

# **META-ANALYSIS OF MULTIPLE OUTCOMES: FUNDAMENTALS AND APPLICATIONS**

by

**Khajak Ishak**

Department of Epidemiology and Biostatistics

McGill University, Montreal

A thesis submitted to McGill University in  
partial fulfillment of the requirements of  
the degree of Doctor of Philosophy.

August 2006

© Khajak Ishak, 2006



Library and  
Archives Canada

Published Heritage  
Branch

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque et  
Archives Canada

Direction du  
Patrimoine de l'édition

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*  
ISBN: 978-0-494-27792-8

*Our file* *Notre référence*  
ISBN: 978-0-494-27792-8

#### NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

#### AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

\*\*  
**Canada**

## ABSTRACT

Meta-analyses often consider the effect of a treatment on multiple, possibly related outcomes. Typically, summary estimates are derived from outcome-specific meta-analyses. Alternatively, a joint analysis can be conducted with a multivariate meta-analysis model, which also quantifies the correlation between the outcomes. This dissertation presents findings from analyses examining issues pertaining to the accuracy of the multivariate approach and its application to meta-analyses of longitudinal studies.

Correlations measured in multivariate meta-analysis models can provide added insight about the treatment and disease. To be useful, however, the measured correlations must reflect the underlying *biological* relationships between treatment effects. I demonstrate, however, that correlations measured across studies may often be distorted by associations between the endpoints or random errors affecting outcomes within studies in a similar way. Thus, correlations measured in multivariate meta-analyses can be misleading.

To properly weight the contribution of each study, the variance of sampling distributions of each estimate is fixed to its observed value. In the multivariate case, however, the sampling distribution also involves covariances between effect estimates on the different outcomes. These are rarely available and must, therefore, be approximated from external information. I evaluated the impact of errors in these approximations on estimates of the parameters of the model in a simulation study. Summary effects and heterogeneity were estimated accurately, but the correlation parameter was prone to possibly large biases and lacked precision.

Longitudinal studies often report treatment effects measured at different times. Multivariate meta-analyses can account for the correlations inherent to this type of data. Alternatively, random-effects can be specified to capture the marginal correlations. I contrasted these and the standard time-specific meta-analysis approaches using data from a review of studies of deep brain stimulation.

Multivariate models, and to a slightly lesser extent the random-effects models, provided better fit and more precise estimates in the interval with fewest observations. These models were also less affected by an apparently outlying observation. This suggests a potential borrowing of information from estimates at other times.

This work builds on research about the potential advantages and limitations of joint meta-analyses of multiple outcomes.

## RÉSUMÉ

Les méta-analyses d'essais cliniques considèrent souvent plusieurs résultats thérapeutiques qui peuvent être corrélés. Typiquement, des estimés sommaires de l'effet de la thérapie sont dérivés séparément par méta-analyses différentes pour chaque résultat thérapeutique. Alternativement, une analyse commune peut être effectuée avec un modèle multidimensionnel qui mesure également la corrélation entre les résultats. Cette dissertation présente des résultats d'études examinant la fiabilité de la méthode multidimensionnelle et son application pour la méta-analyse d'études longitudinales.

Les corrélations mesurées dans les modèles multidimensionnelles de méta-analyse peuvent fournir des informations supplémentaires au sujet du traitement et de la maladie qui ne seraient pas disponibles autrement. Pour être utile, cependant, les corrélations mesurées doivent refléter les associations biologiques existant entre les effets du traitement. Or, mon analyse indique que les corrélations entre les effets mesurés dans les études différentes peuvent souvent être biaisées par des associations entre les résultats thérapeutiques ou des erreurs aléatoires affectant les effets mesurés d'une manière semblable.

Pour pondérer correctement la contribution de chaque étude, la variance des distributions échantillonnale (vraisemblance) de chaque évaluation est fixée à sa valeur observée. Dans le cas multidimensionnel, cependant, la distribution échantillonnale comprend également les covariances entre les estimations des effets. Ceux-ci sont rarement disponibles et doivent donc être approximés en utilisant de l'information externe. J'ai évalué l'impact des erreurs dans ces approximations sur l'estimation des paramètres du modèle. Les mesures sommaires des effets de la thérapie et l'hétérogénéité entre essais sont généralement estimées sans biais, mais des erreurs importantes sont possibles dans les estimés de la corrélation entre les effets de la thérapie, qui étaient aussi comparativement moins précis.

Les rapports d'études longitudinales publient souvent l'effet de la thérapie mesuré à plusieurs occasions. Les modèles multidimensionnels de méta-analyse peuvent tenir compte de ces corrélations inhérentes à ce type de données. Alternativement, des effets aléatoires (modèles à effets mixtes) peuvent être spécifiés pour refléter les corrélations marginales. J'ai comparé ces deux types de modèles avec l'approche plus typique de méta-analyser les données mesurées à chaque occasion séparément en utilisant des données d'un examen des études de la stimulation profonde du cerveau. Le modèle multidimensionnel et, à un moindre niveau, celui à effets mixtes sont mieux ajustés aux données et produisent des estimés plus précis dans l'intervalle avec moins d'observations. Ces modèles sont aussi moins affectés par une observation apparemment extrême. Ceci suggère un emprunt potentiel d'information entre les évaluations effectuées à différentes occasions.

Mes travaux contribuent à la recherche sur les avantages et les limitations potentielles des méthodes de méta-analyse commune des résultats thérapeutiques.

## PREFACE

### FORMAT OF THE THESIS

This is a manuscript-based dissertation, and as such, consists of a collection of three articles prepared from research conducted to address the objectives of the thesis. Requirements for the preparation of manuscript-based theses can be found at <http://www.mcgill.ca/gps/programs/thesis/guidelines/preparation/>. Each article is presented as a separate chapter in this document. A preamble is included at the start of each chapter to describe the rationale and motivation for the paper and how it relates to the general objectives of the thesis. Since space and format are somewhat restricted in manuscripts, additional material from analyses that were not reported in detail in the papers is included in appendix at the end of each chapter when necessary. Findings from the three papers are related and discussed in the last chapter. References from all chapters are listed at the end of this document. Tables and figures of results are placed following the text of the manuscript in each chapter.

### CONTRIBUTIONS OF AUTHORS

The topic of this dissertation was chosen in conjunction with my supervisor, Dr. Robert Platt. The questions examined in the manuscripts comprising this thesis were identified following my review of the literature and refined in discussion with Dr. Platt.

I developed the core ideas for each paper, designed the simulations and formulated the specifications for the analyses for each paper, programmed and carried them out, summarized and interpreted the findings and wrote the manuscripts. Dr. Platt provided guidance in developing the simulation and analytic strategies, interpreting findings and reviewed the manuscripts. As members of my thesis committee, Dr. Lawrence Joseph and Dr. James Hanley contributed to discussions of issues pertaining to the development of analyses and reviewed the manuscripts. Dr. Jaime Caro provided the data used in the third manuscript (Meta-analysis of Longitudinal Studies).

## **ACKNOWLEDGEMENTS**

I would like to sincerely thank Dr. Robert Platt for his guidance and skillful supervision of my research. He was always available for meetings, even during his sabbatical, and advised me on both my studies and career. I am also grateful to Dr. Lawrence Joseph and Dr. James Hanley, who as members of my thesis committee contributed significantly to discussions that determined the course of my project and carefully reviewed drafts of my manuscripts. I feel privileged to have had the opportunity to work with and learn from them all.

I would also like to thank Dr. Jaime Caro for providing the data needed for one of the manuscripts and for advising me on various aspects of my studies. In fact, my decision to pursue doctoral studies was very much inspired by the advice and encouragement I received from Dr. Caro and Dr. Hanley, for which I am truly grateful, as this has been a most enriching experience. I wish to express my gratitude to all the professors who taught me - their enthusiasm and passion was very stimulating.

In addition, I am thankful for funding I received from the Natural Sciences and Engineering Research Council in the form a graduate scholarship, which facilitated the timely completion of this thesis.

On a more personal note, I cannot say enough to thank my wife, Daphne, for all her support and encouragement during the course of my studies. Her patience, compassion and optimism were invaluable.

Finally, I dedicate this dissertation to my parents, Levon and Sonya, for teaching me the value of hard work and sacrifice by example.

# TABLE OF CONTENTS

<b>INTRODUCTION.....</b>	<b>1</b>
1.1 BACKGROUND .....	1
1.2 OBJECTIVES OF THESIS.....	3
<b>LITERATURE REVIEW.....</b>	<b>6</b>
2.1 BRIEF INTRODUCTION TO META-ANALYSIS.....	6
2.1.1 <i>Systematic Review of the Literature</i> .....	7
2.1.2 <i>Measures of Effect and Other Data of Interest</i> .....	8
2.2 META-ANALYSIS OF A SINGLE OUTCOME .....	9
2.2.1 <i>Effect-Based (Approximate) Methods</i> .....	10
The DerSimonian-Laird Model .....	10
Other Effect-based Approaches.....	16
2.2.2 <i>Event-Based (Exact) Methods</i> .....	17
2.3 JOINT META-ANALYSIS OF MULTIPLE OUTCOMES .....	18
2.3.1 <i>Multivariate Meta-Analysis Models</i> .....	19
Model Specification.....	19
Estimation of Parameters.....	21
2.3.2 <i>Applications of the Multivariate Normal Model</i> .....	23
Control-Rate Regression with Multivariate Models .....	23
Survival Rate Ratios over Time.....	26
Multiple Outcomes .....	29
2.3.3 <i>Other Models for Multivariate Meta-Analysis</i> .....	32
2.3.4 <i>Meta-analysis of Longitudinal Studies</i> .....	33
2.3.5 <i>Benefits and Reliability of Multivariate Methods</i> .....	34
Statistical Convenience.....	34
Added Insight and Multivariate Hypotheses.....	35
Correlations in Multivariate Meta-Analyses.....	35
Improved Accuracy and Precision.....	36
Publication Bias.....	37
2.4 OBJECTIVES OF THESIS .....	37
<b>(MANUSCRIPT I) CORRELATIONS IN MULTIVARIATE META-ANALYSIS: WHAT ASSOCIATIONS ARE BEING MEASURED?.....</b>	<b>38</b>
PREAMBLE .....	38
SUMMARY .....	40
3.1 INTRODUCTION.....	40

3.2 CONCEPTUAL MODEL OF CORRELATIONS .....	43
3.2.1 <i>The Disease Process</i> .....	44
3.2.2 <i>The Effect of Treatment</i> .....	45
3.2.3 <i>Between-Study Variability</i> .....	46
3.2.4 <i>Example</i> .....	47
3.3 METHODS .....	48
3.3.1 <i>Correlation between Marginal Log-ORs</i> .....	49
3.3.2 <i>Choice of Parameter Values</i> .....	50
3.3.3 <i>Analyses</i> .....	52
3.4 RESULTS .....	52
3.5 DISCUSSION .....	55
3.6 TABLES AND FIGURES .....	59
3.7 APPENDIX: ADDITIONAL MATERIAL .....	66
A3.7.1 <i>Tables Corresponding to Figures in Manuscript</i> .....	66
A3.7.2 <i>Results from Simulations Based on Log-Odds</i> .....	71
<b>(MANUSCRIPT II) IMPACT OF APPROXIMATING OR IGNORING WITHIN-STUDY COVARIANCES IN MULTIVARIATE META-ANALYSES .....</b>	<b>77</b>
PREAMBLE .....	77
SUMMARY .....	79
4.1 INTRODUCTION .....	79
4.2 MULTIVARIATE META-ANALYSIS .....	82
4.2.2 <i>Multivariate Data from Meta-Analyses</i> .....	82
4.2.2 <i>Multivariate Model</i> .....	83
4.2.3 <i>Impact of Errors in Approximating Within-Study Covariances (<math>S_{12}</math>)</i> .....	84
Estimation .....	84
Impact of Errors .....	84
4.3 SIMULATION OF MULTIVARIATE META-ANALYSIS DATA .....	86
4.3.1 <i>Simulation of Study Results</i> .....	86
4.3.2 <i>Simulation Scenarios</i> .....	88
4.3.3 <i>Approximating Within-Study Covariances</i> .....	89
4.4 ANALYSIS AND COMPARISON MEASURES .....	90
4.5 RESULTS .....	91
4.6 DISCUSSION .....	95
4.7 TABLES .....	100
4.8 APPENDIX: ADDITION MATERIAL .....	106
A4.8.1 <i>Description of Some Estimation Procedures</i> .....	106
A4.8.2 <i>Sensitivity Analyses</i> .....	107

Description of Scenarios.....	107
Results.....	108
<b>(MANUSCRIPT III) META-ANALYSIS OF LONGITUDINAL STUDIES .....</b>	<b>113</b>
PREAMBLE .....	113
SUMMARY .....	115
5.1 INTRODUCTION.....	115
5.2 MODELS FOR LONGITUDINAL META-ANALYTIC DATA .....	117
5.2.1 <i>General Linear Mixed Model for Longitudinal Data</i> .....	117
5.2.2 <i>Mixed Models for Meta-Analysis of Longitudinal Data</i> .....	119
5.2.3 <i>Specification of Correlation Structure</i> .....	121
Random Study-Effect .....	122
Random Time-Effects.....	122
Random Within- and Between Study Time Effects.....	124
5.2.4 <i>Implementation of the Model</i> .....	125
5.3 CASE STUDY.....	125
5.3.1 <i>Deep-Brain Stimulation in Parkinson's Disease</i> .....	125
5.3.2 <i>Models</i> .....	126
5.3.3 <i>Specification of Covariance Matrices</i> .....	128
5.4 RESULTS.....	129
5.5 COMPARISON OF MODELS .....	133
5.6 DISCUSSION.....	135
5.7 FIGURES AND TABLES .....	139
<b>DISCUSSION AND CONCLUSION.....</b>	<b>146</b>
6.1 RELIABILITY OF MULTIVARIATE META-ANALYSIS .....	146
6.2 APPLICATIONS OF MULTIVARIATE META-ANALYSIS .....	148
6.3 ADVANTAGES OF A MULTIVARIATE APPROACH .....	150
6.4 LIMITATIONS OF ANALYSES .....	152
6.5 SUGGESTIONS FOR FUTURE RESEARCH .....	154
6.6 SUMMARY OF CONCLUSIONS.....	155
<b>REFERENCES.....</b>	<b>157</b>

## LIST OF TABLES

**Note:** The first digit in table numbers indicates the chapter in which they appear.  
Table numbers ending in (A) identify tables appearing in Appendix.

<b>Table 3.1</b>	Values of inputs used to create scenarios for simulations from the conceptual model.	59
<b>Table 3.2(A)</b>	Correlations between marginal log odds-ratios when the events are dependent, effects of the treatment are independent and between-study random-effects on log-odds and log-ORs are not correlated.	66
<b>Table 3.3(A)</b>	Correlations between marginal log odds-ratios when the events are independent, effects of the treatment are dependent and between-study random-effects on log-odds and log-ORs are not correlated.	67
<b>Table 3.4(A)</b>	Correlations between marginal log odds-ratios when the events and effects are dependent and between-study random-effects on log-odds and log-ORs are not correlated.	68
<b>Table 3.5(A)</b>	Correlations between marginal log odds-ratios when events and treatment effects are dependent and between-study random-effects on log-odds and log-ORs are correlated.	69
<b>Table 3.6(A)</b>	Correlations between marginal log odds-ratios when the treatment has an effect on only one of the endpoints, events are dependent and between-study random-effects on log-odds and log-ORs are correlated.	70
<b>Table 3.7(A)</b>	Correlations between marginal log odds-ratios when the events are dependent, effects of the treatment are independent and between-study random-effects on log-odds of treated and control groups are not correlated.	72

<b>Table 3.8(A)</b> Correlations between marginal log odds-ratios when the events are independent, effects of the treatment are dependent and between-study random-effects on log-odds of treated and control groups are not correlated.	73
<b>Table 3.9(A)</b> Correlations between marginal log odds-ratios when the events and effects are dependent and between-study random-effects on log-odds of treated and control groups are not correlated.	74
<b>Table 3.10(A)</b> Correlations between marginal log odds-ratios when events and treatment effects are dependent and between-study random-effects on log-odds of treated and control groups are correlated.	75
<b>Table 3.11(A)</b> Correlations between marginal log odds-ratios when the treatment has an effect on only one of the endpoints, events are dependent and between-study random-effects on log-odds of treated and control groups are correlated.	76
<b>Table 4.12</b> Median (5 <sup>th</sup> , 95 <sup>th</sup> percentiles) of estimates obtained from 300 simulated meta-analyses from scenario 1a, where we assume $(S_{11}^2, S_{22}^2) = 2(D_{11}^2, D_{22}^2) = (10.0, 5.0)$ and $\rho_S = \rho_D = 0.25$ .	100
<b>Table 4.13</b> Median (5 <sup>th</sup> , 95 <sup>th</sup> percentiles) of estimates obtained from 300 simulated meta-analyses from scenario 1b, where we assume $(S_{11}^2, S_{22}^2) = 2(D_{11}^2, D_{22}^2) = (10.0, 5.0)$ , and $\rho_S = 2\rho_D = 0.5$ .	101

<b>Table 4.14</b>	Median (5 <sup>th</sup> , 95 <sup>th</sup> percentiles) of estimates obtained from 300 simulated meta-analyses from scenario 2a, which is based on scenario 1b ( $(S_{11}^2, S_{22}^2) = 2(D_{11}^2, D_{22}^2) = (10.0, 5.0)$ , and $\rho_S = 2\rho_D = 0.5$ ), with sample sizes reduced by a factor of 5 (i.e., five times smaller).	102
<b>Table 4.15</b>	Median (5 <sup>th</sup> , 95 <sup>th</sup> percentiles) of estimates obtained from 300 simulated meta-analyses from scenario 2b, which is based on scenario 1b ( $(S_{11}^2, S_{22}^2) = 2(D_{11}^2, D_{22}^2) = (10.0, 5.0)$ , and $\rho_S = 2\rho_D = 0.5$ ), with sample sizes reduced by a factor of 10 (i.e., 10 times smaller).	103
<b>Table 4.16</b>	Median (5 <sup>th</sup> , 95 <sup>th</sup> percentiles) of estimates obtained from 300 simulated meta-analyses from scenario 3a, which is based on scenario 1b ( $(S_{11}^2, S_{22}^2) = 2(D_{11}^2, D_{22}^2) = (10.0, 5.0)$ , and $\rho_S = 2\rho_D = 0.5$ ), with between-study variances reduced by a factor of 5 (i.e., 5 times smaller).	104
<b>Table 4.17</b>	Median (5 <sup>th</sup> , 95 <sup>th</sup> percentiles) of estimates obtained from 300 simulated meta-analyses from scenario 3b, which is based on scenario 1b ( $(S_{11}^2, S_{22}^2) = 2(D_{11}^2, D_{22}^2) = (5.0, 2.5)$ , and $\rho_S = 2\rho_D = 0.5$ ), with between-study variances reduced by a factor of 10 (i.e., 10 times smaller)	105
<b>Table 4.18(A)</b>	Median (5 <sup>th</sup> , 95 <sup>th</sup> percentiles) of estimates obtained from 300 simulated meta-analyses based scenario 1b ( $(S_{11}^2, S_{22}^2) = 2(D_{11}^2, D_{22}^2) = (10.0, 5.0)$ , and $\rho_S = 2\rho_D = 0.5$ ). Some studies may not report both outcomes. We assume that the probability that a study does not report one of the outcomes is 0.50; given that a study does not report one of the outcomes, the probability that results for event A is not reported is 0.25.	110

<b>Table 4.19(A)</b> Median (5 <sup>th</sup> , 95 <sup>th</sup> percentiles) of estimates obtained from 300 simulated meta-analyses of five studies based on scenario 1b ( $(S_{11}^2, S_{22}^2) = 2(D_{11}^2, D_{22}^2) = (10.0, 5.0)$ , and $\rho_S = 2\rho_D = 0.5$ ).	111
<b>Table 4.20(A)</b> Median (5 <sup>th</sup> , 95 <sup>th</sup> percentiles) of estimates obtained from 300 simulated meta-analyses based on scenario 1b ( $(S_{11}^2, S_{22}^2) = 2(D_{11}^2, D_{22}^2) = (10.0, 5.0)$ , and $\rho_S = 2\rho_D = 0.5$ ), where between-subject variances and correlation are allowed to vary randomly from one subject to another.	112
<b>Table 5.21</b> Correlation/covariance matrix of observed estimates of effect. Diagonal elements are the variances, the upper diagonal shows the correlations and the lower diagonal shows the covariances between the observed estimates.	143
<b>Table 5.22</b> Summary effect and between-study variance estimates and 95% confidence intervals from various meta-analysis models.	144
<b>Table 5.23</b> Summary effect and between-study variance and 95% confidence intervals of between-study variance parameters from meta-regressions including mean disease duration and mean baseline UPDRS-motor scores of patients in each study.	145

## LIST OF FIGURES

**Note:** The first digit in figure numbers indicates the chapter in which they appear.

- |                   |                                                                                                                                                                                                                                                                                                                                                         |    |
|-------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figure 3.1</b> | Sufficient component cause diagrams depicting various forms of dependence between events. Each event is assumed to arise from two possible pathways, each represented by a pie diagram. Each slice represents a component cause, which, combined with the remaining component causes of that pathway, are sufficient to produce the onset of the event. | 60 |
| <b>Figure 3.2</b> | Correlations between marginal log odds-ratios when the events are dependent, effects of the treatment are independent and between-study random-effects on log-odds and log-ORs are not correlated.                                                                                                                                                      | 61 |
| <b>Figure 3.3</b> | Correlations between marginal log odds-ratios when the events are independent, effects of the treatment are dependent and between-study random-effects on log-odds and log-ORs are not correlated.                                                                                                                                                      | 62 |
| <b>Figure 3.4</b> | Correlations between marginal log odds-ratios when the events and effects are dependent and between-study random-effects on log-odds and log-ORs are not correlated.                                                                                                                                                                                    | 63 |
| <b>Figure 3.5</b> | Correlations between marginal log odds-ratios when the treatment has an effect on only one of the endpoints, events are dependent and between-study random-effects on log-odds and log-ORs are correlated.                                                                                                                                              | 64 |

- Figure 3.6** Correlations between marginal log odds-ratios when events and treatment effects are dependent and between-study random-effects on log-odds and log-ORs are correlated. 65
- Figure 5.7(a)** Effect estimates by months 3 and 6 with corresponding 95% confidence intervals, and as summary estimates from univariate random-effects models. The size of circles reflects the precision of estimates. 139
- Figure 5.8** Observed effects of DBS of the STN by months 3, 6, 12 and long term and summary effect estimates. Connected points arise from the same study; unconnected points are from studies reporting an effect estimate at a single time. The size of circles reflects the precision of estimates. The arrow points to an apparently extreme observation, which was removed to assess its impact on summary estimates (Figure 5.9). 141
- Figure 5.9** Summary estimates of effect with all observations (solid lines) and when an apparently *outlying* observation in the long-term interval is removed (dashed line). 142

## ABBREVIATIONS

<b>AIC</b>	Akaike's information criterion
<b>AL</b>	Attachment Level
<b>BVN</b>	Bivariate Normal
<b>CORR</b>	Correlation
<b>COV</b>	Covariance
<b>CSR</b>	Cumulative Survival Ratio
<b>DBS</b>	Deep Brain Stimulation
<b>EM</b>	Expectation-Maximization
<b>ETS</b>	Environmental Tobacco Smoke
<b>GLMM</b>	Generalized Linear Mixed Models
<b>GLS</b>	Generalized Least Squares
<b>LRD</b>	Lower Respiratory Disease
<b>ML</b>	Maximum Likelihood
<b>MVN</b>	Multivariate Normal
<b>OR</b>	Odds Ratio
<b>PD</b>	Probing Depth
<b>REML</b>	Restricted or Residual Maximum Likelihood
<b>SAT</b>	Scholastic Aptitude Test
<b>SAT-M</b>	Scholastic Aptitude Test - Math
<b>SAT-V</b>	Scholastic Aptitude Test - Verbal
<b>SE</b>	Standard Error
<b>STN</b>	Sub-thalamic Nucleus
<b>TVN</b>	Trivariate Normal
<b>UPDRS</b>	Unified Parkinson's Disease Rating Scale
<b>VAR</b>	Variance

# *Chapter 1*

## **INTRODUCTION**

### **1.1 BACKGROUND**

In clinical and epidemiological research, meta-analyses are used to pool results from independent experimental or observational studies examining the effect of a new treatment (or exposure). These studies typically examine a series of outcomes that are relevant to the disease in question. Usually, separate meta-analyses are carried out to derive an aggregated measure of the association between the treatment and each outcome independently<sup>1</sup>. Examples of this approach are easy to find in the literature; a representative case is a recent review of clinical trials comparing old (diuretics and  $\beta$ -blockers) and new (calcium channel blockers and ACE inhibitors) antihypertensive medications by Staessen et al.<sup>2</sup>. The authors performed separate analyses to measure the effect of each type of drug on blood-pressure, all cause mortality, cardiovascular mortality, all cardiovascular events, fatal and non-fatal strokes, fatal and non-fatal myocardial infarctions, and congestive heart failure.

In many instances, the endpoints measured in a study will be correlated; this may occur, for instance, when the endpoints measure different aspects of a common disease or endpoints share common risk factors. Multiple endpoints are recognized as part of the broader problem of multiplicity (other examples being repeated measurements and subgroup analyses)<sup>3,4</sup> in the analysis of clinical trials. Analyzing the outcomes separately (and, thus, ignoring the correlations that exist between them) is known to possibly inflate the overall type-I error rate.

The same concerns have been raised in the context of meta-analyses<sup>1</sup> of the effect of a treatment on multiple outcomes. It has been suggested that a *joint* or multivariate meta-analysis of the outcomes should be conducted to account for the correlations. This approach was also expected to improve the accuracy and precision of estimates<sup>5</sup>, allow joint inferences about the outcomes (e.g., multivariate tests)<sup>5</sup>, and provide added insight from the estimation of the magnitude of the correlations between outcomes<sup>6</sup> across studies.

Standard meta-analysis models<sup>7</sup> can be readily extended to handle multivariate outcomes. The unit of analysis becomes the vector of effect estimates for all outcomes and random-effects for each of these are assumed to arise from a multivariate normal distribution. Conditional on the random-effects, the vector of observed estimates are assumed to have a multivariate normal sampling distribution.

Several examples of multivariate meta-analyses have appeared in the literature, with applications like meta-analyses of treatment effects on two or more related endpoints<sup>1,5,9</sup>, occurrence rates of a single endpoint in treated and control groups at different times<sup>6,10</sup>, and in evaluating the relationship between baseline risk and treatment effect<sup>11-13</sup>. These applications have demonstrated the benefits of multivariate meta-analyses with regards to the added insight they can provide, but have not shown any improvements in accuracy or precision of summary estimates compared to standard outcome-specific meta-analyses.

One of the difficulties involved in implementing multivariate meta-analyses is the specification of the covariance matrix of sampling distributions, which define the likelihood of the model. These are typically set to the covariance matrix of the observed estimates and assumed known without error (to weight the relative influence of each study on summary estimates, as with variances in standard meta-analyses). Although the variances of estimates (i.e., diagonal elements of the covariance matrix) are usually reported, the covariances between effect

estimates for different outcomes are rarely available either directly or via calculations from the joint distribution.

This was recognized as a potential limitation in the earliest applications of multivariate models<sup>1,8</sup> and recommendations were made to change reporting practices to allow correct application of the method in the future; however, this advice has had little impact to date. Within-study covariances were available in only one<sup>5</sup> of the multivariate meta-analyses mentioned above. In all other applications, the covariances were specified by either approximating them from external estimates of the correlation between the outcomes<sup>1,8</sup>, assuming independence (i.e., setting covariances to 0)<sup>6,12</sup>, using approximation techniques<sup>10</sup> or incorporating a common within-study correlation nuisance parameter in the model<sup>9</sup>. In the latter case, concern was raised about potential identifiability problems, especially when only a small portion of the included studies reported both outcomes being examined. Thus, approximating the covariances using external information can reduce the *burden* on estimation. It is not clear, however, what impact errors in such approximations may have on estimates of the parameters of the model.

## 1.2 OBJECTIVES OF THESIS

The general objective of this thesis was to assess the reliability of multivariate meta-analyses and to explore new applications. More specifically, the goal was to assess the reliability of the method in measuring the correlation between treatment effects for different outcomes. I was also interested in the impact of inaccurate approximations of within-study covariances, particularly in estimating summary effects and their correlations. Finally, I examined the use of multivariate models in meta-analyses of longitudinal studies, where estimates of effect are reported at various times. These objectives are described in more detail below. Analyses addressing each of these are reported in separate manuscripts, which are presented in the following chapters.

