	0.9700	0.9834	1.0117	0.9746
$SDROR$	0.0711	0.0587	0.0571	0.0556	0.0650	0.0702

n = number of samples
$\overline{ROR}$ = mean ratio of the odds ratios
$SDROR$ = standard deviation of the ratio of odds ratios
D_{X_1} = difference in X_1
D_{X_2} = difference in X_2

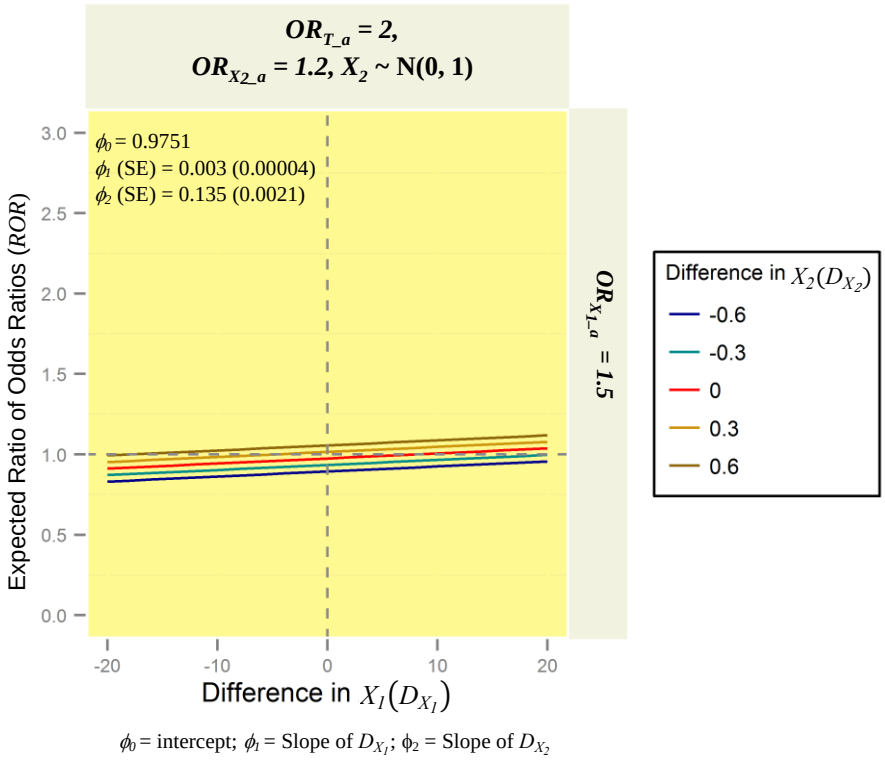
Figure 5-6 depicts the relationship between ROR and the degree of imbalance in both covariates by plotting the expected value of ROR, obtained from the regression model, against D_{X_1} at a given level of D_{X_2}, reflecting a large (D_{X_2} = -0.6 and 0.6), moderate (D_{X_2}= -0.3 and 0.3), or no imbalance (D_{X_2} = 0) in either direction. ROR when both covariates were balanced, ϕ_0, is 0.975, represented by the intersection between the red and dotted lines. ϕ_1 is the slope of D_{X_1}. A steeper regression line thus corresponds to a greater change in ROR for each percentage unit increase in D_{X_1}. ϕ_2 is the slope of D_{X_2} and shows the change in ROR for every unit change in D_{X_2}. Each colour regression line in Figure 5-6 reflects a different D_{X_2} value. A higher value of ϕ_2 gave a wider distance between the regression lines. When a regression model was fitted to the data, ROR increased by 0.003 for every percent increase in the difference in D_{X_1}, and 0.135 for every unit change in the difference in D_{X_2}.

To illustrate further how ROR is affected by unbalanced covariates, Table 5-10 reports the expected ROR, expected unadjusted odds ratio of treatment, and decomposition of the percentage difference between the unadjusted and adjusted results for the same scenario as Figure 5-6, according to different levels of imbalance in X_1 and X_2. There was some variation in ROR, which ranged from 0.9046 to 1.045. The unadjusted odds ratio was at most 9.45% less or 4.56% greater than the adjusted odds ratio (the unadjusted odds ratio ranged from

1.8092 to 2.0912), depending on the level of imbalance in each covariate. For example, when both covariates were balanced (D_{X_1} = 0 and D_{X_2} = 0), the unadjusted odds ratio was 2.49% lower than adjusted odds ratio (1.95 vs 2) due to noncollapsibility. When the mean of X_2 in the treatment group was 0.3 units less than in the control group (D_{X_2}= -0.3), then the unadjusted odds ratio was reduced further by 4.05% due to the imbalance between the two treatment groups (OR_{T_u} = 1.87) (second row of Table 5-10). In contrast, when the mean of X_1 in the treatment group was 0.3 units greater than in the control group (D_{X_2}= 0.3), the unadjusted odds ratio was biased upwards by 4.05%, favouring the treatment effect. The unadjusted odds ratio was thus slightly larger than the adjusted odds ratio even after taking into account the negative effect of noncollapsibility (OR_{T_u} = 2.03). ROR and the corresponding percentage differences continued to vary according to the combination of the imbalance in X_1 and X_2.

Table 5-11 reports the relationship between ROR and the level of imbalance by sample size. Little difference was found in most of the regression parameters. The slope of D_{X_2} (ϕ_2) was greater when the sample size was 5,000 than when a small sample size was used, but the effect of this difference was small when the unit of measure was considered.

Figure 5-6 **Regression line of the expected ratio of the unadjusted and adjusted odds ratios against the difference in X_1 at selected differences in X_2 when OR_{T_a} = 2, $OR_{X_1\_a}$ = 1.5, R = 50%, P_{X_1} = 50%, $OR_{X_2\_a}$ = 1.2, and σ^2 = 1**



ϕ_0 = intercept; ϕ_1 = Slope of D_{X_1}; ϕ_2 = Slope of D_{X_2}

Table 5-10 Expected ratio of the odds ratios, unadjusted odds ratios, and percentage differences due to noncollapsibility and imbalance when OR_{T_a} = 2, $OR_{X_1\_a}$ = 1.5, R = 50%, P_{X_1}= 50%, $OR_{X_2\_a}$ = 1.2, σ^2 = 1, and sample size = 100

D_{X_1}*	D_{X_2}*	ROR	OR_{T_u}	% difference due to noncollapsibility	% difference due to imbalance	Total % difference
0	**0**	0.9751	1.9502	-2.49%	0%	-2.49%
0	**-0.3**	0.9346	1.8692	-2.49%	-4.05%	-6.54%
0	**0.3**	1.0156	2.0312	-2.49%	4.05%	1.56%
-10	**0**	0.9451	1.8902	-2.49%	-3.00%	-5.49%
-10	**-0.3**	0.9046	1.8092	-2.49%	-7.05%	-9.54%
-10	**0.3**	0.9856	1.9712	-2.49%	1.05%	-1.44%
10	**0**	1.0051	2.0102	-2.49%	3.00%	0.51%
10	**-0.3**	0.9646	1.9292	-2.49%	-1.05%	-3.54%
10	**0.3**	1.0456	2.0912	-2.49%	7.05%	4.56%

* D_{X_1} and D_{X_2}= treatment – control

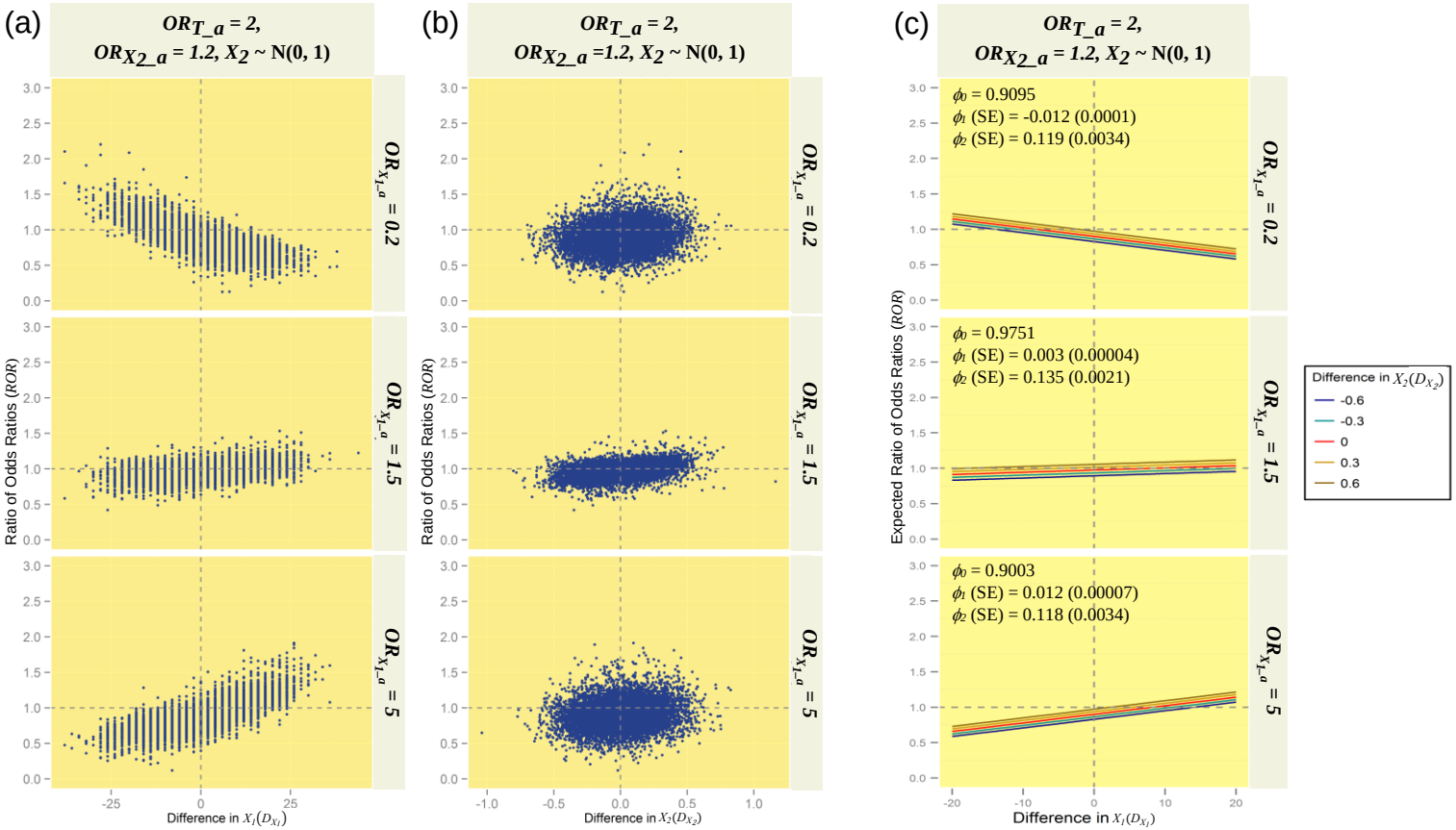
Table 5-11 Estimated intercept (ϕ_0), slope of the imbalance in X_1 (D_{X_1}), and slope of the imbalance in X_2 (D_{X_2}) by sample size when OR_{T_a} = 2, $OR_{X_1\_a}$ = 1.5, R = 50%, P_{X_1}= 50%, $OR_{X_2\_a}$ = 1.2, and σ^2 = 1

Sample size	ϕ_0	$\phi_1 (SE)$	ϕ_2 (SE)
100	0.9751	0.0031 (0.00004)	0.1350 (0.0021)
200	0.9818	0.0034 (0.00003)	0.1530 (0.0016)
500	0.9853	0.0036 (0.00002)	0.1592 (0.0012)
1,000	0.9864	0.0037 (0.00002)	0.1633 (0.0010)
5,000	0.9870	0.0039 (0.00003)	0.1710 (0.0016)

5.3.5.2 Varying the binary covariate (X_1) prognostic strength

Figure 5-7 shows the relationship between ROR and the difference in (a) X_1 and (b) X_2 at different levels of binary covariate prognostic strength (X_1) and fixed X_2. Figure 5-7(c) depicts the expected values from the regression analysis. Figure 5-7(a) shows an inverse linear relationship between ROR and D_{X_1} when X_1 has a negative prognostic effect on the outcome. The slope of X_1, ϕ_1, was -0.012 when $OR_{X_1\_a}$ = 0.2. However, the relationship was positive when the prognostic impact was positive (slope = 0.012 when $OR_{X_1\_a}$ = 5). Figure 5-7(b) shows that ROR and D_{X_2} had a similar relationship to that between ROR and D_{X_1}. ϕ_2 was slightly higher when $OR_{X_1\_a}$ = 1.5 compared with $OR_{X_1\_a}$ = 0.2 or 5, although this increase was small. The values of ROR seen in Figure 5.7(b) were more variable when X_1 was highly prognostic.

Figure 5-7 **Scatter plots of the ratio of the unadjusted and adjusted odds ratios against the difference in (a) X_1 and (b) X_2 between the treatment and control groups, and (c) the fitted regression lines for ROR on the difference in X_1 by the level of difference in X_2 at varying X_1 prognostic strength (OR_{T_a} = 2, $OR_{X_1\_a}$ = 1.5, P_{X_1} = 50%, R = 50%, $OR_{X_2\_a}$ = 1.2, σ^2 = 1, and sample size = 100)**

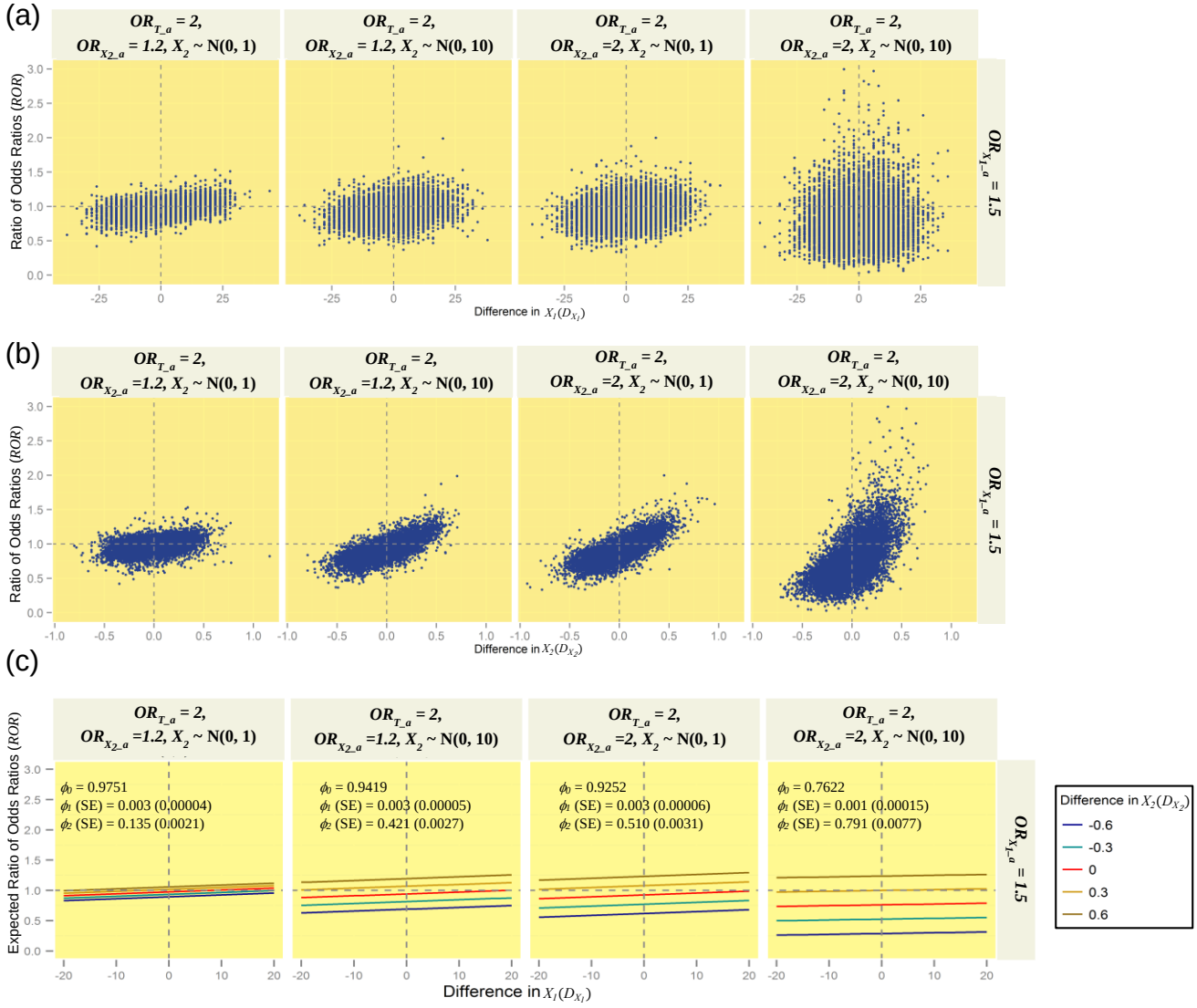


5.3.5.3 Varying the continuous covariate (X_2) prognostic strength

The analysis in Figure 5-7 was repeated for the continuous variable and is shown in Figure 5-8. The relationship between ROR and D_{X_1} was relatively unchanged by changes in the strength of the continuous variable ($OR_{X_2\_a}$). However, the relationship between ROR and D_{X_2} changed as the prognostic strength of X_2 increased. This was reflected in the regression analysis results: the slope of D_{X_2} was steeper as the odds ratio of X_2 increased (ϕ_2 = 0.135 to 0.791 in Figure 5-8(c)). The distance between the regression lines for each D_{X_2} increased as X_2 became highly prognostic.

As the prognostic strength of X_2 increased, both the percentage difference between the unadjusted and adjusted odds ratios due to noncollapsibility and due to imbalance in X_2 increased. For example, when both covariates were balanced, the unadjusted odds ratio decreased by 1.5% (OR_{T_u} = 1.97) when OR_{T_a} = 2 and σ^2 = 1, and by 23.8% (OR_{T_u} = 1.52) when OR_{T_a} = 2 and σ^2 = 10. However, when the mean of X_2 was 0.3 units higher in the control than in the treatment group, the unadjusted odds ratio decreased by a further 1.35% when OR_{T_a} = 2 and σ^2 = 1, and by a further 7.91% when OR_{T_a} = 2 and σ^2 = 10.

Figure 5-8 **Scatter plots of the ratio of the unadjusted and adjusted odds ratios against the difference in (a) X_1 and (b) X_2 between the treatment and control groups, and (c) the fitted regression lines for ROR on the difference in X_1 by the difference in X_2 at different X_1 prognostic strengths (OR_{T_a} = 2, $OR_{X_1\_a}$ = 1.5, P_{X_1} = 50%, R = 50%, $OR_{X_2\_a}$ = 1.2, σ^2 = 1, and sample size = 100)**

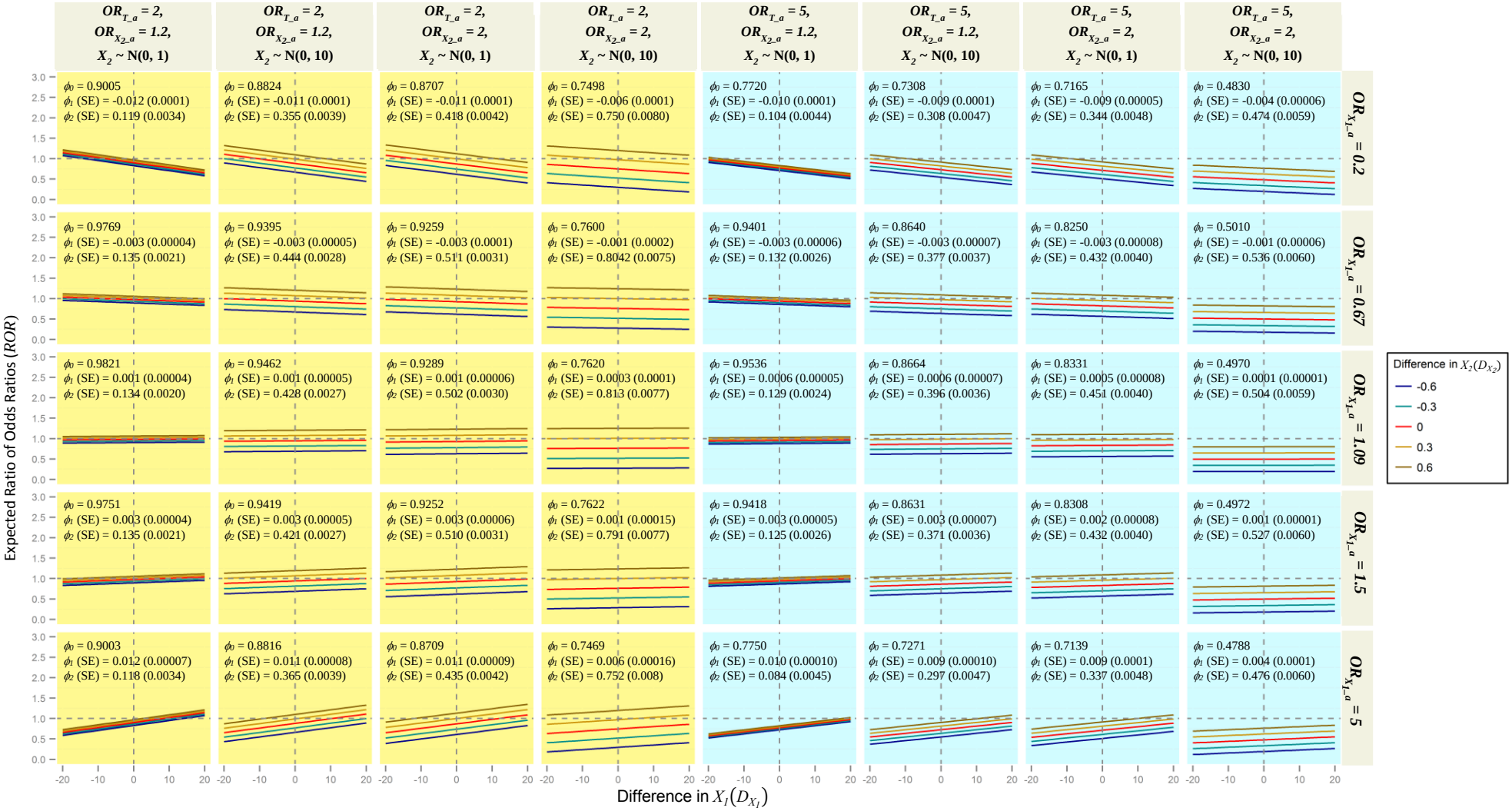


5.3.5.4 Overall results

Figure 5-9 presents the expected values of ROR against D_{X_1} and D_{X_2} for all of the scenarios investigated. The results show that the intercept was greater when OR_{T_a} = 2 than when OR_{T_a} = 5. This is consistent with the results reported in Chapter 3. The slope of D_{X_1}, ϕ_1, was mostly unaffected by changes in the strength of the adjusted treatment effect. The slope of D_{X_2}, ϕ_2, was greater when OR_{T_a} = 2 (yellow panels) than when OR_{T_a} = 5 (blue panels). This result suggests that the difference between the unadjusted and adjusted treatment odds ratios due to noncollapsibiity was greater with a very large treatment effect than a large treatment effect. However, the difference due to an imbalance in the continuous covariate was more apparent with a large treatment effect than a very large treatment effect.

The relationship between ROR and the difference in X_1 and X_2 followed a similar pattern when the event rate and sample size differed (Appendix H).

Figure 5-9 Regression line of the expected ratio of the unadjusted and adjusted odds ratios against the difference in X_1 at selected differences in X_2 for all of the scenarios



5.4 Summary

This chapter has reported the results of a resampling study that drew samples from population data sets without enforcing that both covariates be balanced. The approach followed a similar rationale to the simple randomisation method used in a randomised trial. Both covariates were balanced in only a small percentage of the samples drawn. Although the percentage of samples with balanced covariates increased with sample size, only a quarter of the samples of size 5,000 had two balanced covariates.

Two previous studies reported the percentage imbalance in their simulations. Chu et al (2012) reported that around 15% of their simulated samples of size 100 had a balanced binary covariate, using a threshold difference of 1%. This percentage increased to approximately 50% when they used a sample size of 4,000. Thompson et al (2014) used data from two large trials for their simulations. They found that the percentage of balance in each covariate varied from 56% to 97% when using an overall event rate of 63% and a sample size of 600. These studies both reported a higher percentage of samples with balanced covariates than the simulation study reported in this chapter, as this study required both covariates to be balanced and used more stringent thresholds for balance. However, this simulation study led to a conclusion consistent with those in the published studies: trials are likely to suffer from covariate imbalance if they do not use a restricted method to ensure covariate balance across treatment groups.

The logistic regression analyses performed in this chapter showed that sample size did not affect the difference between adjusted and unadjusted analysis, regardless of whether all of the samples or only those with balanced covariates were used. Highly similar ROR values were obtained for all of the scenarios. The findings are consistent with those of Thompson et al (2014), although they used a different approach to assess the impact of sample size.

This study also found that covariate imbalance affected ROR, dependent on the direction and magnitude of the imbalance. Fitting an ROR regression line on D_{X_1} and D_{X_2} allowed the percentage difference between the unadjusted and adjusted odds ratios to be decomposed into the difference due to noncollapsibility and the difference due to covariate imbalance. Steyerberg, Bossuyt and Lee (2000) also attempted to quantify how much of the difference in the treatment effect was attributable to imbalance and stratification, but did not provide a scientific rationale for how their method was derived. Pang and Kaufman (2013) quantified the two types of difference using a marginal structure model approach.

A very large treatment effect influenced the impact of covariate adjustment due to noncollapsibility more than a large treatment effect. In contrast, a large treatment effect influenced covariate imbalance more than a very large treatment effect.

CHAPTER 6: IMPACT OF COVARIATE ADJUSTMENT ON META-ANALYSIS OF RANDOMISED INTERVENTION STUDIES FOR BINARY OUTCOMES

6.1 Introduction

A simple unadjusted meta-analysis that does not adjust for additional prognostic factors is the most common approach for estimating overall treatment effects from multiple studies. The review reported in Chapter 2 found that the prognostic factors used in the analysis of clinical trials were often not reported in publications, and that the results were often presented in an aggregate form. Conducting a meta-analysis on studies that have adjusted for prognostic factors is complex due to the differences in how each study conducts its analysis. Differences in the sample size and randomisation method can affect the distribution of the covariates between the treatments (Steyerberg et al, 2000; Zhang et al, 2005). Each study can also measure or define the covariates differently. For example, different trials may define an ethnic group such as ‘Asian’ differently. Studies carried out in the UK commonly associate the qualifier ‘Asian’ with people of South Asian origin, whereas studies carried out in Sweden are likely to use ‘Asian’ to refer to those from all Asian countries, including the Middle East.

Missing data adds a further layer of complexity to the process. A meta-analysis may have missing baseline covariates at both the study level, when a study does not have access to or does not collect a particular variable, and at the patient level, when a study is missing the data for a collected variable from certain patients. Even when individual participant data (IPD) are available, the methods available for adjusting covariates are varied (Simmonds et al, 2005). Meta-analysis of IPD can be carried out using a one-stage or two-stage approach. A two-stage approach calculates the pooled treatment effect by combining the individual treatment effects obtained from each study. A one-stage approach obtains the pooled treatment effect by fitting a single model to all of the IPD from all of the trials and handling the variation between the trials as either a fixed or random effect (Goldstein et al, 2000; Debray et al, 2013).

The purpose of this study is to explore the impact of covariate adjustment on meta-analysis results. IPD from the Perinatal Antiplatelet Review of International Studies (PARIS) study (Askie et al, 2007) were used. Meta-analysis was carried out with and without adjustment for baseline covariates and the treatment responses were compared. This chapter addresses the following specific aims:

1. Assess the impact of covariate adjustment on meta-analysis conducted under a one-stage or two-stage analysis framework.

2. Use cumulative meta-analysis to assess the impact of covariate adjustment as additional studies are added to the meta-analysis one by one.

6.2 Methods

6.2.1 Motivating example

Pre-eclampsia is a complication in pregnancy characterised by high blood pressure and proteinuria. The disorder increases the risk of severe poor outcomes for both the mother and baby, including eclampsia and death. Antiplatelet therapy is a class of drugs, such as aspirin, that can prevent blood clots and may therefore prevent or delay the development of pre-eclampsia. The PARIS systematic review was conducted to determine whether antiplatelet agents can be used for the primary prevention of pre-eclampsia. It included IPD from 31 RCTs of pre-eclampsia primary prevention (32,217 women and 32,819 babies). It included trials that randomised women at risk of pre-eclampsia to either antiplatelet therapy (such as low-dose aspirin or dipyridamole, or any combination of antiplatelet agents) or placebo. The pre-specified main outcomes were (1) a diagnosis of pre-eclampsia (hypertension with new onset proteinuria at or beyond 20 weeks' gestation), (2) death in utero or death of the baby before discharge from hospital, (3) preterm birth at less than 34 weeks' gestation, (4) infant who was small for their gestational age at birth, and (5) pregnancy with a serious adverse outcome (PSAO). The full details and results of the project are

published elsewhere (Askie et al, 2007; The Perinatal Antiplatelet Review of International Studies Collaboration Steering Group, 2005).

The PARIS data were chosen for this investigation because they contained a large number of studies with different sample sizes and treatment effects, as well as covariates with varying prognostic strengths and correlations between the covariates. This wide variation of studies aids in understanding how the difference between unadjusted and adjusted analysis varies between studies, and how this difference affects meta-analysis.