## **1. Correlations in Multivariate Meta-Analyses: What Associations are Being Measured?**

The estimation of correlations is an important advantage of multivariate meta-analyses. These provide added insight about the relationship between the outcomes across studies. In some applications<sup>11,13</sup>, the correlation estimates were used to measure the relationship between baseline risk and effect size. A negative correlation between short- and long-term survival rates following surgery suggested the presence of a “survival of the fittest” phenomenon in a multivariate meta-analysis of Arends et al.<sup>6</sup>. One can imagine other situations where these correlations may be useful. For example, one might be interested in the association between adverse events and efficacy of a treatment. A negative association would suggest that patients who suffer adverse events are less likely to respond, while smaller correlations would indicate that factors associated with efficacy are not the same or correlated to those affecting adverse events.

To be useful, however, these correlations must reflect the true underlying relationships between the effects of the treatment on the various endpoints, and not only the “ecologic” correlations between estimates across studies. I examined the associations reflected in correlations measured in a multivariate meta-analysis of two dichotomous endpoints by formulating a conceptual model of how such data can arise, highlighting various sources of correlations, and assessed how well these would be captured in a joint meta-analysis of the outcomes.

## **2. What is The Impact of Approximating or Ignoring Unknown Within-Study Covariances in Multivariate Meta-Analyses?**

One of the main challenges to the proper use of multivariate meta-analysis models is the lack of reporting of within-study covariances between effect estimates. These must be assumed known without error to ensure

identifiability of estimates of the effect and heterogeneity parameters. This was recognized as a potential limitation in the earliest applications of multivariate models<sup>1,8</sup> and recommendations were made to change reporting practices to allow correct application of the method; however, there appears to have been little change in that regard. Therefore, unknown covariances are usually approximated using external information about the outcomes. It is unclear, however, how errors in these approximations may affect estimates of the parameters of the model. This question was explored in a simulation study where I compared findings from multivariate meta-analyses where within-study covariances were known to those where these were either ignored, overestimated or underestimated.

### **3. Meta-Analysis of Longitudinal Data: An Application of Multivariate Methods**

A special type of multivariate data arises in meta-analyses of studies where the treatment effect is measured at various times. Multivariate models can be used in this situation to jointly analyze data from various time points to account for the correlations that are inherent to this type of data and gain insight about the trend in effects over time. This was illustrated with a meta-analysis of the effect of deep-brain stimulation (DBS) on motor skills in patients with Parkinson's disease. Effect estimates were collected at 3, 6, 12 months and long-term (>12 months). I adapted the multivariate meta-analysis model for this application and contrasted results to those obtained from other approaches such as standard time-specific meta-analyses, which ignore the correlations in the data, and linear mixed models, which account for correlations through shared random-effects. The multivariate approach differs from the latter in that it models within- and between-study correlations separately, while the mixed models capture the *total* or *marginal* correlations.

## *Chapter 2*

# LITERATURE REVIEW

In this chapter, I review and describe the multivariate models that have been proposed to meta-analyze multiple endpoints jointly and discuss the various applications of these models to illustrate their usefulness and limitations. I begin, however, with an overview of the standard methods for meta-analyses a single endpoint, with particular attention to the most popular approaches since the multiple-endpoints models are an extension of these. Ultimately, the focus will be on the joint meta-analysis of two dichotomous endpoints; therefore, much of the descriptions will be from this perspective. However, the models that are described below also apply for continuous measures of effect and can be readily generalized to more than two outcomes.

### **2.1 BRIEF INTRODUCTION TO META-ANALYSIS**

Meta-analysis can be loosely described as the quantitative aggregation of findings from different sources. Although the idea dates back to the beginning of the last century<sup>14</sup>, the term was coined by Glass<sup>15</sup> in 1976 who described it as a “systematic quantitative alternative to narrative literature reviews that enhances the scientific rigor of the review process” (quoted from Cornell and Mulrow<sup>16</sup>). DerSimonian and Laird<sup>7</sup> refer to meta-analysis as “the statistical analysis of a collection of analytic results for the purpose of integrating the findings.” In current practice, an important distinction is made between *systematic reviews*, which consist of the collection of evidence from various sources using a rigorous and methodical approach, and *meta-analyses*, which refer to the aggregation of evidence gathered in systematic reviews<sup>17</sup>. In fact, it is emphasized that not all systematic reviews are appropriate for meta-analysis<sup>7</sup>.

### **2.1.1 SYSTEMATIC REVIEW OF THE LITERATURE**

Meta-analyses must begin with a thorough review of the literature. A number of issues must be taken into account at this stage to ensure the validity of the final results. These are reviewed here without much discussion, as the focus of this project is on the statistical models used in meta-analysis; these and other important aspects of systematic reviews are discussed by Normand<sup>18</sup>.

- Care must be taken to clearly define the objective(s) and question(s) that will be answered by the meta-analysis. This requires specifying treatment and outcome(s) of interest.
- Inclusion and exclusion criteria should be set to guide the selection of the studies to be analyzed. These criteria might pertain (but are not limited) to
  - The design of the studies: e.g., Should both observational and experimental studies be considered? Should trials that did not assign treatment randomly be considered? Should studies of “poor quality” be included (although, caution is advised in selecting based on quality, due to the subjectivity involved).
  - The study populations: e.g., Should studies on restricted populations, like older patients only, be included?
  - Outcome definitions: e.g., if the outcome can be measured in more than one way, a decision should be made on which one(s) will be considered.
- A broad search should be conducted to identify potential studies from the published and unpublished (grey) literature, abstracts from conferences, theses, bibliographies of identified studies and other sources to ensure complete coverage of the existing information. Searches restricted to published studies are prone to publication bias, since studies showing a strong or statistically significant result tend to have better chances of being

published. Omission of non-significant or weak results will yield biased results that exaggerate the effects being measured.

### **2.1.2 MEASURES OF EFFECT AND OTHER DATA OF INTEREST**

The unit of meta-analysis is the estimate of effect reported in each study; this may be measured in a number of different ways, depending on the outcome being examined and, ultimately, depending on the way the data were analyzed and reported in each study. When the response is a continuous measure (e.g., blood pressure, lipid levels, scores on questionnaires, etc), the treatment effect may be reported as a difference in mean response (i.e., difference in mean change from baseline levels or scores) between the treated and comparison groups. In some applications, it is useful to convert the observed effect estimates to an *effect size*<sup>19</sup> - that is, the observed differences are standardized, so that they are scale-free and expressed in terms of standard deviations from zero. This allows easy comparison of results for outcomes that may be measured on different scales, but are not as easily interpreted as the original values.

When the outcome is dichotomous (e.g., occurrence or diagnosis of a disease, death or some other event of interest), the treatment effect is quantified in terms of measures of comparisons of risk. The odds-ratio is perhaps the most commonly used measure; however, analyses may also be based on relative risk or absolute risk reduction (difference) estimates. It should be noted, however, that absolute risk reduction may not always be the ideal effect measure to summarize as they are subject to greater heterogeneity<sup>20,21</sup>. For instance, even when the relative risk is stable, different reductions in risk would be observed in populations with differing baseline risks<sup>22</sup>. Thus, only studies based on very similar populations should be combined. Special methods are required to summarize risk differences (as well as risk ratios)<sup>23</sup>. Hazard ratios or rate (i.e., events per person-year) ratios may also be used when the original data are measured as time-to-event and analyzed using survival analysis methods.

In addition to effect estimates, it is important to extract other information from the reviewed studies. For instance, the variance (or standard error) of the estimates,

which reflect the precision of the estimates should be recorded. These are used in the analysis to weight the observations so that more precise estimates have greater influence in the derivation of pooled estimates. Furthermore, characteristics of the studies may explain why estimates of treatment effect differ across studies. For instance, features of the study population (e.g., average age of subjects, prevalence of risk factors at baseline) or the study itself (e.g., year in which it was conducted, design aspects like whether randomization or blinding was employed) may be relevant to explain heterogeneity and should be recorded.

## 2.2 META-ANALYSIS OF A SINGLE OUTCOME

The goal of meta-analysis is to aggregate effect estimates from the included studies to obtain a single pooled estimate. There are two general classes of models that are used: the first, typically referred to as a “fixed-effects” analysis assumes that the effect being estimated is homogeneous across studies; the second, so-called “random-effects” approach incorporates the possibility that there may be some (random) variability in the effect being estimated by each study (unrelated to sampling – i.e., even if the sample size of each study was *infinite*). In most applications, both fixed- and random-effects models are fitted and compared using statistical tests of heterogeneity.

Suppose a meta-analysis is being conducted on  $n$  studies with the aim of quantifying the effect of an intervention (or exposure) on a dichotomous endpoint. Denote by  $n_i^t$  and  $n_i^c$  the size of the treated and control groups, respectively, of the  $i^{\text{th}}$  study. The number of events observed is given by  $r_i^t$  and  $r_i^c$  in the treated and control groups, respectively.

A number of approaches are available for the meta-analysis of such data. For the descriptions that follow, I’ve grouped these methods into two categories: **effect-based** (or approximate) methods, in which the units of analysis are *estimates of effect* derived from the event counts (e.g., odds ratios, risk ratios, risk differences, etc), and what may be called **event-based** (or exact) methods, in which the units of analysis are *event counts* in the treated and control groups, rather than estimates

of effect. In both approaches, the odds ratio  $\left(OR_i = \frac{r_i^t / (n_i^t - r_i^t)}{r_i^c / (n_i^c - r_i^c)}\right)$  is typically the

measure of effect (or association) that is summarized. Effect-based methods are based on approximate likelihoods for the observed effect estimates, while event-based methods model the exact distributions of event counts.

### 2.2.1 EFFECT-BASED (APPROXIMATE) METHODS

The estimates of log odds-ratios from each study, denoted by  $y_i = \log(OR_i)$ , are the unit of meta-analysis in approximate methods. The variances of these are also incorporated in the analyses as weights to allow larger studies to exert more influence in the pooled result. The estimated variances of the estimates are given

by  $s_i^2 = \text{var}(y_{ki}) = \frac{1}{r_i^t} + \frac{1}{n_i^t - r_i^t} + \frac{1}{r_i^c} + \frac{1}{n_i^c - r_i^c}$ ; these are observed from each

study or can easily be calculated from the available information and assumed to be estimated without error in the analyses. The variance cannot be calculated if no events are observed in the treated group since  $r_i^t$  would be 0 (the same would occur if all patients in the control group fail). In such instances, a small value (e.g., 0.5) must be added to event counts to approximate the variance.

#### *The DerSimonian-Laird Model*

DerSimonian and Laird<sup>7</sup> proposed a general likelihood-based approach that is adaptable to a variety of outcomes. The method is based on the assumption that the sampling distribution of the unit of analysis (i.e., effect estimate) can be approximated by a normal (Gaussian) distribution; the observed standard-error of the estimates from each study is used to derive the variance of the sampling distributions. This method remains one of the most popular in meta-analyses since it is easily implemented with standard software and can be applied to data that might naturally be expected to have normal sampling distributions (e.g., reduction in blood-pressure or risk difference) or data that can be transformed to meet the normality assumption (e.g., log of odds-ratio or log of hazard-ratio).

**Fixed-Effect Analysis.** Meta-analyses typically begin with a model that assumes that the  $y_i$  are estimates of a common underlying true log odds-ratio, denoted by  $\theta$ . In other words, it assumes that the studies included in the meta-analysis are homogeneous with respect to the effect of the treatment. Then, a so-called *fixed-effect* model can be used to estimate  $\theta$ ; the model can be written:

$$y_i = \theta + \varepsilon_i$$

where  $\varepsilon_i \sim N(0, s_i^2)$ , or equivalently  $y_i \sim N(\theta, s_i^2)$ , and the variances are assumed known and taken from the studies (i.e., the true variances are assumed to be *perfectly* estimated by the observed variances). This assumption is necessary in this model, to ensure identifiability of estimates.

The log-likelihood function for  $\theta$  is given by:

$$l(\theta; \tilde{y}, \tilde{s}^2) \propto \sum_{i=1}^n \frac{(y_i - \theta)^2}{s_i^2}$$

where  $\tilde{y}$  and  $\tilde{s}^2$  are  $n \times 1$  vectors of the observed log odds-ratios and their variances. The maximum likelihood estimate of  $\theta$  is then given by:

$$\hat{\theta} = \frac{\sum_{i=1}^n w_i \times y_i}{\sum_{i=1}^n w_i},$$

where  $w_i = \frac{1}{s_i^2}$ . Thus,  $\hat{\theta}$  is a weighted average of the observed estimates with

the inverse of the variances serving as weights. The variance of the estimate is given by  $\text{var}(\hat{\theta}) = \left( \sum_{i=1}^n w_i \right)^{-1}$  and a 95% confidence interval can be constructed based on an assumption of (asymptotic) normality of the estimate:  $\hat{\theta} \pm 1.96 \sqrt{\text{var}(\hat{\theta})}$ .

**Heterogeneity.** In practice, assuming a common (homogeneous) effect across studies may not be appropriate. It is, therefore, important to assess the tenability

of the homogeneity assumption. Cochran's  $Q$  statistic<sup>24</sup> is often employed for this purpose, where  $Q = \sum_{i=1}^n w_i (y_i - \hat{\theta})^2$  and  $Q \sim \chi^2_{n-1}$ . High values of the statistic suggest departure from the assumption of a common effect and suggest a poor fit of the fixed-effects model. This test often suffers from weak power for detecting heterogeneity, however, especially when there are few studies in the meta-analysis, or when estimates are imprecise<sup>18,25</sup>. Therefore, failing to reject homogeneity does not preclude the possibility of weak between-study variability. Alternative methods have been proposed recently by Higgins and Thompson<sup>26</sup> who advocate quantifying the degree of heterogeneity rather than relying on a statistical test. Baujat et al.<sup>27</sup> propose graphical methods to examine heterogeneity. Before discussing this further, I review methods used to incorporate heterogeneity.

**Random-Effects Analysis.** The usual (and recommended<sup>28</sup>) approach for incorporating between-study variability is to include an (additive) random term to the fixed effect model described above. This is commonly called a *random-effects* model and can be written as follows:

$$y_i = \theta + \delta_i + \varepsilon_i$$

where  $\delta_i \sim N(0, \tau^2)$  and  $\varepsilon_i | \delta_i \sim N(0, s_i^2)$ , and  $\tau^2$  is an unknown parameter measuring the degree of heterogeneity to be estimated from the data. As in the fixed-effects analysis, within-study variances must be assumed to be known without error to ensure identifiability of the model parameters.

An equivalent expression of this model is:

$$\theta_i \sim N(\theta, \tau^2), \text{ and}$$

$$y_i | \theta_i \sim N(\theta_i, s_i^2).$$

This allows the true log odds-ratio ( $\theta$ ) to vary from one study to another ( $\theta_i = \theta + \delta_i$ ). Using a random term to capture the excess variability implies,

however, that parameters from different studies are assumed to vary in a random or non-systematic way (but subject to a common distribution).

**Estimation of Parameters.** The parameters are typically estimated by maximizing the likelihood function of  $\theta$  and  $\tau^2$ . This is derived from the marginal distribution of the observed estimates ( $y_i$ ), which, in the random effects model, is  $N(\theta, s_i^2 + \tau^2)$ . Therefore, the likelihood function is:

$$l(\theta, \tau^2 | \tilde{y}, \tilde{s}^2) \propto \sum_{i=1}^n \left[ \frac{(y_i - \theta)^2}{s_i^2 + \tau^2} \right] + \log \sum_{i=1}^n (s_i^2 + \tau^2)^{-1}.$$

The maximum likelihood (ML) estimates of  $\theta$  and  $\tau^2$  are given as the solution to the following equations:

$$\hat{\theta} = \frac{\sum_{i=1}^n w_i(\hat{\tau}^2) y_i}{\sum_{i=1}^n w_i(\hat{\tau}^2)},$$

$$\hat{\tau}^2 = \frac{\sum_{i=1}^n w_i(\hat{\tau}^2) ((y_i - \hat{\theta})^2 - s_i^2)}{\sum_{i=1}^n w_i(\hat{\tau}^2)}$$

where  $w_i(\hat{\tau}^2) = \frac{1}{s_i^2 + \hat{\tau}^2}$ . These may be solved using iterative procedures like the

Newton-Raphson or EM algorithm. The latter consists of starting with an initial estimate for  $\tau^2$ , and obtaining an estimate for  $\theta$ ; this is then used to obtain a new estimate for  $\tau^2$ . The process is iterated until the estimates converge to stable values.

The variance of  $\hat{\theta}$  is given by  $\text{var}(\hat{\theta}) = \left( \sum_{i=1}^n w_i(\hat{\tau}^2) \right)^{-1}$ , which resembles the variance of the fixed effect estimate but will always be greater, due to the incorporation of heterogeneity between studies (i.e.,  $\hat{\tau}^2$  in the denominator of the

weights). Inferences are made assuming (asymptotic) normality of the estimate. The variance of  $\hat{\tau}^2$  is obtained from the inverse of the observed Fisher information matrix.

Alternatively, one may use the restricted or residual maximum likelihood (REML) approach, which modifies the likelihood to account for the joint estimation of  $\theta$  and  $\tau^2$ . In this case, the likelihood function is:

$$l(\theta, \tau^2 | \tilde{y}, \tilde{s}^2) \propto \sum_{i=1}^n \left[ \log(s_i^2 + \tau^2) + \frac{(y_i - \theta)^2}{s_i^2 + \tau^2} \right] + \log \sum_{i=1}^n (s_i^2 + \tau^2)^{-1}.$$

The estimate of  $\tau_k^2$  is given as the solution to:

$$\hat{\tau}^2 = \frac{\sum_{i=1}^n w_i (\hat{\tau}^2) \left( \frac{n}{n-1} (y_i - \hat{\theta})^2 - s_i^2 \right)}{\sum_{i=1}^n w_i (\hat{\tau}^2)},$$

while that of  $\theta$  remains the same as above<sup>18</sup>. Estimation proceeds as with the ML approach described above using the Newton-Raphson algorithm. As can be seen from the formula, when there are relatively few studies in the meta-analysis, the REML estimates for  $\tau^2$  will tend to be greater than the ML estimates due to

the  $\frac{n}{n-1}$  factor.

DerSimonian and Laird<sup>7</sup> describe a non-iterative estimate for  $\tau^2$  based on the method of moments and the  $Q$  statistic. This is given by:

$$\hat{\tau}^{2*} = \max \left\{ 0, (Q - (n-1)) \div \left[ \sum_{i=1}^n w_i - \frac{\sum_{i=1}^n w_i^2}{\sum_{i=1}^n w_i} \right] \right\};$$

an estimate for  $\theta$  can be obtained by fixing  $\tau^2$  to  $\hat{\tau}^{2*}$  in the likelihood, and

maximizing  $l(\theta, \hat{\tau}^{2*} | \tilde{y}, \tilde{s}^2)$ . This yields  $\hat{\theta}^* = \frac{\sum_{i=1}^n w_i(\hat{\tau}^{2*}) y_i}{\sum_{i=1}^n w_i(\hat{\tau}^{2*})}$  with variance given by  $\text{var}(\hat{\theta}^*) = \left( \sum_{i=1}^n w_i(\hat{\tau}^{2*}) \right)^{-1}$ ; however, no variance estimate is available for  $\hat{\tau}^{2*}$ .

Other approaches that may be used include profile likelihood<sup>29</sup>, weighted least-squares<sup>5</sup>, as well as Bayesian methods<sup>18</sup>. The latter incorporates additional uncertainty about the population parameters ( $\theta_k$  and  $\tau_k^2$ ) through prior distributions.

**Returning to Heterogeneity.** Once a random-effects model is fitted, one can reassess the presence of heterogeneity. First, the estimate of  $\tau^2$  quantifies the spread of the effect across studies. Thus, one can use clinical judgment to determine the importance of the observed heterogeneity instead of relying on statistical significance alone. In fact, some of the measures proposed by Higgins and Thompson<sup>26</sup> are based on the estimate of  $\tau^2$ . Second, one can compare the fixed and random effects models directly with a likelihood-ratio test to determine which of the two is better suited to the data.

It has been argued that heterogeneity in meta-analyses “is usual rather than exceptional”<sup>30</sup>. In fact, Brockwell and Gordon<sup>31</sup> recommend adopting a random-effects model, irrespective of what is suggested by tests of homogeneity. They showed that random-effects estimates tended to have better coverage (especially when the profile likelihood method was used).

Others have criticized the assumption that study-specific effects arise from a normal distribution as “rather simplistic”<sup>32</sup>. Indeed, at least part of the observed heterogeneity may be systematic and explainable by differences in study design, study populations or other factors. It is therefore recommended to explore and explain heterogeneity to gain more insight from the analysis<sup>33,34</sup>. Glasziou and

Sanders<sup>35</sup> distinguish between effect modification (different effects in different types of populations, for example) and “artifactual” variation (caused by differences in compliance levels, blinding methods, etc). Meta-regression models can be used to ascertain sources of between-study variability by adding study-level covariates (e.g., average age of patients, proportion male, etc.) to describe the true population treatment effect (as this part of the model explains between-study variability). In the random-effects model, this would be given by  $\theta_i \sim N(X_i\beta, \tau^2)$ , and  $y_i|\theta_i \sim N(\theta_i, s_i^2)$ , where  $\beta$  is the vector of unknown regression parameters. Care must be taken to ensure that the observations are properly weighted; this and other issues pertaining to the use of meta-regression models are discussed in more detail elsewhere<sup>5,36</sup>.

### *Other Effect-based Approaches*

Whitehead and Whitehead describe a general parametric approach for meta-analyses effect estimates for various types of outcome measures<sup>37</sup>. Berkey et al.<sup>38</sup> proposed a regression-based approach, and Hardy and Thompson<sup>39</sup> describe how to use profile likelihoods to derive confidence intervals to account for ignoring the uncertainty involved fixing within-study variances. Turner et al.<sup>40</sup> present a multi-level approach for the meta-analysis with individual data.

As mentioned previously, the model described in the previous section may be fitted with a Bayesian approach<sup>18,41</sup>. This requires setting prior distributions on  $\theta_k$  in the fixed-effect model and  $\theta$  and  $\tau^2$  in the random-effects model. Inferences are drawn based on the posterior distributions of the parameters, which reflect the combined information from prior knowledge and the data included in the meta-analysis. With conjugate priors (normal for  $\theta$  and gamma for  $1/\tau^2$ ) closed form solutions exist for the posterior distributions<sup>42</sup>. Otherwise, the models are quite easily estimated using Markov Chain Monte Carlo algorithms as implemented in BUGS<sup>43</sup>; the posterior distributions are then described empirically (moments and percentiles being estimated through simulation).

## **2.2.2 EVENT-BASED (EXACT) METHODS**

The assumption of normality of estimates of the log odds-ratio underlying the DerSimonian and Laird<sup>7</sup> approach is likely to be tenable in many situations but may not be adequate in others. For instance, studies with empty cells (i.e., if no events were observed in the treated or control groups) can not be used in the analysis since the variance of the observed estimate would not be defined. Furthermore, the approach fixes the variance of residuals to observed values, thereby assuming these are known without error (which underestimates the true uncertainty of the estimates).

An alternative approach is to use exact methods that model the observed event counts. The simplest exact approach is perhaps to treat the studies as a series of 2 x 2 tables and derive a pooled odds-ratio using the Mantel-Haenszel approach<sup>44</sup>. Although tests are available to assess the presence heterogeneity, there is no clear way of incorporating between-study variability in the estimates. Also, only categorical (or categorized) covariates can be handled by this method, and this, only through stratification and pooling. That is, the effect of the covariates is not measured explicitly. More general modeling frameworks are available, however. The Bayesian binomial model proposed by Smith et al.<sup>32</sup> and the generalized linear mixed model (GLMM) approach of Platt et al.<sup>45</sup> are two prominent examples. Both approaches accommodate full regression models so that covariates may be included to explain between-study heterogeneity.

The Bayesian binomial method<sup>32</sup> models the number of events in each arm of a study as binomial random variables with unknown probabilities of *success*; these are reformulated in terms of an *average* probability and treatment effect parameters. A Normal distribution is assumed for the random treatment effect (on the logit scale) and priors are specified as appropriate. The performance of this method was found to be generally consistent with standard techniques for fixed-effects analyses, but notable deviations were observed for the mixed-effects models.

The GLMM approach<sup>45</sup> is built on the assumption of a non-central hypergeometric distribution for the number of events in the treated group, conditioning on the total number of events, and is an extension of the non-central hypergeometric approach to the fixed-effects problem<sup>46</sup>. A penalized quasi-likelihood<sup>47</sup> is used to approximate the likelihood of the data. Estimates are obtained with the Newton-Raphson algorithm. Other methods could be used to maximize the likelihood, but as with other GLMM problems, direct maximum likelihood estimation is not feasible. Results based on simulations revealed that estimates from GLMM were generally consistent with those from standard techniques, except when a relatively strong treatment effect was assumed along with moderate heterogeneity and a rare outcome. In this case, the GLMM estimate was less biased and had better coverage probabilities.

In both of these approaches, the likelihoods (conditional on random-effects) are based on the distribution of the event counts. The relative strengths of studies are inherently reflected in the summary estimates since studies with high event counts (which likely arise from larger studies) will be more *influential*. Thus, no assumptions are made about the parameters of the distribution (as with the variances of the likelihoods of effect estimates in approximate methods). Implementing models with exact distributions is more conveniently accomplished with Bayesian methods, which also allow predicting expected effects in a new trial from the posterior distribution<sup>32</sup>.

### **2.3 JOINT META-ANALYSIS OF MULTIPLE OUTCOMES**

Meta-analyses often consider a number of outcomes that are relevant in the context of the question being studied. These outcomes are typically correlated, however, possibly due to a common underlying biological mechanism from which the endpoints arise, common modes of action of the treatment for the various endpoints or even simply due to the fact that measurements from the same study will tend to be more similar than those from different studies. In other cases, the outcomes may be repeated measurements of the same endpoint at different times or subgroups.

Usually, a separate analysis is performed for each outcome<sup>1,10</sup> and conclusions are drawn independently for each. A joint analysis is possible, however, using multivariate approaches that incorporate the correlations between the outcomes. In the following sections, I describe a general multivariate model that is used for meta-analyses of multiple outcomes and present some applications that illustrate the benefits and flexibility of these methods. Multivariate meta-analyses allow testing joint hypotheses about outcomes and are expected to offer potential gains in precision<sup>5</sup>. Only the more general random-effects model is presented, since the fixed-effects model can be obtained as a special case of the latter.

### 2.3.1 MULTIVARIATE META-ANALYSIS MODELS

#### *Model Specification*

A joint analysis of  $k$  outcomes can be performed using the framework of multivariate (or multi-variable) regression. The method that is usually employed is a multivariate extension of the standard univariate meta-analysis model described above<sup>7</sup>. In the multivariate context, the unit of meta-analysis is the  $k$ -dimensional vector of observed estimates  $\tilde{y}_i = (y_{i1}, y_{i2}, \dots, y_{ik})$  in the  $i^{\text{th}}$  study (for  $i=1,2,\dots,n$ ). Similarly,  $\tilde{\theta}$  denotes the  $k \times 1$  vector of effects on the different outcomes and by  $\tilde{\delta}_i$  and  $\tilde{\varepsilon}_i$  the vector of random-effects and residuals for the  $i^{\text{th}}$  study.

Using the notation from previous sections, a multivariate normal random-effects model can be specified by:

$$\tilde{y}_i = \tilde{\theta} + \tilde{\delta}_i + \tilde{\varepsilon}_i.$$

For the analysis of two outcomes ( $k=2$ ), this can be expressed as:

$$\begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} + \begin{pmatrix} \delta_{1i} \\ \delta_{2i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix}$$

$$\text{where } \tilde{\delta}_i \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, D = \begin{pmatrix} d_1^2 & d_{12} \\ d_{12} & d_2^2 \end{pmatrix}\right) \text{ and } \tilde{\varepsilon}_i \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, S_i = \begin{pmatrix} s_{1i}^2 & s_{12i} \\ s_{12i} & s_{2i}^2 \end{pmatrix}\right)$$

(MVN denotes “multivariate normal”). The diagonal elements of the random-effects covariance matrix  $D$  reflect the degree of heterogeneity in the effects estimated by each study, while  $d_{12}$  represents the covariance between the true measures of effect from two different studies. The model reduces to a fixed-effects model by setting the variance components of  $D$  to 0. The diagonal elements in  $S_i$  are the variances of the observed estimates, while  $s_{12i}$  denotes the covariance between estimates in a given study. As in single-outcome models, the within-study covariance matrix is extracted from the reviewed manuscripts and assumed known in the analysis. It should be noted, however, that the covariance component ( $s_{12i}$ ) is not always reported and may have to be obtained by contacting the authors of the studies or approximations may have to be made based on information from external sources.

The model can readily be extended to incorporate covariates, in which case, the regression model can be written as:

$$\tilde{y}_i = X_i \tilde{\beta} + \tilde{\delta}_i + \tilde{\varepsilon}_i$$

where  $X_i$  is a  $k \times p$  matrix of covariates and  $\beta$  is a  $p \times 1$  vector of parameters. By default,  $X_i$  would include one or more indicators to identify outcomes in  $\theta_i$  and derive summary estimates for each outcome or differences in estimates between outcomes. If an intercept is also included, the outcome term measures the difference in effect for the two outcomes and the intercept measures the pooled estimate of the “reference” outcome. That is, in the case of two outcomes, setting

$X_i = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$  would produce a pooled-estimate for the second endpoint (given by

the intercept estimate) and the difference between pooled estimates of the first and

second endpoints. On the other hand, setting  $X_i = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  and not allowing

intercepts in the model would produce a separate pooled-estimate of effect for

each outcome. Other covariates can also be included in  $X_i$ . In this specification covariates that are added to the model are applied to all outcomes in  $\tilde{y}_i$ . Furthermore, by default, this specification allows a single common effect of covariates for each outcome; that is, covariates are assumed to have the same effect on all outcomes, unless an interaction between the covariate and the outcome indicator is also included. In fact, it may be sensible to always include these interactions, particularly when the outcomes are measured on (very) different scales. This is done more conveniently with software (like BUGS<sup>43</sup>) that allow specifying separate regressions for each outcome which can therefore also include different covariates for each of the outcomes.

### *Estimation of Parameters*

The parameters of the model may be estimated using a generalized least-squares (GLS) procedure<sup>1,5,8,10</sup> or with maximum likelihood or residual (restricted) maximum likelihood approaches<sup>5,6,12</sup>; more recently, similar models have been fitted by Bayesian methods<sup>9</sup>, which are easily handled in BUGS<sup>43</sup>.

Likelihood-based estimation is easily done with standard software like the SAS System, using the PROC MIXED procedure. Van Houwelingen et al.<sup>12</sup> describe how to specify code to fit these types of models. The likelihood is constructed from the marginal distributions of the observed data, which is

$$MVN(\tilde{x}'_i \tilde{\beta}, V_i = D + S_i) \equiv \left( \frac{1}{2\pi} \right)^{n/2} |V_i|^{-1/2} \exp \left[ -\frac{1}{2} (\tilde{y}_i - \tilde{x}'_i \tilde{\beta})' V_i^{-1} (\tilde{y}_i - \tilde{x}'_i \tilde{\beta}) \right],$$

where  $|V_i|$  is the determinant of  $V_i$ . As in the univariate case, the within-study covariance matrix ( $S_i$ ) is usually treated as known to ensure that estimates are identifiable. The notation can be extended to represent the joint distribution of the entire data set. Let  $K$  be the total number of observations collected from the reviewed studies (e.g., if all studies report all  $k$  outcomes, then  $K = n \times k$ ). Let  $V$  be the  $K \times K$  block-diagonal matrix with blocks consisting of the  $V_i$ ;  $Y$  denotes the  $K \times 1$  vector of  $\tilde{y}_i$  vectors; and,  $X$  represents the  $K \times p$  matrix of covariates

formed by the  $\tilde{x}_i$ . Assuming that observations from different studies are independent, the likelihood function can be written as follows:

$$L(\tilde{\beta}, D) \propto |V|^{-1/2} \exp\left[-\frac{1}{2}(Y - X\tilde{\beta})' V^{-1} (Y - X\tilde{\beta})\right].$$

For a given value of  $D$  (and, hence,  $V$ ), one can estimate  $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y$  and  $\hat{Y} = X(X'V^{-1}X)^{-1}X'V^{-1}Y$ . The residual  $(Y - X\tilde{\beta})$  can then be estimated by  $r = Y - X(X'V^{-1}X)^{-1}X'V^{-1}Y$ ; replacing this in the likelihood yields

$$L(D) \propto |V|^{-1/2} \exp\left[-\frac{1}{2}(r'V^{-1}r)\right],$$

which now only depends on  $D$ .

An alternative approach is to use the restricted likelihood, which does not condition on  $V$  to estimate  $\beta$  during the estimation process. Rather, it begins by transforming  $Y$  to have a mean of 0 before constructing the likelihood; that is, the likelihood is constructed from the transformed random variable  $Z = Y - X(X'V^{-1}X)^{-1}X'V^{-1}Y$ . In this case, the likelihood is given by:

$$L(D) \propto |V|^{-1/2} \exp\left[-\frac{1}{2}(Z'V^{-1}Z)\right] \times |X'V^{-1}X|^{-1/2}.$$

For both the full and residual likelihood methods, Newton-Raphson or other iterative algorithms may be used to derive the estimate.

GLS estimates of the parameters may be found by minimizing the weighted sum-of-squares  $(Y - X\tilde{\beta})' V^{-1} (Y - X\tilde{\beta})$ . Berkey et al.<sup>5</sup> describe how this may be done using an EM algorithm, alternating between estimating  $\tilde{\beta}$  and  $D$ , conditional on an estimate of the other parameter. As before, for a given value of  $D$  (and, hence,  $V$ ),  $\tilde{\beta}$  is estimated by  $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y$ ; conditional on a value of  $\beta$ , an estimate of  $D$  is obtained from

$$\hat{D} = \frac{1}{n-p} (r^*)' (r^*) - \frac{1}{n} \sum_{i=1}^n S_i,$$

where  $r^*$  is an  $n \times k$  matrix of residuals with rows corresponding to studies and columns to outcomes; the  $i^{\text{th}}$  row is given by  $(Y_i - X_i \tilde{\beta})'$ . The estimates of  $D$  can be thought of as the difference between the average *total* variance and the average within-study variance. The process can be started with a fixed effect estimate of  $\tilde{\beta}$ , from which one can derive an initial estimate of  $D$  using the above equation. This is then used to update the estimate of  $\tilde{\beta}$ . The process is iterated until changes in estimates of all parameters are within some range of tolerance.

### 2.3.2 APPLICATIONS OF THE MULTIVARIATE NORMAL MODEL

Multivariate models have been applied in a variety of ways, with different objectives, in the context of meta-analysis in the recent literature. In addition to being used for the joint analysis of two or more different end-points, these models have also been employed to analyze a single endpoint with the aim of correcting potential biases that may arise with standard approaches, or as a practical alternative to performing separate analyses when the outcome is measured at different times. These applications are discussed in more detail below.

#### *Control-Rate Regression with Multivariate Models*

One of the earliest applications of multivariate methods was proposed by van Houwelingen et al.<sup>11</sup> in which a pooled estimate of the treatment effect is derived by meta-analyzing the joint distribution of event rates in the intervention and control groups observed in each study, as opposed to combining the effect estimates directly, as is done in the standard approaches. The motivation for this approach was to assess the relationship between baseline risk and treatment effects across studies, since standard approaches (i.e., meta-regression with the control rate as a variable) can produce biased results. This is discussed in more detail below, following a description of the multivariate model that was used.

If one is interested in the pooled (log) odds-ratio as the measure of effect, the unit of analysis would be the vector of log-odds in the treated (T) and control (C) groups, denoted by  $y'_i = (\hat{\omega}_{Ti} \quad \hat{\omega}_{Ci})$ , which has covariance matrix  $S_i$  observed from the studies. Since the groups within a study are independent, they assume the covariance (off-diagonal element of  $S_i$ ) to be zero.

A multivariate normal random-effects model is specified, assuming that, conditional on a true log-odds vector for the  $i^{\text{th}}$  study,  $\omega'_i = (\omega_{Ti} \quad \omega_{Ci})$ ,  $y_i | \omega_i \sim MVN(\omega_i, S_i)$ , where MVN denotes the multivariate (bivariate, in this case) normal distribution. The random-effects also have a multivariate normal distribution:  $\omega_i \sim MVN(\omega, \Sigma)$ , where  $\omega' = (\omega_T \quad \omega_C)$  and  $\Sigma = \begin{bmatrix} \Sigma_{TT} & \Sigma_{TC} \\ \Sigma_{CT} & \Sigma_{CC} \end{bmatrix}$ . One can readily extend the model to include covariates that might explain between-study variability through the usual specification described above. The model can be estimated by maximum likelihood using standard software and variances of the estimates can be obtained from the covariance matrix of the estimates.

An estimate of the treatment effect can be obtained in one of two ways: first, an intercept only model can be fitted, with a different intercept for treated and control groups; the difference in intercepts yields the log-odds ratio estimate and the covariance matrix of the estimates can be used to obtain its variance. A more direct approach is to include an indicator for the treated group as a predictor of the true log-odds ( $\omega$ ); in this case, the intercept provides an estimate of the log-odds in the control group and the parameter for the treatment indicator is the pooled estimate of the treatment effect.

There are at least two advantages to this approach: first, an estimate of both the event rate in the reference population and the treatment effect are obtained from a single run of the model; second, it allows the inclusion of results from non-comparative studies (provided it is appropriate to do so). The primary purpose of the model is control-rate regression, however, which refers to an approach used in meta-analysis where the event rate in the control (or reference) group is used to

explain heterogeneity observed in the treatment effect on a given outcome across studies<sup>12,48</sup>. This allows one to infer whether the effect of the treatment is constant across populations of varying risk levels.

Standard univariate meta-regression methods, with the control-rate (log-odds) as a covariate of the log-odds ratio can produce biased results. Van Houwelingen et al.<sup>12</sup> summarize the reasons for this as follows:

- i. The estimated slope between the control rate (or log-odds) and the treatment effect (log-odds-ratio) will be attenuated, since the control rate estimate is subject to measurement error (or sampling variability);
- ii. Furthermore, measurement error in the control rate estimate is negatively correlated with the measurement error associated with the log-odds-ratio estimate. This can then offset the observed relationship between the control-rate and treatment effect.

To avoid these biases, van Houwelingen et al.<sup>12</sup> suggest using the bivariate random-effects model. This approach, in a sense, separates the model into two components: the *structural* component, given by the random-effects distribution, and the *measurement* component, given by the distribution (or likelihood) of the observed estimates. Inference about the treatment effect is based on variability and correlations arising in both levels of the model, as described above. To eliminate the influence of correlations in measurement errors, however, inferences about the relationship between the control rate and treatment effect are based solely on the structural component of the model. Thus, the slope between the log-odds-ratios and the control log-odds can be obtained as follows:

$$\beta = \frac{\text{cov}(\omega_T - \omega_C, \omega_C)}{\text{var}(\omega_C)} = \frac{\text{cov}(\omega_T, \omega_C) - \text{var}(\omega_C)}{\text{var}(\omega_C)} = \frac{\Sigma_{TC}}{\Sigma_{CC}} - 1$$

so that,  $\hat{\beta} = \frac{\hat{\Sigma}_{TC}}{\hat{\Sigma}_{CC}} - 1$ . One can also calculate the slope between the log-odds in

the two groups, which provides an alternative measure of the treatment effect with a slope close to unity indicating no effect.

The degree of heterogeneity explained by the baseline rate is derived as the proportional reduction between the unconditional and conditional (on control log-odds) variance of the treatment effect between different studies. These can also be derived from the random-effects variances; the unconditional variance is given by:

$$\text{var}(\omega_T - \omega_C) = \text{var}(\omega_T) + \text{var}(\omega_C) - 2\text{cov}(\omega_T, \omega_C) = \Sigma_{TT} + \Sigma_{CC} - 2\Sigma_{TC}$$

while the conditional variance is give by:

$$\text{var}(\omega_T - \omega_C | \omega_C) = \text{var}(\omega_T | \omega_C) = \Sigma_{TT} - \frac{\Sigma_{TC}^2}{\Sigma_{CC}}.$$

The latter is derived from the conditional distribution of the log-odds among the treated.

### *Survival Rate Ratios over Time*

In studies with relatively long follow-up, results are commonly reported at various points in time. A particular instance of this is when the outcome is reported in terms of a survival rate and the effect of the intervention is summarized as a rate or hazard ratio. When the effect is constant over time (i.e., the hazards are proportional), the meta-analysis can be based on effect estimates derived over the full study period. Otherwise, the effect must be estimated at different times to fully describe the benefits of treatment. With standard approaches, this requires a separate analysis for each time point of interest, which does not allow formal inference about the pattern of changes in the treatment effect over time. Alternatively, a multivariate model may be used to analyze data from all times jointly.

**Survival after bone-marrow transplant.** Dear<sup>10</sup> used a *fixed-effects* multivariate model, similar in design to the bivariate model of van Houwelingen et al.<sup>11</sup> to meta-analyze the difference in disease-free survival between patients with acute myelogenous leukemia receiving a bone-marrow transplant compared with chemotherapy (control group). The unit of analysis was the survival proportion (and standard error) in each of the groups, measured at yearly intervals over the five years following treatment. This allowed the author to include historical controls in the meta-analysis – that is, data from studies that only examined the effect of chemotherapy – which adds power and precision to the analysis.

A linear model was used, with identity link, with treatment type and time as covariates and an interaction between the two to test whether rate difference is constant over time. Although the model did not include random-effects, an indicator identifying each study was added to the model along with interactions between these and the treatment group indicator to capture and test for variability in effects between studies. No significant difference was detected between studies, which supported the appropriateness of using historical controls.

As described above, multivariate models require the covariance matrix of observations from the same study. While the standard errors of the estimated survival rates were available, the correlations between observations from a given group of a study at two different times were not reported. Dear<sup>10</sup> describes a method where the covariances are estimated as a function of survival rates and recommends that this be done iteratively, in the course of the estimation of other parameters, as opposed to estimating them from the observed rates, to avoid situations where correlations would be found to be one when the survival rate did not change from one time to the next.

**Survival after carotid endarterectomy.** A similar example was treated by Arends et al.<sup>6</sup> in a meta-analysis comparing the effect of carotid endarterectomy (a surgical procedure to remove atherosclerotic plaque) to prevent cerebral infarction (stroke) or death with conventional therapy. The complication, in this case, is that the surgery is associated with an increased mortality risk in the first month

following the procedure but is expected to improve prognosis for patients thereafter. Therefore, a fair comparison would have to properly separate the risk periods following surgery. To do so, the authors summarize the outcomes from each study into three components: 1) the rate (events per person-year) of the composite endpoint of death or stroke in the conventional therapy group (denoted by  $x$ ); 2) the post-surgery risk (proportion) of death or stroke in the group receiving carotid endarterectomy (denoted by  $y$ ); 3) the long-term rate (events per person-year) of deaths or strokes in the group receiving surgery (denoted by  $z$ ). A cumulative survival rate ratio at any given time ( $CSR(t)$ ) is derived to compare survival in the two groups from these components, making the assumption that the rates ( $x$  and  $z$ ) are constant over time, so that the survival rates can be obtained from the exponential distribution, as follows:

$$CSR(t) = \frac{(1-y)\exp(-zt)}{\exp(-xt)}.$$

The denominator is the proportion of control patients alive stroke-free at time  $t$ , derived from an exponential distribution with rate (units: year<sup>-1</sup>)  $x$ . In the numerator,  $(1-y)$  represents the probability of surviving the surgery, and  $\exp(-zt)$  is the conditional probability of being alive stroke-free at time  $t$  among patients who survived the surgery, derived from an exponential distribution with rate  $z$ . Thus,  $CSR(t)$  represents the ratio of survival probabilities in the surgical and control groups at any time  $t$ .

Standard meta-analytic approaches would require that the CSR be calculated for each study at each time of interest and perform a separate analysis for each of these. This is inconvenient in terms of both the added time and effort required to run separate models, but also with respect to the limited opportunities for formal inference about time patterns. Instead, the authors fitted a hierarchical trivariate normal model for a joint analysis of the log-rate in the conventional group, the log-odds during the acute post-surgical period and the log-rate after the first month. This analysis provides a pooled estimate of these parameters as well as

the covariance matrix of the estimates. Thus, a function can be derived that predicts the CSR at any time with the corresponding confidence intervals.

As in the prior example, the covariance between observations from the same study is not known. The authors assume the estimates are independent and set the covariances to zero, even though some correlation is likely to exist due to similar measurement errors and since the acute risk and long-term rate in the surgical group were evaluated from the same set of subjects.

### *Multiple Outcomes*

The applications described so far illustrate the practicality and flexibility of multivariate methods for the meta-analysis of a single, possibly repeated, outcome. The examples showed how a single multivariate model yielded the necessary information to make the same inferences as separate analyses, and, in addition, provided a framework in which to model time trends and test the significance of interactions, which was not possible with outcome-specific analyses.

In many situations, the effect of the treatment is measured in terms of two or more different outcomes that are correlated. Some deal with this issue by aggregating the outcomes into a single measure, or performing a joint analysis in which the outcomes are assumed independent<sup>1</sup>, but most commonly the outcomes are analyzed separately<sup>1,10</sup>. It has been argued, however, that a joint analysis that takes into account the correlations is preferable, for a variety of reasons. First, outcome-specific analyses may lack power and precision<sup>10</sup>; furthermore, a joint meta-analysis allows one to draw unified conclusions about the treatment effect, and test hypotheses that are not otherwise possible (e.g., is the effect of treatment the same on both outcomes?)<sup>1</sup>; a joint analysis is also expected to improve the accuracy and precision of estimates<sup>5</sup>; in some situations, it may also be interesting to quantify the correlation between the outcomes, to provide added insight about the effect of the intervention<sup>6</sup>.

Raudenbush et al.<sup>1</sup> used a multivariate model to meta-analyze the effect of coaching on Scholastic Aptitude Test (SAT) scores in the math (SAT-M) and

verbal (SAT-V) components. They standardized the effect estimates from the studies to be scale-free. Thus, they were able to test whether the effect of coaching and, particularly, the duration of coaching, was the same on SAT-M and SAT-V scores. They employed a fixed-effects model and estimated its parameters by GLS. The within-study covariances were not reported, so the authors used an external estimate of the correlation between SAT-M and SAT-V and the observed variances to estimate the covariance in each study. They acknowledged this as a limitation and recommended that authors change reporting practices and include the covariance of the estimates in publications.

Berkey et al.<sup>8</sup> extended the model of Raudenbush et al.<sup>1</sup> to keep the effect measures in their original scales and to accommodate multiple study arms from each study. Pre vs. post treatment differences in the number of tender joints, erythrocyte sedimentation rates and grip strength were analyzed following the use of two second-line drugs for rheumatoid arthritis compared to placebo. The unit of analysis was the vector of mean changes (pre vs. post) in outcomes in each arm of each study. As mentioned previously, with this approach, it is not required that all studies have examined the same treatment groups. They too were faced with the difficulty of unreported covariance estimates and proceeded, as in the previous example, to use correlation estimates from one of the studies to derive covariances for the others. After performing sensitivity analyses on these approximations, they concluded that the overall conclusions were not seriously impacted by the choice of correlation matrix. They noted, however, that assuming complete independence between observations lead to moderate changes in the estimates of some of the parameters and caused a slight increase in their standard errors, but the comparison between the two drugs were not affected.

Berkey et al.<sup>5</sup> further extended the multivariate models described above to include random-effects and implemented a maximum likelihood estimation procedure. The model was applied in a meta-analysis of five studies comparing surgical and non-surgical treatments of gum disease with respect to the probing depth (PD) and attachment level (AL). The studies adopted a “split-mouth” design, whereby patients received the surgical procedure on one side of their mouth and non-

surgical therapy on the other. Therefore, the effect of surgery on PD and AL was measured in each patient and the unit of meta-analysis was the mean difference observed in each study. Standard errors and covariances between estimates were reported in all five studies. The year in which the study was conducted was included as a covariate. The authors compared the random and fixed-effects models and multivariate and outcome-specific models based on the results of the meta-analysis as well as with a simple simulation study. As expected, the random effects-models produced estimates with larger variance, but were found to have better coverage probabilities. Estimates from multivariate models were almost identical to those from outcome-specific models and had very similar precision. Multivariate models were found to be slightly more sensitive (i.e., lower type-I error rate) in detecting a non-null effect of study year on PD and AL in multivariate tests. Korolija et al.<sup>49</sup> applied Berkey's models in a similar application to compare oncological surgical procedures.

A Bayesian multivariate model was developed by Nam et al.<sup>9</sup> to meta-analyze studies of the effect of environmental tobacco smoke (ETS) on the incidence of asthma and lower respiratory disease (LRD) among children. A total of 59 studies were included: 27 reported data on LRD only, 24 reported on asthma only and the remaining eight reported on both outcomes. The average age of patients, proportion male or female, year of study, country, type of smoking (parental vs. household), and whether the reported ORs were adjusted were considered as covariates of the effect estimates. As in other examples, the within-study covariances were not reported in any of the studies. Rather than seeking external estimates of correlations or assuming independence, the authors estimate these from the data by specifying the covariance matrix as  $SE_i \times C \times SE_i$ , where  $SE_i$  is a 2x2 diagonal matrix of the standard errors of the estimates which serve to scale the common covariance matrix C. The authors acknowledge that there is a potential problem of identifiability in separating out within and between-study variances and correlations but state that this should not affect estimates of effect parameters.

Findings from outcome-specific and multivariate Bayesian models were almost identical. The authors attribute this to the fact that there were only eight studies that reported both outcomes, so that there was little information about the joint distribution of the outcomes. To create a situation where “borrowing strength” from the joint distribution would be more important, they omitted the studies reporting LRD only, so that all information about LRD would come from studies reporting both outcomes. Only a slight difference was found in the point estimate and a small reduction in the variance of the estimate. Furthermore, estimates of the correlation between the effects of ETS on the two outcomes were very imprecise, with confidence intervals covering almost the full range of possible values.

### **2.3.3 OTHER MODELS FOR MULTIVARIATE META-ANALYSIS**

The models and applications described so far are all based on (approximate) multivariate normality of estimates of effect and rely on fixing within-study covariances to weight observations. Recently, the Bayesian binomial model of Smith et al.<sup>32</sup> was extended by Lu and Ades<sup>50</sup> to perform meta-analyses in situations where multiple treatments are available and compared in studies. For instance, in this context, studies may have examined and reported results from three or more active treatments (A, B and C), but some may report only pairs of the treatments. The goal of the meta-analysis is to derive summary measures of *direct* comparisons of all treatments (A vs. B, B vs. C and A vs. C) from a joint analysis of all trials instead of separate meta-analyses for comparisons of each pair. Thus, if few studies examined A and C, for example, a joint analysis can borrow strength from studies including comparisons between A vs. B and B vs. C as these provide indirect comparisons of A vs. C.

The model that is proposed assumes a binomial distribution for event counts observed in each of  $k$  arms of the studies. The probabilities of the events in each arm ( $p_{ij}$ ) are parameterized in terms of an average log-odds  $\mu_i = \frac{1}{k} \sum_{j=1}^k \log it(p_{ij})$  and log-odds-ratios  $\delta_{ij} = \log it(p_{ij}) - \log it(p_{il})$  (for  $j = 2, \dots, k$ ). (Note that this

reduces to the Smith et al. model when  $k = 2$ ). The vector of log-odds-ratios ( $\delta_i$ ) are assumed to arise from a multivariate normal distribution (i.e., they are allowed to vary across studies), but  $\mu_i$  is assumed fixed. The covariance matrix of  $\delta_i$  incorporates and describes the correlations between treatment effects and may be assumed common to all studies or allowed to vary. Prior distributions must be set for  $\mu_i$ ,  $\delta_i$  and covariance matrix of the distribution to log-odds-ratios.

### 2.3.4 META-ANALYSIS OF LONGITUDINAL STUDIES

Multiple outcomes arise in longitudinal studies, which are commonly used in epidemiological research to track the effect of a treatment or exposure over time. These studies typically involve a series of measurements of the response variable at pre-determined intervals. Treatment effects can then be described by estimates calculated at various times (corresponding to the measurement times in the study). Alternatively, longitudinal data are sometimes analyzed using summary measures<sup>51</sup> (e.g., mean or slope of response values for each participant), in which case the treatment effect is expressed in terms of (relative or absolute) differences in the summary measure (e.g., difference in mean rate of change (slope) of blood pressure among treated and controls).

A few examples of meta-analyses of longitudinal data have appeared recently<sup>52-54</sup>. Two of the applications were meta-analyses to combine results from two trials using patient level data<sup>52,53</sup>. Lopes et al.<sup>52</sup> used a Bayesian model with mixtures of multivariate normal distributions to allow more flexible distributions for the random-effects. Farlow et al.<sup>53</sup> used a longitudinal mixed model but do not describe it in detail.

In most instances, however, patient level data are not available and meta-analyses must be based on effect estimates at different times reported in publications or obtained from the authors. This was the case in the meta-analysis of Maas et al.<sup>54</sup>. They describe a linear mixed model for the meta-analysis of estimates of change in measures of mental health in infants at different ages. The model that was used is similar to mixed models used for the analysis of longitudinal patient data in a

single study<sup>55,56</sup> with a slight modification to weight the variance of residuals by the precision of the estimates. These models do not necessarily assume multivariate distributions for residuals and random effects to account for correlations in the data. Rather, correlations are captured through shared random-effects. For instance, assuming independent normal distributions for observations within subjects but allowing a random intercept for each subject leads to a compound symmetry correlation structure between observations within subjects<sup>56</sup> (i.e.,  $\text{cor}(y_{ij}, y_{ik}) = \rho$ ). Alternatively, if response is modeled as a linear function of time, the slope of this relationship can also be allowed to vary across patients. Random intercepts and slopes may have independent distributions or be assumed to be correlated (i.e., have a multivariate distribution). Maas et al.<sup>54</sup> examined both random-intercept and random-time effects (since time was defined as categorical) models.

By assuming independence between outcomes within-studies, these models capture the total (within- and between-study) or *marginal* correlations through the shared random-effects. Factors that lead to correlations at the two levels may be different, however; for instance, “artifactual” sources of heterogeneity<sup>35</sup> between studies (e.g., allocation or blinding methods) can induce similarity in effects within-studies. A full multivariate method like those used in meta-analyses of multiple outcomes separates within- and between study correlations, thus allowing a more meaningful interpretation of the correlations being estimated.

### **2.3.5 BENEFITS AND RELIABILITY OF MULTIVARIATE METHODS**

The applications described above illustrate the versatility and benefits of using the multivariate approach over standard, outcome-specific meta-analyses. This section describes some of the advantages and challenges of performing multivariate meta-analyses as well as issues related to the reliability of the method

#### *Statistical Convenience*

For the analysis of a single endpoint, the multivariate approach may provide more information than outcome specific analyses with less effort. For instance, it is

possible to obtain pooled estimates of the reference rate (i.e., in the control group) and the treatment effect simultaneously. Historical controls or studies that may have employed different comparators can be included in analyses conducted with multivariate models. These also provide a framework for broader inference, allowing tests of hypotheses about the stability of effects over time or other interactions that can not be tested with independent models. The multivariate model is also helpful in overcoming the limitations of univariate models in assessing the effect of the reference risk on the effect estimate.

#### *Added Insight and Multivariate Hypotheses*

The multivariate approach is also efficient for the analysis of multiple (different) outcomes, and is intuitively more appealing than univariate methods, as it incorporates the various dimensions of the treatment effect and presents a unified assessment. Furthermore, this method is technically more appropriate as it accounts for the correlation between estimates from the same study; with a random-effects model, it also measures the correlation between the treatment effects on the different outcomes across studies.

#### *Correlations in Multivariate Meta-Analyses*

The multivariate model explicitly accounts for within- and between-study correlations between the outcomes, which can help to isolate different sources of correlations. Within-study covariance matrices are usually held fixed to observed values, however. In all but one<sup>5</sup> of the applications described above, the within-study covariances were not reported, and hence, unknown in the analysis. This was recognized as a limitation in the earliest applications<sup>1,8</sup> and recommendations were made to investigators to change reporting practices to allow the correct application of multivariate methods; there appears to be little change in this regard, however. Some got around the difficulty by using external estimates of the correlation between outcomes<sup>1,8</sup>, assuming independence (i.e., setting covariances to 0)<sup>6,12</sup>, using approximation techniques<sup>10</sup> or estimating from the data<sup>9</sup>. In the latter case, concern was raised about potential identifiability problems, especially when only a small portion of the included studies reported

both outcomes being examined. Thus, approximating within-study covariances might help lessen the burden by reducing the number of parameters to estimate, but ignores the uncertainty inherent to the approximations. Furthermore, the impact of errors in approximations on summary estimates and other parameters of the model is not clear and has not been formally studied.