6.2.2 Selection of the outcome for this investigation

Four of the five pre-specified outcomes in the PARIS meta-analysis were observed in few participants, resulting in low overall event rates. The simulation studies reported in Chapters 3 and 4 suggested that the difference between adjusted and unadjusted results is maximised when event rates are close to 50%. A composite outcome was thus chosen as the outcome for this investigation, PSAO, defined as the occurrence of any of the four main outcomes (i.e. pre-eclampsia, foetal/baby death before discharge, delivery < 34 weeks' gestation, or a small-for-gestational-age infant) or maternal death. Twenty-four of the 31 studies had data for all four main outcomes. The observed event rates for this outcome ranged from 6% to 45% in the included trials (Table 6-1).

Table 6-1 Event rates for the PARIS pre-specified main outcomes recorded for each trial (ordered according sample size)

Trial	Country	Sample size	Pre-eclampsia	Foetal/baby death before discharge	Delivery < 34 weeks' gestation	Small-for-gestational-age infant	Maternal death	Pregnancy with a serious adverse outcome
Kincaid-Smith et al (1995)	Australia	21	24%	10%	19%	6%	0%	43%
Vainio et al (2002)	Finland	90	13%	0%	3%	4%	0%	17%
Morris et al (1996)	Australia	102	11%	0%	3%	24%	0%	32%
Michael and Walters (1993)	Australia	110	13%	0%	7%	6%	0%	20%
Rogers et al (1999)	China	193	10%	0.5%	2%	10%	0%	18%
Uzan I: Uzan et al (1989)	France	230	13%	6%	13%	17%	0%	32%
Byaruhanga et al (1998)	Zimbabwe	256	18%	7%	11%	0%	0%	25%
Uzan II: Uzan et al (1991)	France	315	11%	5%	12%	13%	0%	29%
Hermida et al (2003)	Spain	341	10%	0%	2%	14%	0%	20%
Yu et al (2003)	UK	560	19%	2%	13%	23%	0%	39%
Hauth et al (1993)	USA	606	4%	0.3%	3%	5%	0%	10%
ECPPA (1996)	Brazil	1,091	13%	7%	7%	9%	0%	24%
Caritis et al (1998)	USA	2,539	17%	5%	17%	9%	0.08%	36%
Sibai et al (1993)	USA	3,135	3%	2%	4%	5%	0.07%	12%
ERASME: Subtil et al (2003)	France, Belgium	3,294	2%	1%	2%	2%	0%	6%
Rotchell et al (1998)	Barbados	3,647	5%	4%	6%	0.7%	0.03%	12%
Golding (1998)	Jamaica	6,275	5%	8%	7%	0%	0.05%	13%
CLASP (1994)	International*	8,021	12%	2%	6%	3%	0.01%	18%

* Argentina, Australia, Belgium, Canada, Germany, Hong kong, Israel, Malaysia, Netherlands, New Zealand, Russia, Spain, Sweden, United Arab Emirates, UK, and USA.

6.3 Selection and assessment of baseline covariates

The covariates used in this investigation came from two sources. The covariates measured at randomisation that were specified as risk factors for pre-eclampsia in the main PARIS report were included (Askie et al, 2007). These covariates are whether the women were hypertensive at baseline, previous hypertensive disorders of pregnancy, diabetes, renal disease, multiple pregnancies, maternal age, a previous small-for-gestational-age infant, parity, and the type of hypertension at trial entry.

Recently published studies and guidelines were also reviewed (Ota et al, 2014; Bilano et al, 2014; National Institute for Health and Care Excellence, 2013; Royal College of Obstetricians and Gynaecologists, 2013) to identify additional risk factors. As PSAO is a composite outcome, the search focused on the two most common outcomes, pre-eclampsia and small-for-gestational-age infants. The covariates were selected based on both the information provided from these sources and the amount of non-missing data available.

Information on each of the covariates was obtained from each study to determine which covariates were included as stratification factors in randomisation, whether the covariates were evenly distributed across the treatment groups, and if they were part of the inclusion and exclusion criteria. A logistic regression was carried out to examine the empirical association between each covariate and the

composite PSAO. A covariate was considered to have a statistically significant association with PSAO if the corresponding P-value was less than 0.05.

6.4 Methodology

6.4.1 Methods of meta-analysis

6.4.1.1 Two-stage meta-analysis

The first step of the two-stage approach is to obtain an estimated treatment effect for each study. These estimates are then combined to form a weighted mean across all of the studies:

$$\hat{\theta}^* = \frac{\sum \omega_i \hat{\theta}_i}{\sum \omega_i} \qquad \text{(Eq. 6-1)}$$

where $\hat{\theta}^*$ is the pooled treatment effect, $\hat{\theta}_i$ is the estimated effect size for study i, and ω_i is the weight assigned to each study, computed as:

$$\omega_i = \frac{1}{v_i + \tau_i^2} \qquad \text{(Eq. 6-2)}$$

where v_i and τ_i^2 denote the within- and between-study variances, respectively. Eq. 6-1 is referred to as a fixed effect model when τ_i^2 = 0 and as a random effects model when τ_i^2 > 0. Different methods for estimating τ_i^2 have been proposed (Sidik and Jonkman, 2006). The most commonly used approach is DerSimonian and Laird's (1986) method of moments.

6.4.1.2 One-stage meta-analysis

IPD data from multiple studies have a hierarchical structure; the participants are the first level and are nested within studies, which form the second level. A one-stage meta-analysis approach can thus be conducted using multilevel modelling techniques that incorporate all IPD in a single analysis (Goldstein et al, 2000; Turner et al, 2000).

A one-stage fixed effect model can be modelled using a generalised linear model with a logit link function:

$$logit(p_{ij}) = \sum_{i=1}^{M}(\alpha_i I_i) + \theta T_{ij} \qquad \text{Model 1}$$

where p_{ij} denotes the probability of an event for participant j in study i ($i = 1, \ldots, M$), T_{ij} denotes the allocated treatment, α_i is the risk of the control group in each study, I_i is the indicator term for study i, and θ is the log odds ratio of the treatment effect.

Turner et al (2000) proposed a method for one-stage random effects models:

$$logit(p_{ij}) = \sum_{i=1}^{M}(\alpha_i I_i) + \theta_i T_{ij} \qquad \text{Model 2}$$

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

where the treatment effect, θ_i, varies across trials. It follows a normal distribution around the overall effect θ with heterogeneity τ^2.

The fixed effect model defined by Model 1 can extended to adjust for covariates:

$$logit(p_{ij}) = \sum_{i=1}^{M}(\alpha_i I_i) + \theta T_{ij} + \sum_{k=1}^{K}(\gamma x_{ij})_k \qquad \text{Model 3}$$

where γ_k is the log odds ratio for the k^{th} covariate, x_k.

The covariate-adjusted random effects model is defined as:

$$logit(p_{ij}) = \sum_{i=1}^{M}(\alpha_i I_i) + \theta_i T_{ij} + \sum_{k=1}^{K}(\gamma_i x_{ij})_k \qquad \text{Model 4}$$

$$\begin{bmatrix} \theta_i \\ \gamma_{i1} \\ \vdots \\ \gamma_{ik} \end{bmatrix} \sim MVN\left(\begin{bmatrix} \theta \\ \gamma_1 \\ \vdots \\ \gamma_k \end{bmatrix}, \begin{bmatrix} \tau_\theta^2 & 0 & \cdots & 0 \\ 0 & \tau_{\gamma_1}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \tau_{\gamma_k}^2 \end{bmatrix}\right)$$

where $\tau_\theta^2, \tau_{\gamma_1}^2, \ldots, \tau_{\gamma_k}^2$ are the between-study variances corresponding to the treatment and each covariate, respectively.

6.4.1.3 Cumulative meta-analysis

Cumulative meta-analysis adds one study at a time to the analysis, ordered by a study characteristic (e.g., year of publication, sample size, or quality) (Lau,

Schmid and Chalmers, 1995). This method can be used to assess how the results of covariate adjustment affect the pooled odds ratio. It is particularly useful for identifying when a treatment indicates clinical effectiveness and for assessing the presence of bias or heterogeneity in a meta-analysis.

6.4.1.4 Data analysis

Data from each study was modelled using a logistic regression with and without adjustment for the selected covariates. As mentioned in Chapter 5, logistic regression is the most commonly used method for the analysis of binary outcomes. However, it can produce biased results and exaggerate the effect size at a low event rate, when separation is an issue.

Separation occurs when the binary outcome variable can be perfectly separated by a single covariate or a non-trivial linear combination of covariates (Albert and Anderson, 1984). When separation occurs, no events are observed in one of the two groups; events are perfectly separated from non-events. Separation occurs more often when the event rate is low or the sample size is small (Heinze, 2006). The analysis was repeated using Firth's correction to avoid separation (Heinze and Schemper, 2002). Firth's correction is a penalised maximum likelihood estimation that reduces the level of bias. Thus, two unadjusted and two adjusted estimates of treatment effect were obtained for each study.

Each of the estimates was meta-analysed using the two-stage fixed method and DerSimonian and Laird random effects method. Heterogeneity in the random effects model was assessed using I^2 and τ^2. I^2 is a measure of the percentage of the variability in the treatment estimates that is attributable to heterogeneity between studies (Higgins and Thompson, 2002). τ^2 is the underlying between-study variability. ROR, defined as the unadjusted odds ratio divided by the adjusted odds ratio, was calculated for each study. The ratio of the pooled odds ratios ($RPOR$), defined as the pooled unadjusted ratio divided by the pooled adjusted odds ratio, was calculated for the pooled results.

For the one-stage approach, the multilevel model was estimated using the adaptive Gaussian quadrature method for numerical integration (Pinheiro and Bates, 1995; Pinheiro and Chao, 2006). The choice of the integration points for the estimation determines the accuracy of the results in this method. Seven integration points were specified for Models 1 to 3, to give accurate results. Only three integration points were used for Model 4 because there were convergence problems when more integration points were specified. This method is available in the Stata program *meqrlogit*.

Cumulative meta-analysis was carried out using the sample size of each study to determine the order that the studies should be added to the analysis. This analysis allowed a comparison of the trends in the evolution of the pooled treatment effect between the adjusted and unadjusted analyses.

All of the analyses were conducted using Stata version 13 and R version 3.1.1.

6.5 Results

6.5.1 Selection of covariates

Maternal age (years), parity, body mass index (kg/m^2), diagnosis of chronic hypertension (yes/no), diagnosis of chronic kidney disease (yes/no), and diagnosis of Type I or Type II diabetes (yes/no) were the most commonly cited risk factors for pre-eclampsia or small-for-gestational-age infants in the published literature (Table 6-2). Data for each of these risk factors were missing for at least some trials in the PARIS review data set. None of the trials in the PARIS data set included body mass index data.

Figure 6-1 displays the percentage of the randomised participants (mothers) for which each trial in the PARIS review collected data on each risk factor for. The colour density of each cell reflects the amount of data available for each risk factor from each study. The darker the colour, the less data are available (or the more missing data) for that risk factor in that trial. For example, the CLASP study recorded 100% of the participant values for seven of its ten covariates, 98% of the participant values for multiple pregnancies, 58% of the participant values for prior hypertensive disorder during pregnancy, and only 0.04% of the participant values for previous pregnancy-induced hypertension (CLASP, 1994).

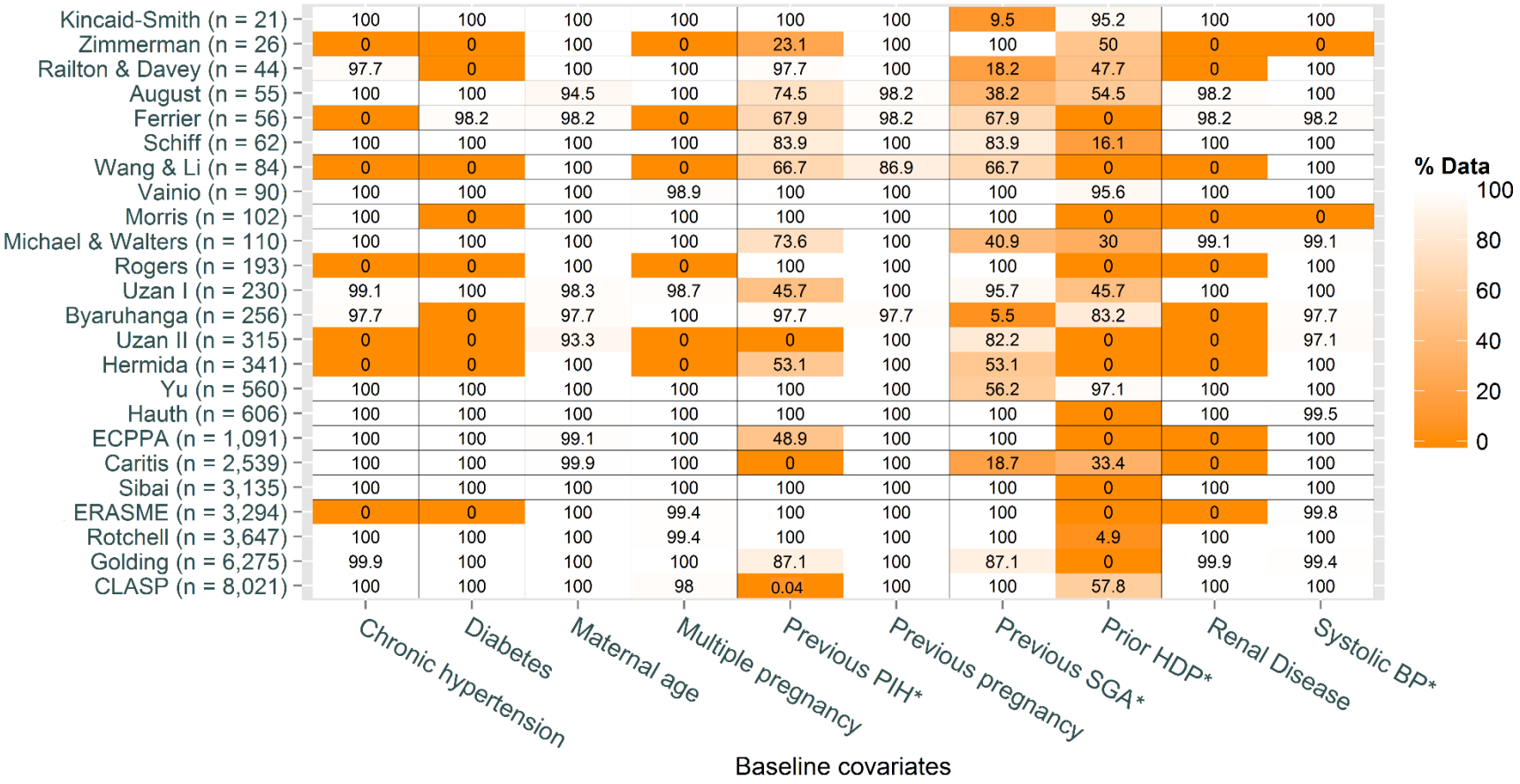
Table 6-2 Known risk factors for pre-eclampsia or small-for-gestational-age infants by source

Risk factors	Pre-eclampsia			Small-for-gestational-age infants	
	The National Institute for Health and Care Excellence guideline	Bilano et al (2014)	Askie et al (2007)	Royal College of Gynaecology guideline	Ota et al (2014)
Maternal age (years)	✓	✓	✓	✓	✓
Anaemia		✓			✓
Antenatal care visit		✓			
Autoimmune disease	✓		✓		
Body mass index	✓	✓		✓	
Chronic hypertension	✓	✓	✓	✓	✓
Chronic kidney disease	✓	✓	✓	✓	
Diet				✓	
Education		✓			
Exercise				✓	
Family history of pre-eclampsia	✓				
Gestational diabetes		✓			
Gestational hypertension			✓		
Hypertensive disease in previous pregnancy	✓		✓		
In vitro fertilisation				✓	
Marital status					✓
Maternal substance exposure (smoking or cocaine use)				✓	
Multiple pregnancy	✓		✓		
Other medical conditions					✓
Parity	✓	✓	✓	✓	✓
Paternal small gestational age			✓	✓	
Pregnancy interval	✓			✓	
Previous small-for-gestational-age infant or stillbirth				✓	
Previous pre-eclampsia				✓	
Type I or II diabetes	✓		✓	✓	
Urinary tract infection or pyelonephritis		✓			

All of the studies were missing data for at least one covariate. Eighteen of the 24 studies had at least 95% of the data for at least six of the covariates. The covariate that was missing data in the most studies was prior hypertensive disorder of pregnancy. This covariate was missing for 70% of all participants. Seventy-five percent of the studies were missing information on prior hypertensive disorder of pregnancy for more than half of their participants. Whether diagnoses of diabetes or renal disease had been made were also largely missing, with 38% and 46% of all participants missing data for these variables, respectively. Zimmerman et al (1997) and Morris et al (1996) were missing data for at least 50% of their participants for over half of the covariates.

The variables that had the least amount of missing data and are known to be strong prognostic factors for the primary outcome were selected for this investigation. The four covariates selected were maternal age (years), previous pregnancy (yes/no), systolic blood pressure (mmHg), and chronic hypertension (yes/no). Sixteen of the 24 studies (n = 26,425; 84.8% of mothers) included data on these four covariates, so were included in the analysis (Appendix J).

Figure 6-1 Heat map of the availability of the baseline covariates in each trial



* PIH: pregnancy-induced hypertension; SGA: small for gestational age; HDP: hypertensive disorder in pregency; Systolic BP: systolic blood pressure

6.5.2 Data description

Published trial reports and protocols were examined to determine whether the selected covariates were used as inclusion or exclusion criteria within each study. Four studies (Schiff et al, 1989; Yu et al, 2003; Hauth et al, 1993; Sibai et al, 1993) excluded participants with pre-existing hypertensive disorders, such as chronic hypertension (Table 6-3). Michael and Walters (1993) and Sibai et al (1993) only included participants with a systolic blood pressure within a specific threshold. Hauth et al (1993) excluded women with chronic hypertension, women older than 28 years, and women who had previously been pregnant.

The studies generally poorly reported how they conducted randomisation. Nearly all of the studies did not mention which randomisation method they used or whether they considered any risk factors during the randomisation process. CLASP (1994) mentioned the use of a minimisation algorithm for treatment allocation, but did not specify which of the minimisation factors were used.

Table 6-3 Reported inclusion/exclusion criteria and randomisation details from trial reports and protocols on the 16 PARIS studies

Study	SBP (mmHg)	Maternal age (years)	Chronic hypertension	Previous pregnancy	Randomisation method
Kincaid-Smith					Not mentioned
Railton and Davey					Not mentioned
August			Inclusion		Not mentioned
Schiff			Exclusion	Nulliparous	Not mentioned
Vainio			Inclusion		Not mentioned
Michael and Walters	SBP > 139				Not mentioned
Uzan I					Stratified by centre, whether had one or two previous poor outcomes
Byaruhanga			Inclusion		Not mentioned
Yu			Exclusion		Fixed blocks, but no mention of stratification
Hauth		< 28 years	Exclusion	Nulliparous	Not mentioned
ECPPA			Inclusion		Not mentioned
Caritis			inclusion as long as no multiple problems		Not mentioned
Sibai	SBP < 135		Exclusion	Nulliparous	Not mentioned
Rotchell					Not mentioned
Golding				Multiparous	Not mentioned
CLASP					Used minimisation, but minimisation factors not presented

SBP: systolic blood pressure
Yellow: exclusion criteria
Green: inclusion criteria

Table 6-4 presents the summary measures of the four covariates by treatment group. Despite of the lack of details reported, both continuous covariates (maternal age and systolic blood pressure) were reasonably balanced across the treatment groups. In contrast, there was considerable imbalance in the binary

covariates (previous pregnancy and chronic hypertension). The imbalance was more apparent in the smaller studies. For example, Railton and Davey (1998) allocated more multiparous participants but fewer pre-existing chronic hypertensive participants to the control group than to the treatment group (93% vs 78% and 28% vs 50%, respectively). As expected, the imbalances disappeared as the study size increased.

One study (Hauth et al, 1993) did not have any participants with a history of previous pregnancy, three studies had nearly all participants with a history of previous pregnancy (Uzan et al, 1989; Kincaid-Smith et al, 1995; August et al, 1994) and four studies ((Schiff et al, 1989; Yu et al, 2003; Hauth et al, 1993; Sibai et al, 1993) had no participants with chronic hypertension, as they used these covariates as inclusion/exclusion criteria. For each study, these covariates were included in the adjusted analysis, but did not contribute to the prediction of the outcome in the logistic regression models: the treatment effect remained unchanged whether or not these covariates were included in the analysis. Sibai et al (1993) and Michael and Walters (1993) used the baseline systolic blood pressure as an inclusion threshold, but the distribution of systolic blood pressure in these two studies was not noticeably different from that in the rest of the studies. The standard deviation of maternal age recorded by Hauth et al (1993) was noticeably smaller than in the other studies because they only included younger women.

The Pearson correlation coefficient for each covariate pair was calculated for each study. (Table 6-5). Maternal age was consistently positively correlated with systolic blood pressure, previous pregnancy and chronic hypertension, with correlation coefficients ranging from 0.03 to 0.50. The direction of the relationship between the other covariates varied. The correlation between previous pregnancy and chronic hypertension was negative in five studies, did not exist or was negligibly small in four studies, and was positive in two studies. The pattern of these variations did not appear to be associated with the size of the trial. The prognostic strength of these covariates also varied between studies (Appendix K). Although the majority of the covariates had moderate prognostic strength, previous pregnancy was strongly prognostic for PSAO (odds ratio = 8) in Schiff (1989). Eleven of the 16 studies (69%) had a statistically significant association between PSAO and at least one of the four selected covariates, based on P-value < 0.05 (values highlighted in red in Table 6-4).

Table 6-4 Descriptive statistics of the four selected covariates in each of the 16 PARIS studies*

Study	N randomised		Systolic blood pressure mean (SD†)		Maternal age mean (SD†)		Previous pregnancy N (%)		Chronic hypertension N (%)	
	Control	AA†	Control	AA†	Control	AA†	Control	AA†	Control	AA†
Kincaid-Smith	11	10	125 (8)	124 (12)	28.8 (3.8)	30.8 (4.6)	10 (91%)	10 (100%)	1 (9%)	4 (40%)
Railton and Davey	14	28	130 (15)	132 (18)	28.4 (6.9)	30.8 (6.8)	13 (93%)	22 (78%)	4 (28%)	14 (50%)
August	27	23	124 (16)	120 (13)	32.0 (6.5)	32.6 (6.0)	22 (82%)	23 (100%)	24 (89%)	18 (78%)
Schiff	29	33	111 (10)	111 (11)	28.3 (4.2)	28.3 (4.9)	4 (14%)	6 (18%)	0 (0%)	0 (0%)
Vainio	43	43	132 (16)	131 (16)	30.1 (5.9)	30.6 (6.3)	33 (77%)	28 (65%)	13 (30%)	16 (37%)
Michael and Walters	55	53	141 (13)	141 (16)	27.7 (5.9)	27.1 (6.2)	46 (84%)	41 (77%)	18 (33%)	24 (45%)
Uzan I	71	153	117 (16)	121 (14)	28.4 (4.6)	29.3 (4.3)	71 (100%)	153 (100%)	42 (59%)	87 (57%)
Byaruhanga	124	122	127 (22)	126 (18)	28.0 (5.0)	27.4 (5.1)	118 (95%)	115 (94%)	20 (16%)	17 (14%)
Yu	280	280	112 (9)	111 (9)	28.5 (6.2)	27.6 (6.4)	129 (46%)	117 (42%)	0 (0%)	0 (0%)
Hauth	301	302	109 (8)	108 (8)	19.7 (2.7)	19.5 (2.6)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
ECPPA	527	510	125 (20)	127 (20)	27.1 (7.2)	27.6 (7.5)	251 (48%)	281 (55%)	230 (44%)	247 (48%)
Caritis	1,247	1,253	116 (15)	117 (15)	26.8 (6.3)	26.6 (6.3)	1,017 (82%)	1,018 (81%)	413 (33%)	405 (32%)
Sibai	1,501	1,485	106 (11)	105 (11)	20.6 (4.8)	20.3 (4.4)	371 (25%)	340 (23%)	0 (0%)	0 (0%)
Rotchell	1,824	1,823	114 (10)	114 (11)	24.3 (5.9)	24.4 (5.8)	1,025 (56%)	1,019 (56%)	8 (0.4%)	8 (0.4%)
Golding	3,108	3,122	107 (12)	106 (12)	20.3 (4.2)	20.4 (4.2)	408 (13%)	396 (13%)	10 (0.3%)	10 (0.3%)
CLASP	4,008	4,013	124 (17)	124 (17)	28.6 (5.3)	28.7 (5.3)	3,088 (77%)	3,085 (77%)	843 (21%)	826 (20%)

* Values highlighted in red correspond to significant risk factors on the outcome for that study

† SD = standard deviation; AA = Antiplatelet agent

Table 6-5 Study-specific Pearson correlation coefficients for each of the 16 PARIS studies

Study	n	SBP / MatAge	SBP / PrevPreg	SBP / Chronic HT	MatAge / PrevPreg	MatAge / Chronic HT	PrevPreg / Chronic HT
Kincaid-Smith	21	0.149	-0.115	0.516	0.096	0.415	0.125
Railton and Davey	42	0.040	0.081	-0.271	0.253	0.191	0.000
August	50	0.101	-0.159	0.490	0.034	0.264	-0.146
Schiff	62	0.165	0.165	-0.002			
Vainio	86	0.275	-0.028	0.560	0.339	0.25	-0.139
Michael and Walters	108	0.215	-0.092	0.026	0.236	0.214	0.056
Uzan I	224	0.153		0.003		0.101	
Byaruhanga	246	0.158	-0.132	0.200	0.227	0.238	-0.256
Yu	560	0.107	0.0001		0.336		
Hauth	603	0.069					
ECPPA	1,037	0.352	0.319	0.483	0.456	0.346	0.504
Caritis	2,500	0.234	-0.015	0.520	0.206	0.344	0.0003
Sibai	2,986	0.092	0.067		0.15		
Rotchell	3,647	0.187	0.039	0.106	0.503	0.099	0.017
Golding	6,232	0.129	0.024	0.046	0.212	0.047	-0.013
CLASP	8,021	0.153	-0.037	0.452	0.136	0.147	-0.153

SBP: systolic blood pressure; MatAge: maternal age; PrevPreg: previous pregnancy; Chronic HT: chronic hypertension

6.5.3 Results of meta-analysis using a two-stage model

6.5.3.1 Study-specific treatment effects

There was considerable between-study variation in the percentage of women with PSAO, with estimates ranging from 10% (Hauth et al, 1993) to 45% (Railton and Davey, 1988). The estimated treatment effects from the unadjusted and adjusted analyses using maximum likelihood logistic regression (OR_{unadj_ML} and OR_{adj_ML}, respectively) and Firth's correction (OR_{unadj_Firth} and OR_{adj_Firth}, respectively) are presented in Table 6-6. The odds ratios obtained using Firth's correction were in general lower than the conventional values obtained from maximum likelihood logistic regression when the odds ratios were greater than 1, and higher when they were less than 1. The penalised method thus gave a more conservative result than the conventional method. The most obvious correction to the odds ratio was observed in Schiff (1989), for which both the adjusted odds ratios and SEs were corrected upwards (i.e., OR_{adj_ML} (SE_{adj_ML}) = 0.0064 (0.0125) vs OR_{adj_Firth} (SE_{adj_Firth}) = 0.0275 (0.0404)). The difference between the two methods' estimates diminished as study size increased. Overall, the treatment effects were more apparent, i.e. larger effects (smaller ORs), in the smaller studies than in the larger studies.

The adjusted treatment effects from both methods were further from the null (i.e., no treatment effect) than the unadjusted results for most of the studies. As expected, the adjusted errors had larger SEs than the unadjusted estimates.

Table 6-6 Unadjusted and adjusted treatment effects and corresponding standard errors for each of the 16 PARIS studies

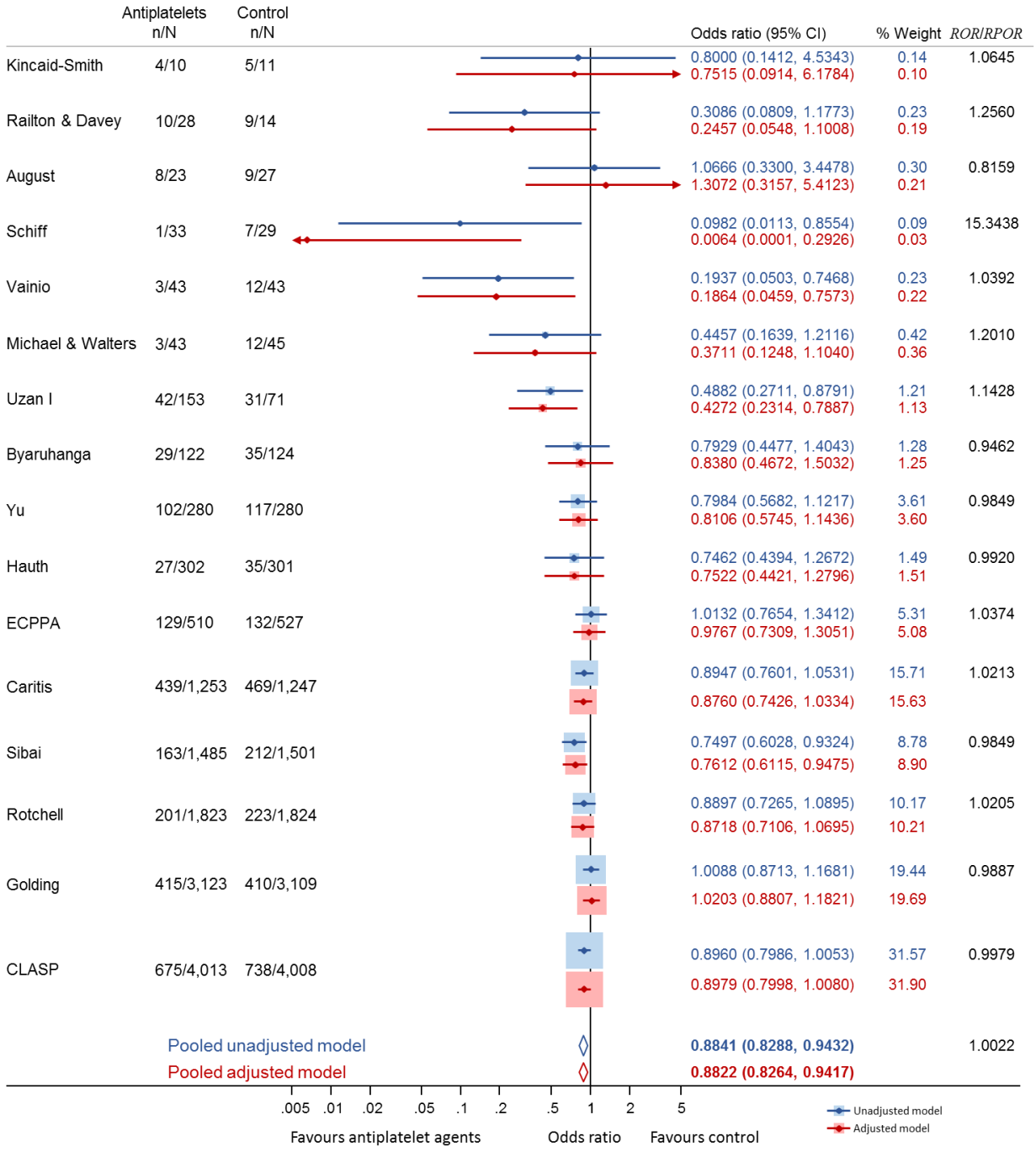
Trial	Sample size (event rate)	OR_{unadj_ML}* (SE_{unadj})	OR_{adj_ML}* (SE_{adj_ML})	ROR_{ML}*	OR_{unadj_Firth}† (SE_{unadj_Firth})	OR_{adj_Firth}† (SE_{adj_Firth})	ROR_{Firth}†
Kincaid-Smith	21 (43%)	0.8000 (0.7081)	0.7515 (0.8078)	1.0645	0.8182 (0.6903)	0.8016 (0.7534)	1.0207
Railton and Davey	42 (45%)	0.3086 (0.2108)	0.2457 (0.1880)	1.2560	0.3286 (0.2171)	0.2986 (0.2106)	1.1005
August	50 (34%)	1.0666 (0.6385)	1.3072 (0.9476)	0.8159	1.0679 (0.6240)	1.2200 (0.8191)	0.8753
Schiff	62 (13%)	0.0982 (0.1084)	0.0064 (0.0125)	15.3438	0.1385 (0.1295)	0.0275 (0.0404)	5.0437
Vainio	86 (17%)	0.1937 (0.1333)	0.1864 (0.1333)	1.0392	0.2178 (0.1415)	0.2234 (0.1470)	0.9749
Michael and Walters	108 (19%)	0.4457 (0.2274)	0.3711 (0.2064)	1.2010	0.4616 (0.2298)	0.4098 (0.2166)	1.1264
Uzan I	224 (33%)	0.4882 (0.1465)	0.4272 (0.1336)	1.1428	0.4901 (0.1461)	0.4354 (0.1342)	1.1256
Byaruhanga	246 (26%)	0.7929 (0.2312)	0.8380 (0.2498)	0.9462	0.7954 (0.2305)	0.8422 (0.2472)	0.9444
Yu	560 (39%)	0.7984 (0.1385)	0.8106 (0.1423)	0.9849	0.7990 (0.1384)	0.8122 (0.1420)	0.9837
Hauth	603 (10%)	0.7462 (0.2016)	0.7522 (0.2039)	0.9920	0.7493 (0.2010)	0.7557 (0.2030)	0.9915
ECPPA	1,037 (25%)	1.0132 (0.1450)	0.9767 (0.1445)	1.0374	1.0132 (0.1448)	0.9769 (0.1440)	1.0372
Caritis	2,500 (36%)	0.8947 (0.0744)	0.8760 (0.0738)	1.0213	0.8947 (0.0744)	0.8763 (0.0738)	1.0210
Sibai	2,986 (13%)	0.7497 (0.0834)	0.7612 (0.0850)	0.9849	0.7502 (0.0834)	0.7618 (0.0849)	0.9848
Rotchell	3,647 (12%)	0.8897 (0.0920)	0.8718 (0.0909)	1.0205	0.8899 (0.0920)	0.8720 (0.0908)	1.0205
Golding	6,232 (13%)	1.0088 (0.0755)	1.0203 (0.0766)	0.9887	1.0088 (0.0754)	1.0202 (0.0766)	0.9888
CLASP	8,021 (17%)	0.8960 (0.0526)	0.8979 (0.0530)	0.9979	0.8961 (0.0526)	0.8977 (0.0529)	0.9982

* OR_{unadj_ML}= unadjusted odds ratio (maximum likelihood logistic regression); SE_{unadj_ML} = standard error of unadjusted odds ratio (maximum likelihood logistic regression); OR_{adj_ML}= adjusted odds ratio (maximum likelihood logistic regression); SE_{adj_ML} = standard error of adjusted odds ratio (maximum likelihood logistic regression); ROR_{ML} = radio of odds ratios (maximum likelihood logistic regression)

† OR_{unadj_Firth} = unadjusted odds ratio (Firth's logistic regression); SE_{unadj_Firth} = standard error of unadjusted odds ratio (Firth's logistic regression); OR_{adj_Firth} = adjusted odds ratio (Firth's logistic regression); SE_{adj_Firth} = standard error of adjusted odds ratio (Firth's logistic regression); ROR_{Firth} = radio of odds ratios (Firth's logistic regression)

The smaller studies had greater ROR (unadjusted to adjusted) variability than the larger studies. The trials with sample sizes under 500 had ROR values ranging from 0.82 to 15.34. The Schiff (1989) ROR with a treatment effect from the adjusted analysis was substantially larger than from the unadjusted result (ROR = 15.34; the unadjusted odds ratio was about 15 times higher than the adjusted odds ratio). This study showed the most extreme beneficial effect of antiplatelet agents. This effect was even more apparent after adjusting for the covariates (OR_{unadj_ML} = 0.0982 and OR_{adj_ML} = 0.0064). The impact of covariate adjustment remained high when Firth's correction was used (ROR = 5.04), but was much less dramatic. The dramatic effect of covariate adjustment may have resulted from the study's small sample (n = 62) and the number of adjusted covariates that had a relatively low event rate (13%). The ROR varied much less in larger trials of at least 300 participants (from 0.98 to 1.04) (Figure 6-2).

Figure 6-2 Forest plot of the unadjusted and adjusted odds ratios of pregnancy with a serious adverse outcome, calculated using a two-stage fixed effect model, showing the 16 PARIS studies



6.5.3.2 Fixed effect meta-analysis

The effect of covariate adjustment on the magnitude of the SE of the treatment effect was proportional to the size of the study. The SE was larger in the smaller studies and smaller in the larger studies. The weight assigned to the smaller studies in the meta-analysis therefore decreased and the weight assigned to the larger studies increased, as weight is proportional to SE (Figure 6-2). As reported in the original publication of this study (Askie et al, 2007), there was an overall reduction in the odds of women experiencing PSAO in the antiplatelet agent group, compared with the placebo (pooled odds ratio = 0.88; 95% CI = 0.82 to 0.96).

The two-stage fixed effect model produced similar unadjusted and adjusted pooled treatment effects (pooled odds ratio = 0.88; 95% CI 0.83 to 0.94 in both cases; Figure 6-2). The pooled adjusted results were almost identical to the pooled unadjusted results ($RPOR$ = 1.0022). The $RPOR$ and ROR calculated using Firth's correction were nearly identical to those calculated using the conventional method (Appendix L). The pooled odds ratio (95% CI) for both the unadjusted and adjusted analyses calculated using Firth's correction was 0.88 (95% CI 0.83 to 0.94, $RPOR$ = 1.0020).

6.5.3.3 Random effects meta-analysis

More variability was observed between the treatment effects in the adjusted analysis than in the unadjusted analysis (I^2 = 47.7% vs 38%, respectively) when the two-stage random effects model was used (Figure 6-3). However, the underlying between-study variability was low (τ^2 = 0.019 vs 0.012, respectively). The increase in I^2 was likely due to the increase in the variation of the log odds ratios after adjustment, which was mainly driven by the data from Schiff (1989). The random effects model thus had a higher $RPOR$ than the fixed effect model. The results were consistent with those obtained using Firth's correction (pooled unadjusted odds ratio (95% CI) = 0.85 (0.77, 0.94) vs pooled adjusted odds ratio (95% CI) = 0.84 (0.75 to 0.94), $RPOR$ = 1.0116) (Appendix L).

6.5.4 Results of meta-analysis using a one-stage model

The results obtained from the one-stage analysis are reported in Figure 6-4, alongside the two-stage results. Using the fixed effect model and the one-stage approach, with or without covariate adjustment, gave similar results to the two-stage approach, although the pooled treatment effect in the one-stage results was marginally further away from the null than in the two-stage results. Using the random effects model and the one-stage approach produced pooled odds ratios that were closer to the null than using the two-stage approach. The pooled odds ratios generated from the random effects one-stage approach were similar to those from the fixed effect model. The between-study variability, τ^2, was zero for both models with and without covariate adjustment. As expected, the CIs of

the pooled odds ratios were narrower for the one-stage analysis than for the two-stage analysis.

Figure 6-3 Forest plot of the unadjusted and adjusted odds ratios of pregnancy with a serious adverse outcome, calculated using the two-stage random effects model, showing the 16 PARIS studies

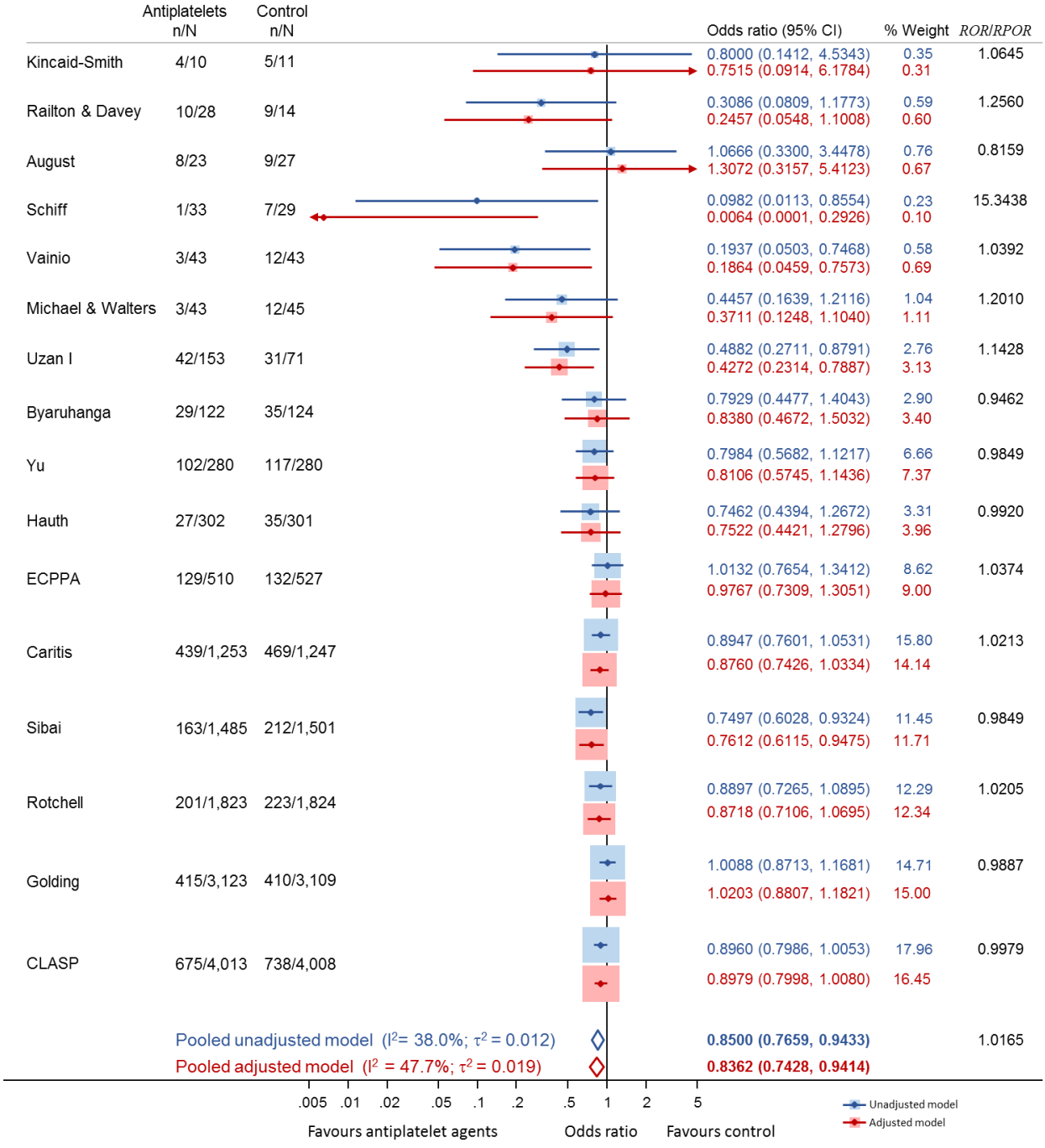
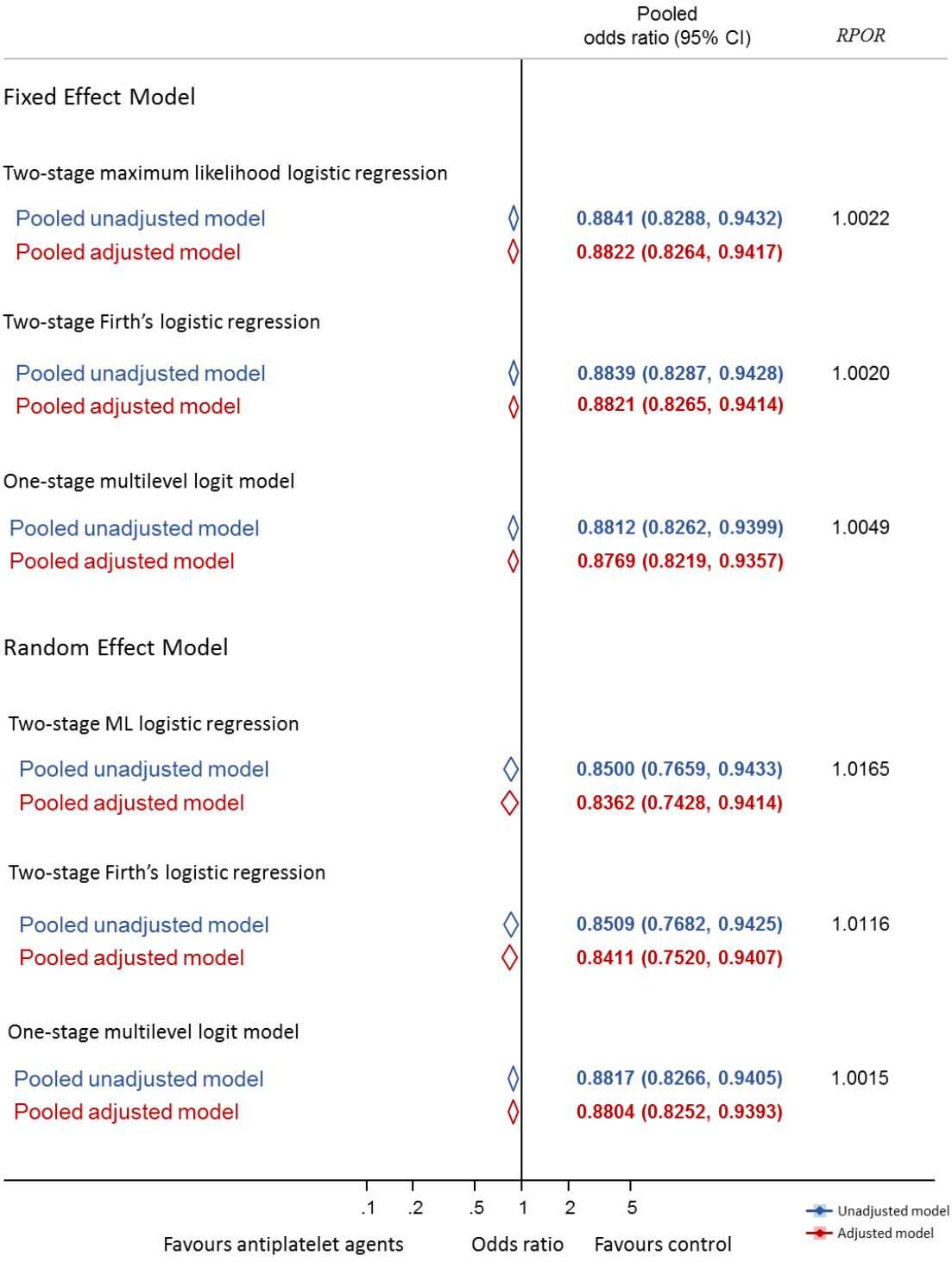


Figure 6-4 Summary of the pooled odds ratios and ratios of pooled unadjusted and adjusted odds ratios calculated using different logistic regression and meta-analysis methods on the 16 PARIS studies



6.5.5 Impact of adjusted analysis in meta-analysis, according to sample size

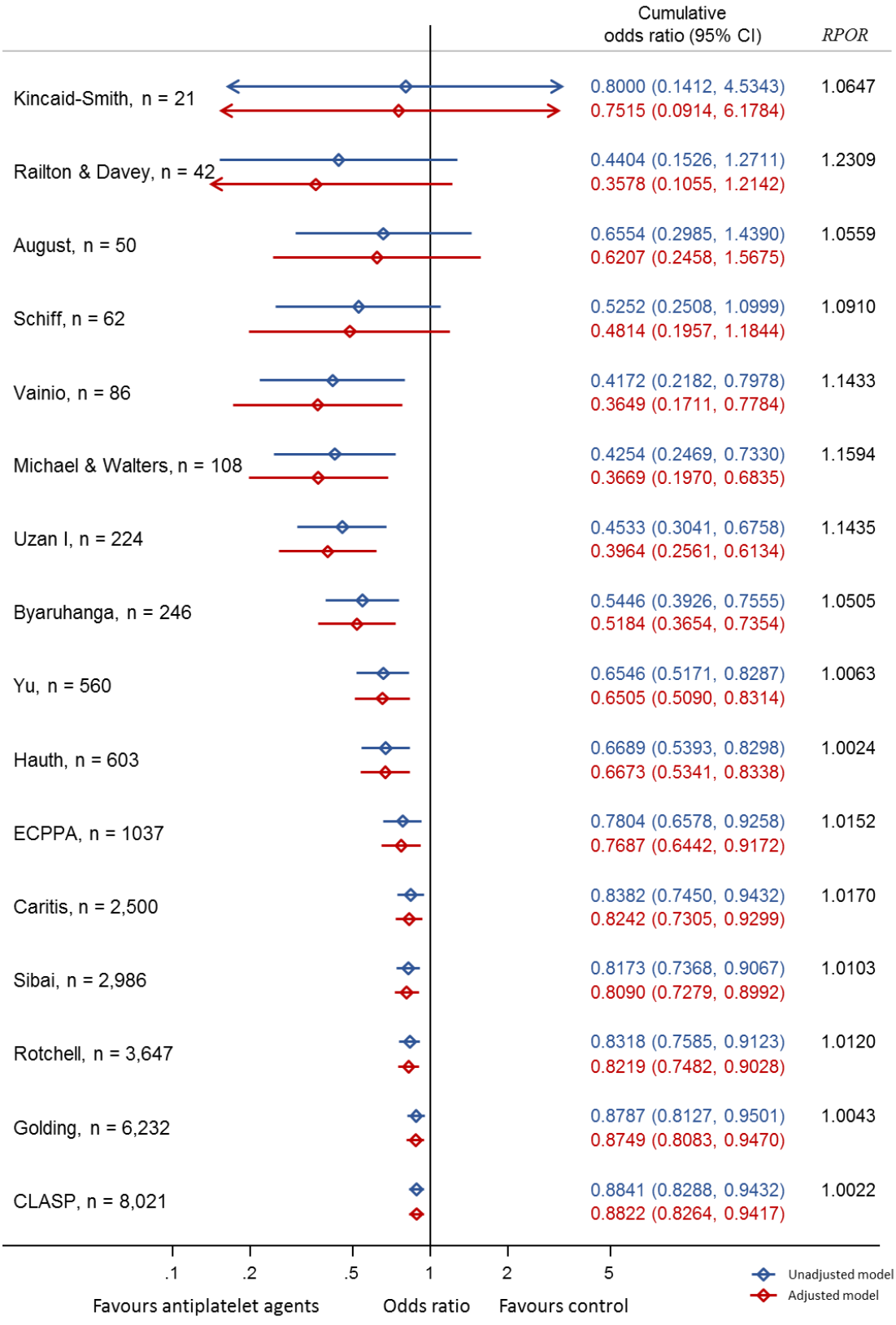
Cumulative meta-analysis was used to explore how $RPOR$ changed when additional studies were added one by one, in order of sample size. As described previously, more extreme treatment effects were observed in studies with small samples. Cumulative meta-analysis showed this phenomenon more clearly. The results of the two-stage fixed effect cumulative meta-analysis are shown in Figure 6-5. Each dot on the plot represents the pooled odds ratio for each study as it was added.

The cumulative pooled odds ratios from the first eight studies (from Kincaid and Smith (1995) to Byaruhanga et al (1998); total sample size = 777) suggest that the effect of the antiplatelet agent significantly reduced the odds of PSAO by 50%, compared with the placebo. This effect size converged to the pooled odds ratio of 0.88 as the larger studies were sequentially added to the meta-analysis. When the meta-analysis included a small number of studies with small samples, the difference between the unadjusted and adjusted treatment effects was magnified. The pooled adjusted odds ratio was 23% lower than the pooled unadjusted odds ratio (i.e., a more beneficial treatment effect was attributed to antiplatelet agents than the placebo) when only two small studies were included in the meta-analysis (pooled unadjusted odds ratio = 0.44 vs pooled adjusted

odds ratio = 0.36; $RPOR$ = 1.23). The difference between the unadjusted and adjusted results diminished as more studies were added.

A cumulative meta-analysis was then conducted using a two-stage random effects model to examine the between-study heterogeneity. The results are presented in Figure 6-6. The $RPOR$ values were noticeably larger than those obtained in the fixed effect cumulative meta-analysis. Early in the cumulative meta-analysis, when few trials had been added, there was an apparent difference in the $RPOR$ values of the fixed effect and random effects cumulative meta-analyses. After five trials had been added to the meta-analyses (from Kincaid-Smith (1995) to Vainio et al (2002)), the fixed effect analysis reported $RPOR$ = 1.14, whereas the random effects analysis reported $RPOR$ = 1.24, a 10% difference between the unadjusted and adjusted results. The random effects model continued to report a larger difference at all points of the analysis, although the magnitude of this difference decreased as more studies were added. Once all of the trials had been added to the analysis, the random effects model reported a final $RPOR$ of 1.0165, 1.4% larger than the fixed effect model's final reported $RPOR$ of 1.0022.

Figure 6-5 Forest plot for a cumulative meta-analysis conducted using the two-stage fixed effect model of the 16 PARIS studies



Between-study heterogeneity estimates were consistently higher in the adjusted analysis than the unadjusted analysis. The values of both I^2 and τ^2 fluctuated widely when only a few small trials were included in the meta-analysis. When only the seven smallest studies were included, the adjusted I^2 was approximately eight times higher than the unadjusted I^2 (I^2 = 34.2% vs 4.4%) and the between-study variance was 14 times higher (τ^2 = 0.224 vs 0.016). This large difference in between-study heterogeneity did not lead to a large difference between the two odds ratios (0.38 vs 0.45, respectively).

The cumulative meta-analysis was then conducted using the one-stage method. The impact on $RPOR$ were less variable than when using the two-stage method. There was less of a difference between the cumulative pooled unadjusted and adjusted odds ratios generated by the one-stage approach than the two-stage approach, and $RPOR$ did not change through the analysis (Figure 6-7 and Figure 6-8). The unadjusted and adjusted treatment effects followed similar patterns to the two-stage results. The between-study variance was not captured in the one-stage results because the values were negligible.

Figure 6-6 Forest plot for a cumulative meta-analysis of the 16 PARIS studies conducted using the two-stage random effects model

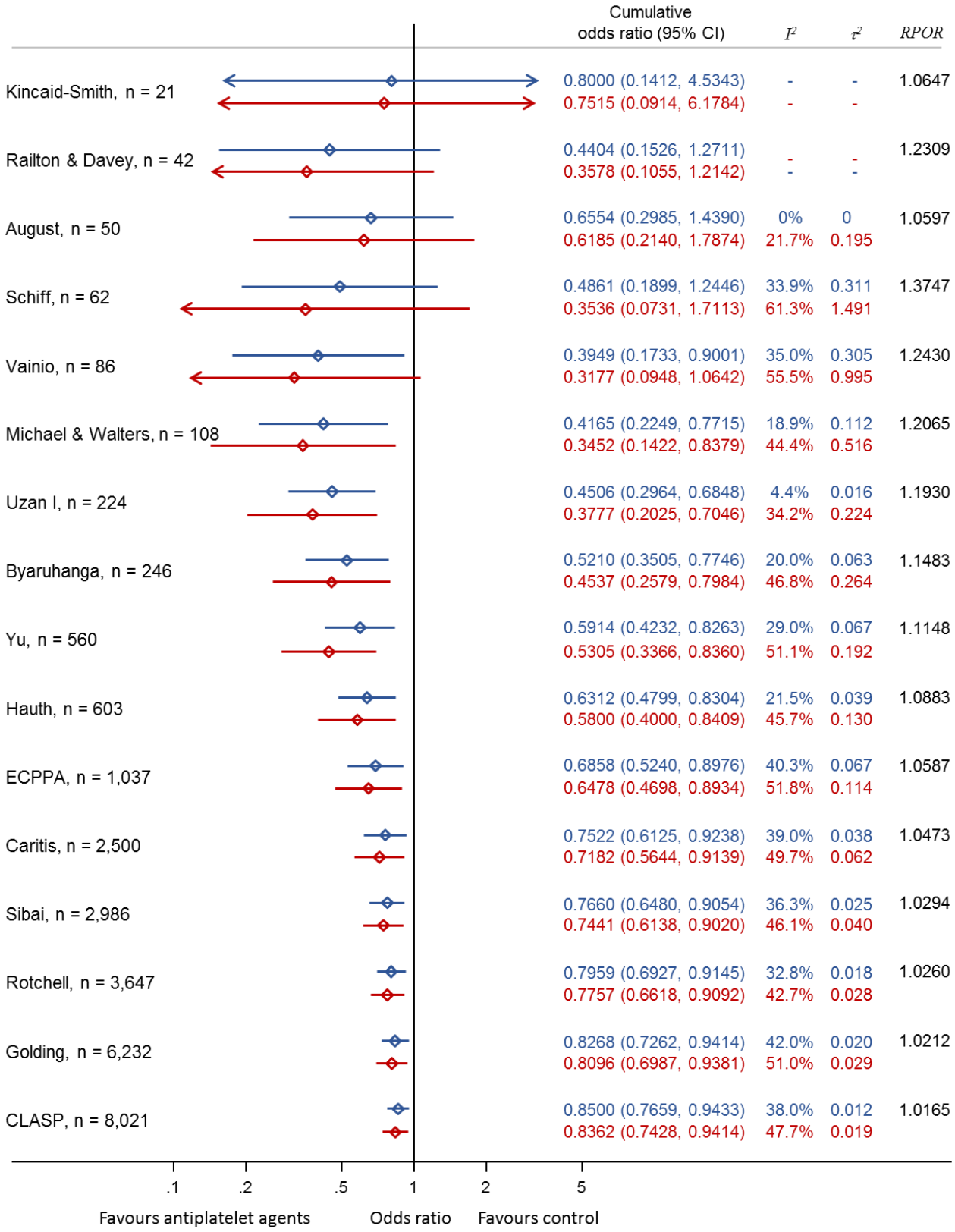


Figure 6-7 Forest plot for a cumulative meta-analysis of the 16 PARIS studies conducted using the one-stage fixed effect model