#### *Improved Accuracy and Precision*

Multivariate methods were initially expected to provide improved accuracy<sup>5</sup> and precision<sup>1,5</sup>, which may in turn lead to added power to detect the overall treatment effect. However, estimates from univariate and multivariate models have been remarkably similar, even when the data were manipulated to create a situation where “borrowing strength” across outcomes would be expected to improve estimates<sup>9</sup>. Considering that only the shared or common “information” about the outcomes can potentially influence the results, a joint analysis does not incorporate new information in deriving summary estimates for each outcome. This may explain why multivariate meta-analyses have not shown very different results.

The benefit of multivariate methods for multiple outcomes would, therefore, appear to be in its practicality (since a single model produces summary estimates for multiple outcomes), and, more importantly, in allowing joint inferences about the outcomes – that is, when multivariate hypotheses are of interest. For instance, Berkey et al.<sup>5</sup> conducted multivariate hypothesis tests about the equality of treatment effects for the different outcomes. Multivariate models also allow the quantification of between-study correlations from the covariance matrix of random-effects. This provides added insight, which is not possible with univariate methods. In the one application where the latter was emphasized, however, correlation estimates lacked precision, to the point of being almost completely uninformative. Thus, it is not clear whether multivariate meta-analyses can be used reliably for this purpose.

### *Publication Bias*

Multivariate meta-analyses can include studies that report all or some of the outcomes of interest. It is important to consider, however, the possibility that some of the outcomes may have been selectively omitted from publication<sup>1,57</sup>; that is, it is possible that those observations are not missing “at random” – a form of publication bias in the multivariate context that could affect inferences about both the treatment effects and their correlations. The usual form of publication bias (studies with weak results being omitted entirely) remains a potential threat in the multivariate case as well.

## **2.4 OBJECTIVES OF THESIS**

The goal of this thesis was to address some of the reliability issues described in the previous section and examine further applications of multivariate models in meta-analyses. More specifically, I was interested in evaluating the reliability of using multivariate meta-analyses to measure correlations between the effects of a treatment on different outcomes, and to determine whether these correlations would reflect the true associations of interest. This can have potentially interesting applications in situations where outcomes are rare, so that power can be gained by measuring the correlation in meta-analyses rather than individual studies.

I also examined the impact of approximating within-study covariances or ignoring (i.e., assuming independence) on estimates of the parameters of the model. Assuming independence within-studies implies that random-effects capture the total correlations, whereas the multivariate model accounts within- and between-study correlations separately. It is not clear whether this can lead to better accuracy or precision of estimates.

Finally, I adapted multivariate models to the meta-analysis of longitudinal studies and compared this approach with more traditional mixed model specifications as well as standard time-specific meta-analyses.

## *Chapter 3*

### **(MANUSCRIPT I)**

# **CORRELATIONS IN MULTIVARIATE META-ANALYSES: WHAT ASSOCIATIONS ARE BEING MEASURED?**

## **PREAMBLE**

This article reports on findings from a simulation study aimed at assessing the reliability of using multivariate meta-analyses to infer about the correlation between effects of a treatment on two or more outcomes. Multivariate meta-analyses produce summary effect estimates for multiple outcomes from a joint model which accounts for within-study correlations between observations and measures between-study correlations between outcomes. The latter is only available from multivariate models, and therefore presents a potentially important advantage over standard outcome-specific meta-analyses. For instance, it may be used to determine whether response to a treatment (efficacy) is related to lower or higher occurrences of adverse events (safety). Since adverse events are often rare, measuring this association in a meta-analysis that combines several studies may be more powerful than from a single study.

The units of meta-analyses are effect estimates reported in publications (or obtained from authors), which can be thought of as *ecologic* measures, in the sense that they may hide underlying patient-level relationships. Thus, the objective was to determine whether correlations measured across studies would accurately reflect the true dependence between treatment effects. To address this

objective, a conceptual model relating various sources of correlations was developed to provide a framework for simulations.

The goal of this study was not to simulate meta-analyses and assess the performance of multivariate models as these would depend on the number of included studies as well as the size of these studies. Rather, the focus was on understanding how the various associations and dependencies in the conceptual model affect the correlations between *true* treatment effects across a large number of studies. Therefore, the measured correlations represent the true value of the correlations estimated in multivariate meta-analyses. Thus, inconsistencies observed at this level would only be magnified in meta-analyses which would be subject to further distortions due to limited number of studies, incomplete reporting, imprecise effect estimates, etc.

This article will be submitted to Statistics in Medicine.

# **CORRELATIONS IN MULTIVARIATE META-ANALYSES: WHAT ASSOCIATIONS ARE BEING MEASURED?**

**Khajak Ishak<sup>1</sup>, Robert W. Platt<sup>1,2</sup>, Lawrence Joseph<sup>1,3</sup>, James A. Hanley<sup>1,4</sup>**

<sup>1</sup>Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada;  
<sup>2</sup>The Montreal Children's Hospital Research Institute, McGill University, Montreal, Canada; <sup>3</sup>Division of Clinical Epidemiology, Royal Victoria Hospital, Department of Medicine, Montreal, Canada; <sup>4</sup>Division of Clinical Epidemiology, Montreal General Hospital, Department of Medicine, Montreal, Canada.

---

## **SUMMARY**

When multiple outcomes are of interest in a meta-analysis of the effect of a treatment or exposure, multivariate models can be used to perform a joint analysis to account for the correlation between the outcomes. This method is also useful to quantify the correlation between treatment effects on the outcomes. We examined whether the correlations measured from aggregated meta-analytic data reflect the relationships of interest, or whether these are distorted by other factors like correlated errors within studies that might induce similarity between observed outcomes. The focus of our analysis was not the estimation of correlations, but rather the relationships captured by the true correlations underlying each study. We simulated studies of the effect of a treatment on two dichotomous endpoints, incorporating various sources of correlation. Correlations observed across different scenarios in which we varied the strengths of the associations of interest were very similar; thus, the true relationship between treatment effects could not always be inferred accurately.

**Keywords:** Meta-analysis, multiple-outcome, correlation, multivariate.

---

### **3.1 INTRODUCTION**

Multivariate meta-analysis refers to a modeling approach in which two or more outcomes are analyzed jointly to derive summary estimates of treatment effects across studies from a single model, as opposed to conducting separate meta-analyses for each outcome. Several examples of such models have appeared in the literature, with applications like meta-analyses of treatment effects on two or more related endpoints<sup>1,5,8,9</sup>, occurrence rates of a single endpoint in treated and

control groups at different times<sup>6,10</sup>, and in evaluating the relationship between baseline risk and treatment effect<sup>11,12,13</sup>.

The model that is most commonly used is a multivariate extension of the DerSimonian-Laird model<sup>7</sup>, based on multivariate normal distributions to model the joint likelihoods of the observed estimates, typically with random-effects also arising from a multivariate normal distribution. While the model can also be fitted with fixed effects only (the first implementations were<sup>1,8,10</sup>), recent applications<sup>5,6,8,9,11-13</sup> have favored the random-effects approach. This is currently the trend in standard univariate applications as well, perhaps because heterogeneity in meta-analysis is increasingly seen as “usual rather than exceptional”<sup>30</sup>. In fact, some<sup>28,31</sup> have recommended that estimates from random-effects models be used irrespective of what is suggested by tests of homogeneity, since these often have very low power, and using a fixed-effects model when effects in fact vary across studies can lead to interval estimates that are too narrow.

In the multivariate context, the random-effects have a particularly important role since the covariance matrix of their distribution, which is estimated in the model, can be used to quantify the strength of the relationship between the outcomes. This provides added insight about the disease or treatment being studied that is not possible in univariate analyses. In other instances, the covariances between random-effects may be needed to derive the *correct* variance of measures derived as (linear) functions of the outcomes. Thus, estimating the correlations between outcomes is, at least theoretically, an important advantage of the multivariate approach.

A nice illustration was given by Arends et al.<sup>6</sup> in a meta-analysis of the risk of death or recurrence of stroke following carotid endarterectomy compared to conservative treatment with medication in studies of high risk patients. The surgery is associated with an acute risk of recurrence and mortality. The occurrence of the endpoint (death or recurrence of stroke) was, therefore, measured separately in the first month after surgery (by a failure probability) and

the subsequent follow-up period (by a rate per person-year). The overall event rate was used for the conservative arm. These parameters also define the survival probability in each arm at any given time (based on an assumption of stable hazards over time). The outcome of interest was the cumulative survival ratio (CSR) of the surgery and conservative treatment at various times. The standard approach would require a separate meta-analysis for each time point. With a multivariate analysis, however, the short-term risk and long-term rate in the surgical group and the rate in the conservative arm are analyzed jointly and an equation is derived for the (log) CSR as a function of time and treatment groups. Having the covariance matrix of the distribution of the risk and rates across studies (i.e., random-effects) allowed them to include all inherent variability into the estimates of the (log) CSR while accounting for the dependence between its components. The covariance matrix was also used to measure the correlations between the acute and long-term risks in the surgical group. Arends et al.<sup>6</sup> found a negative correlation, from which they deduced that there may be a type of “survival of the fittest” phenomenon occurring in patients undergoing the surgery.

There may also be interest in the relationship between the effects of a treatment (or exposure) on two or more endpoints. For instance, Nam et al.<sup>9</sup> estimated the correlation between the effect of environmental tobacco smoke on asthma and its effect on lower respiratory disease in a meta-analysis of a series of observational studies. One can imagine other situations where these correlations may be useful. For example, one might want to assess the association between adverse events and efficacy of a treatment. A negative association would suggest that patients who suffer adverse events are less likely to respond; weak correlations would indicate that response is not much affected by adverse events.

To be useful, however, these correlations must reflect the true underlying relationships between the effects of the treatment on the various endpoints at the level of the individual, and not only the “ecologic” correlations between estimates across studies. That is, the correlations should capture more than associations caused by errors that might affect measurements of the various endpoints in a similar way within each study – what Glasziou and Sanders<sup>35</sup> refer to as

“artifactual” causes of variation (e.g., differences in compliance levels, blinding and allocation methods, etc). Furthermore, one must also distinguish dependence between endpoints from dependence between treatment effects. For instance, myocardial infarction (MI) and stroke are related endpoints since patients who suffer one of these events are usually at higher risk for the other. However, a treatment that reduces the risk of stroke in some patients will not necessarily help prevent MIs in these patients. In other words, patients may respond differently to a treatment for one endpoint compared to another.

Our objective was to explore how these various sources of dependence are reflected in the correlations observed from aggregate meta-analytic data. That is, we aimed to determine whether the correlations of interest (i.e., between treatment effects) can be adequately identified or whether these are indistinguishable from other sources of correlations. The focus is not on the statistical performance of multivariate models in estimating correlations, but rather in describing the underlying conceptual model from which these correlations may arise. Thus, our analyses aimed to measure the *true* value of the correlation parameter estimated in multivariate meta-analyses by simulating a hypothetical population of studies from which those included in a meta-analysis would be selected. We measure the correlation between the effects of a treatment on two endpoints measured across studies in scenarios where we vary the degree of dependence in treatment effects and other factors; for instance, we consider endpoints that are highly dependent in the control population but the effects of the treatment are only weakly associated.

We describe the conceptual model depicting the various sources of correlations in the next section, followed by a description of the simulations we performed. Subsequent sections present and discuss the results.

### **3.2 CONCEPTUAL MODEL OF CORRELATIONS**

We considered multivariate meta-analyses of the effect of a treatment on two dichotomous endpoints, denoted A and B. To replicate the way meta-analytic data would arise, we formulated a conceptual model consisting of three parts; the

first describes the disease process, the second depicts the action of the treatment, and the third represents the execution of studies in this population. Correlations arise at each of these stages, but it is not clear how each would affect the correlation measured between treatment effect estimates in a multivariate meta-analysis.

### 3.2.1 THE DISEASE PROCESS

We use sufficient component cause diagrams<sup>58,59</sup> to represent how events A and B may be related for each individual in the source population (Figure 3.1). Suppose that each of the two events A and B can arise from two pathways, which we denote by  $A_1, A_2$ , and  $B_1, B_2$ . Each pathway is represented by a pie diagram, and each piece of each pie is a component cause that, together with its complement causes, is sufficient to produce the onset of the event. These component causes may be binary (present/absent) or continuous measures of intensity or severity of risk factors. We consider A and B to be independent when they do not share any component causes (Figure 3.1(a)). Conversely, sharing at least one component cause implies dependence between the events in the sense that the occurrence of one of the events informs us about the chances of the other. For instance, Figure 3.1(b) and (c) depict potentially moderate and strong correlations between the events: in (b), each of the pathways share a single common component cause ( $a_2$  and  $b_5$ ), while in (c), event B shares many of the component causes of A. An extreme form of dependence can arise when one of the events is a component cause of the other, as in Figure 3.1(d).

In general, the degree of dependence is not only determined by the number of shared component causes but also by the relative frequency of each pathway in the reference population. Furthermore, the dependence may be negative when the presence of a shared component cause is necessary for A to occur, while B can only occur if it is absent. For simplicity, we assume here that shared component causes are required to be present for both events A and B to occur.

We can represent the relationship between A and B mathematically with conditional probabilities. Suppose A occurs in the reference population with

probability  $P[A]$ , and depending on whether A occurred, event B can occur with probability  $P[B|A]$  or  $P[B|\text{not } A]$ . Events A and B are independent if  $P[B|A] = P[B]$  and dependent otherwise. The magnitude of the difference between  $P[B|A]$  and  $P[B|\text{not } A]$  reflect the strength of the dependence between A and B. In the extreme case where A is a component cause of B (Figure 1(d)),  $P[B|\text{not } A] = 0$ .

### 3.2.2 THE EFFECT OF TREATMENT

In our representation of the disease process, the treatment acts by decreasing the probability of occurrence (if these are binary) or reducing the intensity of one or more of the component causes. The treatment has a strong effect if it acts on many component causes, possibly on different pathways, or if it acts on a single cause of a very common pathway (i.e., that accounts for a substantial proportion of the occurrences of the event). The effects on the two endpoints will be dependent or correlated when the treatment acts on one or more shared component causes of A and B. Ultimately, the strength of the relationship will depend on the *importance* (i.e., commonness or degree of involvement in different pathways) of the common causes that are affected by the treatment.

While it is likely often the case, dependence between events need not imply dependence between treatment effects. For example, in Figure 1(b), events A and B are dependent since  $a_2$  and  $b_5$  are common causes. If the treatment acts on  $a_1$  and  $b_3$  only, however, the treatment effects will be independent, in the sense that the prevention of A would not inform us about the chances of preventing B. Effects would also be independent when the treatment has no effect on one of the outcomes. The converse is not true, however. Dependence between treatment effects would necessarily imply some dependence between endpoints, since the treatment must act on one or more common, or at least indirectly related, component causes of the endpoints.

We represent dependence between treatment effects mathematically by specifying odds-ratios (OR) for the three probabilities defining the disease process: we denote these  $\text{OR}[A]$ ,  $\text{OR}[B|A]$  and  $\text{OR}[B|\text{not } A]$ . The effects are independent when  $\text{OR}[B|A] = \text{OR}[B|\text{not } A]$  and related when  $\text{OR}[B|A] \neq \text{OR}[B|\text{not } A]$ . The

relative magnitude of the difference between these conditional ORs reflects the strength of the relationship between treatment effects.

### 3.2.3 BETWEEN-STUDY VARIABILITY

The previous sections describe the mechanism of the disease and treatment in a specific reference population. Studies conducted on *this* population would likely vary with respect to methodological design, measurement methods, study populations, and various other factors, including aspects of the general quality of the study, which might affect the odds ratios estimated from each study. These differences can potentially be major (e.g., randomized vs. observational study) or relatively minor (e.g., inclusion criteria of minimum age of 45 in one study versus 40 in another) and can alter the effects estimated in a given study in a systematic way or simply introduce random error or noise.

To represent this mathematically, we let  $\varphi$  denote the vector of log-odds of events A, B|A and B|not A in the reference population, and  $\psi$  denotes the vector of corresponding true log-ORs. Differences between studies can cause both systematic and random variation in these parameters across studies. We can write this more formally, for the  $j^{\text{th}}$  study, as follows:

$$\begin{aligned}\varphi_j &= \varphi + \alpha X + \delta_j \\ \psi_j &= \psi + \beta Z + \varepsilon_j,\end{aligned}$$

where  $\alpha$  and  $\beta$  represent the systematic effects of vectors of factors  $X$  and  $Z$  affecting the baseline risks (log-odds) and treatment effects (log-ORs), respectively;  $\delta_j$  and  $\varepsilon_j$  are vectors of random errors that distort the parameters across studies. It is plausible to expect that errors affecting one of the components of  $\varphi_j$  or  $\psi_j$  might have a similar impact on the others. Thus, the errors are likely to be correlated, thereby making the true parameters of a given study more similar than those from different ones.

We note that sampling or between-patient variability is not reflected in this formulation. It is only meant to represent how the true event probabilities and treatment effects may vary across studies. In reality, there would also be variability in these probabilities and effects between individuals. For example, presence (or absence) of component causes, or risk factors, would make the probability of the event higher (or lower) from one subject to another. Similarly, presence of component causes affected by the treatment will affect the subject's response. This variation may be explainable (or systematic) for known or measured component causes, but may also cause random variability (in the sense of frailty) when causes are unknown. This type of noise may also affect endpoints in a similar way (and so induce correlations), depending on the relationships between endpoints and effects. This would have an impact on the estimates, but not on the true values of the parameters for each study, which is the focus of our analysis.

### 3.2.4 EXAMPLE

To illustrate the conceptual model, consider the example of the effect of a new antihypertensive agent on preventing MIs and strokes. Hypertension is an established risk factor for both of these outcomes<sup>60</sup>, which implies that hypertension is a component cause of one or more pathways for MI and stroke. Thus, the risk (probability) of one event is related to the risk of the other, since high (or normal) blood pressure implies a high (or low) risk for both (all else being equal). Therefore, the occurrence of MI in a given patient informs us that this patient is at higher risk of stroke (i.e.,  $P[\text{Stroke}|\text{MI}] > P[\text{Stroke}|\text{not MI}]$ ).

Similarly, if the anti-hypertensive agent is effective, an improvement in blood pressure would imply a reduction in the risk of both outcomes in some patients. That is, the effect of the treatment on the risk of stroke is correlated with its effect on the risk of MI. For instance, among the treated, stroke is more likely to occur if an MI has occurred, as this suggests that treatment has not been effective for such patients (i.e.,  $\text{OR}[\text{Stroke}|\text{MI}] > \text{OR}[\text{Stroke}|\text{not MI}]$ ).

Trials examining the new agent will tend to vary with respect to criteria used to define the study population (e.g., minimum age, co-morbidities, duration of disease, etc.) and methodology (e.g., allocation of patients, blinding, measurement methods, etc). These factors can induce both systematic and random variation in the baseline risk of the population of the study and the effects being estimated in each study. It is likely, however, that these factors would affect the two outcomes in a similar way; for instance, a study in a population with a high baseline risk of MI, is likely to also have a high risk of stroke, and vice-versa. The same would apply for the effect of the treatment; a study that is *prone* to find a strong effect is likely to do so for both outcomes.

### 3.3 METHODS

A multivariate meta-analysis of the effect of the treatment on events A and B is typically based on the marginal log-ORs for these events observed within each study (e.g., Nam et al.<sup>9</sup>). The unit of analysis is the vector of observed effect estimates, denoted  $y'_j = (y_{A_j} \quad y_{B_j})$ , for study  $j = 1, 2, \dots, n$ . A bivariate normal (BVN) random-effects model is specified as follows:

$$\theta_j \sim BVN \left\{ \theta = \begin{pmatrix} \theta_A \\ \theta_B \end{pmatrix}, D = \begin{pmatrix} d_A^2 & d_{AB} \\ d_{AB} & d_B^2 \end{pmatrix} \right\}$$

$$y_j | \theta_j \sim BVN \{ \theta_j, S_j \},$$

where  $\theta$  is the vector of true treatment effects,  $\theta_j$  is the effect estimated in the  $j^{\text{th}}$  study,  $D$  is the covariance matrix of the random-effects, and  $S_j$  is the within-study covariance matrix. The correlation between outcomes would be derived from the components of  $D$  ( $d_{AB}/(d_A d_B)$ ).

The objective of our analysis was to assess how the dependence between endpoints, treatment effects and variability (or random effects) between studies would be reflected in the correlation between marginal log-odds ratios measured across studies. We were particularly interested to see how, and to what extent the

true dependence in treatment effects is captured in this measure. We are not concerned here with the performance of the model in estimating these correlations accurately in meta-analyses, which would depend on the number of studies and their sample sizes. Rather, we are interested in how the relative strengths of the relationships in the underlying *processes* are reflected in the *true* correlations between effects in different studies. Therefore, instead of deriving the correlations from parameter estimates in simulated meta-analyses, we obtained the *true* values of the correlations by simulating the higher level parameters (i.e., the first level of the model -  $\theta_j \sim BVN\{\theta, D\}$ ): true underlying event rates and ORs of each study) for a hypothetical population of studies within which meta-analyses would be carried out.

### **3.3.1 CORRELATION BETWEEN MARGINAL LOG-ORs**

We derive the correlation between the true marginal log-ORs ( $\theta_j$ ) across studies empirically, by simulating studies from the conceptual model described in Section 3.2. Replicates were obtained as follows:

- 1) Fix values of event probabilities  $P = (P[A], P[B|A], P[B|not A])$  and treatment effects  $OR = (OR[A], OR[B|A], OR[B|not A])$ . The strength of the dependence between events and treatment effects was controlled by manipulating the conditional probabilities and ORs of event B. Alternatively, dependence between events and treatment effects may be controlled by simulating a common underlying risk factor. In this case, the strength of the associations between events and effects would depend on the incidence and prevalence of the common risk factor, its implication in pathways for each outcome and the strength of the treatment in “blocking” the risk factor. The formulation of conditional probabilities and odds ratios captures the ultimate impact of the common risk factor on the correlation between events and effects and provides a more intuitive representation of the correlations, all the while limiting the number of parameters to manipulate in the simulations.

2) In general, the underlying event rates and treatment effects of a given study are subject to systematic and random variation, but for simplicity, we assume there are no systematic differences. To impose a correlation between the random errors, we generated the log-odds ( $\varphi_j$ ) and log-ORs ( $\psi_j$ ) from trivariate normal (TVN) distributions. That is, for the  $j^{\text{th}}$  study,

we obtain  $\varphi_j \sim TVN(\varphi, V^\varphi)$  and  $\psi_j \sim TVN(\psi, V^\psi)$  where  $V^\varphi$  and  $V^\psi$  represent  $3 \times 3$  covariance matrices controlling the spread of the errors. We specified these by fixing the variances (diagonal elements) and the correlations, and derived the covariances (off-diagonal elements).

3) Derive the underlying true event rates in the treated and control groups for

the  $j^{\text{th}}$  study,  $P_j^T$  and  $P_j^C$ , from  $\varphi_j$  and  $\psi_j$ , where  $P_j^C = \frac{\exp(\varphi_j)}{1 + \exp(\varphi_j)}$  and  $P_j^T = \frac{\exp(\psi_j + \varphi_j)}{1 + \exp(\psi_j + \varphi_j)}$ . (The equations are applied to each component of

vectors  $\varphi_j$  and  $\psi_j$ ). From these we can derive the marginal probability of event B in each group:  $P[B] = P[B | A]P[A] + P[B | \text{not } A](1 - P[A])$ .

4) Derive the underlying marginal log-OR of A and B ( $\theta_j = (\theta_{jA} \quad \theta_{jB})$ ) in each study from the probabilities of events A and B. That is,

$$\theta_{jA} = \log\left(\frac{P^T[A]/(1 - P^T[A])}{P^C[A]/(1 - P^C[A])}\right) \text{ and } \theta_{jB} = \log\left(\frac{P^T[B]/(1 - P^T[B])}{P^C[B]/(1 - P^C[B])}\right).$$

### 3.3.2 CHOICE OF PARAMETER VALUES

We varied the presence and strength of dependence between events, treatment effects and random-effects and derived the correlation between marginal log-ORs following the process described in the previous section. We were interested specifically in the following scenarios:

0. Independent events, independent treatment effects and uncorrelated random-effects;

1. Dependent events, independent treatment effects and uncorrelated random-effects;
2. Dependent effects, but independent events and uncorrelated errors;
3. Dependent events and effects, but uncorrelated random-effects;
4. Dependent events and effects with correlated random-effects.

Within each of these, we varied the commonness and dependence between events, the strength and dependence between treatment effects, the variability and correlation of random errors on the log-odds and log-ORs. The specific values used for each of these inputs are summarized in Table 3.1. For example, to create scenarios where the events were common, we set  $P[A] = 0.5$  and  $P[B|A] = 0.75$ . We set  $P[B|\text{not } A] = 0.75$  to make the endpoints independent,  $P[B|\text{not } A] = 0.5$  to create a moderate dependence and  $P[B|\text{not } A] = 0.25$  for strong dependence. To maintain comparable levels of dependence when the events were moderately common, we set probabilities by halving the common rates, so that the ratio  $P[B|A]/P[B|\text{not } A]$  was the same in both cases.

We set the treatment effect parameters in a similar way, varying the strength of the effect but maintaining a comparable degree of dependence between strong and moderate effect sizes. In addition to the values shown in Table 3.1, we also considered situations where the treatment had no effect on either outcome ( $\text{OR}[A] = \text{OR}[B|A] = \text{OR}[B|\text{not } A] = 1$ ) and where it had no effect on event B ( $\text{OR}[B|A] = \text{OR}[B|\text{not } A] = 1$ ).

We assumed that there was between-study variability in log-odds and log-ORs in all scenarios but we varied the presence or absence of correlations between these random-effects. In practice, it is unlikely that these errors would be independent; we examined this possibility in our simulations to isolate the impact of the other sources of dependence. For convenience, we assumed that the variance and correlations of between-study random-effects have the same magnitudes for the log-odds and log-ORs. The variance of the conditional log-odds and log-ORs of B were set to be twice as large as those of A to ensure that the variance of the

marginal log-odds and log-ORs for A and B were similar. Assuming equal variances led to marginal variances that were notably smaller for B.

### 3.3.3 ANALYSES

For each scenario, we generated 100,000 studies from the conceptual model following the steps described above. We denote the marginal log-ORs for A and B of the  $j^{\text{th}}$  replicate (study) by  $\theta_j = (\theta_{jA} \quad \theta_{jB})$ ,  $j = 1, 2, \dots, 100,000$ . We approximated the *true* correlations from these, using the standard formula:

$$Corr = \frac{\sum_j (\theta_{jA} - \bar{\theta}_A)(\theta_{jB} - \bar{\theta}_B)}{\sqrt{\sum_j (\theta_{jA} - \bar{\theta}_A)^2 \sum_j (\theta_{jB} - \bar{\theta}_B)^2}},$$

where  $\bar{\theta}_A$  and  $\bar{\theta}_B$  are the mean log-ORs measured from the draws. We consider correlations below 0.25 to be indicative of weak associations, 0.25 – 0.50 to be indicative of moderate associations and correlations above 0.50 to be indicative of strong associations. These thresholds are somewhat arbitrary, but were defined a priori to provide a framework for comparisons.

The goal of the analysis was to determine whether the correlations between marginal effects (OR[A] and OR[B]) measured across studies can be used to determine the strength of underlying relationship between treatment effects. To do so, we assessed the consistency between the underlying and measured correlations. For example, the correlations would be misleading if we observe a correlation below 0.25 when treatment effects are highly dependent. We also examined how the correlations varied in relation to the commonness of the event or strength of the treatment.

## 3.4 RESULTS

We first examined the scenario where all parameters (events, effects and random-effects) were independent. As expected, no correlation was observed between the

marginal log-ORs. Values ranged between -0.02 and 0.02 for differing effect sizes, commonness of events and degree of between-study variability.

We then considered scenarios where only the events were dependent. Figure 3.2 displays the correlations observed between the marginal log-ORs across studies when varying the frequency and strength of the dependence between the events, the size of the treatment effect and the degree of between study variability. Since the effects are assumed to be independent, we expected the correlations measured in these scenarios to be close to 0. However, even when the treatment had no effect on either endpoint and, hence, effects were independent, we observed relatively high correlations (approximately 0.50) between the marginal log-ORs when the events were common and highly dependent. The correlations were around 0.25 when the events were moderately common and highly dependent, or common and moderately dependent, and dropped by about half in the scenario where the events were moderately common and moderately dependent.