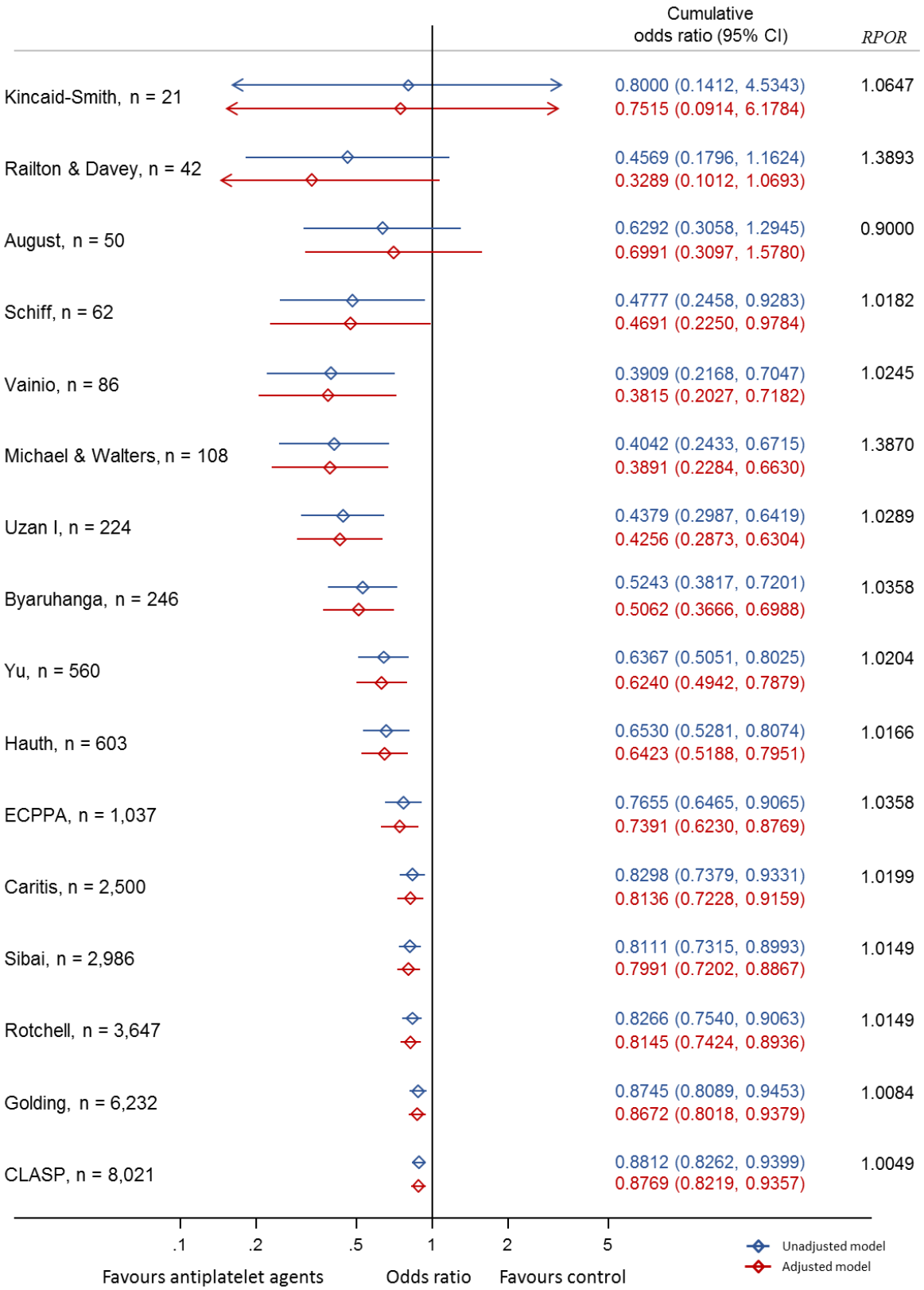
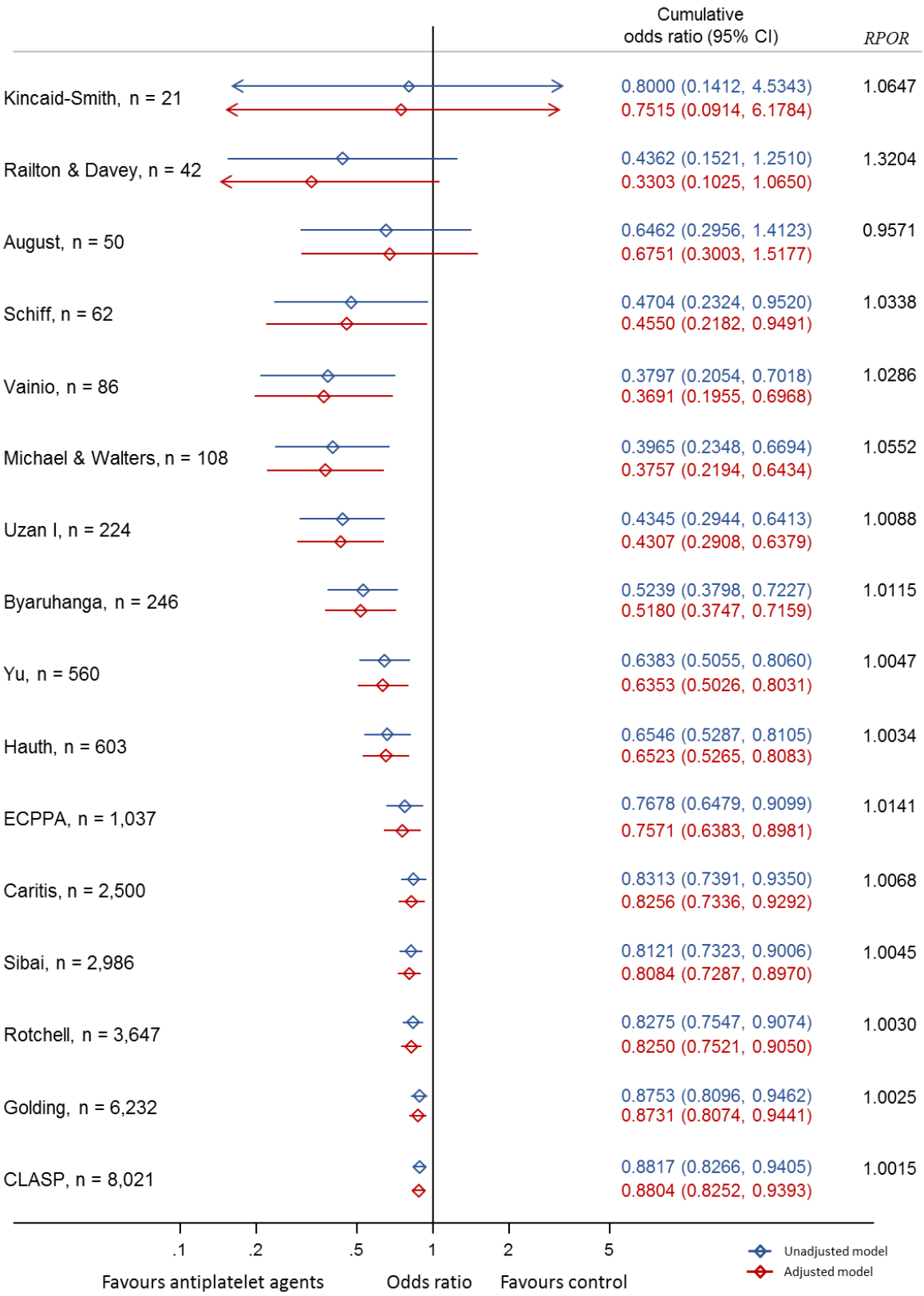


Figure 6-8 Forest plot for a cumulative meta-analysis of the 16 PARIS studies conducted using the one-stage random effects model



6.6 Summary

In this chapter, the impact of covariate adjustment on meta-analysis of RCTs was explored using the PARIS IPD as an illustrative example. The findings from the meta-analysis showed that adjusting for baseline covariates had little effect on the overall treatment effect in the PARIS data. Random effects models tended to increase the between-study variability when adjusted analysis was used, especially when the meta-analysis included a small number of small studies. The PARIS data showed little between-study variability. Therefore the random effects results were similar to the results obtained from the fixed effect model. This result was obtained whether maximum likelihood logistic regression or Firth's method was used.

The one-stage and two-stage approaches gave similar results. However, the pooled results generated by the one-stage approach had less between-study variance and slightly narrower CIs than the results generated by the two-stage method. The two-stage method using random effects gave a slightly higher $RPOR$ and odds ratios that were further away from the null than the one-stage method.

The overall conclusion is robust and consistent with the results of the original PARIS analysis (Askie et al, 2007). Covariate adjustment probably did not greatly affect the treatment effect because the analysis included many studies, most with event rates less than 30%. Additionally, most of the covariates were

moderately prognostic, which implies that adjusted analysis would have a less profound effect.

A cumulative meta-analysis was conducted to show how covariate adjustment affected overall pooled results as studies were added to the meta-analysis in order of sample size. The resulting $RPOR$ values showed that when only a few small studies were included in the meta-analysis, the pooled adjusted results yielded a larger treatment effect than the unadjusted results. The smaller studies yielded consistently higher treatment effects than the larger studies, indicating the possibility of publication bias. The adjusted results may therefore have been further exaggerated.

Many common risk factors were identified in the published literature and guidelines. However, not all were included in this investigation, as data were not available in the included trials. Studies may not have recorded these factors as they were not considered risk factors in earlier studies. Studies may not have included the risk factors that apply to pre-eclampsia and small for gestational age infants, as these outcomes were not the primary outcomes in these studies. Finally, some studies may have collected the missing data and simply not included it in their main publications. However, as these factors were not part of the main PARIS analysis, the PARIS group would not have requested the data. The covariates included in the adjusted analysis were chosen on a pragmatic basis to ensure that the maximum amount of data was included in the analysis.

This is likely to be the approach that most meta-analysts use in practice. The adjusted analysis therefore included a different number of observations from the original PARIS analysis. The results were consistent between the two, most likely due to the large number of studies involved in the review.

Although most of the PARIS studies did not report how they carried out randomisation, covariate imbalance was not severe in any of the studies. This is, however, not necessarily always the case. The effects of the magnitude of imbalance within a trial and the number of trials with unbalanced covariates on the pooled treatment effect is not known.

The two-stage and one-stage approaches in this investigation gave similar results. Stewart et al (2012) also used the PARIS IPD to compare one- and two-stage methods with and without treatment-covariate interactions and reported similar results. Debray et al (2013) compared different one- and two-stage random effects IPD meta-analysis methods when exploring factor-outcome associations. They demonstrated that different methods using different estimation algorithms yielded different pooled results and that the choice of method affected which factors were shown to be associated with the outcome of interest, especially when random effects models were used. Stewart et al (2012) and Debray et al (2013) agreed that a two-stage approach is usually sufficient if a meta-analysis contains many studies, each with a large percentage of events; otherwise the method is less statistically efficient than the one-stage

model. In contrast, the one-stage approach is likely to be more robust if there are few studies, but may encounter convergence issues if the specified model or variance-covariance matrix is too complex.

Debray et al (2013) also advocated that the choice of method be pre-specified. However, in practice, it is hard to know in advance whether a one-stage model will converge. This problem was encountered by Stewart et al (2012), Debray et al (2013), and this investigation.

In summary, these results suggested that covariate adjustment has little impact on the overall results of meta-analysis if the analysis includes many studies. It is likely to have some impact if few studies are included and if these studies have small samples. However, the resulting increase or decrease in the adjusted treatment effect may be confounded by publication bias.

CHAPTER 7: CONCLUSIONS AND FURTHER WORK

7.1 Introduction

Should covariate adjustment be the primary analysis method for the analysis of clinical trial data? Despite the ongoing developments in clinical trials methodology, this question is still hotly debated by the researchers who carry out the day-to-day analysis of randomised clinical trials. This thesis has described a detailed investigation of the effect of using covariate adjustment to analyse individual clinical trials and meta-analyses. This final chapter summarises the key findings and implications of the research presented. The limitations of the study and areas for further research are discussed.

7.2 Summary of research

This study began by assessing and comparing the quality of reporting adjusted analysis before and after the 2001 revision of the CONSORT Statement (Chapter 2). Few studies were found to report carrying out adjusted analyses, and the details were generally poorly reported. Although there has been a slight improvement in reporting since the 2001 revision of the CONSORT Statement, much room for improvement in reporting the results of adjusted analysis remains.

There is still little evidence on the use and reporting of adjusted analysis in meta-analyses of RCTs. Many meta-analysis studies using IPD for a binary outcome

have since been carried out, but most have involved observational studies. Thomas, Radji and Benedetti (2014) recently found that, among 13 IPD meta-analysis studies that included RCTs, only two adjusted for baseline covariates. Their results suggest that covariate adjustment remains uncommon, even in IPD meta-analysis of RCTs. The authors concluded that covariate adjustment is reported poorly and that use of the method is inadequately justified. A recently published guideline, PRISMA-IPD, aims to improve the quality of the reporting of meta-analysis of IPD studies (Stewart et al, 2015). The guideline is suitable for all types of study design, but it does not emphasise the baseline covariate information that is required for adjusted analysis.

The effects of a broad spectrum of controlled scenarios can be examined using simulations to compare unadjusted and adjusted analysis in RCTs. Chapters 3-5 presented simulation studies performed to identify the factors that affect covariate adjusted analysis in different clinical trial scenarios. Simulations were carried out using results that were simultaneously adjusted for two covariates. The ROR was used to compare the adjusted treatment effects with the corresponding results without any covariate adjustment. This metric describes the relative difference between the unadjusted and adjusted odds ratios.

In Chapter 3, population data were generated under different scenarios based on combinations of factors, such as the overall event rate, distributions of the binary and continuous covariates, prognostic strength of the covariates on the outcome, and size and direction of the treatment effects. The data sets were generated under the assumption that the covariates were not correlated with each other and that their distributions were perfectly balanced across the two treatment groups. It was found that if the covariates were not adjusted for, the treatment effects were consistently closer to no treatment effect than the corresponding adjusted treatment effect. Adjusted analysis had a non-linear effect that depended on a combination of the tested factors.

The effect of each factor was tested individually. The distribution of the binary covariate had the smallest effect on the adjusted results. The greatest effect was observed in the difference between a continuous covariate with large and small variance, even with identical prognostic strength. A covariate with large variance had a greater difference between its unadjusted and adjusted standardised odds ratios than a covariate with less variance. The extent of the variability in a continuous covariate can be affected by the type of participants included in an RCT. For example, if a continuous covariate, such as age, is subjected to a narrow inclusion criteria, then the variability of this covariate will be small and its prognostic strength will also be small. The spread of the baseline covariate is thus likely to be associated with the inclusion criteria.

The effect of covariate adjustment was not influenced by the direction of the relationships between the covariates and outcome when both covariates were completely independent. This scenario occurred when both covariates had a positive or negative effect, or a mixture of positive and negative effects, on the outcome.

The largest relative difference found between the unadjusted and adjusted results was around 50%. It occurred when both event rates and the prevalence of the binary covariate were 50% and the adjusted odds ratios of the treatment and the two covariates were very large. However, there was usually little difference between the unadjusted and adjusted results when the covariates had moderate treatment effects or prognostic strength, which corresponds to a typical clinical trial scenario.

The simulation study reported in Chapter 4 found that a correlation between the two covariates affected the adjusted results. The direction and strength of the correlation increased or decreased the difference between the unadjusted and adjusted results. A strong negative correlation between the covariates led to a smaller difference between the unadjusted and adjusted treatment effects than a weak negative correlation between the covariates. A strong positive correlation, in contrast, generated a larger difference between the unadjusted and adjusted results. Again, the extent of the correlation coefficients between the two covariates had little effect on the adjusted analysis when the adjusted

odds ratios of the treatment and covariates were moderately related to the outcome.

Chapter 5 presented a simple approach for quantifying the difference between unadjusted and adjusted treatment effects due to conditioning (i.e., noncollapsibility) and covariate imbalance. The reported resampling study found that study sample size did not affect the difference between the unadjusted and adjusted results, whether the covariates were balanced across the treatment groups or not. However, covariate imbalance at randomisation resulted in bias in the unadjusted results. The magnitude and direction of the bias depended on the extent of the covariate imbalance between the two treatment groups. For example, when the mean of the covariate was higher in the treatment group than in the control group, then the difference between the unadjusted and adjusted treatment effects increased and biases formed towards favouring the treatment group.

Chapter 6 assessed the effect of adjusted analysis using the PARIS (Askie et al, 2005) IPD data set as a real-world example. This data set includes a large number of studies conducted using different approaches. As consistent results were found across the approaches, the pooled treatment effect was unlikely to be affected by adjustments of the baseline covariates. The adjusted results increased the between-study heterogeneity, giving a wider CI. Further assessment using cumulative meta-analysis suggested that adjusted analysis can be detrimental when a meta-analysis includes a small number of small

studies and when the publication bias is present, exaggerating the treatment effect even further.

7.3 Implications for reporting clinical trials

There was no evidence of change in the reporting of adjusted analysis results five years after the 2001 revision of the CONSORT Statement. The overall quality of the reporting of adjusted analysis and adherence to the CONSORT recommendations remained low. Although the unclear reporting of results does not necessarily reflect poor research conduct, there is clear evidence that reporting quality is associated with bias in the estimation of treatment effects (Chan and Altman, 2005a; Pildal et al, 2007).

Researchers should fully report their rationale for covariate adjustment, methods of analysis and choice of covariates for adjustment so that readers can assess whether the adjusted analysis was carried out correctly. This information should be transparently available in trial reports. Trial reports should include unadjusted and adjusted results. They should state which analysis represents the primary analysis and whether the adjusted analysis was pre-specified in the protocol. If researchers report adjusted (or unadjusted) odds ratios of covariates on the outcome and the correlation coefficients between covariates, then meta-analysts can estimate the corresponding unadjusted (or adjusted) treatment effects using the simulation approach described in this thesis. Trial reports seldom report correlation coefficients between covariates when conducting adjusted or unadjusted analysis. This could be added as a

recommendation in a reporting guideline, such as the CONSORT Statement, which will aid in conducting sensitivity analyses in meta-analysis.

7.4 Implications for analysing randomised clinical trials

Many researchers seek guidance on whether adjusted or unadjusted analysis should be the pre-specified primary analysis when preparing their statistical analysis plan. Most recently published studies have focused on the advantage of gaining statistical power when performing adjusted analysis (Hernández, Steyerberg and Habbema, 2004; Tuner et al, 2012; Roozenbeek et al, 2009; Lingsma, Roozenbeek, and Steyerberg, 2010; Gray, Bath and Collier, 2009), particularly when covariate imbalance occurs (Kahan et al, 2014). Studies have focused particularly on clinical trials that fail to reach the original recruitment target. However, covariate imbalance is less common than might However, covariate imbalance is less common than may be expected because most clinical trials today use restricted randomisation methods, such as stratified randomisation and minimisation, to ensure that covariates are balanced across treatment groups (Hewitt and Torgerson, 2006; Kahan and Morris, 2012). When covariates are balanced across treatment groups, typical clinical trials will not gain much efficiency from using adjusted analysis.

Treatment effect estimates from unadjusted and adjusted analyses are not directly comparable; unadjusted analysis gives population-averaged estimates of treatment effect, whereas adjusted analysis assesses subject-specific

estimates. It is therefore important that results are reported clearly so that the treatment effect can be interpreted correctly. This argument is most pertinent for the analysis of RCTs with non-continuous outcomes, because the treatment effect estimate will change if covariates are included in the analysis (Hauck, Anderson and Marcus, 1998).

Increasing study power should therefore not be the sole reason for carrying out adjusted analysis. Researchers should also take into account the use of a restrictive randomisation method. For example, Kahan and Morris (2012) found bias in SEs for treatment effect and a slight reduction in power if adjusting for stratification or minimisation factors was ignored, although they found that this method was uncommon in clinical trials. The meaning of the treatment effect also changes if adjustment is carried out. For example, unadjusted results may be more relevant in a trial that has a broad spectrum of inclusion criteria and is large enough to be generalisable to the general population.

The adjusted analysis method used should be carefully chosen. If the study sample size is small or the event rate is either very high or very low, then exact or penalised methods should be used, such as exact logistic regression or by applying Firth's correction. The rationale, choice of covariates, method of adjustment, and details of reporting should all be explicitly stated in the protocol and statistical analysis plan.

7.5 Implications for meta-analysis of randomised clinical trials

The current Cochrane handbook for systematic reviews of interventions (http://handbook.cochrane.org/) gives little guidance on covariate adjustment in meta-analysis. Should the Cochrane Collaboration continue to use unadjusted results? This is a more complex issue than a standalone RCT, as it is not always possible to carry out a covariate adjustment in a meta-analysis. Researchers considering carrying out covariate adjustment in meta-analysis of RCTs using IPD should identify the prognostic factors for adjustment and determine whether they are available and measured consistently across all of the included studies. These factors and the analysis method should ideally be pre-specified. However, not all covariates are collected for all trials, and trials may have different inclusion and exclusion criteria.

The results in Chapter 6 indicated that covariate adjustment may have little effect on meta-analysis, especially when the analysis includes many studies. Smaller studies have a greater likelihood of publication bias. The results of an adjusted analysis in a meta-analysis of a small number of studies with small sample sizes should therefore be interpreted with caution. Sensitivity analysis can also be used to assess whether adjustment will affect the final results, by setting a maximum increase in the odds ratio had there been any adjustment (Altman et al, 2012), or simply omitting small studies from the analysis.

7.6 Limitations and further work

7.6.1 Adjusted analysis in clinical trials

The simulation methods used here can only be applied to two covariates. Generalising simulation methods to more than two covariates may be too computationally intensive, especially when taking into account correlations between covariates.

This study focused on the difference between unadjusted and adjusted results in different simulated clinical trial scenarios for a binary outcome with two covariates. The results obtained are not yet usable in future meta-analyses. Future work should extend the simulation, using a similar approach to derive the corresponding SE for the adjusted and unadjusted odds ratios. The unadjusted odds ratio could then be estimated from the adjusted odds ratio using the SE (or vice versa), even if IPD are not available. Published papers contain most of the parameter information required for simulation. Estimation could then be carried out using simulation. It would be particularly useful to then develop a web-based estimation tool that can calculate adjusted and unadjusted results and the associated SE from a given set of parameters.

Many studies have evaluated the effect of covariate adjustment in time-to-event data (Gail, 1984; Hernández, Eijkemans and Steyerberg, 2006; Todd and Sahdra, 2001). Some have focused solely on the increase in power when performing adjusted survival analysis. Little attention has been paid to how the

level of censoring or survival model used influence the effect of adjusted analysis. Royston and Parmar (2011) recently used the restricted mean survival time to estimate the treatment effect when the assumption of proportional hazards was violated. Further research is required to determine whether the effect of adjusting a covariate on this new treatment effect measure is similar to the effect of adjusted analysis for a continuous outcome in conventional survival analysis.

Further research is also needed for dealing with small studies that contain too many covariates to adjust for. For example, if a small study has many minimisation variables and it is infeasible to adjust for all of the covariates, then a risk score could be developed for minimisation. The analysis could adjust for this risk score rather than for the covariates, thus reducing the burden of including too many covariates in the analysis. Further work is required to determine whether adjusting for the risk score has the same effect as adjusting for an individual risk factor.

The ratio of odds ratios was used as the sole measure of the difference between unadjusted and adjusted results in this thesis. Other measures to explain the impact of covariate adjustment, such as the use of pseudo-R^2, could also be explored in future work.

7.6.2 Adjusted analysis in meta-analysis of clinical trials

The analysis of the PARIS IPD identified challenges for adjusted analysis in meta-analyses of clinical trials. Meta-analyses are prone to missing data, as a proportion of observations may not be available in a study, or all of the covariates considered in the meta-analysis may be completely missing from a study. Adding covariates to an analysis increases the amount of missing data. Missing data were excluded from the research presented in this thesis. Although the overall conclusion was not affected because most of the missing data were from smaller studies, some covariates were not included in the analysis simply because they were not available.

Missing data will be a continuing problem in meta-analysis. This problem has been addressed elsewhere (The Fibrinogen Studies Collaboration 2009; Burgess et al, 2013; Resch-Rigon et al, 2013; Jolani et al, 2015) in the context of adjusting confounding factors in observational studies. A more effective approach would be to ensure that all important covariates are collected in every study. A consensus is therefore required on the core prognostic factors for all medical conditions. This approach is similar to the Core Outcome Measures in Effectiveness Trials (COMET) Initiative (http://www.comet-initiative.org), which collects core outcome sets for specific conditions that should be reported in all clinical trials, allowing the results of the trials to be compared and combined. Similar databases can be developed for prognostic factors.

Although covariate imbalance leads to bias in the estimation of treatment effects, the degree of imbalance in PARIS was small and is likely to cancel out the bias. Investigation is required to confirm whether this is the case.

As mentioned earlier, adjusted analysis should be carried out if covariates are included in randomisation. This introduces a challenge for meta-analyses, as different covariates may be stratified or minimised in randomisation in each included study. How this variability will affect the pooled treatment effect and the corresponding SE is unknown.

Further research is needed to provide insights into how covariate adjustment can affect meta-analysis results under different scenarios, particularly the presence of large between-study heterogeneity. Simulation methods or similar strategies (such as that used in Chapter 6) should be used to examine this and other real-world examples.

In conclusion, improving the reporting of adjusted analysis will help subsequent meta-analysis if appropriate information is presented. The impact of covariate adjustment is likely to be small in most clinical trials. The impact become apparent only in extreme scenarios. Similarly, meta-analyses containing a large number of studies are not greatly affected by covariate adjustment. This thesis has demonstrated how poor reporting of adjusted analysis can render a study unusable for meta-analysis in clinical decision making. The recommendation presented here have the potential to reduce some of the current waste in

biomedical research, helping to ensure that all well conducted studies can find their place in the decision-making literature. With adequate reporting, it is possible to compute the adjusted or unadjusted treatment effects using simulation approach even only aggregate information is available. This work can also reassure meta-analysts that making adjusted analysis in the studies they work with is unlikely to affect the overall conclusion of the pooled treatment effect, provided those studies do not include extreme scenarios.

APPENDICES

AND

REFERENCES

APPENDIX A: PubMed search strategy for randomised controlled trials published in December 2000

1. Randomized controlled trial [pt]
2. Controlled clinical trial [pt]
3. Randomized controlled trials [mh]
4. Random allocation [mh]
5. Double blind method [mh]
6. Single blind method [mh]
7. Cross-over studies [mh]
8. Multicentre study [pt]
9. #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8
10. #9 NOT (animal [mh] NOT human [mh])
11. Limit 00/12/01 – 00/12/31

PubMed search strategy for randomised controlled trials published in December 2006:

1. Randomized controlled trial [pt]
2. Controlled clinical trial [pt]
3. Randomized controlled trials [mh]
4. Random allocation [mh]
5. Double blind method [mh]
6. Single blind method [mh]
7. Cross-over studies [mh]
8. Multicentre study [pt]
9. #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8
10. #9 NOT (animal [mh] NOT human [mh])
11. Limit 06/12/01 – 06/12/31

pt = publication type ; mh = medical subject headings

APPENDIX B: R code for generating population data sets

a) Function that generates a series of intercept values for different parameters

```
FunctionSimInt_c <- function(btx, bx1, bx2, sx2, pz, sim, sampsize) {
    K1 <- round((sampsize/2)*(1-pz),0)
    K2 <- round((sampsize/2)*(pz),0)
    Tx <- rep(c(0,1), c((sampsize/2), (sampsize/2)))  ## e.g. treatment
    X1 <- rep(c(0,1,0,1), c(K1, K2, K1, K2)) ##  binary covariate
    X2 <- c(rnorm(K1, 0,sx2), rnorm(K2, 0,sx2), rnorm(K1, 0,sx2), rnorm(K2, 0,sx2))
    intercept <- seq(-6, 6, 0.001)
    beta <- cbind(intercept, btx, bx1, bx2)
    con <-1
    X <- cbind(con, Tx, X1, X2)
    linpred <- apply(beta, 1, function(x) X%*%x)
    prob <- apply(linpred, 1, function(x) exp(x)/(1 + exp(x)))
    proportion  <-  replicate(
                n = sim,
                expr = {
                        runis <- matrix(runif(sampsize*length(intercept),0,1),
                        nrow=length(intercept), ncol=sampsize)
                         apply(runis < prob,1,mean)
                 }
    )
    names(proportion) <- "event_rate"
    stepd.result <- cbind(apply(proportion, 1, mean), intercept)
    return(stepd.result)
}
```

b) Function that generates a population data set for a given scenario, perfectly balancing two uncorrelated covariates across two groups

```
FunctionSimMaster <- function(h1,h2, h3, h4, h5, h6, h7, h8, h9, h10, h11) {
    initval <- round(h1, 0)  # set initial value
    eventinit <- round(h2, 0)  # set eventrate initial value
    tx <- round(h3, 3)  # OR(treatment)
    x2 <- round(h4, 3)  # OR(X2)
    Vx2 <- round(h5, 3)  # variance of X2
    x1 <- round(h6, 3)  # OR(X1)
    px1 <- round(h7, 2)  # proportion when X1=1
    intercept <- round(h8,2)  # intercept start value
    event.rate <- round(h9,2)  # overall event rate
    ii <- round(h10,0)  # counter for each scenario
    n <- round(h11,0)  # total sample size
    T1 <- round((n/2)*(1-px1),0)
    T2 <- round((n/2)*(px1),0)
    tol.val <- 0.01
    Sx2 <- round(sqrt(Vx2), 3)
    timer <- 1
    filecount <- 1
    logor.X1 <- round(log(x1), 3)
    logor.Tx <- round(log(tx), 3)
    logor.X2 <- round(log(x2),3)
    k1 <- 0
    k2 <- 0
    Tx <- rep(c(0,1), c((n/2), (n/2)))
    X1 <- rep(c(0,1,0,1), c(T1, T2, T1, T2))
    X2 <- round(c(rnorm(T1, 0,Sx2), rnorm(T2, 0,Sx2), rnorm(T1, 0,Sx2), rnorm(T2, 0,Sx2)), 4)
    set.seed(initval)
    group <- rep(c(1,2,3,4), c(T1, T2, T1, T2))
    m.found <- 0
    con <-1
    X <- cbind(con, Tx, X1, X2)
    beta <- c(intercept, logor.Tx, logor.X1, logor.X2)
    linpred <- X%*%(beta)
    prob <- exp(linpred)/(1 + exp(linpred))
     while (m.found < 4) {
          m.found <- 0
          k1 <- k1+1
          k2 <- k2+1
          set.seed(initval* eventinit + k1)
          Y <- rbinom(length(prob), 1, prob)
          fitall <- glm(Y ~ Tx + X1 + X2, family = binomial)
          e.Tx <- round(summary(fitall)$coefficient[2,1],digits = 3)
          e.X1 <- round(summary(fitall)$coefficient[3,1],digits= 3)
          e.X2 <- round(summary(fitall)$coefficient[4,1], digits = 3)
          if (round(mean(Y),digits=1) == event.rate) {m.found <- m.found+1}
          if (e.X1 >= logor.X1-tol.val & e.X1 <= logor.X1+tol.val) {m.found <- m.found+1}
          if (e.Tx >= logor.Tx-tol.val & e.Tx <= logor.Tx+tol.val) {m.found <- m.found+1}
          if (e.X2 >= logor.X2-tol.val & e.X2 <= logor.X2+tol.val) {m.found <- m.found+1}
          if (k1 >= 1000) {
                eventinit <- eventinit + 9
                k1 <- 1
                tol.val <- tol.val + 0.005
    }
```

```
            flush.console()
            cat("For=",ii,k1, " ", k2, tol.val, intercept, "Tx=", e.Tx , "X1=", e.X1, "X2=", e.X2  , "\n")
        }
        master <- data.frame(group, X,Y)
        names(master)=c("Group", "con", "Tx", "X1", "X2", "Y")
        return(master)
}
```

APPENDIX C: Empirical correlations between covariates in three representative clinical trials

Spine Stabilisation Trial (SST)

	Age	Smoking	Litigation	Owestry disability index at baseline
Smoking	-0.0383			
Litigation	-0.0751	-0.0132		
Owestry disability index at baseline	-0.0794	0.1026	-0.1346	
Clinical state	0.0195	-0.0381	0.0092	0.0957

Arterial Revascularisation Trial (ART)

	Age	Male	Ejection fraction	Diabetes
Male	-0.1111			
Ejection fraction	0.0449	0.0361		
Diabetes	0.0313	-0.0562	0.0326	
Body mass index	-0.0943	-0.0197	0.0489	0.1322

Memantine for dementia in adults with Down's syndrome study (MeaDowS)

	Age	Female	DAMES* score	Dementia status
Female	0.0271			
DAMES* score	-0.3728	0.0637	0.1105	
Dementia status	0.3005	-0.0809	0.0366	-0.2697

* Down's syndrome attention, memory, and executive function scales

APPENDIX D: R code for generating population data sets with specific covariance between two covariates

Function that generates a population data set for a given scenario, perfectly balancing two covariates across two groups with specific covariance between the two covariates

```
FunctionSimMaster <- function(h1,h2, h3, h4, h5, h6, h7, h8, h9, h10, h11, h12) {
        initval <- round(h1, 0)  # set initial value
        eventinit <- round(h2, 0)  # set eventrate initial value
        tx <- round(h3, 3)  # OR(treatment)
        x2 <- round(h4, 3)  # OR(X2)
        vx2 <- round(h5, 3)  # variance of X2
        x1 <- round(h6, 3)  # OR(X1)
        px1 <- round(h7, 2)  # proportion when X1=1
        intercept <- round(h8,2)  # intercept start value
        event.rate <- round(h9,2)  # overall event rate
        ii <- round(h10,0)  # counter for each scenario
        n <- round(h11,0)  # total sample size
        covx1x2 <- round(h12,4)
        vTx <- 0.25
        vx1 <- px1*(1-px1)
        T1 <- round((n/2)*(1-px1),0)
        T2 <- round((n/2)*(px1),0)
        tol.val <- 0.01
        Sx2 <- round(sqrt(vx2), 3)   # standard deviation of X2
        timer <- 1
        filecount <- 1
        logor.X1 <- round(log(x1), 3)
        logor.Tx <- round(log(tx), 3)
        logor.X2 <- round(log(x2),3)
        k1 <- 0
        k2 <- 0
        parms<-c(n/2, T2,0)
        names(parms)<-c("Tx","X1","X2")
        cov.parms<-matrix(0,nr=3,nc=3)
        cov.parms[lower.tri(cov.parms,T)]<-c(vTx, 0.0, 0.0, vx1, covx1x2, vx2)
        temp<-t(cov.parms)
        temp[(lower.tri(cov.parms,T))]<-c(vTx, 0.0, 0.0, vx1, covx1x2, vx2)
        cov.parms<-temp
        dimnames(cov.parms)<-list(c("Tx","X1","X2"),c("Tx","X1","X2"))
        set.seed(initval)
        data1 <-mvrnorm(n,mu=parms,Sigma=cov.parms)
        Tx <- ifelse(data1[,1] < median(data1[,1]), 1, 0)
        X1 <- ifelse(data1[,2] < quantile(data1[,2], 0.2), 1, 0)
        X2 <- data1[,3]
        group <- rep(c(1,2,3,4), c(T1, T2, T1, T2)) ## subgroup indicator for sampling later
```

```
        m.found <- 0
        con <-1
        X <- cbind(con, Tx, X1, X2)
        beta <- c(intercept, logor.Tx, logor.X1, logor.X2)
        linpred <- X%*%(beta)
        prob <- exp(linpred)/(1 + exp(linpred))
         while (m.found < 4) {
                m.found <- 0
                k1 <- k1+1
                k2 <- k2+1
                set.seed(initval* eventinit + k1)
                Y <- rbinom(length(prob), 1, prob)


                fitall <- glm(Y ~ Tx + X1 + X2, family = binomial)
                e.Tx <- round(summary(fitall)$coefficient[2,1],digits = 3)
                e.X1 <- round(summary(fitall)$coefficient[3,1],digits= 3)
                e.X2 <- round(summary(fitall)$coefficient[4,1], digits = 3)
                if (round(mean(Y),digits=1) == event.rate) {m.found <- m.found+1}
                 if (e.X1 >= logor.X1-tol.val & e.X1 <= logor.X1+tol.val) {m.found <- m.found+1}
                if (e.Tx >= logor.Tx-tol.val & e.Tx <= logor.Tx+tol.val) {m.found <- m.found+1}
                if (e.X2 >= logor.X2-tol.val & e.X2 <= logor.X2+tol.val) {m.found <- m.found+1}
                if (k1 >= 1000) {
                        eventinit <- eventinit + 9
                        k1 <- 1
                        tol.val <- tol.val + 0.005
                }

                flush.console()
                cat("For=",ii,k1, " ", k2, tol.val, intercept, "Tx=", e.Tx , "X1=", e.X1, "X2=", e.X2  ,
        "\n")

        }
        master <- data.frame(group, X,Y)
        names(master)=c("Group", "con", "Tx", "X1", "X2", "Y")
        return(master)

}
```

APPENDIX E: Odds ratios and ratios of odds ratios at varying levels of correlation at given scenarios

a) $OR_{T\_a} = 5$, $\boldsymbol{OR}_{X_2\_a} = 2$, $\sigma 2 = 1$, $\boldsymbol{OR}_{X_1\_a} = 0.2$ and 5, and $\boldsymbol{P}_{X_1} = 20\%$

Event rate (R)	OR / ROR	$OR_{X_1\_a}$ = 0.2 Correlation between X_1 and X_2 (ρ)					$OR_{X_1\_a}$ = 5 Correlation between X_1 and X_2 (ρ)				
		-0.3	-0.1	0	0.1	0.3	-0.3	-0.1	0	0.1	0.3
10%	OR_{T_u}	4.5172	4.5654	4.6099	4.6446	4.6838	4.4540	4.4123	4.3527	4.2952	4.0422
	OR_{T_a}	4.9818	4.9779	5.0204	5.0254	4.9555	4.9525	4.9709	4.9888	5.0354	5.0249
	ROR	0.9067	0.9171	0.9182	0.9242	0.9450	0.8993	0.8876	0.8725	0.8530	0.8044
30%	OR_{T_u}	3.9944	4.0776	4.0927	4.1791	4.3103	4.2043	3.9793	3.8698	3.8594	3.5265
	OR_{T_a}	4.9799	5.0013	4.9764	5.0173	5.0369	4.9878	4.9818	5.0038	5.0264	4.9674
	ROR	0.8021	0.8153	0.8224	0.8329	0.8557	0.8429	0.7988	0.7734	0.7678	0.7099
50%	OR_{T_u}	3.6969	3.7958	3.8997	3.9888	4.1239	4.1425	3.9773	3.8880	3.8301	3.6645
	OR_{T_a}	4.9838	5.0329	5.0153	5.0138	5.0309	4.9853	4.9848	4.9928	5.0078	4.9774
	ROR	0.7418	0.7542	0.7776	0.7956	0.8197	0.8309	0.7979	0.7787	0.7648	0.7362
70%	OR_{T_u}	3.5999	3.7792	3.8783	4.0052	4.2093	4.2182	4.0919	4.0812	4.0297	3.9825
	OR_{T_a}	5.0269	4.9654	4.9570	4.9813	4.9978	4.9699	5.0465	4.9813	4.9689	4.9938
	ROR	0.7161	0.7611	0.7824	0.8040	0.8422	0.8487	0.8108	0.8193	0.8110	0.7975
90%	OR_{T_u}	3.9944	4.2279	4.2862	4.4145	4.5947	4.5631	4.6103	4.5581	4.5064	4.5809
	OR_{T_a}	5.0440	4.9739	4.9590	5.0295	5.0033	4.9471	5.0501	4.9714	4.9843	5.0193
	ROR	0.7919	0.8500	0.8643	0.8777	0.9183	0.9224	0.9129	0.9169	0.9041	0.9127

b) OR_{T_a} = 5, $\boldsymbol{OR}_{X_2\_a}$ = 2, $\sigma 2$ = 1, $\boldsymbol{OR}_{X_1\_a}$ = 0.2 and 5, and $\boldsymbol{P}_{X_1}$ = 50%

Event rate (R)	OR / ROR	$OR_{X_1\_a}$ = 0.2 Correlation between X_1 and X_2 (ρ)					$OR_{X_1\_a}$ = 5 Correlation between X_1 and X_2 (ρ)				
		-0.3	-0.1	0	0.1	0.3	-0.3	-0.1	0	0.1	0.3
10%	OR_{T_u}	2.7462	2.8202	2.9919	3.0165	3.0325	3.1393	3.0093	2.9087	2.7957	2.7256
	OR_{T_a}	4.9823	4.9838	4.9933	5.0224	5.0274	4.9774	4.9729	4.9570	4.9590	4.9843
	ROR	0.5512	0.5659	0.5992	0.6006	0.6032	0.6307	0.6051	0.5868	0.5638	0.5468
30%	OR_{T_u}	2.3333	2.4407	2.5325	2.6235	2.6851	2.7112	2.6174	2.4068	2.4591	2.3535
	OR_{T_a}	4.9993	4.9759	4.9968	5.0314	5.0224	5.0178	5.0440	4.9620	4.9843	5.0254
	ROR	0.4667	0.4905	0.5068	0.5214	0.5346	0.5403	0.5189	0.4851	0.4934	0.4683
50%	OR_{T_u}	2.2712	2.4061	2.4596	2.4841	2.5904	2.5459	2.4851	2.3790	2.3315	2.2728
	OR_{T_a}	4.9794	5.0309	5.0239	5.0334	4.9833	5.0480	5.0163	5.0173	4.9585	4.9674
	ROR	0.4561	0.4783	0.4896	0.4935	0.5198	0.5043	0.4954	0.4742	0.4702	0.4575
70%	OR_{T_u}	2.3658	2.4261	2.5259	2.5738	2.6182	2.7349	2.5687	2.4920	2.4405	2.3041
	OR_{T_a}	5.0254	5.0173	4.9883	5.0183	4.9629	5.0420	5.0098	4.9888	4.9963	4.9565
	ROR	0.4708	0.4836	0.5064	0.5129	0.5276	0.5424	0.5127	0.4995	0.4885	0.4649
90%	OR_{T_u}	2.8185	2.9388	2.9583	2.9695	3.1639	3.1629	2.8855	2.8968	2.8383	2.7475
	OR_{T_a}	5.0143	4.9684	5.0053	4.9639	5.0415	4.9818	4.9575	5.0108	4.9575	4.9535
	ROR	0.5621	0.5915	0.5910	0.5982	0.6276	0.6349	0.5820	0.5781	0.5725	0.5547

c) $OR_{T\_a} = 5$, $\boldsymbol{OR_{X_2\_a}} = 2$, $\sigma 2 = 1$, $\boldsymbol{OR_{X_1\_a}} = 0.2$ and 5, and $\boldsymbol{P_{X_1}} = 20\%$

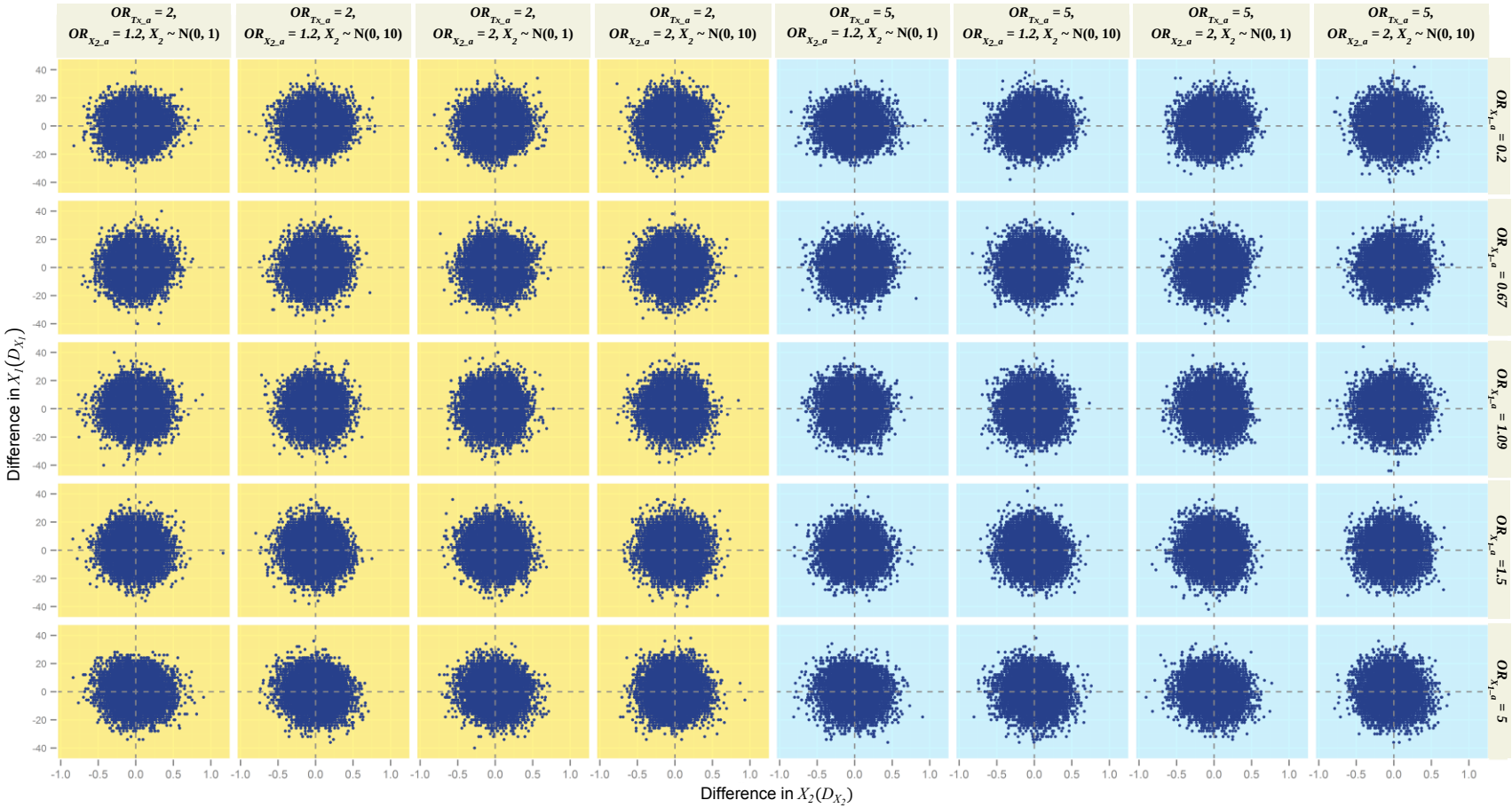
Event rate (R)	OR / ROR	$OR_{X_1\_a}$ = 0.2 Correlation between X_1 and X_2 (ρ)					$OR_{X_1\_a}$ = 5 Correlation between X_1 and X_2 (ρ)				
		-0.3	-0.1	0	0.1	0.3	-0.3	-0.1	0	0.1	0.3
10%	OR_{T_u}	2.9689	3.0033	3.0120	2.9151	3.1003	3.1903	3.0496	2.9728	2.7768	2.7191
	OR_{T_a}	4.9828	5.0279	5.0143	4.9799	5.0299	4.9724	5.0385	5.0078	5.0269	5.0249
	ROR	0.5958	0.5973	0.6007	0.5854	0.6164	0.6416	0.6053	0.5936	0.5524	0.5411
30%	OR_{T_u}	2.5115	2.5088	2.5495	2.5607	2.7333	2.7525	2.6337	2.5236	2.4818	2.3636
	OR_{T_a}	4.9863	4.9580	4.9729	4.9530	5.0158	4.9654	5.0374	4.9555	4.9923	4.9739
	ROR	0.5037	0.5060	0.5127	0.5170	0.5449	0.5543	0.5228	0.5093	0.4971	0.4752
50%	OR_{T_u}	2.3955	2.4305	2.4128	2.4269	2.6138	2.6060	2.5302	2.4359	2.4513	2.3523
	OR_{T_a}	5.0415	5.0415	4.9908	4.9853	4.9883	4.9511	5.0455	4.9808	4.9466	4.9978
	ROR	0.4752	0.4821	0.4835	0.4868	0.5240	0.5263	0.5015	0.4890	0.4955	0.4707
70%	OR_{T_u}	2.3662	2.4618	2.5153	2.5997	2.7167	2.6203	2.6240	2.5348	2.5523	2.4998
	OR_{T_a}	4.9843	5.0083	5.0395	5.0083	5.0043	4.9694	4.9501	5.0460	5.0501	5.0465
	ROR	0.4747	0.4915	0.4991	0.5191	0.5429	0.5273	0.5301	0.5023	0.5054	0.4953
90%	OR_{T_u}	2.6581	2.7977	2.9737	2.9982	3.2018	3.2356	3.0734	2.9654	2.9651	2.9603
	OR_{T_a}	4.9491	5.0198	5.0475	4.9963	5.0078	5.0460	5.0249	5.0073	4.9699	5.0506
	ROR	0.5371	0.5573	0.5891	0.6001	0.6394	0.6412	0.6116	0.5922	0.5966	0.5861

APPENDIX F: R code that samples data from a population data set for a given scenario

```
FunctionSimAnalysis_c <- function(iter= 1, data, h1,h2, h3, h4, h5, h6, h7, h8, h9, loop, ssize) {
        initval <- round(h1, 0)  # set initial value
        eventinit <- round(h2, 0)  # set eventrate initial value
        tx <- round(h3, 3)  # OR(treatment)
        x2 <- round(h4, 3)  # OR(X2)
        Vx2 <- round(h5, 3)  # variance of X2
        x1 <- round(h6, 3)  # OR(X1)
        px1 <- round(h7, 2)  # proportion when X1=1
        timer <- h9
        event.rate <- round(h8,2)  # overall event rate
         n1 <- round(ssize, 0)
         sim <- loop
        s   <- strata(master, stratanames = c("Tx"), size = c(n1/2, n1/2), method = "srswr")
        sY <- sample2$Y
          sample2$Y <- factor(sample2$Y, levels = c(0, 1))
          sample2$Tx <- factor(sample2$Tx, levels = c(0,1))
          sample2$X1 <- factor(sample2$X1, levels = c(0,1))
          desX2.Tx        <- describeBy(sample2$X2, sample2$Tx, mat = TRUE)
          desX2.y         <- describeBy(sample2$X2, sY,  mat = TRUE)
          fitTx           <- summary(glm(Y ~ Tx, data = sample2, family = binomial))
          fitX1           <- summary(glm(Y ~ X1, data = sample2, family = binomial))
          fitX2           <- summary(glm(Y ~ X2, data = sample2, family = binomial))
          fitTxX1         <- summary(glm(Y ~ Tx + X1, data = sample2, family = binomial))
          fitTxX2         <- summary(glm(Y ~ Tx + X2, data = sample2, family = binomial))
          fitall          <- summary(glm(Y ~ Tx + X1 + X2, data = sample2, family = binomial))
         return(list(n1    = n1,
            eventTx1 = table(sample2$Y, sample2$Tx)[2, 2],
            eventTx0 = table(sample2$Y, sample2$Tx)[2, 1],
            event1X1 = table(sample2$Y, sample2$X1)[2, 2],
            event0X1 = table(sample2$Y, sample2$X1)[1, 2],
            ptotal = round(mean(sY), 4),
            meanX2 = round(mean(sample2$X2), 4),
            sdX2 = round(sd(sample2$X2), 4),
            nX1Tx1 = table(sample2$X1, sample2$Tx)[2, 2],
            nX1Tx0 = table(sample2$X1, sample2$Tx)[2, 1],
            meanX2Tx1 = round(desX2.Tx[2, 5], 4),
            sdX2Tx1 =  round(desX2.Tx[2, 6], 4),
            meanX2Tx0 = round(desX2.Tx[1, 5], 4),
            sdX2Tx0 = round(desX2.Tx[1, 6], 4),
            meanX2Y1 = round(desX2.y[2, 5], 4),
            sdX2Y1 = round(desX2.y[2, 6], 4),
            meanX2Y0 = round(desX2.y[1, 5], 4),
            sdX2Y0 = round(desX2.y[1, 6], 4) ,
            coeffTx = round(fitTx$coef[2, 1],4),
            seTx = round(sqrt(fitTx$coef[2, 2]),4),
            coeffX1 = round(fitX1$coef[2, 1],4),
            seX1 =round(sqrt(fitX1$coef[2, 2]),4),
             coeffX2 =round(fitX2$coef[2, 1],4),
             seX2 = round(sqrt(fitX2$coef[2, 2]),4),
             coeffTxa = round(fitall$coef[2, 1],4),
             seTxa = round(sqrt(fitall$coef[2, 2]),4),
```

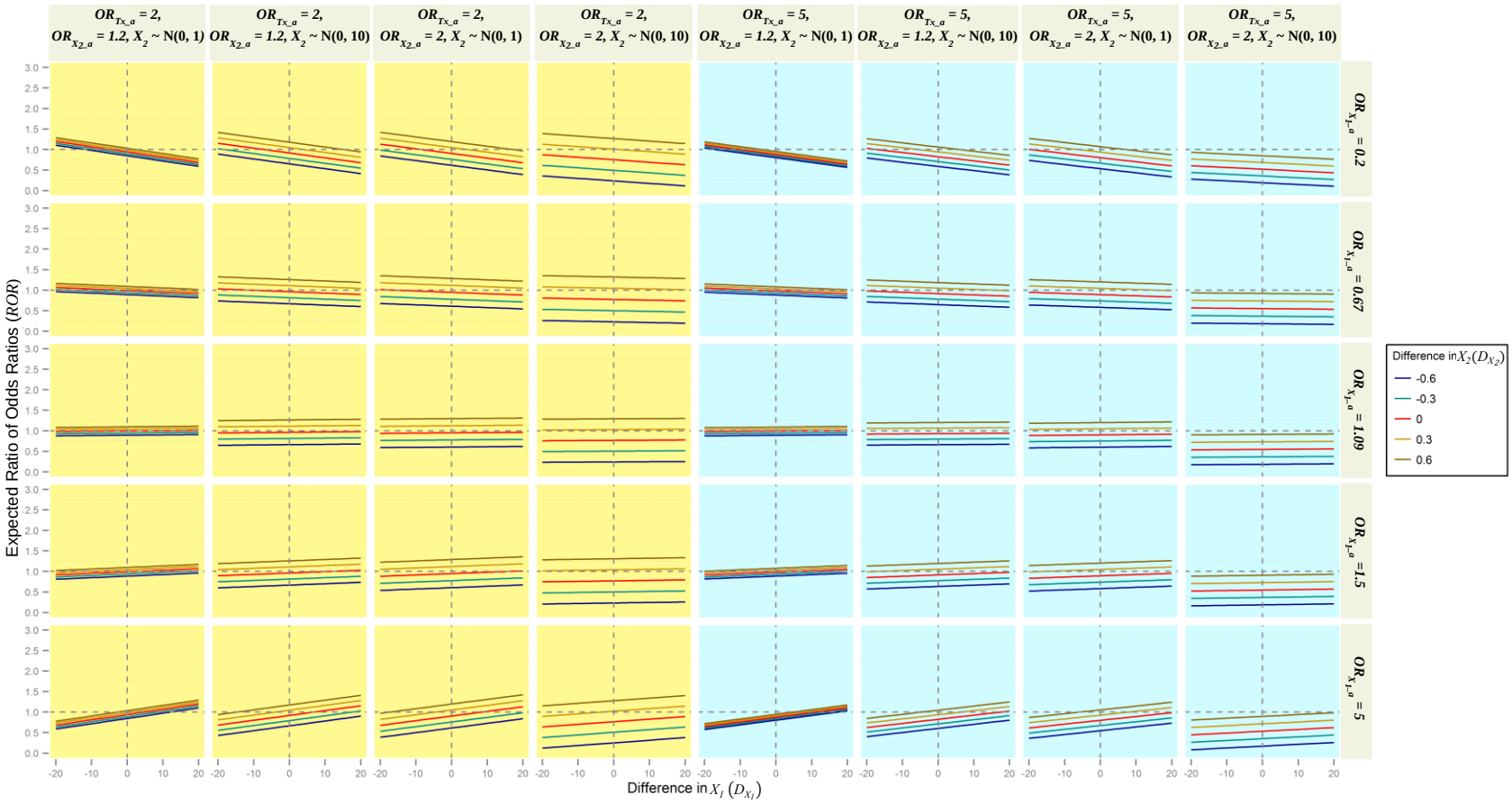
```
                coeffX1a = round(fitall$coef[3, 1],4),
                seX1a = round(sqrt(fitall$coef[3, 2]),4),
                coeffX2a = round(fitall$coef[4, 1] ,4),
                seX2a = round(sqrt(fitall$coef[4, 2]),4),
                coeffTxaX1 = round(fitTxX1$coef[2, 1],4),
                seTxaX1 = round(sqrt(fitTxX1$coef[2, 2]) ,4),
                CoeffTxX1a = round(fitTxX1$coef[3, 1],4),
                seTxX1a = round(sqrt(fitTxX1$coef[3, 2]) ,4),
                coeffTxaX2 = round(fitTxX2$coef[2, 1],4),
                seTxaX2 = round(sqrt(fitTxX2$coef[2, 2]) ,4),
                CoeffTxX2a = round(fitTxX2$coef[3, 1],4),
                seTxX2a =  round(sqrt(fitTxX2$coef[3, 2]) ,4),
                 diffTxa = round((fitall$coef[2,1]- fitTx$coef[2, 1]) ,4),
                diffTxX1 = round((fitTxX1$coef[2,1]- fitTx$coef[2, 1] ),4),
                diffTxX2 = round((fitTxX2$coef[2,1]- fitTx$coef[2, 1] ),4),
                diffcomb = round(((fitTxX2$coef[2,1]- fitTx$coef[2, 1]) +(fitTxX1$coef[2,1]- fitTx$coef[2, 1])) ,4),
                Px1 = px1,
                Totrate = event.rate,
                betaTx = round(log(tx), 3),
                betaX1 = round(log(x1), 3),
                betaX2 = round(log(x2),3)))


}
```

APPENDIX G: Imbalance distribution under each scenario

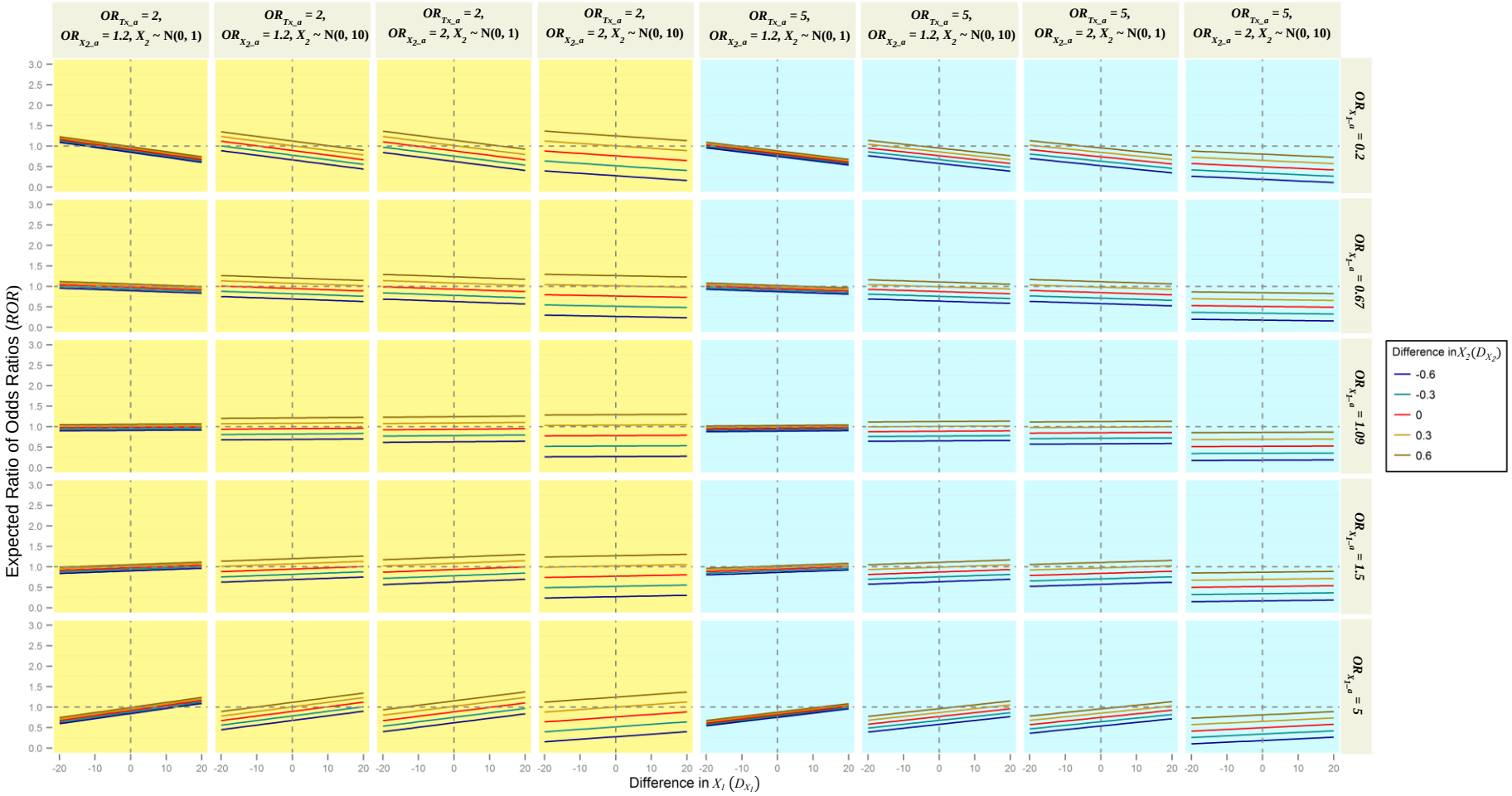


APPENDIX H: Regression line of the expected ratio of unadjusted and adjusted odds ratios against the difference in X_1 by selected values of the difference in X_2 for all scenarios