We observed similar (but slightly lower) correlations in scenarios where the treatment was assumed to have a moderate effect on one or both of the endpoints. Stronger treatment effects lead to weaker correlations between the marginal log-ORs, particularly when only one of the endpoints were affected by treatment. We still found misleadingly high correlations (between 0.25 and 0.50), however, when the endpoints were common and highly dependent. In all scenarios, the results were independent of the degree of between study variability.

Figure 3.3 shows correlations measured in scenarios where treatment effects were related but all other parameters were assumed independent. This may be an unlikely scenario in practice, but we considered it to understand how relationships in true conditional ORs translate to correlations between marginal log-ORs. To correctly reflect the underlying relationships, we would expect to observe correlations between 0.25 and 0.50 when the effects were assumed to be moderately dependent and above 0.50 when they were assumed to be highly dependent. In all cases, however, correlations did not exceed 0.25, even when the effects were highly dependent, suggesting, at best, a moderate relationship

between treatment effects, certainly not reflective of the true strength of dependence in treatment response (I.e.,  $OR[B|A] = 0.90$  while  $OR[B|\text{not } A] = 0.30$ ). As before, the correlations between marginal effects were weaker when the effect of treatment was assumed to be strong. Furthermore, we measured higher correlations when the endpoints were common.

Correlations were more representative of the underlying relationships in the more likely scenario where both events and effects are related, as shown in Figure 3.4. Nonetheless, the frequency and level of dependence of the endpoints often distorted the observed values. For instance, in scenarios where effects were assumed to be moderately related, we would expect the correlations to lie between 0.25 and 0.50. This only occurred, however, if the endpoints were either moderately frequent and highly dependent, or common and moderately dependent. The relationship between the effects was exaggerated when the endpoints were common and highly dependent and underestimated when the endpoints were moderately common and moderately dependent. On the other hand, high levels of dependence between treatment effects were likely to be accurately reflected only when the endpoints were common and highly dependent.

Figure 3.4 highlights the potential for misleading conclusions because of a substantial overlap between the correlations observed between marginal log-ORs across scenarios, when, in fact, the actual level of dependence of the effects were sometimes very different. Furthermore, stronger effect sizes led to consistently lower correlations, so that high levels of dependence would be more difficult to detect in these situations. In all cases, the results were independent of the degree of between study variability.

The results discussed so far have assumed that the random-effects are independent. We found that allowing a correlation between the random-effects of the log-odds of the events had very little influence on correlations between marginal log-ORs. As above, results were also not sensitive to changes in the variance of the log-OR random-effects. Therefore, for ease of presentation, the log-odds random-effects were fixed to be moderately correlated, and the log-OR

random-effects to be highly variable in findings from scenarios involving correlated random-effects. In addition to endpoint and effect parameters, we also varied the correlation between random-effects of the log-ORs in these scenarios.

We first considered situations where the treatment had an effect on only one of the endpoints (Figure 3.5), and, therefore, no dependence could exist between effects. Nevertheless, we observed considerably high correlations that were proportional to the magnitude of correlations between the random-effects. Similarly, in situations where we assumed dependence between effects (Figure 3.6), the observed correlations were noticeably higher than those from previous scenarios (Figure 3.4) and dominated by the correlations between the random-effects of the log-ORs. For example, when the random-effects were highly correlated, the correlations between the marginal log-ORs exceeded 0.75 regardless of the actual degree of dependence between the effects and settings for other parameters of the simulation. Thus, it was impossible to discern the underlying relationships when between-study errors were highly related.

There was also substantial overlap between observed correlations in scenarios where the treatment effects were moderately rather than highly dependent when the random-effects were moderately correlated. Correlations were consistently close to or greater than 0.50, particularly when the endpoints were common and highly related. Thus, moderate dependence between the treatment effects would be difficult to identify. On the other hand, scenarios with weakly related random-effects missed strong relationships between treatment effects when the endpoints were moderately common and moderately related, but captured moderate dependence between the treatment effects more accurately.

### **3.5 DISCUSSION**

An important potential advantage of multivariate meta-analyses is the estimation of the covariance matrix of the outcomes. This is not only useful statistically, as it allows one to derive the correct variance of measures derived from the outcomes, but also to quantify the correlations between outcomes. To provide

useful insight about the disease or treatment being examined, these correlations must reflect more than the ecological relationships between estimates across studies. In fact, perhaps most important is that they mirror the strength of the dependence between the effect of the treatment on the endpoints of the studies as this can inform whether a patient's response to treatment on one endpoint is correlated with their response on another (e.g., whether efficacy is related to safety of new treatments).

Our analysis revealed that correlations measured across studies do not always reflect the strength of the true relationships between treatment effects. Indeed, major differences between some of the scenarios we considered were not always clearly reflected in the correlations measured between marginal log-ORs; for instance, similar correlations were sometimes observed in scenarios with moderately and highly related events or effects, or even for independent and related treatment effects. In particular, the event frequency was a key factor; weaker correlations were consistently observed when the events were less common, even though the level of dependence between events were comparable to those in scenarios with common events.

In general, correlations between random-effects of log-ORs had the greatest influence on the observed correlations, while the dependence between treatment effects had much less impact, despite strong underlying relationships. In fact, very high correlations were almost always indicative of highly correlated errors, but low or moderate levels were more difficult to interpret, since there was considerable overlap in correlations observed across different scenarios. Thus, one might be led to potentially misleading conclusions about the relationship between the events and effects being considered.

Scenarios where the effect of treatment was assumed to be strong consistently yielded lower correlations than when the effect was moderate (other factors held constant). This is because, when the treatment has an effect on the endpoints, not only are the event rates different in the treated and control groups, but the relationship between events is also altered when the magnitude of the effect on

the two endpoints is not the same. Consider, for instance, the scenario where the event is moderately common and moderately dependent in the control group ( $P[A]=0.25$ ,  $P[B|A]=0.375$ ,  $P[B|\text{not } A]=0.25$ ), where event B is 50% more likely to occur when A has occurred ( $P[B|A]/P[B|\text{not } A]=1.5$ ). When the effect has a moderate (but independent) effect ( $\text{OR}[A]=0.6$ ,  $\text{OR}[B|A]=0.9$ ,  $\text{OR}[B|\text{not } A]=0.9$ ), the corresponding rates in the treated group are 0.17, 0.35 and 0.23, respectively and event B remains 52% more likely to occur when A has occurred. If the effect was strong ( $\text{OR}[A]=0.3$ ,  $\text{OR}[B|A]=0.45$ ,  $\text{OR}[B|\text{not } A]=0.45$ ), however, the rates in the treated group are 0.09, 0.21 and 0.13 and the ratio of conditional rates is 63%. This increased discrepancy in the relationship between events in the study groups likely leads to lower correlations between the marginal log-ORs for the two events, which leads to a dilution of the actual correlations between treatment effects.

Higher variability in treatment effects across studies had little impact on results, but correlations between random-effects of the treatment effects were very influential. This may be due to the fact that these are applied on the same scale as the correlations measured across studies (i.e., log-ORs). We examined this by applying random-effects to the log-odds of the treated and control groups instead of applying them to the log-ORs (results shown in appendix, Section A3.7.2). The correlation between random-effects of log-odds in control and treated groups had similar influence on the results in these analyses. We found similar results to those presented in this manuscript when the random-effects were weakly or moderately correlated, but stronger correlations between the random-effects were less influential. The conclusions from the two sets of analyses remain consistent, however.

We made some simplifying assumptions to facilitate the implementation and presentation of our analyses and results. We assumed that there are no systematic differences in the various parameters across studies. In reality, treatment effects or event probabilities may depend on study characteristics. Although we allowed variability in event rates and treatment effects between studies (i.e., that studies may have slightly different reference populations), we assumed that the variability

was only random. Here again, it is possible that there may be systematic differences between these populations. In practice, ignoring systematic variability between studies can affect the correlations measured in meta-analyses; it would, therefore, be important to control for these by including the relevant characteristics in meta-regression models.

Our simulations examined correlations between study-level parameters and did not incorporate patient level variability. In practice, correlations estimated from meta-analyses are subject to additional variability due to sampling error within the studies and will depend on the size of the meta-analysis. Thus, we would expect further distortions from the underlying true values; for instance, even when events, treatment effects and between-study errors are independent, correlation estimates may not be as close to 0 as what we observed.

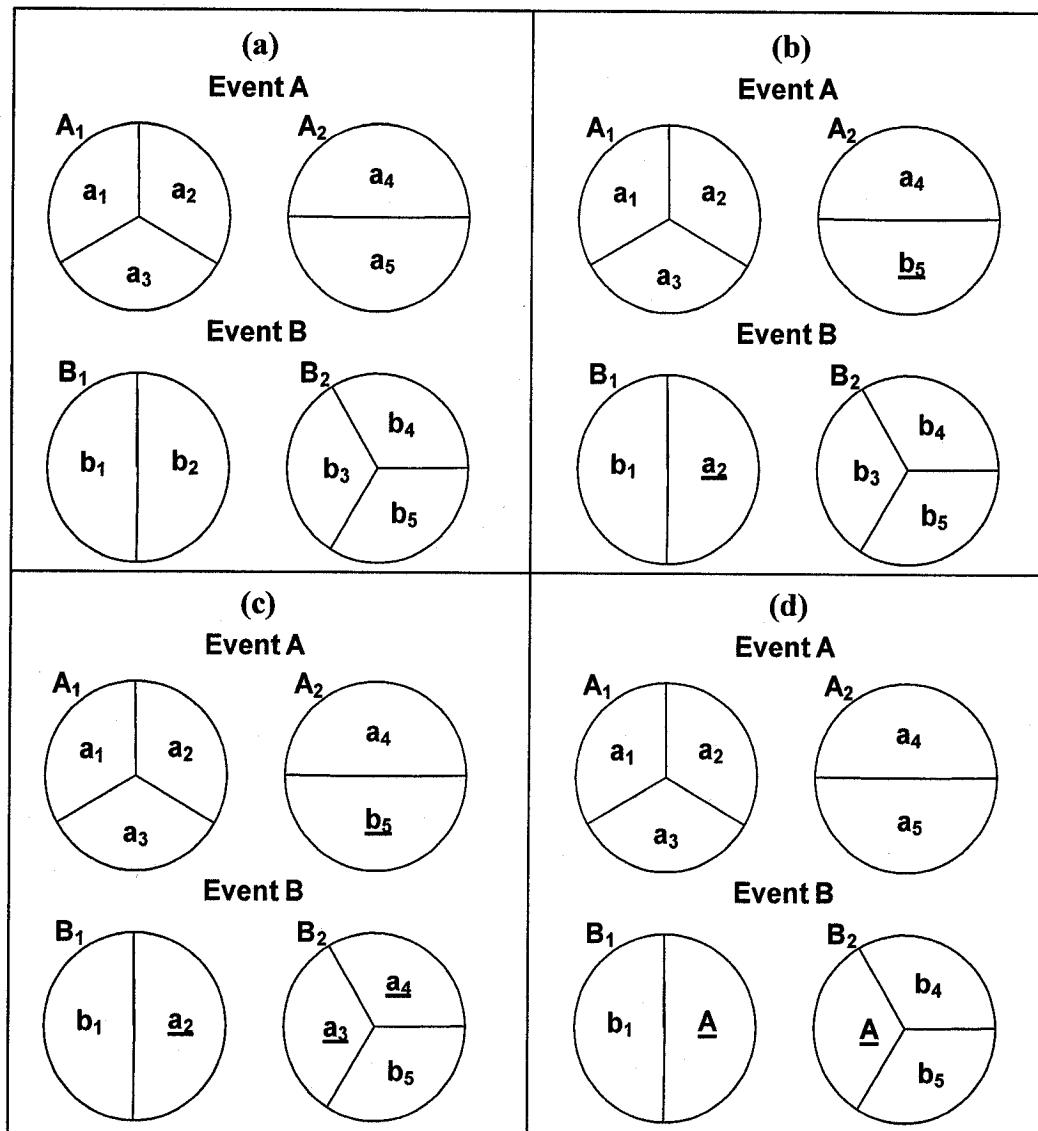
Findings from this study may also have implications for correlations measured in multivariate analyses of patient-level data in a given study. The conceptual model we used can also be interpreted as the underlying process of the disease and treatment within individuals. That is, the underlying risk of the endpoints and response to treatment varies across individuals; and, background factors (random-effects) causing variability between individuals may affect the risks and treatment effects for different endpoints in a similar way. Although the analytical approaches for patient-level data would be different than those used in meta-analyses (particularly for dichotomous endpoints), findings from our study suggest that apparent correlations in responses to treatment for different endpoints within individuals may be misleading if strong associations exist between the endpoints and random-effects.

In summary, marginal effect estimates and multivariate normal models may be inadequate when one wishes to assess the correlation between treatment effects for different outcomes in meta-analyses (or in analyses of patient data, perhaps). Richer data and models that explicitly isolate the various sources of correlations may be more appropriate for this purpose.

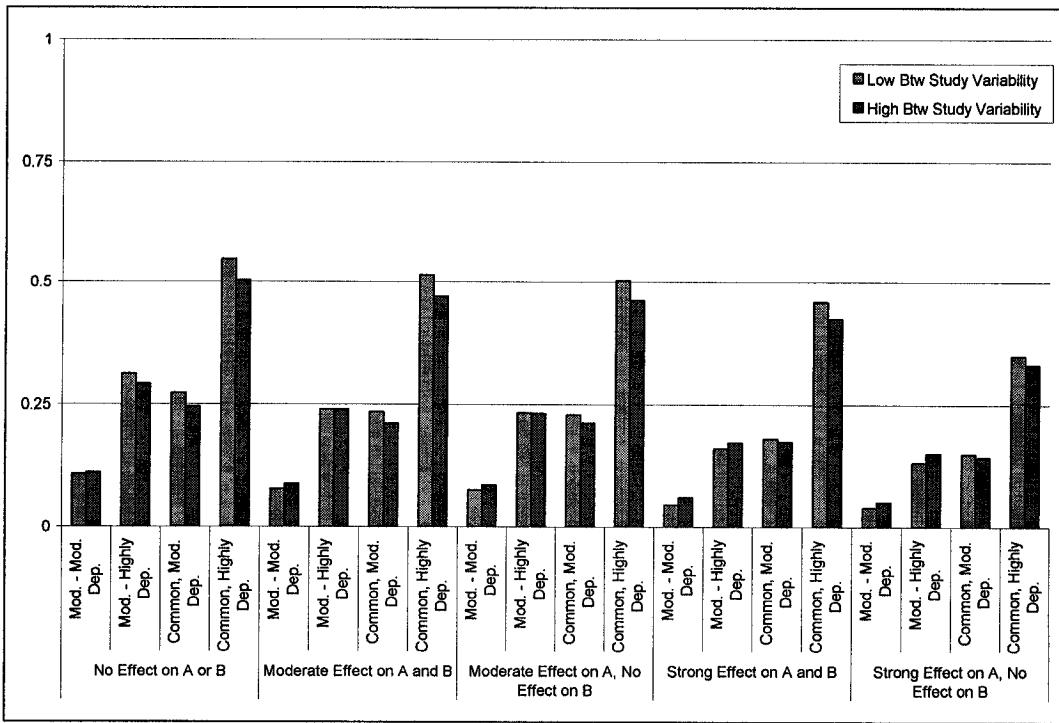
### 3.6 TABLES AND FIGURES

Event Parameters			
Probabilities	P[A]	P[B A]	P[B not A]
<u>Common</u> , Independent	0.50	0.75	0.75
<u>Common</u> , Moderately Dependent	0.50	0.75	0.50
<u>Common</u> , Highly Dependent	0.50	0.75	0.25
<u>Moderate</u> , Independent	0.25	0.375	0.375
<u>Moderate</u> , Moderately Dependent	0.25	0.375	0.25
<u>Moderate</u> , Highly Dependent	0.25	0.375	0.125
<b>Between Study Variance: Var[logit(P)]</b>			
High	0.1	0.2	0.2
Low	0.01	0.02	0.02
<b>Correlation of Errors: Corr[logit(P[x]), logit(P[y])]</b>			
High	0.75	0.75	0.75
Moderate	0.25	0.25	0.25
Low	0.10	0.10	0.10
Treatment Effect Parameters			
Odds Ratios	OR[A]	OR[B A]	OR[B not A]
<u>Strong</u> , Independent	0.30	0.45	0.45
<u>Strong</u> , Moderately Dependent	0.30	0.45	0.30
<u>Strong</u> , Highly Dependent	0.30	0.45	0.15
<u>Moderate</u> , Independent	0.60	0.90	0.90
<u>Moderate</u> , Moderately Dependent	0.60	0.90	0.60
<u>Moderate</u> , Highly Dependent	0.60	0.90	0.30
<b>Between Study Variance: Var[log(OR)]</b>			
High	0.1	0.2	0.2
Low	0.01	0.02	0.02
<b>Correlation of Errors: Corr[log(OR[x]), log(OR[y])]</b>			
High	0.75	0.75	0.75
Moderate	0.25	0.25	0.25
Low	0.10	0.10	0.10

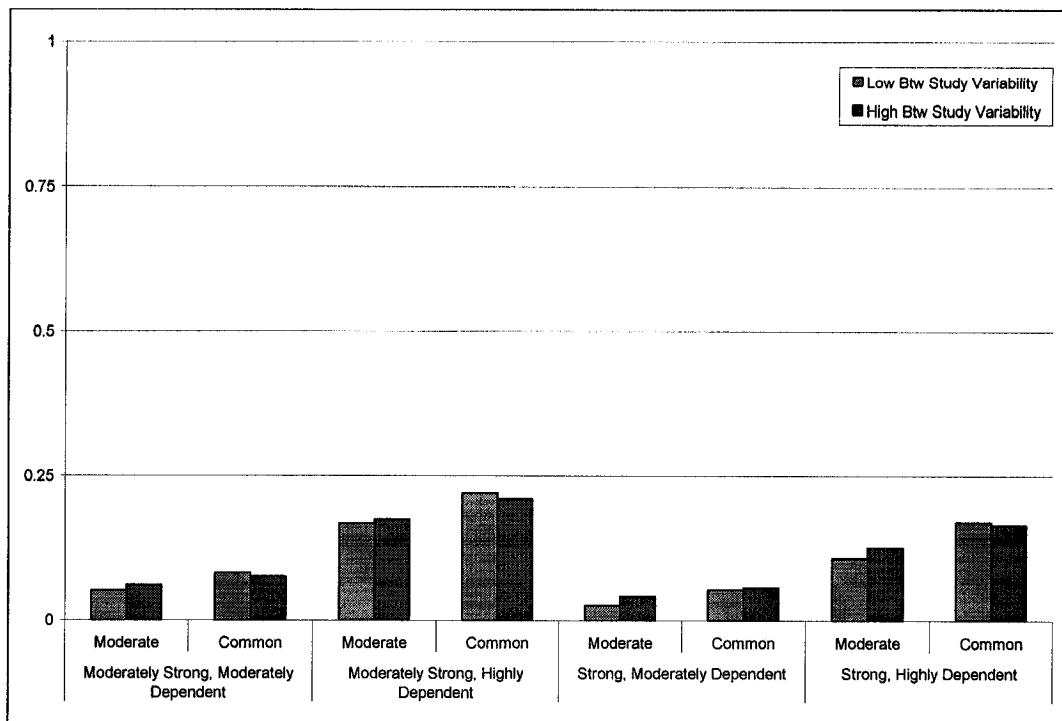
**Table 3.1** Values of inputs used to create scenarios for simulations from the conceptual model.



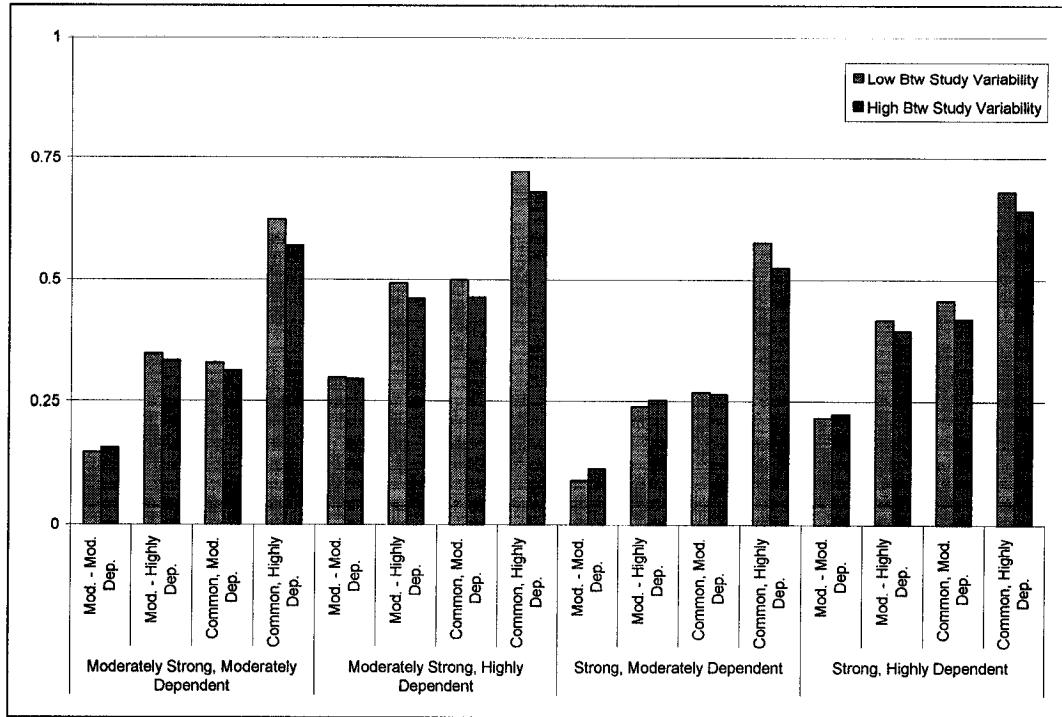
**Figure 3.1** Sufficient component cause diagrams depicting various forms of dependence between events. Each event is assumed to arise from two possible pathways, each represented by a pie diagram. Each slice represents a component cause, which, combined with the remaining component causes of that pathway, are sufficient to produce the onset of the event.



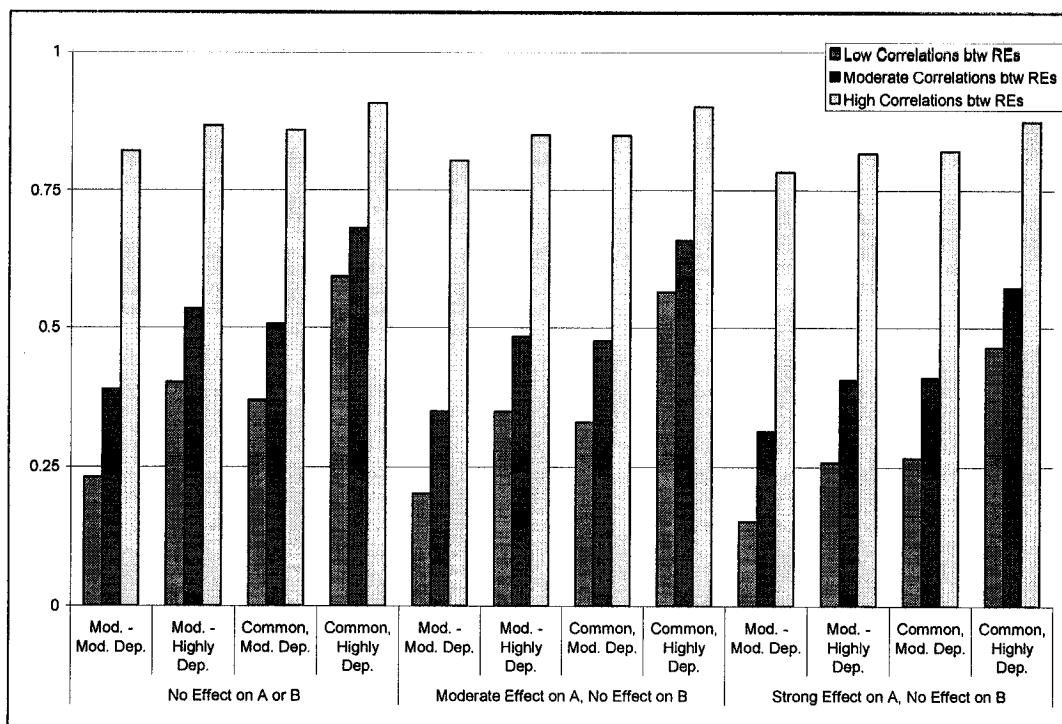
**Figure 3.2** Correlations between marginal log odds-ratios when the events are dependent, effects of the treatment are independent and between-study random-effects on log-odds and log-ORs are not correlated.



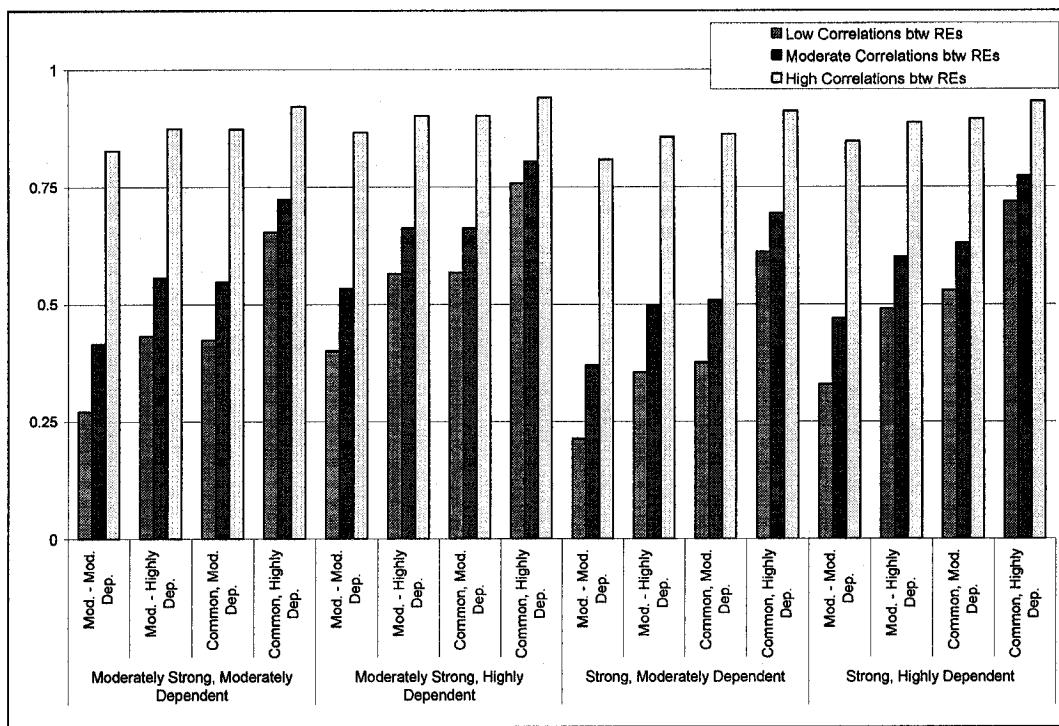
**Figure 3.3** Correlations between marginal log odds-ratios when the events are independent, effects of the treatment are dependent and between-study random-effects on log-odds and log-ORs are not correlated.



**Figure 3.4** Correlations between marginal log odds-ratios when the events and effects are dependent and between-study random-effects on log-odds and log-ORs are not correlated.



**Figure 3.5** Correlations between marginal log odds-ratios when the treatment has an effect on only one of the endpoints, events are dependent and between-study random-effects on log-odds and log-ORs are correlated.



**Figure 3.6** Correlations between marginal log odds-ratios when events and treatment effects are dependent and between-study random-effects on log-odds and log-ORs are correlated.

### 3.7 APPENDIX: ADDITIONAL MATERIAL

#### A3.7.1 TABLES CORRESPONDING TO FIGURES IN MANUSCRIPT

The tables below show the complete set of correlations measured in scenarios examined in the simulations and summarized graphically in the manuscript. The figures in the manuscript do not show results from scenarios where the degree of variability in log-odds and log-ORs were different (low/high, high/low) since these were not very different from and generally lay between those where the two were the same (low/low, high/high).