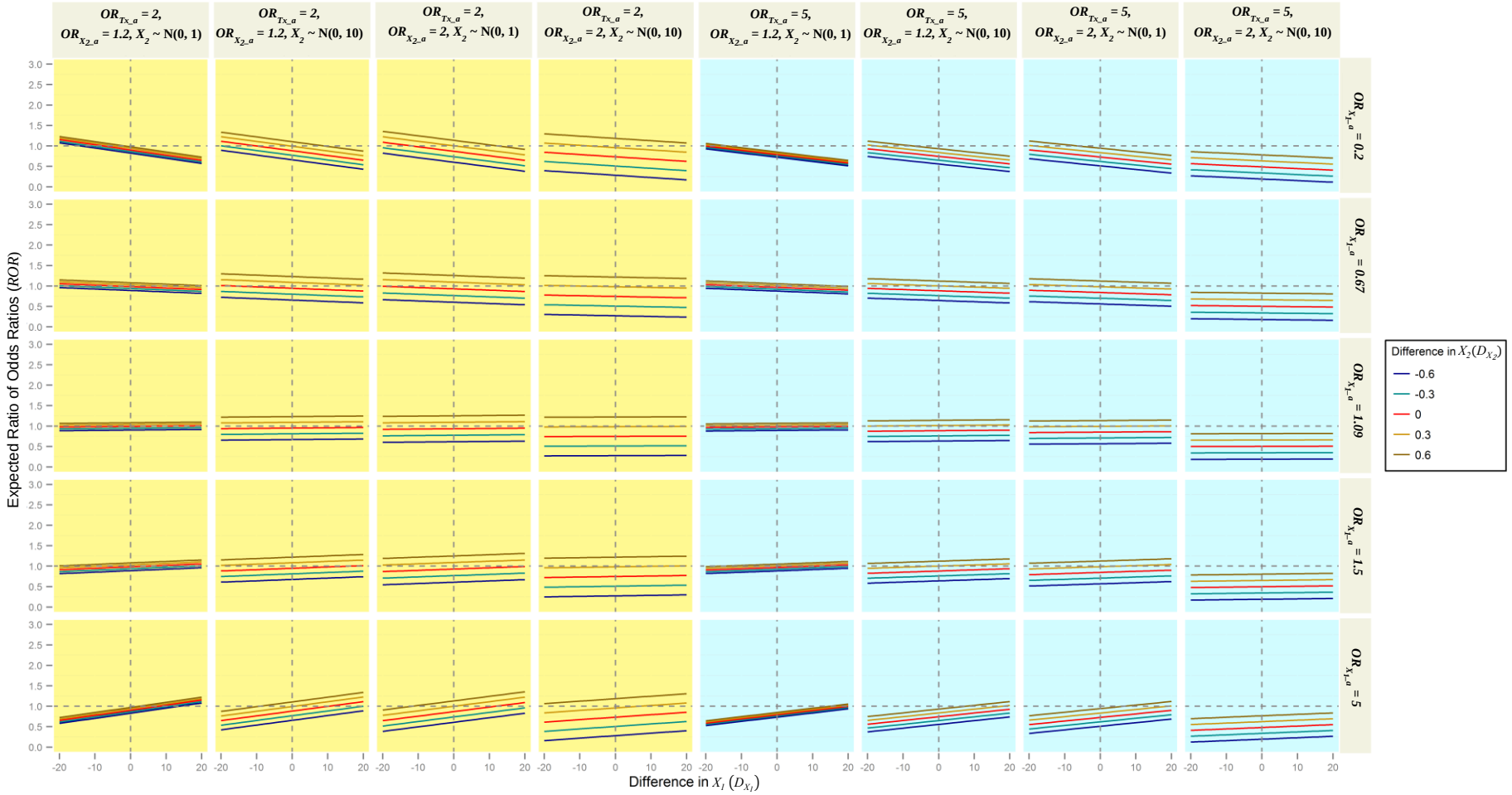
a) Event rate = 20%



b) Event rate = 70%



c) Sample size = 500



APPENDIX J: PARIS studies included in the analysis and their sample sizes

Trial	Original sample size	Sample size (%) included in the analysis	Total number (%) of pregnancies with a serious adverse outcome
Kincaid-Smith et al (1995)	21	21 (100%)	9 (43%)
Zimmerman et al (1997)	26		
Railton and Davey (1998)	44	42 (95.4%)	19 (45%)
August et al (1994)	55	50 (90.9%)	17 (34%)
Ferrier et al (1996)	56		
Schiff et al (1989)	62	62 (100%)	8 (13%)
Wang and Li (1996)	84		
Vainio et al (2002)	90	86 (95.6%)	15 (17%)
Morris et al (1996)	102		
Michael and Walters (1993)	110	108 (98.2%)	21 (19.4%)
Rogers et al (1999)	193		
Uzan I: Uzan et al (1989)	230	224 (97.4%)	73 (32.6%)
Byaruhanga et al (1998)	256	246 (96.1%)	64 (26.0%)
Uzan II: Uzan et al (1991)	315		
Hermida et al (2003)	341		
Yu et al (2003)	560	560 (100%)	219 (39.1%)
Hauth et al (1993)	606	603 (99.5%)	62 (10.3%)
ECPPA (1996)	1,091	1,037 (95.0%)	261 (25.2%)
Caritis et al (1998)	2,539	2,500 (98.5%)	908 (36.3%)
Sibai et al (1993)	3,135	2,986 (95.2%)	375 (12.6%)
ERASME: Subtil et al (2003)	3,294		
Rotchell et al (1998)	3,647	3,647 (100%)	424 (11.6%)
Golding (1998)	6,275	6,232 (99.3%)	825 (13.2%)
CLASP (1994)	8,021	8,021 (100%)	1413 (17.6%)
Total	31,153	26,425 (84.8%)	

APPENDIX K: Prognostic strength of the four covariates on pregnancy with a serious adverse outcome in each PARIS study

Trial: Kincaid-Smith

Baseline variable	Coeff (SE)	OR (SE)	P-value
Treatment	-0.2231 (0.8851)	0.8 (0.7081)	0.8009
Systolic blood pressure	0.0532 (0.0487)	1.0546 (0.0514)	0.2746
Maternal age	0.0249 (0.108)	1.0252 (0.1107)	0.8176
Previous pregnancy	-16.9715 (2399.5448)	0 (0)	0.9944
Chronic hypertension	0.9163 (1.0488)	2.5 (2.622)	0.3823

Trial: Railton and Davey

Baseline variable	Coeff (SE)	OR (SE)	P-value
Treatment	-1.1756 (0.6831)	0.3086 (0.2108)	0.0853
Systolic blood pressure	-0.0163 (0.0198)	0.9838 (0.0195)	0.4104
Maternal age	-0.049 (0.0473)	0.9522 (0.045)	0.3002
Previous pregnancy	-0.5754 (0.8367)	0.5625 (0.4706)	0.4916
Chronic hypertension	0.3365 (0.6274)	1.4 (0.8784)	0.5918

Trial: August

Baseline variable	Coeff (SE)	OR (SE)	P-value
Treatment	0.0645 (0.5986)	1.0666 (0.6385)	0.9141
Systolic blood pressure	0.0359 (0.0218)	1.0366 (0.0226)	0.0997
Maternal age	-0.1112 (0.0531)	0.8948 (0.0475)	0.0365
Previous pregnancy	0.7916 (1.1606)	2.2069 (2.5613)	0.4952
Chronic hypertension	-0.1823 (0.8003)	0.8334 (0.667)	0.8198

Trial: Schiff

Baseline variable	Coeff (SE)	OR (SE)	P-value
Treatment	-2.3206 (1.1043)	0.0982 (0.1084)	0.0356
Systolic blood pressure	-0.0109 (0.0361)	0.9892 (0.0357)	0.7628
Maternal age	-0.1623 (0.092)	0.8502 (0.0782)	0.0776
Previous pregnancy	2.0794 (0.8292)	7.9997 (6.6334)	0.0121
Chronic hypertension	-	-	-

Trial: Vainio

Baseline variable	Coeff (SE)	OR (SE)	P-value
Treatment	-1.6412 (0.6884)	0.1937 (0.1333)	0.0171
Systolic blood pressure	0.0139 (0.0178)	1.014 (0.018)	0.4335
Maternal age	0.0035 (0.0471)	1.0035 (0.0473)	0.9403
Previous pregnancy	0.5855 (0.6946)	1.7959 (1.2474)	0.3993
Chronic hypertension	0.6672 (0.5777)	1.9488 (1.1258)	0.2481

Trial: Michael and Walters

Baseline variable	Coeff (SE)	OR (SE)	P-value
Treatment	-0.8082 (0.5103)	0.4457 (0.2274)	0.1132
Systolic blood pressure	-0.0451 (0.0173)	0.9559 (0.0165)	0.0092
Maternal age	0.0162 (0.0399)	1.0163 (0.0406)	0.6858
Previous pregnancy	0.448 (0.6774)	1.5652 (1.0603)	0.5084
Chronic hypertension	0.4463 (0.4902)	1.5625 (0.7659)	0.3626

Trial: Uzan I

Baseline variable	Coeff (SE)	OR (SE)	P-value
Treatment	-0.717 (0.3001)	0.4882 (0.1465)	0.0169
Systolic blood pressure	0.0184 (0.0096)	1.0186 (0.0098)	0.0555
Maternal age	0.0492 (0.0329)	1.0504 (0.0346)	0.1344
Previous pregnancy	-	-	-
Chronic hypertension	0.2483 (0.2911)	1.2818 (0.3731)	0.3937

Trial: Byaruhanga

Baseline variable	Coeff (SE)	OR (SE)	P-value
Treatment	-0.232 (0.2916)	0.7929 (0.2312)	0.4263
Systolic blood pressure	0.0106 (0.0069)	1.0107 (0.007)	0.1238
Maternal age	0.0812 (0.0286)	1.0846 (0.031)	0.0046
Previous pregnancy	0.6902 (0.7829)	1.9941 (1.5612)	0.378
Chronic hypertension	0.371 (0.3859)	1.4492 (0.5592)	0.3363

Trial: Yu

Baseline variable	Coeff (SE)	OR (SE)	P-value
Treatment	-0.2252 (0.1735)	0.7984 (0.1385)	0.1942
Systolic blood pressure	0.0067 (0.0096)	1.0067 (0.0097)	0.4849
Maternal age	0.0286 (0.0138)	1.029 (0.0142)	0.0388
Previous pregnancy	-0.2818 (0.1757)	0.7544 (0.1325)	0.1087
Chronic hypertension	-	-	-

Trial: Hauth

Baseline variable	Coeff (SE)	OR (SE)	P-value
Treatment	-0.2928 (0.2702)	0.7462 (0.2016)	0.2785
Systolic blood pressure	0.0135 (0.0174)	1.0136 (0.0176)	0.4386
Maternal age	-0.0355 (0.0515)	0.9651 (0.0497)	0.4913
Previous pregnancy	-	-	-
Chronic hypertension	-	-	-

Trial: ECPPA

Baseline variable	Coeff (SE)	OR (SE)	P-value
Treatment	0.0131 (0.1431)	1.0132 (0.145)	0.9271
Systolic blood pressure	0.0268 (0.0037)	1.0272 (0.0038)	0
Maternal age	0.0185 (0.0098)	1.0187 (0.01)	0.0589
Previous pregnancy	0.3947 (0.1448)	1.4839 (0.2149)	0.0064
Chronic hypertension	0.6391 (0.1449)	1.8948 (0.2746)	0

Trial: Caritis

Baseline variable	Coeff (SE)	OR (SE)	P-value
Treatment	-0.1113 (0.0832)	0.8947 (0.0744)	0.1809
Systolic blood pressure	0.0154 (0.0028)	1.0155 (0.0028)	0
Maternal age	0.0098 (0.0066)	1.0098 (0.0067)	0.1379
Previous pregnancy	-0.4775 (0.1042)	0.6203 (0.0646)	0
Chronic hypertension	0.0853 (0.0883)	1.089 (0.0962)	0.3339

Trial: Sibai

Baseline variable	Coeff (SE)	OR (SE)	P-value
Treatment	-0.2881 (0.1113)	0.7497 (0.0834)	0.0096
Systolic blood pressure	0.0176 (0.0049)	1.0178 (0.005)	0.0003
Maternal age	0.0197 (0.0115)	1.0199 (0.0117)	0.0869
Previous pregnancy	0.0286 (0.1289)	1.029 (0.1326)	0.8247
Chronic hypertension	-	-	-

Trial: Golding

Baseline variable	Coeff (SE)	OR (SE)	P-value
Treatment	0.0088 (0.0748)	1.0088 (0.0755)	0.9064
Systolic blood pressure	0.0197 (0.003)	1.0199 (0.0031)	0
Maternal age	0.0206 (0.0086)	1.0208 (0.0088)	0.0159
Previous pregnancy	0.229 (0.1051)	1.2573 (0.1321)	0.0294
Chronic hypertension	0.7848 (0.5178)	2.192 (1.135)	0.1296

Trial: Rotchell

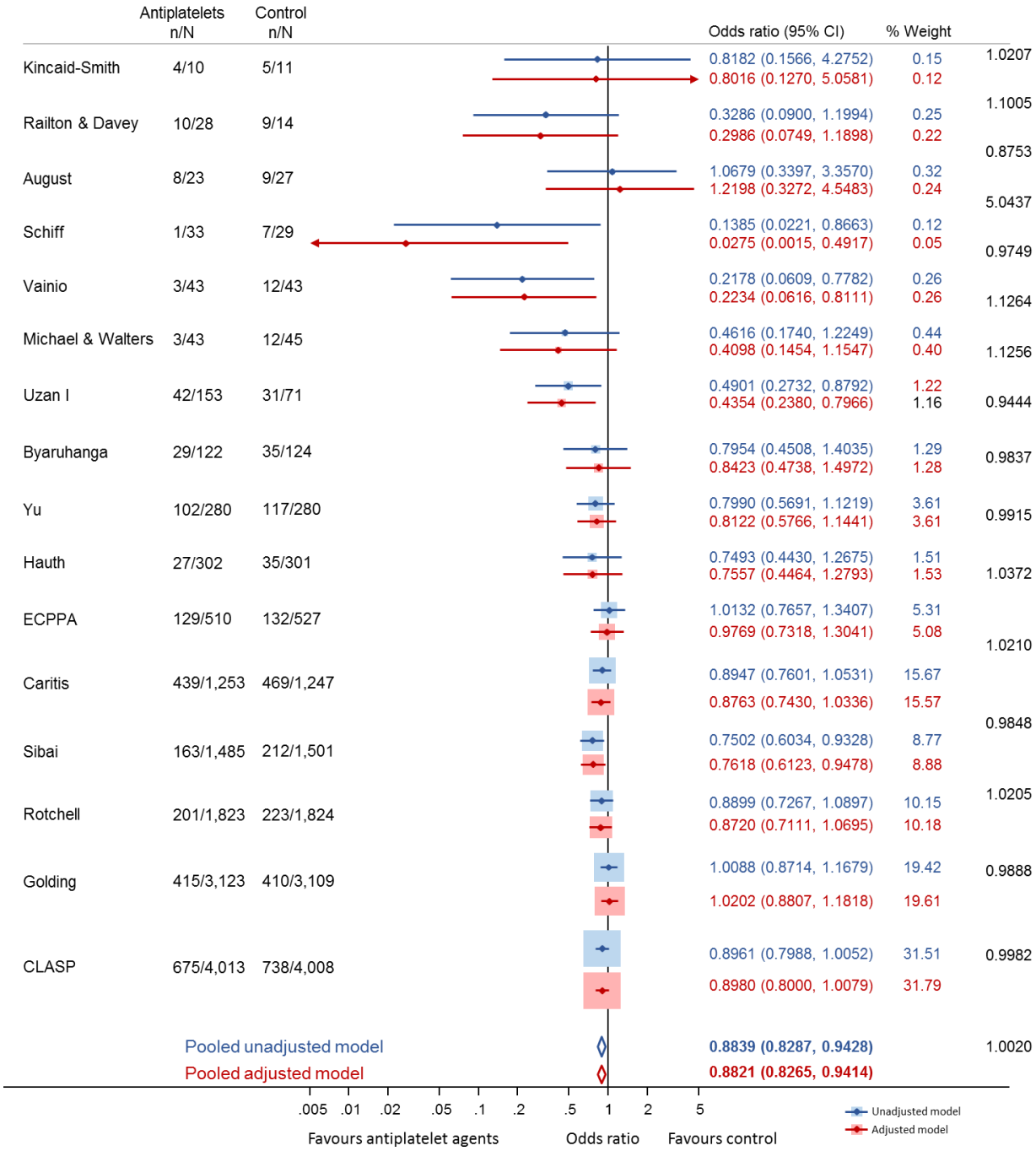
Baseline variable	Coeff (SE)	OR (SE)	P-value
Treatment	-0.1169 (0.1034)	0.8897 (0.092)	0.2584
Systolic blood pressure	0.0311 (0.0048)	1.0316 (0.005)	0
Maternal age	0.0317 (0.0085)	1.0322 (0.0088)	0.0002
Previous pregnancy	-0.0069 (0.104)	0.9931 (0.1033)	0.9473
Chronic hypertension	0.5651 (0.6426)	1.7596 (1.1307)	0.3792

Trial: CLASP

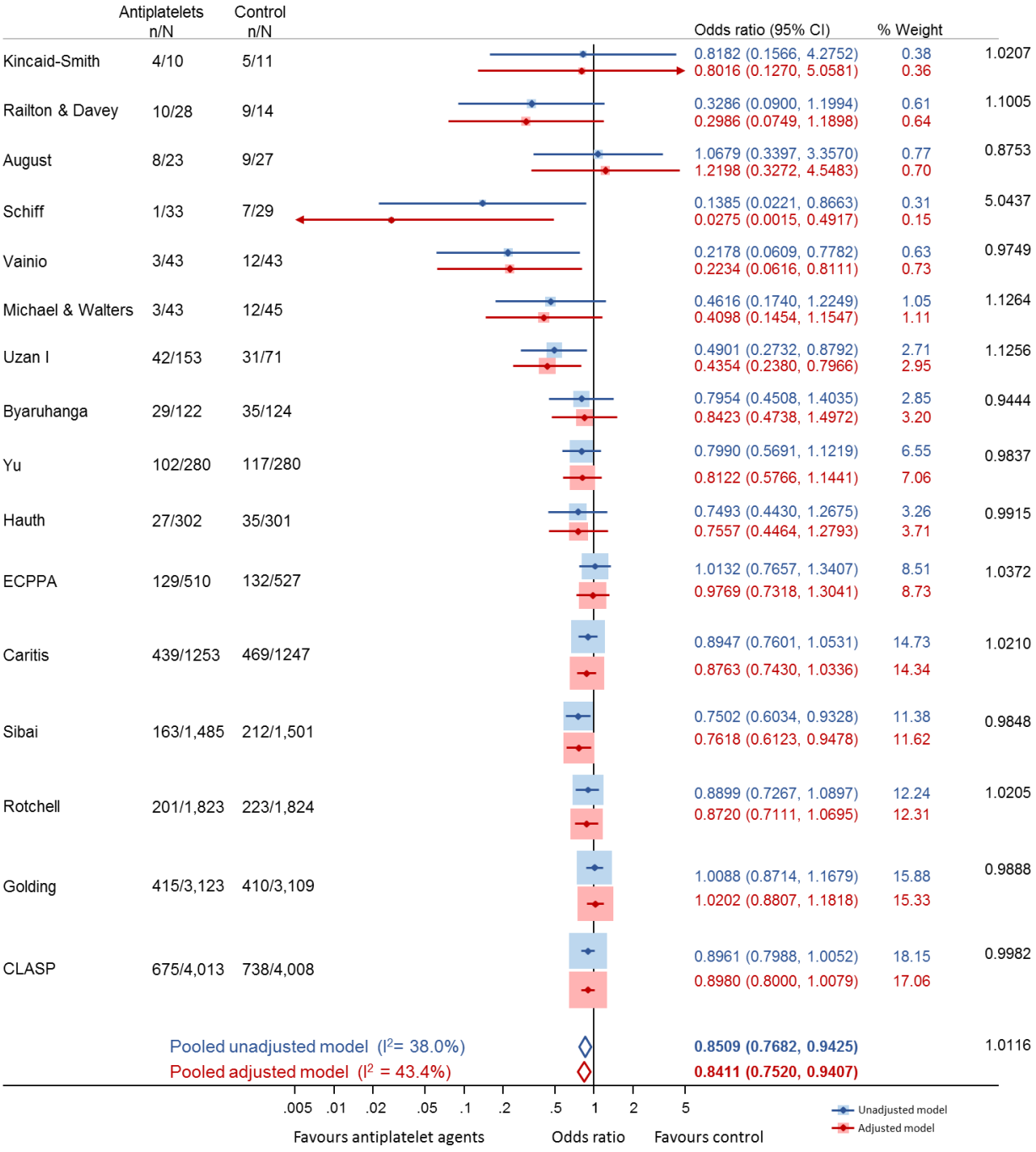
Baseline variable	Coeff (SE)	OR (SE)	P-value
Treatment	-0.1098 (0.0587)	0.896 (0.0526)	0.0612
Systolic blood pressure	0.0132 (0.0017)	1.0133 (0.0017)	0
Maternal age	0.0153 (0.0055)	1.0154 (0.0056)	0.0056
Previous pregnancy	0.0516 (0.0702)	1.053 (0.0739)	0.4628
Chronic hypertension	0.4447 (0.0673)	1.56 (0.105)	0

APPENDIX L: Forest plots of unadjusted and adjusted odds ratios of pregnancy with a serious adverse outcome using Firth's correction on the 16 PARIS studies

a) Two-stage fixed effect model



b) Two-stage random effects model



APPENDIX M: RELEVANT PUBLICATION BY THE THESIS AUTHOR

Yu et al (2010) Reporting on covariate adjustment in randomised controlled trials before and after revision of the 2001 CONSORT statement: a literature review. *Trials* 11:59

Yu *et al. Trials* 2010, **11**:59
http://www.trialsjournal.com/content/11/1/59

TRIALS

REVIEW Open Access

Reporting on covariate adjustment in randomised controlled trials before and after revision of the 2001 CONSORT statement: a literature review

Ly-Mee Yu*[1], An-Wen Chan[2], Sally Hopewell[1], Jonathan J Deeks[3] and Douglas G Altman[1]

Abstract

Objectives: To evaluate the use and reporting of adjusted analysis in randomised controlled trials (RCTs) and compare the quality of reporting before and after the revision of the CONSORT Statement in 2001.

Design: Comparison of two cross sectional samples of published articles.

Data Sources: Journal articles indexed on PubMed in December 2000 and December 2006.

Study Selection: Parallel group RCTs with a full publication carried out in humans and published in English

Main outcome measures: Proportion of articles reported adjusted analysis; use of adjusted analysis; the reason for adjustment; the method of adjustment and the reporting of adjusted analysis results in the main text and abstract.

Results: In both cohorts, 25% of studies reported adjusted analysis (84/355 in 2000 vs 113/422 in 2006). Compared with articles reporting only unadjusted analyses, articles that reported adjusted analyses were more likely to specify primary outcomes, involve multiple centers, perform stratified randomization, be published in general medical journals, and recruit larger sample sizes. In both years a minority of articles explained why and how covariates were selected for adjustment (20% to 30%). Almost all articles specified the statistical methods used for adjustment (99% in 2000 vs 100% in 2006) but only 5% and 10%, respectively, reported both adjusted and unadjusted results as recommended in the CONSORT guidelines.

Conclusion: There was no evidence of change in the reporting of adjusted analysis results five years after the revision of the CONSORT Statement and only a few articles adhered fully to the CONSORT recommendations.

Introduction

Adjusted Analysis in Randomised Controlled Trials

The randomised controlled trial (RCT) is widely accepted as the 'gold standard' design for comparing the effects of health care interventions. Randomisation aims to prevent bias in the allocation of patients to treatment and produce unbiased estimates of treatment effects, but it does not guarantee comparability, particularly in small trials. Adjustment for baseline covariates in the analysis of an RCT is less common than in epidemiological studies. There are four main reasons to consider covariate adjustment methods in RCTs [1-5]: first, to correct for imbalances in baseline prognostic covariates despite randomisation; second, to increase power by modelling the variability in outcome explained by relationships with highly prognostic covariates; third, to obtain treatment effect estimates that would be more closely relevant to individual patients than to an average population; and finally to account for features of study design in the analysis, such as covariates that are used in stratified randomisation. Guidelines suggest that adjusted analysis, including methods of adjustment and choice of covariates, should be pre-specified in the trial protocol [6-8]. In practice, however, adjustment may be done only when baseline imbalance is seen in some covariates [9,10].

* Correspondence: ly-mee.yu@csm.ox.ac.uk
[1] Centre for Statistics in Medicine, University of Oxford, Wolfson College Annexe, Linton Road, Oxford, UK
Full list of author information is available at the end of the article

CONSORT Guidance on Adjusted Analysis

The CONSORT Statement, first published in 1996 and revised in 2001, provides recommendations for reporting parallel groups RCTs. It has received considerable support and has been endorsed by many journals and editorial groups worldwide. While briefly mentioned in the 1996 version, the 2001 revision elaborated the recommendations for reporting of adjusted analysis. This includes specification of the rationale for any adjusted analysis, statistical methods used, and clarification of the choice of variables used for adjustment. When reporting results, CONSORT recommends reporting both unadjusted and adjusted analyses, and stating whether the adjusted analysis was planned. However, information on the extent and quality of such practices in published papers is lacking.

In this study, we carried out a systematic review of two cohorts of publications indexed in PubMed to determine the use and reporting of adjusted analysis in RCTs. We also compared the quality of reporting before and after the revision of the CONSORT Statement in 2001.

Methods

Study selection

This review included two cohorts: (1) articles published in December 2000 and indexed in PubMed, as previously identified by Chan et al [11,12]; (2) a newly identified cohort of articles indexed in December 2006 in PubMed (as of 22 March 2007). Both cohorts were identified by searching PubMed using the extended version of Phase 1 of the Cochrane Highly Sensitive Search Strategy for trials [13]. The abstracts of the search results for December 2006 were screened by one of the authors (LY). Based on the abstract, all articles that were obviously not trials were excluded. The full text of all remaining articles was fully reviewed (LY) to assess their eligibility.

We included in this review RCTs of parallel group design with a full publication carried out in humans and published in English. Articles published as a letter or brief communication, and articles reporting phase I or pilot studies were excluded. We also excluded studies that did not provide sufficient information on statistical analysis or did not perform any formal comparison between treatment groups.

Defining adjusted analysis

We identified all trial outcomes that were explicitly reported to have undergone adjusted analysis for comparisons between randomised groups in either the Methods or Results section of the article. We sought mention of the statistical analysis of the treatment effect accounting for covariates or an explicit statement that some results were adjusted. Analyses that used multiple regression methods to identify prognostic variables or risk factors were not defined as adjusted analysis.

Data extraction

Information on trial characteristics and all outcomes were extracted from the 2006 articles using the same definitions as those in the 2000 cohort [13]. Briefly, the primary outcome reported in the articles was defined if it was explicitly specified in the article, an outcome used in the power calculation, or a main outcome described explicitly in the primary study objectives. Multi-center involvement was defined as data being collected from more than one study site; sample size was defined as the total number of participants randomised in the study.

To maintain independence of observations, we selected one outcome for each trial if more than one outcome underwent adjusted analysis. We selected the outcome according to the following hierarchy: (1) it was a prespecified primary outcome; (2) the sample size of the trial was based on this outcome; or (3) it had most information on adjusted analysis reported in the article. If more than one outcome was equally reported within an article, then the outcome was chosen at random.

For articles in both cohorts reporting adjusted analysis we assessed the types of analysis reported explicitly in the Methods and Results sections. Articles were classified as reporting unadjusted analysis, adjusted analysis, both, or unspecified/unclear. We also recorded the reason for adjustment, the method of adjustment, and details of the covariates used in the analysis. We assessed whether the unadjusted or adjusted results, such as summary statistics, confidence intervals (CI) or standard error (SE) within group, treatment effect, CI/SE of treatment effect, and the corresponding P-value were reported in the main text and abstract. If results reported in the abstract were not clear, we referred to the main text for type of analysis used.

We also evaluated whether the reporting of adjusted analyses adhered to the 2001 CONSORT guidelines. For the 2006 cohort, we assessed whether articles were published in a CONSORT endorsing journal based on the journals' 'Instruction to Authors' (assessed June 2008). Data regarding trial characteristics were extracted by two reviewers (LY and SH), while outcome and adjusted analysis information were extracted by a single reviewer (LY).

Data Synthesis and Analysis

Frequency of adjusted analysis was expressed as the proportion of trials that reported using adjusted analysis. Comparisons of trial characteristics and adherence to the 2001 CONSORT Statement between 2000 and 2006 were carried out by Chi-square test for categorical data or Fisher's exact test if expected counts were less than five, and Mann-Whitney test for continuous data. Percentage

difference and corresponding precision based on 95% confidence intervals (CI) were calculated to quantify the change in reporting between 2000 and 2006. Similar analyses were used for comparisons of trial characteristics between trials that did or did not report adjusted analysis within each cohort. Data were analyzed using Stata 9 (Stata Corporation, College Station, TX, USA) and a P-value of less than 0.05 was considered to indicate statistical significance.

Results

Characteristics of trials

In total, 1735 citations were identified from December 2006 and 616 articles were included. Full details of included and excluded articles are shown in Figure 1. Of the 519 articles retrieved from 326 journals in 2000 and 616 from 316 journals in 2006, 355 and 421 parallel group studies were included in this review, respectively (Figure 2). A significantly lower proportion of articles specified the primary outcome in 2000 than 2006 (51% vs 65%, respectively; P < 0.0001). In both years, most studies were characterized by two study arms (74% for 2000 vs 78% for 2006), a single study centre (both 66%), and publication in specialty journals (both 91%) (Table 1). The average sample size and number of trial outcomes were similar in both years and most reported outcomes were continuous (about 70%). Fewer studies performed stratified randomisation in 2000 than in 2006 (16% vs 20%, respectively; P = 0.1).

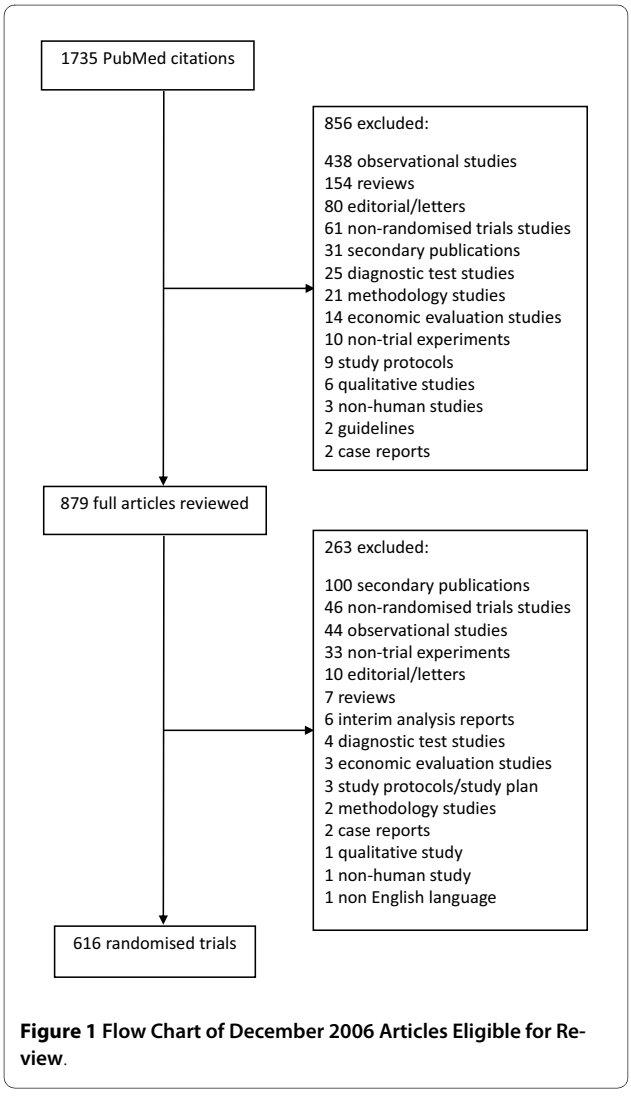


Figure 1 Flow Chart of December 2006 Articles Eligible for Review.

Eighty four articles (24%) and 113 articles (27%) in 2000 and 2006, respectively, reported adjusted analyses performed on at least one outcome in the Methods, Results, or both sections.

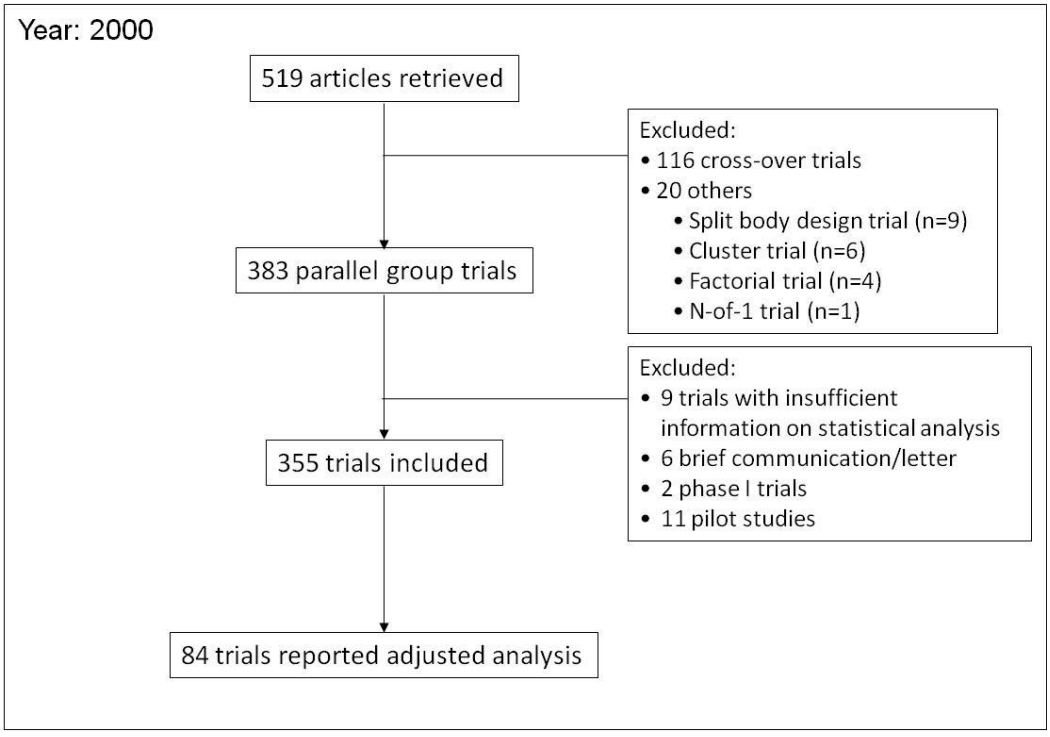
Characteristics of trials that did or did not report adjusted analysis

There was a marked difference in the characteristics of studies that did or did not report adjusted analysis in both cohorts. A higher proportion of articles reporting adjusted analysis had specified primary outcomes, involved multiple centers, had performed stratified randomisation, and were published in general journals. Trials with adjusted analysis recruited more participants and had fewer outcomes (Figure 3).

Consistency of analysis reported between Methods and Results in articles reported adjusted analysis

Among the adjusted analyses articles, 79 and 109 articles had a statistical methods section in 2000 and 2006, respectively. For the outcome selected from each trial, we examined the consistency of the type of analysis reported in the Methods and Results sections. In 2000, 43 out of 79 articles (54%) explicitly specified adjusted analyses were used in the Methods and had subsequently reported them in the Results. Discrepancies between the information in Methods and Results sections were found in 36 articles (46%). For example, two articles had specified adjusted analysis in the Methods but reported only unadjusted results in the Results and 24 (30%) articles did not specify clearly the type of analysis used in the Results section.

In 2006, the consistency of the type of analysis reported in the Method and Results sections increased to 69% (74/109) (P = 0.06; Difference [95% CI] = 13.5% [-0.6% to 27.5%]), while there was a reduction in the proportion of articles that did not specify clearly the type of analysis used in the Results section for the selected outcome (19/109 = 17%) (P = 0.04; Difference [95% CI] = -12.9 (-25.3 to -0.6)). Three articles specified adjusted analysis in the Methods but reported only unadjusted results in the Results. We contacted the authors of the 34 articles with an inconsistency between the Methods and Results section but only three responded.



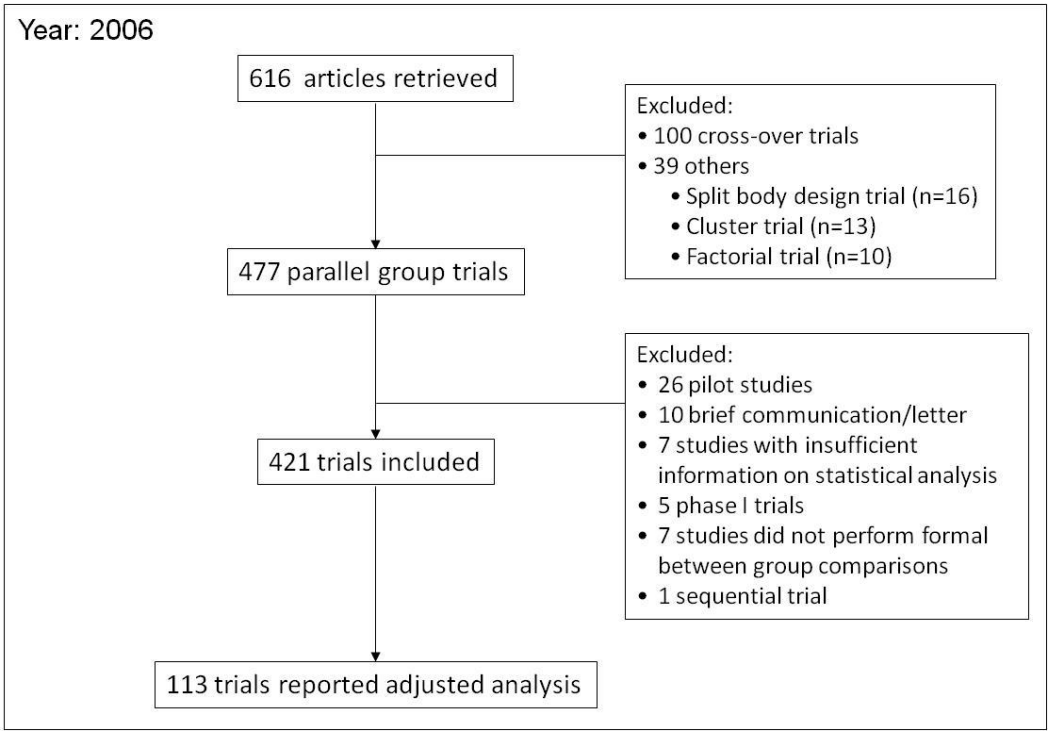


Figure 2 Flowchart diagram of articles retrieved and included in the review.

Table 1: Characteristics of articles of parallel group randomized trials by year of publication

Year of Publication	2000 (n = 355)	2006 (n = 421)	% difference† (95% CI)	P-value
Outcome specification				
Primary	180 (50.7%)	273 (64.9%)	14.1 (7.2 to 21.0)	<0.0001
Unspecified	175 (49.3%)	148 (35.1%)		
Centre involved*				
Multiple centres	119 (34.1%)	138 (33.7%)	-0.4 (-7.2 to 6.3)	0.9
Single centre	230 (65.9%)	272 (66.3%)		
Number of intervention groups	261 (73.5%)	328 (77.9%)		0.2
2	57 (16.1%)	64 (15.2%)		
3	37 (10.4%)	29 (6.9%)		
> 3				
Performed stratified randomisation	56 (15.8%)	85 (20.2%)	4.4 (-1.0 to 9.8)	0.1
Sample size				
< 50	116 (32.7%)	129 (30.6%)		
51 - 150	141 (39.7%)	169 (39.9%)		
151 - 300	49 (13.8%)	52 (12.4%)		
301 - 450	20 (5.6%)	27 (6.6%)		
> 450	29 (8.2%)	44 (10.5%)		
Median (10^{th} to 90^{th} percentile)	91 (27 to 394)	80 (28 to 462)		0.7
Journal type				
General medical	31 (8.7%)	36 (8.6%)	-0.1 (3.8 to -4.2)	0.9
Specialty	324 (91.3%)	385 (91.4%)		
Number of outcomes per trial				
Median (range)	15 (1, 131)	14 (1, 372)		0.2
Type of outcomes	(n = 7132)	(n = 8299)		<0.0001
Continuous	4984 (69.9%)	5705 (68.7%)		
Binary	1961 (27.5%)	2357 (28.4%)		
Time-to-event	47 (0.6%)	128 (1.5%)		
Ordinal	140 (2.0%)	98 (1.2%)		
Categorical	0	11 (0.1%)		
Adjusted analysis	84 (23.7%)	113 (26.8%)	3.1 (-2.9 to 9.3)	0.3

* Unclear: 6 for year 2000 and 11 for year 2006

† Percentage difference = percentage in 2006 - percentage in 2000

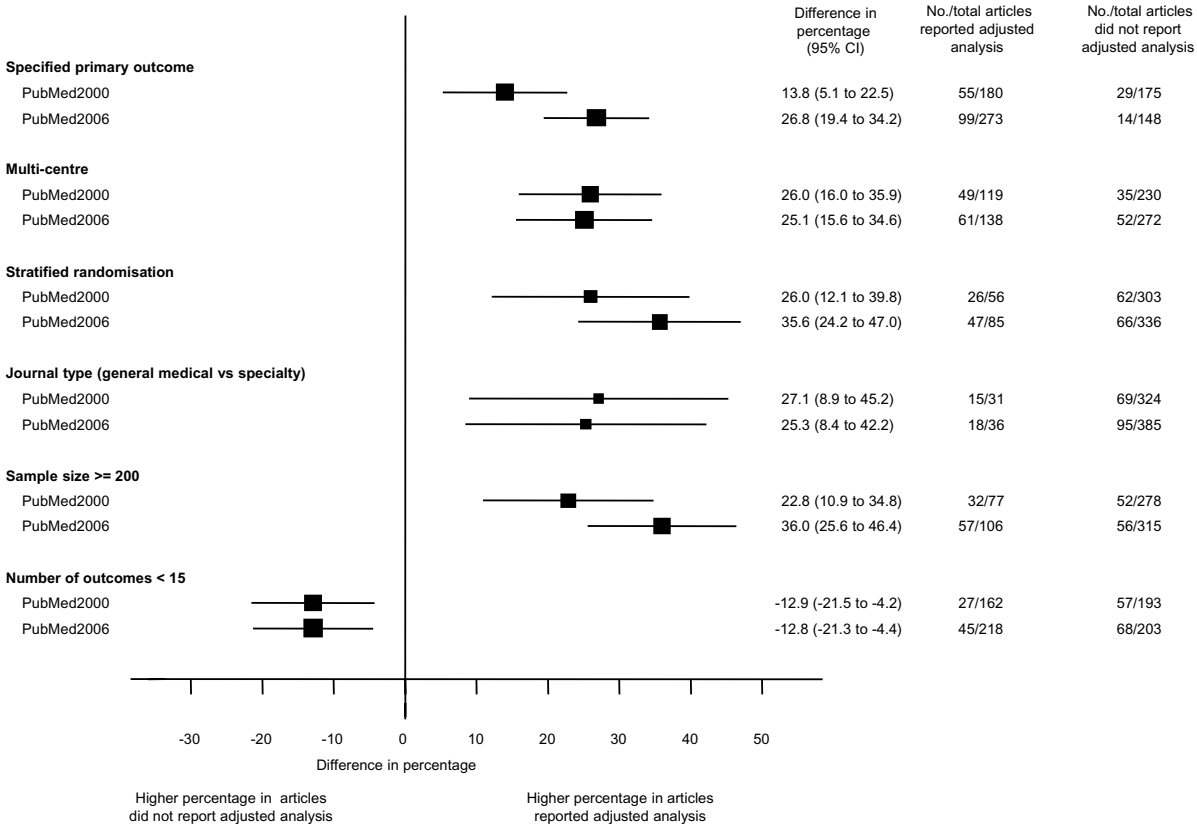


Figure 3 Comparison of characteristics of articles that did or did not report adjusted analysis for trials published in 2000 and 2006.

Details of adjusted analysis

Details of adjusted analysis are summarized in Table 2. In the 2000 cohort, over 90% of articles had carried out adjusted analysis on the primary outcome. Overall, the majority of articles (80%) did not report the reasons for adjustment or how the covariates were selected for adjustment. Of the 78 articles that specified the covariates, nine (12%) included covariates that were collected after randomisation and 16 (20%) did not specify in the Method section the methods used for adjustment. Fewer than half of the articles included all of the stratification factors used at randomisation in the adjusted analysis. Very few (8% in 2000 and 9% in 2006) specified explicitly whether the adjusted analysis was the primary or secondary analysis.

Eighty three articles (99%) in 2000 reported the statistical methods used for adjustment. Since outcomes were predominately continuous, most studies used regression methods (ANCOVA, ANOVA or multiple regressions) for adjustment (Table 3). Binary outcomes and time-to-event data were analysed mainly by logistic regression and Cox regression, respectively. Stratified analyses (e.g. Cochrane-Mantel-Haenszel or Chi-squared analysis) for adjustment were used more often for binary outcomes than other types of outcomes.

In the 2006 cohort, there was no evidence of change in the reporting of the reason for adjustment (30%) and choice of covariates (27%). More trials in 2006 had adjusted for covariates that were believed to be correlated with the outcomes (13% vs 7%) but only two articles explicitly stated that the covariates selected for adjustment were pre-specified. In addition, more covariates were adjusted for than in 2000, especially outcomes collected at baseline, but fewer multi-centre studies had adjusted for centre effect. Use of statistical methods was similar in both cohorts (Table 3).

Reporting of adjusted analysis

Table 4 presents the type of results reported in the Results section and abstract. Fifty four articles in 2000 reported any results of adjusted or unadjusted analysis in the Results section. Of these, 80% reported explicitly the type of analysis used to derive the P-values while just under a half reported estimated treatment effects (e.g. odds ratio or difference between means) and the corresponding confidence intervals. Lack of reporting of

Table 2: Details of adjusted analysis

Year of publication	2000 (n = 84)*	2006 (n = 113)*	% difference (95% CI)	P-value
Performed adjusted analysis on primary outcome†	50 (90.9%)	93 (93.9%)	30.0 (-5.9 to 12.0)	0.5
Reason for adjustment				0.4
Imbalance in covariates	9 (10.7%)	12 (10.6%)		
Prognostic covariates	6 (7.1%)	15 (13.3%)		
Both	0	3 (2.6%)		
Other reasons ‡	3 (3.6%)	4 (3.5%)		
Not mentioned	66 (78.6%)	79 (69.9%)		
Choice of covariates				0.5
All pre-specified	5 (5.9%)	8 (7.1%)		
All suggested by data	12 (14.3%)	20 (17.7%)		
Combination of pre-specified and post hoc	0	3 (2.6%)		
Not mentioned	67 (79.8%)	82 (72.6%)		
Number of covariates adjusted for §				0.02 ‖
1	39 (46.4%)	36 (31.8%)		
2	23 (27.4%)	33 (29.2%)		
3-5	14 (16.7%)	25 (22.1%)		
6-9	2 (2.4%)	12 (10.6%)		
Not mentioned	6 (7.1%)	7 (6.2%)		
Covariate used for adjustment				
Outcome assessed at baseline	33/62 (53.2%)	55/81 (67.9%)	14.7 (-1.4 to 30.7)	0.07
Centre/Country	31/49 (63.3%)	25/61 (41.0%)	-22.3 (-40.6 to -39.9)	0.02
Assessed after randomisation	9/78 (11.8%)	9/107 (8.4%)	-3.1 (-12.0 to 5.7)	0.5
All stratification factors were adjusted for	11/25 (44.0%)	20/46 (43.5%)	-0.5 (-24.7 to 23.6)	1.0
Explicitly specified nature of analysis				
Primary analysis	2	5		
Secondary/ sensitivity analysis	5	5		

Table 2: Details of adjusted analysis (Continued)

Type of outcomes				0.8
Binary	13 (15.5%)	19 (16.8%)		
Continuous	65 (77.4%)	81 (71.7%)		
Ordinal	1 (1.2%)	3 (2.6%)		
Time-to-event	5 (5.9%)	10 (8.9%)		
Adjusted analysis method used was mentioned for specific outcome in the Method section	62/78 ¶(79.5%)	88/109 ¶ (80.7%)	1.2 (-10.4 to 12.9)	0.8

*One adjusted analysis selected per study only

†Number of studies that have specified primary outcomes: Year 2000 = 55 and Year 266 = 99

‡Year 2000: Clinical relevance (n = 1), significant at 3 weeks after randomisation (n = 1), exploring role of baseline variables (n = 1); Year 2006: Mediated treatment effect on outcome (n = 1), related to compliance/adherence of treatment (n = 2), effect of outcome decline over time (n = 1)

§Year 2000: 6 studies did not report number of covariates; Year 2006: 6 studies did not report number of covariates and 1 stated at least 2 covariates

‖ Mann-Whitney test

¶Did not have statistical method section: Year 2000 (n = 6), Year 2006 (n = 5)

results, for the selected outcome, in the abstract was more severe. Over 80% of the articles did not report either the treatment effect or the corresponding confidence interval in the abstract. Even P-values were reported in only 31% of the studies.

Overall, there was an increase in reporting any adjusted results in the abstract in 2006 when compared with the 2000 cohort (Table 4). However, in both years a high percentage of articles which used adjustment did not report any adjusted treatment effect. Only 26/50 (52%) in 2000 and 61/93 (66%) in 2006 reported the results of any treatment comparison (i.e. treatment effect estimate, confidence interval, or P-value) in the abstract. Of these, 50% and 61% reported any adjusted results, respectively, but in both years only 30% presented the adjusted treatment effect. Confidence intervals were rarely provided.

Adherence to the CONSORT guidelines

With regard to how adjusted analysis should be reported according to the revised CONSORT Statement, there was a slight improvement in some items five years after the revision but the overall adherence is still low (Table 5). Although fewer articles in 2000 reported that stratified randomisation was performed, the proportion that adjusted for any stratification variables was in fact higher than in the 2006 cohort (46% in 2000 vs. 35% in 2006).

Reporting of both adjusted and unadjusted results was poor. Only four out of 84 articles and 11 out of 113 articles in 2000 and 2006, respectively, reported both results. Of 21 articles (25%) in 2000 that mentioned both adjusted and unadjusted analyses, seven reported only the unadjusted results because the results were similar for both analyses. Similarly, 27 articles had performed both analyses in 2006, of which two reported the adjusted results and five reported the unadjusted results because both results were similar. In addition, four studies in that cohort had reported that the significance of treatment effect was different from unadjusted analysis after adjusting for covariates.

In 2006, 65 of the 113 (57%) articles that reported adjusted analysis were published in CONSORT-endorsing journals. Among these, 23 (35%) specified the rationale for the adjusted analysis performed compared with 11 of the 48 (23%) articles from journals that did not endorse CONSORT. The number of articles which reported both adjusted and unadjusted results was slightly higher in CONSORT endorsing journals compared to non endorsing journals (seven vs four articles, respectively).

Discussion

Our study provides a comprehensive assessment and comparison of the quality of reporting of adjusted analysis before and after the revision of the CONSORT Statement in 2001. In our review, we found that the characteristics of published reports of parallel group randomised trials indexed in PubMed in 2000 and 2006 were similar, though there was a significant improvement in primary outcome specification in 2006. Only a quarter of randomised trials reported any covariate adjustment analysis. The prevalence of adjusted analysis in our broad cohorts is much lower than the 72% reported in a previous review which was restricted to four high impact general medical journals in 1997 [1] and 64% in a recent

Table 3: Methods used in adjusted analysis

Year of Publication	2000 (n = 84)	2006 (n = 113)
Continuous data	65 (77.4%)	81 (71.7%)
ANOVA/ANCOVA	*50 (76.9%)*	*56 (69.1%)*
Multiple regression method*	*7 (10.8%)*	*19 (23.5%)*
Stratified analysis	*1 (1.5%)*	*0*
Other†	*6 (9.2%)*	*6 (7.4%)*
Not mentioned	*1 (1.5%)*	*0*
Binary data	13 (15.5%)	19 (16.8%)
Logistic regression	*6 (46.1%)*	*11 (57.9%)*
Stratified analysis (Cochrane-Mantel-Haenszel test)	*5 (38.5%)*	*5 (26.3%)*
Other‡	*2 (15.4%)*	*3 (15.8%)*
Ordinal data	1 (1.2%)	3 (2.6%)
Stratified analysis (Cochrane-Mantel-Haenszel test)	*1 (100%)*	*1 (33.3%)*
Nonlinear mixed effect model	*0*	*1 (33.3%)*
Ordinal logistic regression	*0*	*1 (33.3%)*
Time to event data	5 (5.9%)	10 (8.9%)
Cox proportional hazard	*5 (100%)*	*9 (90.0%)*
Stratified log rank test	*0*	*1 (10.0%)*

* Including random effect and mixed effect models

† Including GEE, GLM, ANCOVA for rank data, Zellner seemingly unrelated regression, Poisson model, Van Elteren test

‡ Including non-parametric Generalized mixed effect model, GEE, non-parametric ANCOVA

review conducted by Austin et al [14]. Another review looked at 34 scientific medical journals in 1998 with a high impact factor and reported 31% of articles had specified adjustment for confounding factors [15]. A further study found similar percentage of adjusted analysis in clinical trials of traumatic brain injury [16]. To our knowledge, these three studies are the only previous such studies addressing this issue. By including journals from all specialties, we believe that the frequency of adjusted analysis in our cohorts is representative of the overall randomised trial literature.

We found that analyses specified in the Methods sections did not necessarily reflect how the results reported in the Results section were obtained. Often the method was either not clearly specified or the results were obtained from different analyses from the specified ones. Readers often trust that the results were derived from analyses specified in the Method section. Our findings have shown that further clarification for reporting results is needed; especially in studies involving adjusted analysis.

Although many authors have discussed how adjusting for baseline covariates in the analysis of RCTs can improve the power of analyses of treatment effect and account for any imbalances in baseline covariates [4,5,17-19], the debate on whether this practice should be carried out remains unresolved. Many recommend that the analysis should be undertaken only if the methods of analysis and choice of covariates are pre-specified in the protocol or statistical analysis plan [1,6-8]. Unfortunately, the rationale for adjustment and choice of covariates were missing in most of the articles we reviewed, although there has been an improvement in the overall reporting of adjusted analysis in trial reports published in 2006 compared to 2000. This lack of pre-specification echoes the findings in the recent review carried out by Chan et al [20]. They found that most trials that mentioned adjusted analysis in either the protocol or article had discrepancies between the two (18/28). Among 18 trials with published

Table 4: Presentation of results in the Results section and abstract for studies reporting adjusted analysis

Year of Publication	Results Section			Abstract		
	2000 (n = 54)	2006 (n = 89)	P-value	2000 (n = 71)*	2006 (n = 101)†	P-value
Summary statistics for each group			0.7			0.5
Unadjusted only	42 (78%)	70 (80%)		26 (37%)	45 (44%)	
Adjusted only	6 (11%)	12 (14%)		3 (4%)	5 (5%)	
Both	4 (7%)	3 (3%)		0	0	
None/not clear	2 (4%)	3 (3%)		42 (59%)	51 (51%)	
Confidence interval/SE within group			0.2			1.0
Unadjusted only	12 (22%)	11 (13%)		2 (3%)	2 (2%)	
Adjusted only	6 (11%)	10 (11%)		2 (3%)	4 (4%)	
Both	1 (2%)	0		0 (%)	0	
None/not clear	35 (65%)	67 (76%)		67 (94%)	95 (94%)	
Treatment effect			0.4			0.1
Unadjusted only	5 (9%)	5 (6%)		3 (4%)	5 (5%)	
Adjusted only	17 (31%)	35 (39%)		5 (7%)	19 (19%)	
Both	4 (7%)	12 (13%)		1 (2%)	1 (1%)	
None/not clear	28 (52%)	37 (42%)		62 (87%)	76 (75%)	
Confidence interval/SE of treatment effect			0.6			0.4
Unadjusted only	6 (11%)	5 (6%)		2 (3%)	6 (6%)	
Adjusted only	16 (30%)	24 (27%)		3 (4%)	17 (17%)	
Both	4 (7%)	10 (11%)		1 (2%)	1 (1%)	
None/not clear	28 (52%)	49 (56%)		65 (91%)	77 (76%)	
P-value for treatment effect			0.2			0.2
Unadjusted only	9 (17%)	8 (9%)		9 (13%)	13 (13%)	
Adjusted only	27 (50%)	52 (59%)		13 (18%)	30 (30%)	
Both	7 (13%)	17 (19%)		0	2 (2%)	
None/not clear	11 (20%)	11 (13%)		49 (69%)	56 (55%)	

*13 studies did not report the selected outcome in abstract

† 2 studies did not have abstract and 10 studies did not report the selected outcome in the abstract

adjusted analyses, 12 included covariates that were not pre-specified in the protocol ten of which did not mention any adjusted analysis in the protocol.

Most articles that gave their reason for adjustment or choice of covariates were not in accordance with the guidelines' recommendations [6,7]. Few studies performed and reported the adjusted analysis adequately. For example, where procedures such as stratified randomisation or minimisation methods were used, the analysis without adjustment of stratifying variables could over-estimate the standard error of the treatment effect as well as distort the P-value [21]. Our findings indicate that trials that performed these procedures often did not adjust for stratification/minimisation factors. Furthermore, covariates assessed after randomisation require special attention because their relationship with the study

outcome could be confounded by treatment; a different analytical approach is needed [6,7,22,23]. However, we found that some trials included such covariates in the analyses, as has been documented by others [24,25].