		Events									
		<u>Common,</u> Moderately Dependent		<u>Common,</u> Highly Dependent		<u>Moderate,</u> Moderately Dependent		<u>Moderate,</u> Highly Dependent			
<b>Treatment Effects</b>	<b>Btw- Study Variance</b>	Low	High	Low	High	Low	High	Low	High		
		No Effect on A or B	Low	0.272	0.248	0.547	0.521	0.107	0.110	0.311	0.299
			High	0.265	0.245	0.525	0.503	0.110	0.110	0.303	0.291
Moderate Effect on A and B	<i>Low</i>	0.235	0.191	0.514	0.437	0.077	0.073	0.240	0.206		
		0.233	0.211	0.497	0.470	0.085	0.087	0.242	0.240		
Moderate Effect on A, No Effect on B	<i>Low</i>	0.229	0.185	0.502	0.410	0.075	0.070	0.233	0.206		
		0.222	0.212	0.496	0.462	0.082	0.085	0.235	0.232		
Strong Effect on A and B	<i>Low</i>	0.180	0.127	0.459	0.369	0.045	0.040	0.160	0.117		
		0.186	0.174	0.451	0.423	0.057	0.060	0.177	0.172		
Strong Effect on A, No Effect on B	<i>Low</i>	0.149	0.094	0.348	0.218	0.039	0.031	0.131	0.090		
		0.155	0.142	0.370	0.331	0.047	0.050	0.149	0.150		

**Table 3.2(A)** Correlations between marginal log odds-ratios when the events are dependent, effects of the treatment are independent and between-study random-effects on log-odds and log-ORs are not correlated.

		Events			
		Common		Moderate	
Treatment Effects	Btw-Study Variance	Low	High	Low	High
	No Effect on A or B	Low	0.001 0.000	0.001 0.007	
		High	-0.001 0.000	0.005 0.004	
		Low	-0.004 -0.004	-0.001 -0.007	
		High	-0.008 -0.006	-0.003 0.006	
		Low	0.081 0.064	0.051 0.060	
		High	0.079 0.075	0.058 0.060	
		Low	0.221 0.137	0.167 0.146	
		High	0.236 0.210	0.175 0.174	
		Low	0.002 -0.006	0.000 0.003	
		High	-0.006 -0.015	0.002 0.004	
		Low	0.053 0.035	0.025 0.038	
		High	0.070 0.056	0.039 0.041	
		Low	0.170 0.098	0.107 0.104	
		High	0.185 0.164	0.126 0.125	

**Table 3.3(A)** Correlations between marginal log odds-ratios when the events are independent, effects of the treatment are dependent and between-study random-effects on log-odds and log-ORs are not correlated.

		Events									
		<u>Common,</u> Moderately Dependent		<u>Common,</u> Highly Dependent		<u>Moderate,</u> Moderately Dependent		<u>Moderate,</u> Highly Dependent			
<b>Treatment Effects</b>	<b>Btw- Study Variance</b>	<i>Low</i>	<i>High</i>	<i>Low</i>	<i>High</i>	<i>Low</i>	<i>High</i>	<i>Low</i>	<i>High</i>		
		Moderately Strong, Moderately Dependent	<i>Low</i>	0.328	0.276	0.622	0.556	0.147	0.150	0.347	0.327
			<i>High</i>	0.329	0.313	0.600	0.568	0.152	0.157	0.346	0.334
Moderately Strong, Highly Dependent	<i>Low</i>	0.499	0.362	0.721	0.539	0.299	0.240	0.491	0.368		
	<i>High</i>	0.498	0.463	0.725	0.679	0.302	0.296	0.495	0.461		
Strong, Moderately Dependent	<i>Low</i>	0.268	0.209	0.575	0.520	0.090	0.083	0.240	0.213		
	<i>High</i>	0.271	0.264	0.552	0.524	0.107	0.114	0.261	0.251		
Strong, Highly Dependent	<i>Low</i>	0.456	0.346	0.679	0.533	0.218	0.201	0.416	0.371		
	<i>High</i>	0.449	0.419	0.676	0.641	0.222	0.226	0.412	0.394		

**Table 3.4(A)** Correlations between marginal log odds-ratios when the events and effects are dependent and between-study random-effects on log-odds and log-ORs are not correlated.

		Events							
		<u>Common,</u> <u>Moderately</u> <u>Dependent</u>		<u>Common,</u> <u>Highly</u> <u>Dependent</u>		<u>Moderate,</u> <u>Moderately</u> <u>Dependent</u>		<u>Moderate,</u> <u>Highly</u> <u>Dependent</u>	
		Variance of Random Effects of Log-Odds							
Treatment Effects	Correlation of Random-Effects	Low	High	Low	High	Low	High	Low	High
Moderately Strong, Moderately Dependent	<i>High</i>	0.871	0.864	0.920	0.912	0.827	0.826	0.873	0.870
	<i>Moderate</i>	0.546	0.534	0.724	0.706	0.414	0.418	0.556	0.547
	<i>Low</i>	0.422	0.408	0.654	0.630	0.270	0.270	0.432	0.420
Moderately Strong, Highly Dependent	<i>High</i>	0.901	0.886	0.939	0.924	0.864	0.853	0.901	0.887
	<i>Moderate</i>	0.663	0.636	0.804	0.777	0.533	0.516	0.662	0.636
	<i>Low</i>	0.568	0.540	0.758	0.725	0.399	0.389	0.565	0.546
Strong, Moderately Dependent	<i>High</i>	0.860	0.846	0.911	0.906	0.807	0.801	0.853	0.845
	<i>Moderate</i>	0.508	0.497	0.694	0.677	0.369	0.367	0.498	0.489
	<i>Low</i>	0.375	0.367	0.612	0.595	0.212	0.218	0.353	0.356
Strong, Highly Dependent	<i>High</i>	0.893	0.879	0.931	0.919	0.845	0.840	0.886	0.880
	<i>Moderate</i>	0.630	0.612	0.772	0.749	0.469	0.466	0.601	0.589
	<i>Low</i>	0.530	0.501	0.719	0.692	0.328	0.324	0.490	0.479

**Table 3.5(A)** Correlations between marginal log odds-ratios when events and treatment effects are dependent and between-study random-effects on log-odds and log-ORs are correlated.

		Events							
		<u>Common,</u> Moderately Dependent		<u>Common,</u> Highly Dependent		<u>Moderate,</u> Moderately Dependent		<u>Moderate,</u> Highly Dependent	
		Variance of Random Effects of Log-Odds							
Treatment Effects	Correlation of Random-Effects	Low High		Low High		Low High		Low High	
No Effect on A or B	High	0.859	0.856	0.907	0.903	0.821	0.818	0.867	0.863
	Moderate	0.506	0.494	0.680	0.664	0.388	0.394	0.534	0.521
	Low	0.369	0.362	0.592	0.573	0.231	0.227	0.401	0.398
Moderate Effect on A, No Effect on B	High	0.849	0.841	0.901	0.892	0.804	0.801	0.850	0.844
	Moderate	0.475	0.477	0.658	0.638	0.349	0.353	0.483	0.477
	Low	0.330	0.326	0.564	0.548	0.201	0.193	0.348	0.343
Strong Effect on A, No Effect on B	High	0.821	0.795	0.874	0.843	0.783	0.767	0.817	0.792
	Moderate	0.409	0.392	0.572	0.539	0.313	0.302	0.405	0.392
	Low	0.266	0.248	0.464	0.431	0.152	0.159	0.257	0.247

**Table 3.6(A)** Correlations between marginal log odds-ratios when the treatment has an effect on only one of the endpoints, events are dependent and between-study random-effects on log-odds and log-ORs are correlated.

### A3.7.2 RESULTS FROM SIMULATIONS BASED ON LOG-ODDS

Findings from the main simulations revealed that the correlations between random-effects on treatment effects (log-ORs) had a substantial impact on the correlation between treatment effects measured across studies. We suspected this may be an artifact of the design of the simulations, since the correlation between random-effects were applied on the same scale as that on which correlations were being measured— i.e., log-ORs. We examined this by applying random-effects to the log-odds in treated and control groups, instead of log-ORs directly. Findings from these analyses are summarized in the tables below.

Results from these analyses were consistent with those presented in the manuscript. As expected, a high correlation between random-effects of the log-odds led to slightly weaker correlations than those from the original simulations. The conclusions presented in the manuscript still hold, however, as high correlations in random-effects still caused significant distortions in the correlations measured between treatment effects across studies.

		Events								
		Common, Moderately Dependent		Common, Highly Dependent		Moderate, Moderately Dependent		Moderate, Highly Dependent		
Treatment Effects	Btw- Study Variance	Low	High	Low	High	Low	High	Low	High	
		No Effect on A or B	Low	0.269	0.253	0.552	0.531	0.106	0.118	0.312
Moderate Effect on A and B	High		0.268	0.261	0.531	0.529	0.109	0.114	0.304	0.306
	Low		0.252	0.260	0.538	0.528	0.093	0.113	0.278	0.296
Moderate Effect on A, No Effect on B	High		0.237	0.244	0.508	0.515	0.087	0.095	0.248	0.271
	Low		0.251	0.255	0.536	0.529	0.094	0.110	0.272	0.298
Strong Effect on A and B	High		0.226	0.244	0.505	0.514	0.083	0.097	0.244	0.273
	Low		0.232	0.249	0.510	0.523	0.081	0.104	0.234	0.289
Strong Effect on A, No Effect on B	High		0.195	0.222	0.463	0.491	0.062	0.082	0.190	0.231
	Low		0.211	0.253	0.459	0.516	0.068	0.108	0.218	0.284
	High		0.164	0.205	0.392	0.444	0.051	0.080	0.164	0.221

**Table 3.7(A)** Correlations between marginal log odds-ratios when the events are dependent, effects of the treatment are independent and between-study random-effects on log-odds of treated and control groups are not correlated.

		Events			
		Common		Moderate	
Treatment Effects	Btw-Study Variance	Low	High	Low	High
No Effect on A or B	Low	0.002	-0.001	0.000	0.005
	High	-0.002	0.001	0.005	0.008
Moderate Effect on A, No Effect on B	Low	-0.002	-0.005	0.000	0.004
	High	-0.007	-0.003	0.000	0.007
Moderately Strong, Moderately Dependent	Low	0.046	0.010	0.026	0.009
	High	0.074	0.043	0.052	0.029
Moderately Strong, Highly Dependent	Low	0.116	0.026	0.085	0.022
	High	0.215	0.117	0.159	0.095
Strong Effect on A, No Effect on B	Low	-0.002	-0.001	-0.001	0.000
	High	-0.007	-0.001	0.004	0.002
Strong, Moderately Dependent	Low	0.031	0.007	0.014	0.005
	High	0.063	0.031	0.036	0.021
Strong, Highly Dependent	Low	0.092	0.018	0.056	0.009
	High	0.170	0.095	0.116	0.065

**Table 3.8(A)** Correlations between marginal log odds-ratios when the events are independent, effects of the treatment are dependent and between-study random-effects on log-odds of treated and control groups are not correlated.

		Events							
		<u>Common,</u> Moderately Dependent		<u>Common,</u> Highly Dependent		<u>Moderate,</u> Moderately Dependent		<u>Moderate,</u> Highly Dependent	
<b>Treatment Effects</b>	<b>Btw- Study Variance</b>	Low	High	Low	High	Low	High	Low	High
		Moderately Strong, Moderately Dependent	<i>Low</i>	0.304 0.268	0.593 0.536	0.126 0.119		0.332 0.311	
		Moderately Strong, Highly Dependent	<i>High</i>	0.327 0.296	0.599 0.565	0.145 0.134		0.345 0.329	
			<i>Low</i>	0.394 0.281	0.654 0.551	0.206 0.127		0.413 0.323	
			<i>High</i>	0.480 0.384	0.715 0.630	0.284 0.212		0.483 0.401	
		Strong, Moderately Dependent	<i>Low</i>	0.276 0.262	0.566 0.533	0.102 0.105		0.279 0.296	
		Strong, Highly Dependent	<i>High</i>	0.273 0.264	0.556 0.542	0.106 0.108		0.265 0.275	
			<i>Low</i>	0.369 0.276	0.627 0.541	0.165 0.117		0.365 0.318	
			<i>High</i>	0.436 0.358	0.669 0.607	0.213 0.165		0.404 0.358	

**Table 3.9(A)** Correlations between marginal log odds-ratios when the events and effects are dependent and between-study random-effects on log-odds of treated and control groups are not correlated.

		Events							
		<u>Common,</u> <u>Moderately</u> <u>Dependent</u>		<u>Common,</u> <u>Highly</u> <u>Dependent</u>		<u>Moderate,</u> <u>Moderately</u> <u>Dependent</u>		<u>Moderate,</u> <u>Highly</u> <u>Dependent</u>	
		Variance of Random Effects of Log-Odds							
Treatment Effects	Correlation of Random-Effects	Low	High	Low	High	Low	High	Low	High
Moderately Strong, Moderately Dependent	High	0.843	0.705	0.902	0.807	0.792	0.627	0.848	0.720
	Moderate	0.544	0.527	0.722	0.704	0.413	0.406	0.555	0.544
	Low	0.432	0.469	0.659	0.666	0.280	0.328	0.441	0.481
Moderately Strong, Highly Dependent	High	0.873	0.721	0.921	0.820	0.828	0.649	0.874	0.736
	Moderate	0.652	0.584	0.796	0.742	0.520	0.460	0.652	0.598
	Low	0.564	0.536	0.755	0.717	0.397	0.399	0.566	0.552
Strong, Moderately Dependent	High	0.833	0.698	0.894	0.805	0.776	0.615	0.830	0.707
	Moderate	0.508	0.511	0.696	0.688	0.369	0.379	0.502	0.511
	Low	0.390	0.444	0.620	0.653	0.225	0.295	0.372	0.448
Strong, Highly Dependent	High	0.864	0.713	0.913	0.815	0.808	0.636	0.861	0.728
	Moderate	0.622	0.569	0.768	0.729	0.463	0.430	0.598	0.571
	Low	0.530	0.515	0.718	0.700	0.333	0.364	0.495	0.513

**Table 3.10(A)** Correlations between marginal log odds-ratios when events and treatment effects are dependent and between-study random-effects on log-odds of treated and control groups are correlated.

		Events							
		<u>Common,</u> Moderately Dependent		<u>Common,</u> Highly Dependent		<u>Moderate,</u> Moderately Dependent		<u>Moderate,</u> Highly Dependent	
		Variance of Random Effects of Log-Odds							
Treatment Effects	Correlation of Random-Effects	Low	High	Low	High	Low	High	Low	High
No Effect on A or B	<i>High</i>	0.832	0.701	0.891	0.802	0.786	0.619	0.842	0.713
	<i>Moderate</i>	0.507	0.504	0.682	0.680	0.389	0.391	0.536	0.532
	<i>Low</i>	0.384	0.446	0.604	0.638	0.244	0.311	0.416	0.471
Moderate Effect on A, No Effect on B	<i>High</i>	0.823	0.694	0.885	0.799	0.772	0.614	0.826	0.706
	<i>Moderate</i>	0.478	0.495	0.661	0.666	0.353	0.371	0.489	0.506
	<i>Low</i>	0.349	0.424	0.579	0.629	0.218	0.294	0.368	0.443
Strong Effect on A, No Effect on B	<i>High</i>	0.798	0.679	0.863	0.784	0.754	0.601	0.798	0.690
	<i>Moderate</i>	0.417	0.460	0.586	0.627	0.322	0.344	0.420	0.468
	<i>Low</i>	0.290	0.385	0.488	0.571	0.172	0.270	0.284	0.396

**Table 3.11(A)** Correlations between marginal log odds-ratios when the treatment has an effect on only one of the endpoints, events are dependent and between-study random-effects on log-odds of treated and control groups are correlated.

## *Chapter 4*

### **(MANUSCRIPT II)**

# **IMPACT OF APPROXIMATING OR IGNORING WITHIN-STUDY COVARIANCES IN MULTIVARIATE META-ANALYSES**

## **PREAMBLE**

This second manuscript aimed to evaluate the impact of using poor approximations of within-study covariances when using multivariate meta-analyses models. In standard univariate meta-analyses, observed effect estimates are typically weighted by their precision to allow larger studies to exert greater influence on the summary estimate. This is usually done (in frequentist approaches, at least) by fixing within-study variances in the likelihood to the observed values, which are assumed known without error. (Bayesian meta-analyses are often based on event counts<sup>32</sup>, which inherently reflect the strength of the study). In the multivariate case, however, the likelihood term for each study is a multivariate normal distribution and, therefore, involves a covariance matrix. While studies usually report the variance of effect estimates, the covariance terms of the matrix are rarely reported (or even calculated by the researchers).

External estimates of the correlation are sometimes used to approximate the unknown covariances – i.e., by using some measure of the correlation and multiplying by study-specific standard errors. Another approach is to ignore within-study correlations (i.e., assume independence between observed estimates) but allow a correlation between the random-effects for each outcome. The impact

of potential errors in these approximations on summary estimates and other parameters (e.g., between-study correlations) from multivariate meta-analyses has not been formally evaluated. This was the objective of this paper, with a specific focus on situations where approximations were of poor quality, which would occur when little external information is available about the correlation between outcomes.

The current and previous articles both aimed to evaluate the reliability of using multivariate models. While the first paper considered the use of these models for a specific application (i.e., measuring correlations), the second aimed to provide a more general assessment of the use of the model when within-study covariances are not reported.

This article will be submitted to Statistics in Medicine.

# **IMPACT OF APPROXIMATING OR IGNORING WITHIN-STUDY COVARIANCES IN MULTIVARIATE META-ANALYSES**

**Khajak Ishak<sup>1</sup>, Robert W. Platt<sup>1,2</sup>, Lawrence Joseph<sup>1,3</sup>, James A. Hanley<sup>1,4</sup>**

<sup>1</sup>Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada;

<sup>2</sup>The Montreal Children's Hospital Research Institute, McGill University, Montreal, Canada; <sup>3</sup>Division of Clinical Epidemiology, Royal Victoria Hospital, Department of Medicine, Montreal, Canada; <sup>4</sup>Division of Clinical Epidemiology, Montreal General Hospital, Department of Medicine, Montreal, Canada.

---

## **SUMMARY**

Multivariate meta-analyses are used to derive summary estimates of treatment effects for two or more outcomes of interest from a joint model. In addition to treatment effects, these models also quantify the correlations between outcomes across studies. To be fully specified, the model requires an estimate of the covariance or correlations between outcomes observed in each study. These are rarely available in published reports, so that analysts must either approximate these or ignore the within-study correlations between effect estimates from the same studies. We examined the impact of errors in approximating within-study covariances on the parameters of multivariate models in a simulation study. We found that treatment effect and heterogeneity estimates were not affected by inaccurate approximations, but estimates of the correlation between outcomes were sometimes highly biased. The potential for error is greatest when the covariance between outcomes within- and between-studies are of comparable scale.

**Keywords:** Meta-analysis, multiple-outcome, correlation, bias.

---

## **4.1 INTRODUCTION**

Meta-analyses typically examine the effect of a treatment on a set of outcomes that are relevant in the context of the disease and treatment being studied. The standard practice is to perform separate meta-analyses to obtain summary estimates of the effect of treatment on each outcome. This approach ignores a possible correlation between the outcomes and does not allow inferences about the *overall* effect of the treatment based on all outcomes, or variations in effect across outcomes. Alternatively, a joint analysis is possible with multivariate

(multiple-outcome) models that account for the dependence between observed estimates and quantify the correlation between treatment effects across studies. Thus, multivariate meta-analyses provide added insight the models include parameters describing how the outcomes are jointly distributed.

Several applications of these methods have appeared,<sup>1,5,6,8-12</sup> illustrating the potential flexibility and strength of the multivariate approach. These include meta-analyses of treatment effects on two or more related endpoints<sup>1,5,8,9</sup>, where the objective may be, for example, to estimate the relative sizes of the effects or quantify the correlation between the outcomes. Others<sup>6,10</sup> have meta-analyzed occurrence rates of a single endpoint in treated and control groups at different times; in this context, multivariate models can be used to measure treatment effects at different times without making assumptions about proportionality of effects, and to examine the correlations between the probabilities of the endpoint occurrence over time<sup>6</sup>. Multivariate methods have also been used to evaluate the relationship between baseline risks and treatment effect estimates for a given outcome across studies<sup>11-13</sup>, sometimes called control-rate-regression. Standard methods (such as meta-regression) can produce biased results in this situation since they are affected by correlated measurement errors between estimates from the same study; a hierarchical multivariate model can be used to isolate the structural relationship between baseline risks and treatment effects.

A commonly used multivariate meta-analytic model is a multivariate extension of the DerSimonian and Laird approach<sup>7</sup>. The model includes random-effects for outcomes within studies, which is thought to be more appropriate for meta-analyses<sup>28,30,31</sup>. In the multivariate case, the joint distribution of the random-effects yields estimates of the correlations between outcomes across studies. The joint likelihood of parameters from each study is approximated by a multivariate normal distribution with known covariance matrices obtained from each study, conditional on a vector of random-effects also arising from a multivariate normal distribution. As with the variances of estimates in univariate meta-analyses, the within-study covariance matrices must be assumed known (i.e., without error), to ensure identifiability of the other parameters of the model. The model can be

applied to a set of continuous effect estimates (e.g., differences in means) or log-odds-ratio estimates for dichotomous endpoints or a combination of both.

While the variances of observed estimates are typically reported or can be derived from information in the reviewed publications, the within-study covariances or correlations are rarely available. Indeed, estimates of these quantities are available only if each study includes a joint analysis of the patient data. This is rarely done, however, unless specific questions about the relationship between the outcomes were of interest in the studies. This was recognized as a potential limitation in the earliest applications of multivariate models<sup>1,8</sup> and recommendations were made to change reporting practices to allow correct application of the method in the future; however, this advice has had little impact to date. Within-study covariances were reported in only one<sup>5</sup> of the multivariate meta-analyses mentioned above. In all other applications, the covariances were specified by either approximating them from external information about the correlation between the outcomes<sup>1,8</sup>, assuming independence (i.e., setting covariances to 0)<sup>6,12</sup>, using approximation techniques<sup>10</sup> or incorporating a common within-study correlation nuisance parameter in the model<sup>9</sup>. In the latter case, concern was raised about potential identifiability problems, especially when only a small portion of the included studies report both outcomes being examined.

The impact of approximating or ignoring within-study correlations has not been examined in much detail. Berkey et al.<sup>8</sup> compared results from meta-analyses where within-study covariances were specified from an approximation of the correlation between outcomes (assumed constant across studies) to those from analyses that assume independence in fixed-effects models. They found moderate changes in pooled effect estimates and slight to moderate changes in the corresponding standard errors. We sought to perform a more formal assessment for multivariate meta-analysis models with random-effects in a simulation study. We compared the accuracy and precision of point estimates and coverage probabilities of interval estimates obtained from analyses where within-study covariances are observed to those obtained when these are ignored (i.e., outcomes assumed to be independent) or approximated using external information about the

correlations between the outcomes. We examined the impact of errors made in approximating the covariances on pooled effect estimates and the between-study covariance matrix parameters, with particular attention to estimates of the correlation between outcomes across studies, as this parameter may be of interest in multivariate meta-analyses.

The following section presents the general multivariate meta-analysis model that is commonly employed. We then describe the simulation design and the analyses that were performed on the simulated data. This is followed by a discussion of the results.

## 4.2 MULTIVARIATE META-ANALYSIS

### 4.2.2 MULTIVARIATE DATA FROM META-ANALYSES

Consider a meta-analysis of  $N$  studies examining the effect of a treatment on two endpoints. The treatment effects may be expressed as differences in means (for continuous endpoints) or risks (for dichotomous endpoints), or log transformed odds, risk or hazard ratio estimates. We represent the estimates recorded from the

$i^{\text{th}}$  study by the vector  $y_i = \begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix}$ , for  $i = 1, 2, \dots, N$ . It is possible that one of the

components of  $y_i$  may be missing, however, since not all studies will necessarily report both outcomes. As in univariate meta-analyses, it is necessary to record the variance of the estimates, denoted  $S_{11i}^2$  and  $S_{22i}^2$ . In the multivariate context, however, the models also require the observed correlations or covariance ( $S_{12i}$ ) between outcomes from each study. In short, the covariance matrix of the

estimates,  $S_i = \begin{pmatrix} S_{11i}^2 & S_{12i} \\ S_{12i} & S_{22i}^2 \end{pmatrix}$ , is needed to fully specify the multivariate model.

#### 4.2.2 MULTIVARIATE MODEL

The random-effects multivariate model assumes that the true underlying effect of

the  $i^{\text{th}}$  study,  $\theta_i = \begin{pmatrix} \theta_{1i} \\ \theta_{2i} \end{pmatrix}$ , arise from a bivariate normal distribution (BVN):

$$\theta_i \sim BVN \left\{ \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, D = \begin{pmatrix} D_{11}^2 & D_{12} \\ D_{12} & D_{22}^2 \end{pmatrix} \right\},$$

where  $\theta$  is the vector of true treatment effect parameters, and  $D$  is the covariance matrix of the random-effects; its diagonal elements measure the degree of heterogeneity in the true treatment effects across studies, while  $D_{12}$  represents the covariance between the effects. The correlation between effects is derived from the components of  $D$ , as  $\rho_D = D_{12} / \sqrt{D_{11}^2 \cdot D_{22}^2}$ .

The likelihood function of the  $i^{\text{th}}$  vector of observed estimates is given by a bivariate normal (BVN) distribution:

$$y_i | \theta_i \sim BVN \{ \theta_i, S_i \},$$

where  $S_i$  is assumed exactly known to ensure identifiability of the model. Some studies may report only one of the outcomes. If unobserved values are not imputed, the likelihood term reduces to that corresponding to a univariate normal distribution. It is important to consider, however, the possibility that outcomes may be selectively omitted from publication<sup>1</sup>; that is, that those observations are not missing at random – a form of publication bias in the multivariate context that could affect inferences about both the treatment effects and their correlations.

Study and outcome-specific covariates can easily be incorporated by setting  $\theta_i = X_i \beta$ , where  $X_i$  is a  $2 \times p$  matrix of covariates and  $\beta$  is a  $p \times 1$  vector of parameters. By default,  $X_i$  would include one or more indicators to identify outcomes in  $\theta_i$  and derive summary estimates for each outcome or differences in estimates between outcomes. For instance, setting  $X_i = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  and not an

allowing intercept in the model would produce a separate summary estimate of effect for each outcome. Otherwise, setting  $X_i = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$  would produce a summary estimate for the second endpoint (the intercept estimate) and the difference between summary estimates of the first and second endpoints.

#### **4.2.3 IMPACT OF ERRORS IN APPROXIMATING WITHIN-STUDY COVARIANCES ( $S_{12i}$ )**

To determine the extent to which errors in approximating within-study correlations can impact results from multivariate meta-analyses, we examine the way the parameters of the models are estimated.

##### *Estimation*

Estimation is typically done (from a frequentist viewpoint) by maximum likelihood (ML), restricted or residual maximum likelihood (REML), or generalized least-squares (GLS), in which case distributional assumptions about the data are only required for inference, not estimation. Analytical or closed form solutions exist only in the case where all outcomes are reported in each study and all  $S_i$  are equal<sup>55</sup>. This is unlikely, however, since the  $S_i$  would almost always vary due to differing sample sizes. In general, estimation is carried out with iterative procedures like the expectation-maximization (EM), Newton-Raphson or Fisher scoring algorithms. These are implemented in standard software for mixed models, like PROC MIXED in SAS/STAT software or the *lme* function in *R* (or S-plus); Berkey et al. describe an EM algorithm<sup>5</sup> for GLS and likelihood approaches. Multivariate meta-analysis can also be performed using Bayesian methods<sup>9</sup>, in which case the models can be conveniently implemented in BUGS<sup>43</sup>.

##### *Impact of Errors*

Since closed form parameter estimates do not exist, it is difficult to derive the exact impact of uncertainty about the within-study covariance parameters analytically. We note, however, that the estimation formulae that arise in commonly used algorithms (described in more detail in appendix) involve the

within-study covariance only through the total variance matrix  $V_i = D + S_i$ ; more specifically, the  $S_i$  always appear as part of the inverse of the total covariance matrices ( $V_i^{-1}$ ), where

$$V_i^{-1} = \frac{1}{(D_{11}^2 + S_{11i}^2)(D_{22}^2 + S_{22i}^2) - (D_{12} + S_{12i})^2} \begin{pmatrix} D_{22}^2 + S_{22i}^2 & -(D_{12} + S_{12i}) \\ -(D_{12} + S_{12i}) & D_{11}^2 + S_{11i}^2 \end{pmatrix}.$$

At each iteration, the components of  $D$  are replaced by their current estimated values while components of  $S$  are held fixed, since they are assumed known without error. Therefore, imprecise approximations of the  $S_{12i}$  will distort  $V_i^{-1}$  throughout the estimation procedure. And, since  $V_i^{-1}$  appears in estimators or score equations of  $\beta$  and  $D$ , both estimates of the treatment effects and the heterogeneity covariance matrix components may be affected by errors in specifying within-study covariances.

The *accuracy* of the approximations of the  $S_{12i}$  is clearly an important determinant of the potential extent of errors in estimates from the model. It seems plausible, however, that poor approximations may be more influential in certain situations. We hypothesize that the potential for bias is greater when  $S_{12i}$  is of similar or larger scale than  $D_{12}$ , since, in the opposite situation, the covariance term in  $V_i$  will be dominated by  $D_{12}$ . For instance, suppose the true within-study correlation is 0.50, but we erroneously approximate it to be 0.25; then, in a study where  $S_{11i}^2 = 10$  and  $S_{22i}^2 = 5$ , the covariance would be estimated to be 1.76, instead of 3.54. The impact of this error will be more important when  $D_{12}$  is in the same order or smaller than  $S_{12i}$  (e.g.,  $D_{12} < 5$ ) compared to a situation where  $D_{12}$  is of greater magnitude (e.g.,  $D_{12} > 20$ ).

We must consider, however, that the scale of the within-study covariance matrix is inversely proportional to the size of the study: small studies will likely produce less precise estimates (i.e., large values in  $S_i$ ), while variances (and so covariances) from large studies will tend to be comparatively small. Thus, the condition  $S_{12i} > D_{12}$  is more likely to be met in meta-analyses with small studies.

Alternatively, the same may happen when treatment effects are fairly homogeneous across studies (i.e., components of  $D$  are relatively small).

The goal of our study was to assess the extent of the bias or loss of precision caused by approximating within-study covariances by simulating meta-analyses in scenarios where we vary parameters affecting the relative scale of  $S_{12i}$  and  $D_{12}$ . In each situation, we compare results from meta-analyses based on approximations of differing accuracy. We describe our approach in more detail in the next section.

## 4.3 SIMULATION OF MULTIVARIATE META-ANALYSIS DATA

We simulated data from a hierarchical structure where we first *created* a study by assigning a set of true parameter values and then generating the responses of each patient in the study. Simulating patient-level data was necessary to allow calculation of within-study covariances; this would not be possible, for instance, by simulating observed effect estimates directly as in other simulations<sup>5</sup>. The simulation model is described in more detail below.

### 4.3.1 SIMULATION OF STUDY RESULTS

We simulated data for meta-analyses of the effect of a treatment on **two** correlated continuous outcomes. We assume  $N$  studies are included in the meta-analysis and  $n_i$  subjects are included in the (parallel) treated and control arms of the  $i^{\text{th}}$  study (total sample size =  $2n_i$ ). We denote the *true* response of patients to treatment

(e.g., reduction in blood pressure or lipid levels) by  $\delta' = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}$ . Patients in the

control groups are assumed to receive placebo; for simplicity, we also assume that the disease remains stable, so that no change would be expected in patients'

condition, i.e.,  $\delta^c = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ . Thus,  $\delta'$  also represents the true treatment effect.

Studies are likely to vary with respect to methodological design, measurement methods, study populations and various other factors that might affect the true values of parameters. To emulate this, we added random-effects to generate the true underlying responses of the  $i^{\text{th}}$  study,  $\delta_i^t$  and  $\delta_i^c$ . We assume the variability of treatment effects across studies is given by the  $2 \times 2$  covariance matrix  $D = \text{var}(\delta_i^t - \delta_i^c)$ , so that the covariance of responses in each arm is  $0.5D$ . Within a given study, we expect variability in the responses of different subjects, possibly in a correlated way. We assume the covariance matrix of the two responses for each subject is given by  $0.5S$ .

We generate data for studies from the following steps:

- 1) Define a scenario by fixing values for  $\delta^t$ ,  $D$  and  $S$ . The covariance matrices are specified by setting the variances (diagonal elements) and the correlation between random-effects ( $\rho_D$ ) and within-study responses for the two outcomes ( $\rho_S$ ). We assume the same correlation in treated and control arms; that is, treatment does not alter the association between outcomes (although this may well happen in practice). The covariance component of each matrix (off-diagonal element) can then be derived from the correlations and variances.
- 2) Generate the underlying expected responses,  $\delta_i^t$  and  $\delta_i^c$ , for studies  $i = 1, 2, \dots, N$ , by drawing  $\delta_i^t \sim BVN(\delta^t, 0.5D)$  and  $\delta_i^c \sim BVN(\delta^c, 0.5D)$ . This allows the possibility of a correlation between responses from a given study, which may occur, for instance, if factors that affect one of the outcomes affect the other in a similar (or opposite) way.
- 3) Simulate the responses of subjects in each study. For the  $j^{\text{th}}$  subject of the  $i^{\text{th}}$  study, we obtain  $d_{ij}^t | \delta_i^t \sim BVN(\delta_i^t, 0.5S)$  and  $d_{ij}^c | \delta_i^c \sim BVN(\delta_i^c, 0.5S)$ . We denote by  $d_i^t$  and  $d_i^c$  the  $n_i \times 2$  matrices of responses for all patients in the  $i^{\text{th}}$  study generated from the BVN draws.

4) Calculate estimates of effect and corresponding observed covariance matrix for each study. This is given by  $y_i = \begin{pmatrix} \bar{d}_{1i}^t - \bar{d}_{1i}^c \\ \bar{d}_{2i}^t - \bar{d}_{2i}^c \end{pmatrix}$  and  $S_i = \frac{1}{n_i} (\text{cov}(d_i^t) + \text{cov}(d_i^c))$ , where the covariance matrices of the observed data,  $d_i^t$  and  $d_i^c$ , are calculated empirically, using standard formulae.

Each replication of steps 2 to 4 creates a new set of studies for meta-analysis within the scenario defined in step 1.

#### 4.3.2 SIMULATION SCENARIOS

The relative sizes of within- and between-study covariances are the primary (input) parameters of interest in our simulations. We first examine a set of scenarios we consider to be reflective of most practical situations: we assume the degree of heterogeneity between studies is not so large as to completely dominate within-study variances, while at the same time studies are not so small as to produce highly imprecise estimates. We then consider more extreme cases where we reduce sample sizes and minimize between-study heterogeneity.

We consider meta-analyses of  $N = 25$  studies; this may be considered relatively large compared to most practical situations, but was chosen to ensure the feasibility of the multivariate method. We generated the sample sizes of the two arms of a hypothetical trial by randomly selecting numbers ranging between 50 and 250 to create studies of moderate size ( $n = \{230\ 231\ 160\ 152\ 248\ 116\ 198\ 230\ 158\ 190\ 55\ 209\ 60\ 238\ 88\ 241\ 138\ 177\ 108\ 140\ 54\ 219\ 208\ 131\ 99\}$ ). We assume the true response to treatment is a reduction of 10 units for the first outcome and a reduction of 5 units for the second, i.e.,  $\delta^t = (-10\ -5)$ , while no change is expected among the controls, i.e.,  $\delta^c = (0\ 0)$ . We consider alternative values for these and other parameters in sensitivity analyses (described in more detail below).

We set the between-study heterogeneity covariance matrix of treatment effects to be  $D_{11}^2 = 5.0$ ,  $D_{22}^2 = 2.5$  and  $\rho_D = 0.25$ , so that  $D_{12} = 0.25\sqrt{2.5 \cdot 1.25} = 0.88$ .

Between-patient variances were set to  $S_{11}^2 = 10.0$  and  $S_{22}^2 = 5.0$ , twice the magnitude of between-study variances, since one would expect less variability in aggregated measures (e.g., mean group vs. subject-specific difference). We note

that the variance of observed estimates from the  $i^{\text{th}}$  study will be  $S_{i11}^2 \approx \frac{10.0}{n_i}$  and

$$S_{i22}^2 \approx \frac{5.0}{n_i}.$$

In a first pair of scenarios, we set  $\rho_S = \rho_D = 0.25$  (scenario 1a) and  $\rho_S = 2\rho_D = 0.5$  (scenario 1b). We then reduced the sample size of the studies in scenario 1b, first by a factor of 1/5 so that the size of each arm ranges from 10 to 50, (scenario 2a) and then by 1/10, so that the size of each arm ranges from 5 to 25 (scenario 2b). Next, we examined the potential impact of the relative size of within- and between study variances based on specifications of scenario 1b. We first reduced between-study variances by a factor of 1/5, so that  $D_{11}^2 = 1.0$  and  $D_{22}^2 = 0.5$  (scenario 3a); we further reduced these variance by half (1/10 of the original value), so that  $D_{11}^2 = 0.5$  and  $D_{22}^2 = 0.25$  (scenario 3b).

#### 4.3.3 APPROXIMATING WITHIN-STUDY COVARIANCES

Within-study covariances are usually approximated from an estimate of the correlation between outcomes, denoted  $\rho_S^*$ , and calculating  $S_{12i} = \rho_S^* \times S_{11i} \times S_{22i}$  for each study. In practice, the estimate may be based on expert opinion or observed from a study that may or may not be included in the meta-analysis, or obtained by other means. Clearly, an important determinant of the impact of approximating unobserved covariances is the accuracy of the approximation itself. We consider the following for our simulations:

1. Setting  $\rho_S^* = 0$ ; that is, conducting a multivariate meta-analysis where within-study correlations are ignored.

2. Setting  $\rho_S^* = 1.5 \times \rho_S$ ; that is, the correlations are overestimated by 50%;
3. Setting  $\rho_S^* = 0.5 \times \rho_S$ ; that is, the correlations are underestimated by 50%;
4. Setting  $\rho_S^* = -0.25$ ; the direction and size of the correlation are wrong.

We compare results from meta-analyses using these four approximations to those where within-study covariances were known (the reference, or best-case results). For each scenario, we simulated 300 meta-analyses, each of which was then modeled with the observed and approximated covariances. Thus, five meta-analyses were performed on each set of studies and results were compared with respect to bias and precision.

#### **4.4 ANALYSIS AND COMPARISON MEASURES**

The models were fitted by REML with PROC MIXED using SAS/STAT software (Version 8e), as described by Van Houwelingen et al.<sup>12</sup>. Estimation was carried out by Fisher scoring instead of the standard Newton-Raphson algorithm to avoid non-positive-definite Hessian matrices, which may cause estimation to fail for some meta-analyses<sup>55</sup>. Since within-study variances and covariances must be specified and fixed, the procedure also requires that starting values be provided for the components of the between-study covariance matrix. We derive these by calculating them from the observed data in each meta-analysis.

For each of the five the meta-analysis models (reference + 4 approximations), we recorded the estimates of summary treatment effects on events A and B (i.e.,  $\hat{\beta}$ ), the between-study variances (i.e.,  $\hat{D}_{11}^2, \hat{D}_{22}^2$ ) and the correlation between effects across studies (i.e.,  $\hat{\rho}_D = \hat{D}_{12} / \sqrt{\hat{D}_{11}^2 \hat{D}_{22}^2}$ ) with corresponding standard errors and/or confidence intervals. We compared the performance of the models where within-study covariances were approximated to the reference analysis (covariances known) with respect to the bias and precision of estimates and coverage probabilities of the confidence intervals of the estimates of each parameter. More specifically, we calculated the absolute (estimate – true) and

proportional bias (bias/true) of the estimates from each model within each replication of the scenarios. We quantified the precision of results in terms of the standard-error of the effect estimates ( $\hat{\beta}$ ) and the width of the confidence intervals of variance and correlation estimates.

Findings are summarized in terms of the median, 5<sup>th</sup> and 95<sup>th</sup> percentiles of the distribution of results from the 300 replications. We favored percentiles over means and standard errors, since the latter may be affected by extreme results and could be misleading if the distribution of results from the simulations are skewed. We calculated the coverage probabilities of the confidence intervals of estimates of each parameter as the proportion of runs of each scenario where the confidence interval covered the true value. The accuracy of the estimated coverage probability depends to some extent on the number of replications. If the true coverage probability of a parameter is 95%, with 300 replicates of the scenarios, the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles of the sampling distribution of observable values are 92.5% and 97.5%, respectively. Thus, with 300 replicates, we can infer that observed coverage probabilities below 92% are indicative of coverage that may truly be below nominal values.

## 4.5 RESULTS

Table 4.12 summarizes results obtained from simulations of scenario 1a. There was little (if any) bias in estimates of treatment effects for events A and B when the observed within-study covariances were used (“Known” column). Although negligible, the bias in estimates of the effect on event B appear to be slightly larger (median: 0.52%, 5<sup>th</sup>-95<sup>th</sup> percentiles -12.1%-11.9%) compared to event A (0.08%, -6.9%-8.4%), and slightly less precise (median SE: 0.32 vs. 0.45). Furthermore, the coverage probabilities of interval estimates were closer to the nominal value (95%) for estimates of the effect on A. We attribute this difference to the smaller size of the effect of the treatment on event B.

The potential for bias was much greater in the estimation of heterogeneity ( $D_{11}^2, D_{22}^2$ ) and correlation ( $\rho_D$ ) parameters. Although the medians of the

distributions of the bias were close to 0, the 5<sup>th</sup> and 95<sup>th</sup> percentiles reveal errors ranging from -45.8% to 52.3% (for A) and -42.5% to 52.9% (for B) for the variance parameters. That is, in about 10% of replications, variance estimates were more than 50% biased (in absolute value). The confidence intervals for estimates of  $D_{11}^2$  tended to be twice as large as those of  $D_{22}^2$ , but the coverage probabilities of both were close to 95%.

Similarly, although estimates of the correlation parameter were very accurate “on average” (median bias: 0.005), 5% of replications produced estimates with (absolute) bias greater than approximately 0.32, or about 120%. Thus, one can expect correlation estimates ranging from -0.06 to 0.57, when in fact the true value is 0.25. The confidence intervals of the estimates tended to be fairly wide; the median width was 0.76, which represents 38% of the possible range of values for this parameter and only 91% of interval estimates included the true value.

Meta-analyses where within-study covariances were approximated produced essentially identical results to those where the covariances were observed, regardless of the accuracy of the approximations. Very slight differences were apparent in the medians of bias distributions of estimates of  $\rho_D$ , but the other percentiles of the distributions were comparable. We point out, however, that the size of the median bias in meta-analyses based on “Independent”, “Underestimated Approximation” and “Negative Approximation” approaches (all of which underestimate the true correlation) was consistent with the accuracy of each method. For example, the median was highest for analyses based on negative approximations, the most inaccurate approximation. This pattern was also observed in simulations of the other scenarios, suggesting perhaps that the difference between the medians, however small, is not spurious.

We observed similar results in simulations of scenario 1b (Table 4.13), where the within-study correlation was assumed to be 0.50 – twice as large as the between-study correlation. Results from meta-analyses where within-study covariances were observed were very similar to those from scenario 1a. Approximations of within-study covariance are subject to larger absolute errors in this scenario.

Nonetheless, we observe very little change in the bias and precision of effect and variance parameter estimates from meta-analyses employing approximations compared to the previous scenario. However, we do detect increases in the median bias of correlation estimates from meta-analyses that assumed independence, underestimated correlations or used a negative correlation estimate. The median bias was largest in the latter case, which is the most inaccurate approximation. Surprisingly, meta-analyses overestimating the correlation by 50% produced correlation estimates that were “on average” only 0.6% biased, compared to 1.8% in meta-analyses where correlations were known. On the other hand, underestimating the correlation by the same degree produced more biased results (3.0%). We note, however, that meta-analyses with overestimated approximations were more likely to underestimate within-study correlations (i.e., lower 5<sup>th</sup> percentile), while the other three approaches that underestimate within-study correlations tended to overestimate (i.e., higher 95<sup>th</sup> percentile). This pattern is maintained and becomes more apparent in the next scenarios.

We examined the impact of approximations in meta-analyses of smaller studies in scenarios 2a and 2b. We reduced the sample sizes by 1/5 in scenario 2a (Table 4.14). While the medians of the bias distributions for all parameters increased (in absolute value), we detected less fluctuation in the upper and lower 5<sup>th</sup> percentiles or the precision of estimates. This occurred similarly in analyses using known and approximated covariances. Although the median bias doubled for some of the effect and variance parameter estimates, the magnitude of the error remained very low (< 5%). This was not the case for correlation estimates, however, which tended to be more erroneous: the median bias was about -5.3% when covariances were known and 4.3% when correlations were underestimated. Greater bias was observed in meta-analyses that misspecified the direction of the correlation (median: 17.7%) or assumed independence (11.4%); these approaches generally produced overestimates of the between-study correlations (95<sup>th</sup> percentiles: 133.0%, 126.8%). On the other hand, meta-analyses that overestimated within-study correlations tended to underestimate between-study correlations (median: -11.2%, 5<sup>th</sup> percentile: -131.6%). The confidence intervals of the estimates tended

to be wider than in scenario 1b, which possibly accounts for the slight increase in their coverage probabilities.

We observed larger errors in estimates when sample sizes were further reduced (Table 4.15). Once again, estimates of the correlation parameter were most affected. While changes in the median bias were also apparent for the other parameter estimates, the overall magnitude of the errors remained low and approximating within-study covariances had little impact on estimates. We note a substantial change the median of the bias distribution for correlation estimates from meta-analyses with known covariances: 15.0% compared to -5.3%. We also observe important increases in the median bias from analyses assuming independence (42.9%), underestimating correlations (29.7%) or using negative approximations (56.7%). The 95<sup>th</sup> percentiles of these distributions also increased substantially. The median bias when correlations were overestimated was surprisingly low (1.4%), but the 5<sup>th</sup> percentile of the distribution was significantly higher than those from the other approximations (-175.8% compared to -134.6% or greater). Thus, as in previous scenarios, the upper and lower 5<sup>th</sup> percentiles suggest that analyses using known or overestimated correlations were more likely to underestimate between-study correlations (i.e., had lower 5<sup>th</sup> percentiles), while estimates from the other approaches tended to overestimate this parameter (i.e., had higher 95<sup>th</sup> percentiles).

In scenarios 3a and 3b, we considered the situation where there was little heterogeneity across studies. Compared to scenario 1b, there was little change in the medians of the bias distributions of effect and variance parameter estimates. The range of bias in effect parameter estimates was reduced considerably when between-study variances were reduced by a factor of 1/5 (scenario 3a - Table 4.16), regardless of whether within-study covariances were known or approximated. For example, the upper and lower 5<sup>th</sup> percentiles of the bias in estimates of the effect of treatment on event A declined from -6.9% - 8.4% in scenario 1b, to -3.0% – 3.7% in scenario 3a. Variance parameter estimates were still prone to large bias, however (e.g., from -49.0% to 57.8% for  $\hat{D}_{11}^2$ ). Reducing

the degree of heterogeneity did not make the models more susceptible to bias for the effect and variance parameters, as differences between “Known” and approximated covariance-based meta-analyses remained inconsequential.

Although the data in this scenario are less noisy, correlation estimates tended to be more biased compared to scenario 1b. The precision of estimates declined slightly but the coverage probability of confidence intervals remained at about 0.90. In fact, the previously observed patterns were apparent here as well. The impact of errors in approximations of within-study covariances was more evident in this scenario with respect to differences in both the median and upper and lower 5<sup>th</sup> percentiles of the distributions. For example, half of the estimates from simulations using negative approximations were biased by more than 24.8%, compared to 5.3% in meta-analyses using observed covariances. In general, approximations that underestimated within-study covariances tended to yield correlation estimates that exaggerate the actual association between outcomes across studies, while the approach that overestimated the covariances produced estimates that tended to be below the true value.

We observed further reductions in the bias and standard errors of effect parameter estimates when we fixed between-study variances to 1/10 of the values in scenario 1b (Table 4.17), while correlation estimates tended to be more biased and less precise. In fact, the median biases for correlation estimates were approximately doubled compared to scenario 3a.

## 4.6 DISCUSSION

This paper examined the impact of errors made in approximating unobserved within-study covariances in multivariate meta-analyses on estimates of treatment effects, heterogeneity parameters and the correlation between outcomes across studies. We compared results from meta-analyses based on approximations of within-study covariances to those from the reference analysis where the actual covariances are observed. These were available since the data were simulated but are generally not available in practice. The approximations that were employed in

the simulations were intentionally made to be of poor accuracy to test the models under *difficult* conditions – when little is known about the correlation between outcomes.

Our analyses revealed that treatment effects were reasonably well estimated, even when within-study covariances were poorly approximated. Heterogeneity and between-study correlation parameters were subject to larger biases, however, even when within-study covariances were observed. In fact, estimates of these parameters were relatively less precise and confidence intervals of correlation estimates had coverage probabilities that were consistently around or below 90%. With 300 replications, it is unlikely to observe coverage probabilities below 92.5% if the interval estimates do in fact have 95% coverage; thus, the coverage of confidence intervals of correlation estimates were indeed below the nominal values. We attribute this in part to the relatively small size of the data (25 studies), which is typical in meta-analyses. Furthermore, correlation estimates are prone to more bias since they depend on multiple parameters estimates (covariance between outcomes as well as two variances). We note, however, that depending on the directions, it is possible that errors in these estimates cancel each other out in some cases.

We also confirmed the hypothesis that the impact of errors in estimates would be most apparent in scenarios where within- and between-study covariances were of similar scale (scenarios 2a & b, 3a & b). However, the impact was most apparent in the accuracy of between-study correlation estimates; the precision of these estimates were poor in all analyses, even when covariances were observed, but did not seem to get worse because of the errors in approximations. Treatment effects or heterogeneity parameters were generally similar in both meta-analyses based on observed or approximated covariances. Thus, if only treatment effects are of interest, one can even ignore within-study correlations and assume independence without any significant risk of bias or loss of precision in estimates. In meta-analyses where an estimate of the correlation between outcomes is desired, however, a comparison of the relative magnitude within- and between-study variances in the outcomes (e.g., from univariate random-effects models),

sample sizes of the studies, and crude estimates of the correlation from the recorded data can be used to assess the potential for bias. Like Berkey et al.<sup>8</sup>, we recommend performing sensitivity analyses around the external estimates used to approximate within-study covariances.

We did not examine cases where estimates were more variable within-studies than between, as these are likely cases where effects are fairly homogeneous across studies and random-effects models may not be necessary or adequate. We can deduce from our findings, however, that correlation estimates in such meta-analyses may be susceptible to substantial variability and possibly large biases.

The scenarios reported in this paper were based on a fixed number of studies and assumed that all studies reported both outcomes and that the between-patient variance of the outcome was constant across studies. We examined the potential impact of varying these factors (not reported in the manuscript, but shown in appendix). As would be expected, there was a noticeable loss in precision when fewer studies were included in the meta-analysis or when not all studies reported both outcomes. Here again, estimates of between-study correlations were most affected, with larger biases and interval estimates with poor coverage probabilities. Similarly, allowing between-patient variances to be different in studies also made the correlation parameter difficult to estimate, but did not affect summary estimates of treatment effects or heterogeneity parameters significantly.

The relative magnitude of differences in the bias of correlation estimates from meta-analyses employing approximations were consistent with the relative accuracy of the approximations. Meta-analyses that assumed a negative within-study correlation, the most erroneous approximation, produced the most biased estimates. Results from meta-analyses that overestimated within-study correlations tended to be closest to those from the reference case, but were generally more likely to underestimate the correlation between outcomes across studies. Conversely, meta-analyses where within-study correlations were under-approximated tended to overestimate the between-study correlation. This suggests that, the model may be compensating for the errors in within-study

correlations, by over- or underestimating the correlation across studies to *Maintain* the *total* correlation measured in the data. This *balancing* of within- and between-study correlations explains why summary treatment effect parameters were not affected by the errors in approximation: their estimation relies only on the total covariance matrices ( $V_i=D+S_i$ ), which will tend to be accurate since estimates of  $D$  seem to counter errors in the specification of the  $S_i$ .

Although we simulated meta-analyses of continuous endpoints, our findings should also apply when multivariate models are used for two or more log-odds-ratio, log-rate-ratio or other measures. To obtain the correlation between log odds-ratios for two outcomes, for instance, would require that each study analyze the data with a repeated measures logistic regression for the two outcomes. Our findings suggest, however, that a joint meta-analysis for two log odds-ratios may be carried out reliably without any knowledge of within-study correlations (i.e., setting within-study covariances to 0). More generally, within-study correlations are harder to specify for these measures, since by definition, they apply to populations or groups. That is, it is difficult to interpret within-study correlations between two log odds-ratios in an individualistic sense, in the same way that one can with continuous measures (e.g., the correlation between a patient's blood pressure and lipid levels). The relationship between dichotomous outcomes may be more easily described with probabilistic arguments (e.g., probability of having both outcomes vs. probability of only one). Consequently, multivariate models based on this type of structure may be more useful and provide more information about the relationship between the outcomes in this setting.

In summary, multivariate meta-analysis models appear to be fairly robust to errors made in approximating within-study covariances when only summary effect estimates are of interest. In fact, assuming independence within studies and accounting for the correlations in the data through the random-effects for each outcome can do as well as when within-study covariances are observed. Multivariate meta-analyses do not appear to be as reliable when interest lies in estimating correlations between outcomes; even when within-study covariances were observed, estimates of the correlations were prone to relatively large biases

and lacked precision. The direction of the bias in between-study correlation estimates depended on whether within-study correlations were under- or over-approximated. Furthermore, the correlations measured across studies may not reflect the underlying association between treatment effects (Chapter 3).

## 4.7 TABLES

		<b>Known</b>	<b>Overestimated Approximation</b>	<b>Underestimated Approximation</b>	<b>Independent</b>	<b>Negative Approximation</b>
		$\rho_s = 0.25$	$\rho_s^* = 0.375$	$\rho_s^* = 0.125$	$\rho_s^* = 0$	$\rho_s^* = -0.25$
<b>Effect on A (True = -10)</b>	Bias	-0.01 (-0.84, 0.69)	-0.01 (-0.84, 0.69)	-0.01 (-0.84, 0.69)	-0.01 (-0.84, 0.69)	-0.01 (-0.84, 0.69)
	% Bias	0.08 (-6.9, 8.38)	0.08 (-6.9, 8.39)	0.08 (-6.9, 8.38)	0.08 (-6.9, 8.38)	0.08 (-6.9, 8.38)
	SE	0.45 (0.33, 0.56)	0.45 (0.33, 0.56)	0.45 (0.33, 0.56)	0.45 (0.33, 0.56)	0.45 (0.33, 0.56)
	Coverage	0.95	0.95	0.95	0.95	0.95
<b>Effect on B (True = -5)</b>	Bias	-0.03 (-0.6, 0.6)	-0.03 (-0.6, 0.6)	-0.03 (-0.6, 0.6)	-0.03 (-0.6, 0.6)	-0.03 (-0.6, 0.6)
	% Bias	0.52 (-12.06, 11.92)	0.52 (-12.05, 11.91)	0.52 (-12.05, 11.9)	0.52 (-12.05, 11.9)	0.50 (-12.05, 11.92)
	SE	0.32 (0.24, 0.39)	0.32 (0.24, 0.39)	0.32 (0.24, 0.39)	0.32 (0.24, 0.39)	0.32 (0.24, 0.39)
	Coverage	0.93	0.93	0.93	0.93	0.93
$D_{11}^2$ (True = 5.0)	Bias	-0.03 (-2.29, 2.62)	-0.03 (-2.29, 2.61)	-0.03 (-2.29, 2.62)	-0.03 (-2.29, 2.62)	-0.03 (-2.28, 2.63)
	% Bias	-0.66 (-45.76, 52.32)	-0.62 (-45.83, 52.21)	-0.65 (-45.77, 52.35)	-0.66 (-45.74, 52.42)	-0.66 (-45.68, 52.56)
	CI Width	6.72 (3.72, 10.24)	6.72 (3.71, 10.23)	6.72 (3.72, 10.24)	6.72 (3.72, 10.25)	6.72 (3.72, 10.26)
	Coverage	0.94	0.94	0.94	0.94	0.94
$D_{22}^2$ (True = 2.5)	Bias	-0.02 (-1.06, 1.32)	-0.02 (-1.06, 1.32)	-0.02 (-1.06, 1.32)	-0.02 (-1.06, 1.32)	-0.02 (-1.06, 1.32)
	% Bias	-0.94 (-42.46, 52.88)	-0.96 (-42.49, 52.88)	-0.93 (-42.46, 52.87)	-0.91 (-42.44, 52.87)	-0.82 (-42.41, 52.87)
	CI Width	3.35 (1.98, 5.13)	3.35 (1.97, 5.13)	3.35 (1.97, 5.13)	3.35 (1.97, 5.127)	3.35 (1.98, 5.13)
	Coverage	0.95	0.95	0.95	0.95	0.95
$\rho_D$ (True = 0.25)	Bias	0.01 (-0.31, 0.32)	0.003 (-0.31, 0.31)	0.01 (-0.31, 0.32)	0.01 (-0.31, 0.32)	0.01 (-0.3, 0.32)
	% Bias	1.9 (-124.76, 126.07)	1.25 (-125.35, 125.21)	2.32 (-123.73, 126.8)	2.8 (-122.89, 127.59)	3.72 (-121.28, 129.15)
	CI Width	0.76 (0.56, 0.81)	0.76 (0.56, 0.81)	0.76 (0.55, 0.81)	0.76 (0.55, 0.81)	0.76 (0.55, 0.81)
	Coverage	0.91	0.91	0.91	0.91	0.91

**Table 4.12** Median (5<sup>th</sup>, 95<sup>th</sup> percentiles) of estimates obtained from 300 simulated meta-analyses from scenario 1a, where we assume  $(S_{11}^2, S_{22}^2) = 2(D_{11}^2, D_{22}^2) = (10, 0.5, 0)$  and  $\rho_s = \rho_D = 0.25$ .