Generally, the reporting of adjusted analysis was comparable between the two cohorts we reviewed, which represent trials published before and after the revision of the CONSORT Statement in 2001. Reporting of the main results, such as treatment group summary statistics, treatment effect and confidence intervals, as suggested by CONSORT, were often lacking or unclear in both the Results section and abstract. Such deficiencies could be due to the fact that much more attention has been given to other issues, such as adequacy and transparency of sample size calculation, blinding and randomisation methods, etc, that have already been addressed more often in other systematic reviews [26,27]. Treatment effect estimates from unadjusted and adjusted analyses are not directly comparable because the former gives population-averaged estimates of treatment effect while the latter assesses subject-specific estimates, so it is important that these results are reported clearly so that the treatment effect can be interpreted correctly. This argument is most pertinent in analyses of RCTs with non-continuous outcomes because the treatment effect estimate changes when covariates are included in the analysis [3].

There is little previous evidence about the use and reporting of adjusted analysis in RCTs (19). However, two recent studies reported the impact of selective reporting of adjusted estimates in meta-analyses of observational studies [28,29]. Both studies found that the pooled unadjusted effects differed according to whether studies contributed both adjusted and unadjusted estimates to the meta-analyses or only unadjusted effects. To what extent this lack of clarity in reporting adjusted analyses in RCTs could represent reporting bias that may affect subsequent meta-analyses is unclear. We appreciate that unclear reporting of results does not necessarily reflect poor research conduct, but there is clear evidence suggesting that quality of reporting is associated with bias in the estimation of treatment effect [12,30].

We identified slightly better reporting of key methodological items in CONSORT endorsing as opposed to non CONSORT endorsing journals. However, because there was a time-lag between article publication (December 2006) and when the journal 'Instructions to Authors' were assessed (June 2008) these results should be viewed with some caution. A limitation of this study is that, apart from the trial characteristics for the 2006 cohort, data were extracted by a single reviewer. However, the reviewer revisited the data extraction a few months after the first extraction as a quality assurance procedure. We also used slightly different sampling techniques between

Table 5: Ad herence to the CONSORT recommendations

Year of publication	2000 (n = 84)	2006 (n = 113)	Relative risk (95% CI)	P-value
Have adjusted for any stratification variables*	26 (46%)	30 (35%)	0.76(0.51, 1.14)	0.2
Have specified rationale for any adjusted analysis	18 (21%)	34 (30%)	1.40(0.85, 2.31)	0.2
Have specified statistical method used for adjusted analysis	83 (99%)	113 (100%)	1.0 (0.97, 1.03)	1.0
Have reported results from adjusted analysis only[2]	18 (21%)	29 (26%)	1.20 (0.71, 2.01)	0.5
Have reported results from both adjusted and unadjusted analysis†	4 (5%)	11 (10%)	2.04 (0.67, 6.20)	0.3

* n = 56 for Year 2000 and n = 85 for Year 2006

†Results included summary in each group, effect size, and confidence interval

the two years. The 2000 cohort included all reports of randomised trials published in December 2000 and indexed in PubMed by July 2002 to account for the lag in PubMed indexing. For pragmatic reasons, the 2006 cohort included those trials indexed in PubMed in December (as of March 2007). This meant that we were able to capture our sample of trials within one search but may have missed a small number of trials which were published in December 2006 but indexed in PubMed after March 2007.

In conclusion, there was no evidence of change in the reporting of adjusted analysis results five years after the revision of CONSORT Statement. Furthermore, overall quality of reporting of adjusted analysis and adherence to CONSORT recommendations remain low. The rationale for covariate adjustment, methods of analysis and choice of covariates for adjustment should be fully reported so that readers can assess whether the adjusted analysis has been adequately carried out and, therefore, should be made transparent in the trial reports. Finally, both unadjusted and adjusted results, which analysis represents the primary analysis, and whether the adjusted analysis was pre-specified in the protocol should also be included in the report.

Competing interests
We declare that we have no conflict of interest. SH and LMY are funded by NHS. DGA is funded by Cancer Research UK.

Authors' contributions
L-MY, DGA, JJD contributed to the study design. SH, L-MY and A-WC contributed to the data collection. All authors contributed to the interpretation of results and drafting of the manuscript. L-MY performed the statistical analyses and is the guarantor.

Acknowledgements
We thank Dr. Isabelle Boutron for her helpful comments.

Author Details
[1]Centre for Statistics in Medicine, University of Oxford, Wolfson College Annexe, Linton Road, Oxford, UK, [2]Women's College Research Institute, Department of Medicine, University of Toronto, Canada and [3]Medical Statistics Group/Diagnostic Research Group, Public Health, Epidemiology & Biostatistics, The Public Health Building, The University of Birmingham, Birmingham, UK

Received: 21 January 2010 Accepted: 18 May 2010
Published: 18 May 2010

References

1. Assmann SF, Pocock SJ, Enos LE, Kasten LE: **Subgroup analysis and other (mis)uses of baseline data in clinical trials.** *Lancet* 2000, **355:**1064-9.
2. Altman DG: **Comparability of randomised groups.** *Statistician* 1985, **34:**125-36.
3. Hauck WW, Anderson S, Marcus SM: **Should we adjust for covariates in nonlinear regression analyses of randomized trials?** *Control Clin Trials* 1998, **19:**249-56.
4. Senn SJ: **Covariate imbalance and random allocation in clinical trials.** *Stat Med* 1989, **8:**467-75.
5. Pocock SJ, Assmann SE, Enos LE, Kasten LE: **Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems.** *Stat Med* 2002, **21:**2917-30.
6. ICH E9 Expert Working Group: **Statistical principles for clinical trials: ICH harmonized tripartite guidelines.** *Stats Med* 1999, **18:**1905-42.
7. CPMP: **Points to consider on adjustment for baseline covariates (cpmp/ewp/2863/99).** Emea: London; 2003.
8. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, for the CONSORT Group, *et al.*: **The revised CONSORT statement for reporting randomized trials: explanation and elaboration.** *Ann Intern Med* 2001, **134:**663-94.
9. Yu LM, Alder N, Altman DG, Deeks J, Chan AW: **Adjusting for covariates in randomised clinical trials: a review of published trials.** *Proceedings of the ISCB meeting: 15-19 August, 2004; Leiden* .
10. Altman DG: **Adjustment for covariate imbalance.** In *Encyclopedia of biostatistics* 2nd edition. Edited by: Armitage P, Colton T. Chichester: John Wiley; 2005:1273-8.
11. Chan AW, Altman DG: **Epidemiology and reporting of randomised trials published in PubMed journals.** *Lancet* 2005, **1365:**1159-62.
12. Chan AW, Altman DG: **Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors.** *BMJ* 2005, **333:**753.
13. Robinson KA, Dickersin K: **Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed.** *Int J Epidemiol* 2002, **31:**150-3.
14. Austin PC, Manca A, Zwarenstein M, Juurlink DN, Stanbrook MB: **A substantial and confusing variation exists in handling of baseline covariates in randomized controlled trials: a review of trials published in leading medical journals.** *J Clin Epidemiol* 2010, **63:**142-153.
15. Mullner M, Matthews H, Altman DG: **Reporting on statistical methods to adjust for confounding: a cross-sectional survey.** *Ann Intern Med* 2002, **136:**122-26.
16. Hernández AV, Steyerberg EW, Taylor GS, Marmarou A, Habbema JD, Maas AI: **Subgroup analysis and covariate adjustment in randomized clinical trials of traumatic brain injury: a systematic review.** *Neurosurgery* 2005, **57:**1244-53.
17. Hernandez AV, Eijkemans MJ, Steyerberg EW: **Randomized controlled trials with time-to-event outcomes: how much does pre-specified covariate adjustment increase power?** *Annals of Epidemiology* 2006, **16:**41-48.
18. Hernandez AV, Steyerberg EW, Habbema JD: **Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements.** *J Clin Epidemiol* 2004, **57:**454-60.
19. Negassa A, Hanley JA: **The effect of omitted covariates on confidence interval and study power in binary outcome analysis: a simulation study.** *Contemp Clin Trials* 2007, **28:**242-8.
20. Chan AW, Hróbjartsson A, Jørgensen KJ, Gøtzsche PC, Altman DG: **Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols.** *BMJ* 2008, **337:**a2299.
21. Forsythe AB, Stitt FW: **Randomization or minimization in the treatment assignment of patient trials: validity and power of tests.** In *Technical Report* Issue 28 Health Sciences Computing Facility, University of California, Los Angeles; 1977.
22. Rochon J: **Issues in adjusting for covariates arising postrandomization in clinical trials.** *Drug Inf J* 1999, **33:**1219-1228.
23. Yusuf S, Wittes J, Probstfield J, Tyroler HA: **Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials.** *JAMA* 1991, **266:**93-8.
24. Walraven C van, Davis D, Forster AJ, Wells GA: **Time-dependent bias was common in survival analyses published in leading clinical journals.** *J Clin Epidemiol* 2004, **57:**672-82.
25. Hirji KF, Fagerland MW: **Outcome based subgroup analysis: a neglected concern.** *Trials* 2009, **10:**33.
26. Moher D, Jones A, Lepage L, CONSORT Group (Consolitdated Standards for Reporting of Trials): **Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation.** *JAMA* 2001, **285:**1992-5.
27. Kane RL, Wang J, Garrard J: **Reporting in randomized clinical trials improved after adoption of the CONSORT statement.** *J Clin Epidemiol* 2007, **60:**241-9.
28. Nicholson A, Kuper H, Hemingway H: **Depression as an aetiologic and prognostic factor in coronary heart disease: a meta-analysis of 6362 events among 146 538 participants in 54 observational studies.** *Eur Heart J* 2006, **27:**2763-74.

29. Peters J, Mengersen K: **Selective reporting of adjusted estimates in observational epidemiology studies: reasons and implications for meta-analyses.** *Eval Health Prof* 2008, **31:**370-89.
30. Pildal J, Hróbjartsson A, Jørgensen KJ, Hilden J, Altman DG, Gøtzsche PC: **Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials.** *Int J Epidemiol* 2007, **36:**847-57.

doi: 10.1186/1745-6215-11-59
Cite this article as: Yu *et al.*, Reporting on covariate adjustment in randomised controlled trials before and after revision of the 2001 CONSORT statement: a literature review *Trials* 2010, **11**:59

REFERENCES

Altman, D. G. (1985) Comparability of randomized groups. *The Statistician* 34(1): 125-136.

Altman, D. G. (2005) Adjustment for covariate imbalance. In: Armitage, P. and Colton, T. eds. *Encyclopedia of Biostatistics*. 2nd edition. Chichester: John Wiley: 1273-1278.

Altman, D. G., et al. (2012) Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. *PLoS Medicine* 9(5): e1001216.

Altman, D. G., et al. (2001) The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Annals of Internal Medicine* 134(8): 663-694.

Askie, L. M., et al. (2007) Antiplatelet agents for prevention of pre-eclampsia: a meta-analysis of individual patient data. *Lancet* 369(9575): 1791-1798.

Assmann, S. F., et al. (2000) Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 355(9209): 1064-1069.

August, P., et al. (1994) Sustained release, low-dose aspirin ameliorates but does not prevent preeclampsia (PE) in a high risk population. *9th World Congress of the ISSHP*, Sydney, Australia, 72.

Austin, P. C., et al. (2010) A substantial and confusing variation exists in handling of baseline covariates in randomized controlled trials: a review of trials published in leading medical journals. *Journal of Clinical Epidemiology* 63(2): 142-153.

Bickel, P. J. and Lehmann E. L. (1975) Descriptive statistics for nonparametric models I: introduction. *The Annals of Statistics* 3(5): 1038-1044.

Bilano, V. L., et al. (2014) Risk factors of pre-eclampsia/eclampsia and its adverse outcomes in low- and middle-income countries: a WHO secondary analysis. *PLoS One* 9(3): e91198.

Burgess, S., et al. (2013) Combining multiple imputation and meta-analysis with individual participant data. *Statistics in Medicine* 32(26): 4499-4514.

Burton, A., et al. (2006) The design of simulation studies in medical statistics. *Statistics in Medicine* 25(24): 4279-4292.

Byaruhanga, R. N., Chipato, T. and Rusakaniko, S. (1998) A randomized controlled trial of low-dose aspirin in women at risk from pre-eclampsia. *International Journal of Gynaecology and Obstetrics* 60(2): 129-135.

Caritis, S., et al. for the National Institute of Child Health and Human Development Network of Maternal-Fetal Medicine Units. (1998) Low-dose aspirin to prevent preeclampsia in women at high risk. *New England Journal of Medicine* 338(11): 701-705.

Chan, A. W. and Altman, D. G. (2005a) Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *British Medical Journal* 330(7494): 753-759.

Chan, A. W. and Altman, D. G. (2005b) Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet* 365(9465): 1159-1162.

Chan, A. W., et al. (2008) Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *British Medical Journal* 337: a2299.

Chastang, C., Byar, D. and Piantadosi, S. (1988) A quantitative study of the bias in estimating the treatment effect caused by omitting a balanced covariate in survival models. *Statistics in Medicine* 7(12): 1243-1255.

Christensen, E., et al. (1985) Beneficial effect of azathioprine and prediction of prognosis in primary biliary cirrhosis: final results of an international trial. *Gastroenterology* 89(5): 1084-1091.

Chu, R., et al. (2012) Assessment and implication of prognostic imbalance in randomized controlled trials with a binary outcome − a simulation study. *PLoS One* 7(5): e36677.

Collaborative Low-dose Aspirin Study in Pregnancy Collaborative Group (CLASP). (1994) CLASP: a randomised trial of low-dose aspirin for the prevention and treatment of pre-eclampsia among 9364 pregnant women. *Lancet* 343(8898): 619-629.

Committee for Proprietary Medicinal Products. (2015) Guideline on adjustment for baseline covariates in clinical trials. EMA/CHMP/295050/2013.

Cox, D. R. and Snell, E. J. (1989) *Analysis of Binary Data*. 2nd edition. London: Chapman and Hall.

Debray, T. P., et al. (2013) Individual participant data meta-analysis for a binary outcome: one-stage or two-stage? *PLoS One* 8(4): e60650.

Deeks, J. J., et al. (2003) Evaluating non-randomised intervention studies. *Health Technology Assessment* 7(27): 1-173.

DerSimonian, R. and Laird, N. (1986) Meta-analysis in clinical trials. *Controlled Clinical Trials* 7(3): 177-188.

Estudo Colorativo para Prevencao da Pre-eclampsia com Aspirina (ECCPA) Collaborative Group. (1996) ECPPA: randomised trial of low dose aspirin for the prevention of maternal and fetal complications in high risk pregnant women. *British Journal for Obstetrics and Gynaecology* 103(1): 39-47.

Fairbank, J., et al. (2005) Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: the MRC spine stabilisation trial. *British Medical Journal* 330(7502): 1233.

Ferrier, C., North, R. and Kincaid-Smith, P. (1996) Low dose aspirin delays the onset of pre-eclampsia in pregnancies with abnormal uteroplacental circulation. *10th World Congress of the ISSHP*, Seattle, Washington, USA, 151.

Forsythe, A. B. and Stitt, F. W. (1977) *Randomization or minimization in the treatment assignment of patient trials: validity and power of tests*. Technical Report No. 28. Health Sciences Computing Facility, University of California, Los Angeles.

Gail, M. H., Wieand, S. and Piantadosi, S. (1984) Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* 71(3): 431-444.

Golding, J. for the Jamaica Low Dose Aspirin Study Group (1998) A randomised trial of low dose aspirin for primiparae in pregnancy. *British Journal of Obstetrics and Gynaecology* 105(3): 293-299.

Goldstein, H., et al. (2000) Meta-analysis using multilevel models with an application to the study of class size effects. *Applied Statistics* 49(3): 399-412.

Gray, L. J., Bath, P. M. and Collier, T. (2009) Should stroke trials adjust functional outcome for baseline prognostic factors? *Stroke* 40(3): 888-894.

Greenland, S., Robins, J. M. and Pearl, J. (1999) Confounding and collapsibility in causal inference. *Statistical Science* 14(1): 29-46.

Guo, J. and Geng, Z. (1995) Collapsibility of logistic regression coefficients. *Journal of the Royal Statistical Society B* 57(1): 263-267.

Hanney, M., et al. (2012) Memantine for dementia in adults older than 40 years with Down's syndrome (MEADOWS): a randomised, double-blind,

placebo-controlled trial. *Lancet* 379(9815): 528-536.

Hauck, W. W., Anderson, S. and Marcus, S. M. (1998) Should we adjust for covariates in nonlinear regression analyses of randomized trials? *Controlled Clinical Trials* 19(3): 249-256.

Hauck, W. W., et al. (1991) A consequence of omitted covariates when estimating odds ratios. *Journal Clinical Epidemiology* 44(1): 77-81.

Hauth, J. C., et al. (1993) Low-dose aspirin therapy to prevent preeclampsia. *American Journal of Obstetrics and Gynecology* 168(4): 1083-1091; discussion: 91-93.

Heinze, G. (2006) A comparative investigation of methods for logistic regression with separated or nearly separated data. *Statistics in Medicine* 25(24): 4216-4226.

Heinze, G. and Schemper, M. (2002) A solution to the problem of separation in logistic regression. *Statistics in Medicine* 21(16): 2409-2419.

Hermida, R. C., Ayala, D. E. and Iglesias, M. (2003) Administration time-dependent influence of aspirin on blood pressure in pregnant women. *Hypertension* 41(3 Pt 2): 651-656.

Hernandez, A. V., Steyerberg, E. W. and Habbema, J. D. (2004). Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. *Journal of Clinical Epidemiology* 57(5): 454-460.

Hernandez, A. V., Eijkemans, M. J. and Steyerberg, E. W. (2006) Randomized controlled trials with time-to-event outcomes: how much does pre-specified covariate adjustment increase power? *Annals of Epidemiology* 16(1): 41-48.

Hernandez, A. V., et al. (2005) Subgroup analysis and covariate adjustment in randomized clinical trials of traumatic brain injury: a systematic review. *Neurosurgery* 57(6): 1244-1253.

Hewitt, C. E., Torgerson, D. J. and Berger, V. W. (2009) Potential for technical errors and subverted allocation can be reduced if certain guidelines are followed: examples from a web-based survey. *Journal of Clinical Epidemiology* 62(3): 261-269.

Hirji, K. F. and Fagerland, M. W. (2009) Outcome based subgroup analysis: a neglected concern. *Trials* 10: 33.

International Conference on Harmonisation E9 Expert Working Group. (1999) International Conference on Harmonisation Harmonised Tripartite Guideline: Statistical principles for clinical trials. *Statistics in Medicine* 18(15): 1905-1942.

Jackson, D., et al. (2009) Systematically missing confounders in individual participant data meta-analysis of observational cohort studies. *Statistics in Medicine* 28(8): 1218-1237.

Jolani, S., et al. (2015) Imputation of systematically missing predictors in an individual participant data meta-analysis: a generalized approach using MICE. *Statistics in Medicine* 34(11): 1841-1863.

Kahan, B. C., et al. (2014) The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. *Trials* 15: 139.

Kane, R. L., Wang, J. and Garrard, J. (2007) Reporting in randomized clinical trials improved after adoption of the CONSORT statement. *Journal of Clinical Epidemiology* 60(3): 241-249.

Kincaid-Smith, P., et al. (1995) Prevention of pre-eclampsia in high risk women with renal disease: A prospective randomized trial of heparin and dipyridamole. *Nephrology* 1: 297-300.

Lau, J., Schmid, C. H. and Chalmers, T. C. (1995) Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. *Journal of Clinical Epidemiology* 48(1): 45-57.

Lingsma, H., Roozenbeek, B. and Steyerberg, E. (2010) Covariate adjustment increases statistical power in randomized controlled trials. *Journal of Clinical Epidemiology* 63(12): 1391; author reply 92-93.

Michael, C. A. and Walters, B. N. J. (1993) Low-dose aspirin in the prevention of pre-eclampsia: current evaluation. In: Teoh, E. S., Shan Ratnam S. and Macnaughton, M. eds. *Maternal Physiology and Pathology: The Current Status of Gynaecology and Obstetrics Series*. Volume 4. Carnforth: Parthenon Pub Group Ltd: 183-189.

Moher, D., Jones, A. and Lepage, L. (2001) Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *Journal of the American Medical Association* 285(15): 1992-1995.

Molyneux, A. J., et al. (2005) International subarachnoid aneurysm trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2,143 patients with ruptured intracranial aneurysms: a randomised comparison of effects on survival, dependency, seizures, rebleeding, subgroups, and aneurysm occlusion. *Lancet* 366(9488): 809-817.

Morris, J. M., et al. (1996) A randomized controlled trial of aspirin in patients with abnormal uterine artery blood flow. *Obstetrics and Gynecology* 87(1): 74-78.

Mullner, M., Matthews, H. and Altman, D. G. (2002) Reporting on statistical methods to adjust for confounding: a cross-sectional survey. *Annals of Internal Medicine* 136(2): 122-126.

Negassa, A. and Hanley, J. A. (2007) The effect of omitted covariates on confidence interval and study power in binary outcome analysis: a simulation study. *Contemporary Clinical Trials* 28(3): 242-248

National Institute for Health and Care Excellence. (2013) *Hypertension in pregnancy.* [online] Available from: http://www.nice.org.uk/guidance/qs35/chapter/quality-statement-2-antenatal-assessment-of-pre-eclampsia-risk (accessed on 27/5/2015).

Neuhaus, J. M., Kalbfleisch, J. D. and Hauck, W. W. (1991) A comparison of cluster-specific and population-average approaches for analyzing correlated binary data. *International Statistical Review* 59(1): 25-35.

Nicholl, J. (2007) Case-mix adjustment in non-randomised observational evaluations: the constant risk fallacy. *Journal of Epidemiology and Community Health* 61(11): 1010-1013.

O'Connor, C. M., Mentz, R. J. and Whellan, D. J. (2011) Covariate adjustment in heart failure randomized controlled clinical trials: a case analysis of the HF-ACTION trial. *Heart Failure Clinics* 7(4): 497-500.

Ota, E., et al. (2014) Risk factors and adverse perinatal outcomes among term and preterm infants born small-for-gestational-age: secondary analyses of the WHO Multi-Country Survey on Maternal and Newborn Health. *PLoS One* 9(8): e105155.

Pang, M., Kaufman, J. S. and Platt, R. W. (2013) Mixing of confounding and non-collapsibility: a notable deficiency of the odds ratio. *American Journal of Cardiology* 111(2): 302-303.

Pildal, J., et al. (2007) Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *International Journal of Epidemiology* 36(4): 847-857.

Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* 4(1): 12-35.

Pinheiro, J. C. and Chao, E. C. (2006). Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models. *Journal of Computational and Graphical Statistics* 15(1): 58-81.

Pocock, S. J., et al. (2002) Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Statistics in Medicine* 21(19): 2917-2930.

Railton, A. and Davey, D. A. (1988) Aspirin and dypyridamole in the prevention of pre-eclampsia: effect on plasma prostanoids 6-keto PG1a and TXB2 and clinical outcome of pregnancy. *6th World Congress of the ISSHP*, Toronto, Canada, 60.

Resche-Rigon, M., et al. (2013) Multiple imputation for handling systematically missing confounders in meta-analysis of individual participant data. *Statistics in Medicine* 32(28): 4890-4905.

Roberts, C. and Torgerson, D. J. (1999) Understanding controlled trials: baseline imbalance in randomised controlled trials. *British Medical Journal* 319(7203): 185.

Robinson, L. D. and Jewell, N. P. (1991) Some surprising results about covariate adjustment in logistic regression models. *International Statistical Review* 58(2): 227-240.

Rochon, J. (1999) Issues in adjusting for covariates arising postrandomization in clinical trials. *Drug Information Journal* 33(4): 1219-1228.

Rogers, M. S., Fung, H. Y. and Hung, C. Y. (1999) Calcium and low-dose aspirin prophylaxis in women at high risk of pregnancy-induced hypertension. *Hypertension in Pregnancy* 18(2): 165-72.

Roozenbeek, B., et al. (2009) Baseline characteristics and statistical power in randomized controlled trials: selection, prognostic targeting, or covariate adjustment? *Critical Care Medicine* 37(10): 2683-2690.

Rotchell, Y. E., et al. (1998) Barbados Low Dose Aspirin Study in Pregnancy (BLASP): a randomised trial for the prevention of pre-eclampsia and its complications. *British Journal of Obstetrics and Gynaecology* 105(3): 286-292.

Royal College of Obstetricians and Gynaecologists. (2013) *The Investigation and Management of the Small-for-Gestational-Age Fetus*. Green-top Guideline No. 31, 2nd Edition.

Royston, P. and Parmar, M. K. (2011) The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in Medicine* 30(19): 2409-2421.

Saquib, N., Saquib, J., and Ioannidis, J. P. A. (2013). Practices and impact of primary outcome adjustment in randomized controlled trials: meta-epidemiologic study. *British Medical Journal* 347: f4313.

Sawilowsky, S. S. (2002) A measure of relative efficiency for location of a single sample. *Journal of Modern Applied Statistical Methods* 1(1): 52.

Schiff, E., et al. (1989) The use of aspirin to prevent pregnancy-induced hypertension and lower the ratio of thromboxane A2 to prostacyclin in relatively high risk pregnancies. *New England Journal of Medicine* 321(6): 351-356.

Senn, S. J. (1989) Covariate imbalance and random allocation in clinical trials. *Statistics in Medicine* 8(4): 467-475.

Sibai, B. M., et al. for the National Institute of Child Health and Human Development Network of Maternal-Fetal Medicine Units. (1993) Prevention of preeclampsia with low-dose aspirin in healthy, nulliparous pregnant women. *New England Journal of Medicine* 329(17): 1213-1218.

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

Simmonds, M. C., et al. (2005) Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clinical Trials* 2(3): 209-217.

Stewart, G. B., et al. (2012) Statistical analysis of individual participant data meta-analyses: a comparison of methods and recommendations for practice. *PLoS One* 7(10): e46042.

Stewart, L. A., et al. (2015) Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. *Journal of the American Medical Association* 313(16): 1657-1665.

Steyerberg, E. W. and Eijkemans, M. J. (2004) Heterogeneity bias: the difference between adjusted and unadjusted effects. *Medical Decision Making* 24(1): 102-104.

Steyerberg, E. W., Bossuyt, P. M. M. and Lee, K. L. (2000) Clinical trials in acute myocardial infarction: should we adjust for baseline characteristics? *American Heart Journal* 139(5): 745-751.

Stigler, S. M. (1973) The asymptotic distribution of the trimmed mean. *The Annals of Statistics* 1(3): 472-477.

Subtil, D., et al. (2003) Aspirin (100 mg) used for prevention of pre-eclampsia in nulliparous women: the Essai Regional Aspirine Mere-Enfant (ERASME) study (Part 1). *British Journal of Obstetrics and Gynaecology* 110(5): 475-84.

Taggart, D. P., et al. (2010) Randomized trial to compare bilateral vs. single internal mammary coronary artery bypass grafting: 1-year results of the Arterial Revascularisation Trial (ART). *European Heart Journal* 31(20): 2470-2481.

The Perinatal Antiplatelet Review of International Studies Collaboration Steering Group on behalf of the Paris Collaboration. (2005) Antiplatelet agents for prevention of pre-eclampsia and its consequences: a systematic review and individual patient data meta-analysis. *BMC Pregnancy Childbirth* 5(1): 7.

Thomas, D., Radji, S. and Benedetti, A. (2014) Systematic review of methods for individual patient data meta-analysis with binary outcomes. *BMC Medical Research Methodology* 14: 79.

Thompson, D. D., et al. (2014) Covariate adjustment had similar benefits in small and large randomized controlled trials. *Journal of Clinical Epidemiology* [Epub ahead of print] doi: 10.1016/j.jclinepi.2014.11.001.

To, M. J., et al. (2013) Are reports of randomized controlled trials improving over time? A systematic review of 284 articles published in high-impact general and specialized medical journals. *PLoS One* 8(12): e84779.

Todd, S. and Sahdra, M. (2001), A case study of the effect of covariate adjustment in a sequential survival clinical trial. *Journal of Biopharmaceutical Statistics* 11(4): 297-311.

Trowman, R., et al. (2007) The impact of trial baseline imbalances should be considered in systematic reviews: a methodological case study. *Journal of Clinical Epidemiology* 60(12): 1229-1233.

Turner, E. L., et al. (2012a) Covariate adjustment increased power in randomized controlled trials: an example in traumatic brain injury. *Journal of Clinical Epidemiology* 65(5): 474-481.

Turner, L., et al. (2012b) Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Systematic Reviews* 1: 60.

Turner, R. M., et al. (2000) A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine* 19(24): 3417-3432.

Uzan, S., et al. (1989) Idiopathic recurrent fetal growth retardation and aspirin-dipyridamole therapy. *American Journal of Obstetrics and Gynecology* 160(3): 763-764.

Uzan, S., et al. (1991) Prevention of fetal growth retardation with low-dose aspirin: findings of the EPREDA trial. *Lancet* 337(8755): 1427-1431.

Vainio, M., et al. (2002) Low dose acetylsalicylic acid in prevention of pregnancy-induced hypertension and intrauterine growth retardation in women with bilateral uterine artery notches. *British Journal of Obstetrics and Gynaecology* 109(2): 161-167.

Wang, Z. and Li, W. (1996) A prospective randomized placebo-controlled trial of low-dose aspirin for prevention of intra-uterine growth retardation. *Chinese Medical Journal (English)* 109(3): 238-242.

Yu, C. K., et al. (2003) Randomized controlled trial using low-dose aspirin in the prevention of pre-eclampsia in women with abnormal uterine artery Doppler at 23 weeks' gestation. *Ultrasound in Obstetrics and Gynecology* 22(3): 233-239.

Yusuf, S., et al. (1991) Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *Journal of the American Medical Association* 266(1): 93-98.

Zhang, Z., et al. (2005) Importance of baseline distribution of proteinuria in renal outcomes trials: lessons from the reduction of endpoints in NIDDM with the angiotensin II antagonist losartan (RENAAL) study. *Journal of the American Society of Nephrology* 16(6): 1775-1780.

Zimmerman, P., et al. (1997) Effect of low dose aspirin treatment on vascular resistance in the uterine, uteroplacental, renal and umbilical arteries. *European Journal of Ultrasound* 5: 17-30.