		<b>Known</b>	<b>Overestimated Approximation</b>	<b>Underestimated Approximation</b>	<b>Independent</b>	<b>Negative Approximation</b>
		$\rho_S = 0.25$	$\rho_S^* = 0.375$	$\rho_S^* = 0.125$	$\rho_S^* = 0$	$\rho_S^* = -0.25$
<b>Effect on A (True = -10)</b>	<b>Bias</b>	-0.01 (-0.84, 0.69)	-0.01 (-0.84, 0.69)	-0.01 (-0.84, 0.69)	-0.01 (-0.84, 0.69)	-0.01 (-0.84, 0.69)
	<b>% Bias</b>	0.10 (-6.91, 8.43)	0.09 (-6.92, 8.43)	0.10 (-6.91, 8.43)	0.11 (-6.9, 8.43)	0.11 (-6.9, 8.43)
	<b>SE</b>	0.45 (0.33, 0.55)	0.45 (0.33, 0.55)	0.45 (0.33, 0.55)	0.45 (0.33, 0.56)	0.45 (0.33, 0.56)
	<b>Coverage</b>	0.95	0.95	0.95	0.95	0.95
<b>Effect on B (True = -5)</b>	<b>Bias</b>	-0.02 (-0.58, 0.6)	-0.02 (-0.58, 0.6)	-0.02 (-0.59, 0.6)	-0.02 (-0.59, 0.6)	-0.02 (-0.59, 0.6)
	<b>% Bias</b>	0.31 (-12.08, 11.69)	0.31 (-12.08, 11.69)	0.32 (-12.08, 11.71)	0.33 (-12.08, 11.72)	0.35 (-12.08, 11.74)
	<b>SE</b>	0.32 (0.24, 0.39)	0.32 (0.24, 0.39)	0.32 (0.24, 0.39)	0.32 (0.24, 0.39)	0.32 (0.24, 0.39)
	<b>Coverage</b>	0.93	0.93	0.93	0.93	0.93
$D_{11}^2$ <b>(True = 5.0)</b>	<b>Bias</b>	-0.03 (-2.28, 2.6)	-0.03 (-2.28, 2.6)	-0.03 (-2.28, 2.61)	-0.03 (-2.28, 2.62)	-0.03 (-2.28, 2.62)
	<b>% Bias</b>	-0.67 (-45.59, 52.1)	-0.55 (-45.64, 51.96)	-0.68 (-45.58, 52.22)	-0.67 (-45.56, 52.35)	-0.67 (-45.54, 52.48)
	<b>CI Width</b>	6.73 (3.72, 10.22)	6.73 (3.71, 10.21)	6.73 (3.73, 10.23)	6.72 (3.73, 10.24)	6.71 (3.74, 10.25)
	<b>Coverage</b>	0.94	0.94	0.94	0.94	0.94
$D_{22}^2$ <b>(True = 2.5)</b>	<b>Bias</b>	-0.05 (-1.05, 1.34)	-0.05 (-1.05, 1.34)	-0.04 (-1.05, 1.34)	-0.04 (-1.05, 1.34)	-0.04 (-1.05, 1.34)
	<b>% Bias</b>	-1.79 (-42.12, 53.55)	-1.80 (-42.17, 53.56)	-1.75 (-42.11, 53.58)	-1.72 (-42.08, 53.59)	-1.68 (-42.05, 53.6)
	<b>CI Width</b>	3.32 (1.99, 5.16)	3.32 (1.98, 5.16)	3.33 (1.99, 5.16)	3.33 (1.99, 5.16)	3.33 (1.99, 5.158)
	<b>Coverage</b>	0.96	0.96	0.96	0.96	0.95
$\rho_D$ <b>(True = 0.25)</b>	<b>Bias</b>	0.004 (-0.31, 0.32)	0.002 (-0.32, 0.31)	0.01 (-0.31, 0.32)	0.01 (-0.31, 0.33)	0.014 (-0.3, 0.33)
	<b>% Bias</b>	1.78 (-125.92, 127.09)	0.64 (-127.3, 125.32)	2.97 (-124.27, 129.19)	4.25 (-122.33, 130.76)	5.70 (-120.38, 132.08)
	<b>CI Width</b>	0.76 (0.55, 0.81)	0.76 (0.55, 0.813)	0.76 (0.55, 0.81)	0.76 (0.55, 0.81)	0.76 (0.54, 0.83)
	<b>Coverage</b>	0.89	0.89	0.89	0.89	0.90

**Table 4.13** Median (5<sup>th</sup>, 95<sup>th</sup> percentiles) of estimates obtained from 300 simulated meta-analyses from scenario 1b, where we assume  $(S_{11}^2, S_{22}^2) = 2(D_{11}^2, D_{22}^2) = (10.0, 5.0)$ , and  $\rho_S = 2\rho_D = 0.5$ .

		<b>Known</b>	<b>Overestimated Approximation</b>	<b>Underestimated Approximation</b>	<b>Independent</b>	<b>Negative Approximation</b>
		$\rho_S = 0.25$	$\rho_S^* = 0.375$	$\rho_S^* = 0.125$	$\rho_S^* = 0$	$\rho_S^* = -0.25$
<b>Effect on A (True = -10)</b>	Bias	0.02 (-0.78, 0.72)	0.03 (-0.78, 0.73)	0.03 (-0.77, 0.72)	0.03 (-0.77, 0.72)	0.03 (-0.77, 0.73)
	% Bias	-0.24 (-7.22, 7.77)	-0.26 (-7.26, 7.78)	-0.26 (-7.21, 7.74)	-0.27 (-7.23, 7.72)	-0.30 (-7.27, 7.7)
	SE	0.45 (0.34, 0.573)	0.45 (0.34, 0.57)	0.45 (0.34, 0.57)	0.45 (0.34, 0.57)	0.45 (0.34, 0.57)
	Coverage	0.94	0.94	0.94	0.94	0.94
<b>Effect on B (True = -5)</b>	Bias	0.02 (-0.52, 0.51)	0.02 (-0.53, 0.5)	0.02 (-0.53, 0.5)	0.02 (-0.52, 0.51)	0.02 (-0.52, 0.51)
	% Bias	-0.46 (-10.19, 10.49)	-0.46 (-10.05, 10.56)	-0.46 (-10.09, 10.51)	-0.46 (-10.1, 10.44)	-0.45 (-10.12, 10.49)
	SE	0.33 (0.25, 0.40)	0.33 (0.25, 0.40)	0.33 (0.25, 0.40)	0.33 (0.25, 0.40)	0.33 (0.25, 0.40)
	Coverage	0.95	0.95	0.94	0.94	0.95
$D_{11}^2$ (True = 5.0)	Bias	-0.18 (-2.41, 2.84)	-0.19 (-2.43, 2.86)	-0.21 (-2.41, 2.83)	-0.21 (-2.4, 2.82)	-0.20 (-2.39, 2.81)
	% Bias	-3.57 (-48.14, 56.89)	-3.72 (-48.5, 57.13)	-4.16 (-48.28, 56.6)	-4.10 (-48.09, 56.42)	-4.07 (-47.8, 56.29)
	CI Width	7.0 (4.12, 11.13)	6.98 (4.1, 11.12)	7.0 (4.13, 11.14)	6.99 (4.17, 11.12)	7.0 (4.21, 11.10)
	Coverage	0.97	0.97	0.96	0.96	0.96
$D_{22}^2$ (True = 2.5)	Bias	0.02 (-1.14, 1.35)	0.02 (-1.15, 1.34)	0.02 (-1.15, 1.36)	0.01 (-1.16, 1.37)	0.02 (-1.16, 1.38)
	% Bias	0.69 (-45.71, 53.84)	0.93 (-45.94, 53.6)	0.71 (-46.05, 54.46)	0.51 (-46.33, 54.92)	0.60 (-46.54, 55.21)
	CI Width	3.65 (2.13, 5.48)	3.66 (2.12, 5.47)	3.65 (2.13, 5.48)	3.65 (2.12, 5.48)	3.65 (2.12, 5.48)
	Coverage	0.97	0.97	0.97	0.97	0.97
$\rho_D$ (True = 0.25)	Bias	-0.01 (-0.31, 0.29)	-0.03 (-0.33, 0.27)	0.01 (-0.3, 0.3)	0.03 (-0.28, 0.32)	0.04 (-0.26, 0.33)
	% Bias	-5.298 (-124.67, 115.85)	-11.16 (-131.56, 109.75)	4.30 (-118.73, 121.48)	11.44 (-112.98, 126.83)	17.7 (-105.83, 132.96)
	CI Width	0.81 (0.61, 0.88)	0.81 (0.61, 0.88)	0.81 (0.6, 0.88)	0.80 (0.59, 0.87)	0.80 (0.59, 0.87)
	Coverage	0.93	0.93	0.92	0.92	0.91

**Table 4.14** Median (5<sup>th</sup>, 95<sup>th</sup> percentiles) of estimates obtained from 300 simulated meta-analyses from scenario 2a, which is based on scenario 1b ( $S_{11}^2, S_{22}^2 = 2(D_{11}^2, D_{22}^2) = (10, 0.5, 0)$ , and  $\rho_S = 2\rho_D = 0.5$ ), with sample sizes reduced by a factor of 5 (i.e., five times smaller).

		<b>Known</b>	<b>Overestimated Approximation</b>	<b>Underestimated Approximation</b>	<b>Independent</b>	<b>Negative Approximation</b>
		$\rho_S = 0.25$	$\rho_S^* = 0.375$	$\rho_S^* = 0.125$	$\rho_S^* = 0$	$\rho_S^* = -0.25$
<b>Effect on A (True = -10)</b>	<b>Bias</b>	-0.04 (-0.75, 0.75)	-0.05 (-0.74, 0.75)	-0.05 (-0.74, 0.75)	-0.05 (-0.73, 0.75)	-0.05 (-0.74, 0.75)
	<b>% Bias</b>	0.42 (-7.49, 7.49)	0.48 (-7.5, 7.42)	0.51 (-7.5, 7.43)	0.53 (-7.5, 7.34)	0.51 (-7.53, 7.39)
	<b>SE</b>	0.47 (0.37, 0.58)	0.48 (0.37, 0.58)	0.48 (0.37, 0.58)	0.48 (0.38, 0.58)	0.48 (0.38, 0.58)
	<b>Coverage</b>	0.96	0.96	0.96	0.96	0.95
<b>Effect on B (True = -5)</b>	<b>Bias</b>	0.02 (-0.54, 0.57)	0.023 (-0.53, 0.56)	0.024 (-0.52, 0.56)	0.021 (-0.53, 0.57)	0.02 (-0.53, 0.57)
	<b>% Bias</b>	-0.40 (-11.39, 10.76)	-0.47 (-11.27, 10.67)	-0.49 (-11.26, 10.47)	-0.42 (-11.31, 10.52)	-0.42 (-11.34, 10.64)
	<b>SE</b>	0.33 (0.26, 0.42)	0.33 (0.26, 0.42)	0.33 (0.26, 0.42)	0.33 (0.26, 0.42)	0.33 (0.26, 0.42)
	<b>Coverage</b>	0.93	0.93	0.93	0.93	0.93
$D_{11}^2$ <b>(True = 5.0)</b>	<b>Bias</b>	-0.09 (-2.18, 2.75)	-0.08 (-2.23, 2.73)	-0.09 (-2.17, 2.74)	-0.07 (-2.11, 2.75)	-0.07 (-2.08, 2.76)
	<b>% Bias</b>	-1.84 (-43.53, 54.91)	-1.65 (-44.62, 54.53)	-1.782 (-43.33, 54.87)	-1.40 (-42.29, 55.0)	-1.40 (-41.56, 55.11)
	<b>CI Width</b>	7.81 (4.98, 11.60)	7.84 (5, 11.55)	7.80 (4.96, 11.53)	7.80 (4.95, 11.57)	7.80 (4.98, 11.65)
	<b>Coverage</b>	0.96	0.96	0.96	0.96	0.97
$D_{22}^2$ <b>(True = 2.5)</b>	<b>Bias</b>	-0.07 (-1.16, 1.61)	-0.07 (-1.16, 1.6)	-0.08 (-1.15, 1.59)	-0.09 (-1.14, 1.59)	-0.09 (-1.14, 1.58)
	<b>% Bias</b>	-2.67 (-46.49, 64.28)	-2.64 (-46.6, 64.06)	-3.0 (-46, 63.47)	-3.44 (-45.71, 63.42)	-3.62 (-45.47, 63.37)
	<b>CI Width</b>	3.85 (2.43, 6.0)	3.85 (2.46, 6.0)	3.85 (2.46, 6.03)	3.84 (2.46, 6.03)	3.84 (2.45, 6.04)
	<b>Coverage</b>	0.96	0.96	0.96	0.96	0.96
$\rho_D$ <b>(True = 0.25)</b>	<b>Bias</b>	0.04 (-0.38, 0.33)	0.003 (-0.44, 0.29)	0.07 (-0.34, 0.35)	0.11 (-0.3, 0.38)	0.14 (-0.26, 0.42)
	<b>% Bias</b>	14.97 (-152.15, 130.24)	1.38 (-175.75, 115.01)	29.69 (-134.59, 140.45)	42.89 (-118.02, 153.94)	56.73 (-102.69, 169.75)
	<b>CI Width</b>	0.84 (0.63, 0.96)	0.85 (0.65, 0.96)	0.83 (0.61, 0.96)	0.82 (0.59, 0.95)	0.81 (0.58, 0.94)
	<b>Coverage</b>	0.91	0.93	0.90	0.87	0.83

**Table 4.15** Median (5<sup>th</sup>, 95<sup>th</sup> percentiles) of estimates obtained from 300 simulated meta-analyses from scenario 2b, which is based on scenario 1b ( $S_{11}^2, S_{22}^2 = 2(D_{11}^2, D_{22}^2) = (10.0, 5.0)$ ), and  $\rho_S = 2\rho_D = 0.5$ ), with sample sizes reduced by a factor of 10 (i.e., 10 times smaller).

		<b>Known</b>	<b>Overestimated Approximation</b>	<b>Underestimated Approximation</b>	<b>Independent</b>	<b>Negative Approximation</b>
		$\rho_S = 0.25$	$\rho_S^* = 0.375$	$\rho_S^* = 0.125$	$\rho_S^* = 0$	$\rho_S^* = -0.25$
<b>Effect on A (True = -10)</b>	Bias	-0.002 (-0.37, 0.3)	-0.003 (-0.37, 0.3)	-0.004 (-0.37, 0.3)	-0.004 (-0.37, 0.3)	-0.004 (-0.37, 0.3)
	% Bias	0.02 (-3.02, 3.7)	0.03 (-3.03, 3.69)	0.04 (-3.03, 3.7)	0.04 (-3.03, 3.71)	0.04 (-3.03, 3.72)
	SE	0.20 (0.15, 0.26)	0.20 (0.15, 0.26)	0.20 (0.15, 0.26)	0.20 (0.15, 0.26)	0.20 (0.15, 0.26)
	Coverage	0.95	0.95	0.95	0.95	0.95
<b>Effect on B (True = -5)</b>	Bias	-0.01 (-0.26, 0.28)	-0.01 (-0.26, 0.28)	-0.01 (-0.26, 0.27)	-0.01 (-0.26, 0.27)	-0.01 (-0.26, 0.27)
	% Bias	0.23 (-5.5, 5.22)	0.21 (-5.53, 5.23)	0.23 (-5.48, 5.23)	0.24 (-5.46, 5.24)	0.24 (-5.45, 5.27)
	SE	0.15 (0.11, 0.18)	0.15 (0.11, 0.18)	0.15 (0.11, 0.18)	0.15 (0.11, 0.18)	0.15 (0.11, 0.18)
	Coverage	0.92	0.93	0.92	0.92	0.92
$D_{11}^2$ (True = 5.0)	Bias	-0.05 (-0.49, 0.58)	-0.047 (-0.49, 0.58)	-0.049 (-0.49, 0.58)	-0.047 (-0.49, 0.59)	-0.045 (-0.49, 0.59)
	% Bias	-5.13 (-49.01, 57.78)	-4.67 (-48.94, 57.85)	-4.92 (-49, 58.33)	-4.66 (-48.96, 58.58)	-4.49 (-48.89, 58.72)
	CI Width	1.40 (0.82, 2.21)	1.40 (0.82, 2.21)	1.40 (0.81, 2.22)	1.4 (0.81, 2.22)	1.4 (0.81, 2.22)
	Coverage	0.94	0.94	0.94	0.94	0.94
$D_{22}^2$ (True = 2.5)	Bias	-0.005 (-0.21, 0.28)	-0.005 (-0.21, 0.28)	-0.007 (-0.22, 0.28)	-0.006 (-0.22, 0.28)	-0.005 (-0.22, 0.28)
	% Bias	-1.062 (-42.89, 55.87)	-0.94 (-42.92, 55.96)	-1.32 (-43.07, 56.08)	-1.22 (-43.09, 56.19)	-0.94 (-43.1, 56.32)
	CI Width	0.72 (0.45, 1.10)	0.72 (0.45, 1.10)	0.72 (0.45, 1.10)	0.721 (0.45, 1.10)	0.72 (0.45, 1.10)
	Coverage	0.96	0.97	0.96	0.96	0.96
$\rho_D$ (True = 0.25)	Bias	0.01 (-0.34, 0.34)	-0.008 (-0.36, 0.32)	0.03 (-0.33, 0.36)	0.05 (-0.31, 0.37)	0.06 (-0.29, 0.39)
	% Bias	5.34 (-137.5, 135.61)	-3.13 (-142.61, 128.5)	12.96 (-131.62, 143.4)	18.77 (-124.8, 149.99)	24.79 (-115.73, 156.19)
	CI Width	0.80 (0.56, 0.88)	0.81 (0.57, 0.88)	0.80 (0.55, 0.88)	0.79 (0.54, 0.87)	0.79 (0.53, 0.87)
	Coverage	0.91	0.92	0.89	0.89	0.87

**Table 4.16** Median (5<sup>th</sup>, 95<sup>th</sup> percentiles) of estimates obtained from 300 simulated meta-analyses from scenario 3a, which is based on scenario 1b ( $S_{11}^2, S_{22}^2 = 2(D_{11}^2, D_{22}^2) = (10, 0.5, 0)$ , and  $\rho_S = 2\rho_D = 0.5$ ), with between-study variances reduced by a factor of 5 (i.e., 5 times smaller).

		<b>Known</b>	<b>Overestimated Approximation</b>	<b>Underestimated Approximation</b>	<b>Independent</b>	<b>Negative Approximation</b>
		$\rho_S = 0.25$	$\rho_S^* = 0.375$	$\rho_S^* = 0.125$	$\rho_S^* = 0$	$\rho_S^* = -0.25$
<b>Effect on A (True = -10)</b>	<b>Bias</b>	0 (-0.27, 0.21)	0.001 (-0.27, 0.21)	0.001 (-0.27, 0.21)	0.003 (-0.26, 0.21)	0.003 (-0.26, 0.21)
	<b>% Bias</b>	0 (-2.11, 2.67)	-0.01 (-2.14, 2.67)	-0.01 (-2.12, 2.66)	-0.03 (-2.12, 2.65)	-0.03 (-2.11, 2.63)
	<b>SE</b>	0.15 (0.11, 0.19)	0.15 (0.11, 0.19)	0.15 (0.11, 0.19)	0.15 (0.11, 0.19)	0.15 (0.11, 0.19)
	<b>Coverage</b>	0.96	0.95	0.95	0.95	0.95
<b>Effect on B (True = -5)</b>	<b>Bias</b>	-0.01 (-0.19, 0.2)	-0.01 (-0.19, 0.2)	-0.01 (-0.19, 0.2)	-0.01 (-0.19, 0.2)	-0.01 (-0.19, 0.2)
	<b>% Bias</b>	0.22 (-4.01, 3.82)	0.19 (-4.02, 3.79)	0.21 (-4.01, 3.83)	0.19 (-4.02, 3.84)	0.20 (-3.99, 3.84)
	<b>SE</b>	0.11 (0.08, 0.13)	0.11 (0.08, 0.13)	0.11 (0.08, 0.13)	0.11 (0.08, 0.13)	0.11 (0.08, 0.13)
	<b>Coverage</b>	0.91	0.91	0.91	0.91	0.91
$D_{11}^2$ (True = 0.5)	<b>Bias</b>	-0.03 (-0.26, 0.31)	-0.03 (-0.26, 0.31)	-0.03 (-0.26, 0.31)	-0.02 (-0.26, 0.32)	-0.03 (-0.26, 0.32)
	<b>% Bias</b>	-5.39 (-52.36, 61.91)	-5.23 (-52.6, 61.66)	-4.96 (-52.31, 62.48)	-4.79 (-52.02, 63.08)	-5.09 (-51.64, 64.18)
	<b>CI Width</b>	0.76 (0.45, 1.20)	0.76 (0.45, 1.19)	0.76 (0.45, 1.21)	0.76 (0.45, 1.21)	0.76 (0.45, 1.22)
	<b>Coverage</b>	0.96	0.96	0.96	0.96	0.96
$D_{22}^2$ (True = 0.25)	<b>Bias</b>	-0.003 (-0.11, 0.15)	-0.004 (-0.11, 0.15)	-0.005 (-0.11, 0.15)	-0.005 (-0.11, 0.15)	-0.004 (-0.11, 0.15)
	<b>% Bias</b>	-1.12 (-43.93, 60.62)	-1.40 (-43.84, 60.92)	-1.99 (-44.17, 60.27)	-2.08 (-44.9, 60.09)	-1.56 (-45.63, 60)
	<b>CI Width</b>	0.39 (0.25, 0.59)	0.39 (0.25, 0.60)	0.39 (0.25, 0.59)	0.39 (0.25, 0.59)	0.40 (0.25, 0.59)
	<b>Coverage</b>	0.97	0.97	0.96	0.96	0.96
$\rho_D$ (True = 0.25)	<b>Bias</b>	0.02 (-0.38, 0.35)	-0.02 (-0.41, 0.32)	0.06 (-0.34, 0.38)	0.09 (-0.31, 0.42)	0.12 (-0.26, 0.46)
	<b>% Bias</b>	8.99 (-152.63, 140.49)	-8.3 (-162.23, 126.35)	22.7 (-136.6, 151.78)	36.00 (-122.87, 167.85)	47.01 (-105.1, 184.26)
	<b>CI Width</b>	0.85 (0.6, 0.96)	0.86 (0.62, 0.969)	0.84 (0.58, 0.952)	0.83 (0.56, 0.94)	0.82 (0.55, 0.94)
	<b>Coverage</b>	0.91	0.91	0.89	0.86	0.82

**Table 4.17** Median (5<sup>th</sup>, 95<sup>th</sup> percentiles) of estimates obtained from 300 simulated meta-analyses from scenario 3b, which is based on scenario 1b ( $S_{11}^2, S_{22}^2 = 2(D_{11}^2, D_{22}^2) = (5.0, 2.5)$ ), and  $\rho_S = 2\rho_D = 0.5$ ), with between-study variances reduced by a factor of 10 (i.e., 10 times smaller)

## 4.8 APPENDIX: ADDITIONAL MATERIAL

### A4.8.1 DESCRIPTION OF SOME ESTIMATION PROCEDURES

Berkey et al.<sup>5</sup> describe an EM-based procedure for GLS, where the optimizing equations are given by:

$$\hat{\beta}(\hat{D}) = (X'V^{-1}X)^{-1}(X'V^{-1}Y)$$

$$\hat{D}(\hat{\beta}) = \frac{1}{n-p}(\hat{R}'\hat{R}) - \frac{1}{n}\sum_i S_i,$$

where  $X$  is the  $2n \times p$  matrix of stacked covariate matrices  $X_i$ ,  $V$  is  $2n \times 2n$  block-diagonal matrix with blocks consisting of  $V_i = D + S_i$ ,  $Y$  is the  $2n \times 1$  vector made up by stacking the  $y_i$  vectors and  $R$  is the  $n \times 2$  matrix comprised of the stacked residuals,  $\hat{r}_i = (y_i - X_i\hat{\beta})'$ . Estimates of  $\beta$  and  $D$  are used iteratively in turn to estimate the other parameter, until some threshold of convergence is attained.

In likelihood-based estimation, the optimal estimator of  $\beta$ , for a given estimate of  $D$ , is the same as above<sup>55</sup>. The score equation for  $D$  is given by

$$\frac{\partial l(D|\hat{\beta}, y_i, S_i)}{\partial D} = -\frac{1}{2}\sum_{i=1}^N V_i^{-1} - V_i^{-1}\hat{r}_i'\hat{r}_i V_i^{-1} = 0 \text{ for ML estimation; the equation for}$$

REML remains as above, with the addition of the term  $X_i \left( \sum_{i=1}^N X_i' V_i^{-1} X_i \right) X'$  to the summation<sup>55</sup>.  $D$  is estimated differently, depending on the estimating procedure being employed. In the EM algorithm described by Berkey et al<sup>5</sup>, the estimate of  $D$  at a given iteration is:

$$\hat{D}(\hat{\beta}) = \frac{1}{N} \sum_{i=1}^N [\hat{D}V_i^{-1}\hat{r}_i'\hat{r}_i V_i^{-1}\hat{D} + \hat{D} - \hat{D}V_i^{-1}\hat{D}],$$

where  $\hat{D}$  on the right-hand-side are the values obtained in the previous step. In the Newton-Raphson or Fisher scoring algorithms, the *current* estimate has the general form  $\hat{D}_{new} = \hat{D}_{old} + \Delta(\hat{\beta}, \hat{D}_{old}, S_1, \dots, S_N)$ , where  $\Delta(\cdot)$  is an updating

function that depends on previous estimates of the parameters and the within-study covariance matrices and involves the score equation of  $D^{55}$ .

#### A4.8.2 SENSITIVITY ANALYSES

##### *Description of Scenarios*

We performed sensitivity analyses around assumptions that were made or parameters that were held fixed in the primary scenarios to determine whether results would change significantly. We base these again on scenario 1b. We considered the following changes based on scenario 1b:

1. **Partial Reporting of Effects:** Our primary analysis was based on meta-analyses where the two outcomes are reported by all studies. Since this may not always be the case, we removed one of the outcome estimates from randomly selected studies. We considered the possibility that, in studies that report only one of the outcomes: a) A and B are equally likely not to be reported; b) A is more likely to be the missing outcome (with probability 0.75); c) B is more likely to be the missing outcome (with probability 0.75).
2. **Size of meta-analyses:** We reduced the number of studies included in the meta-analysis from 25 to 5.
3. **Homogeneous Within-Study Covariance Matrix:** Our simulation model assumes that the within-study covariance matrix is homogeneous across studies – the observed values vary only due to sampling variation. We manipulated this by adding, for each study, random noise to the true within-study variances and correlation. The noise was generated from a normal distribution with mean 0 and variance equal to 15% of the magnitude of variances and 25% of correlations. Greater variability was assumed for the correlations since these are a function of two variables and since fluctuation in this parameter was expected to be more influential.

## *Results*

In the primary scenarios, we assumed that all studies reported both outcomes, although this will not always be the case in practice. We recreated this in simulations of scenario 1b where only half of studies reported both outcomes. We further assumed that studies reporting only one outcome were less likely to not report endpoint A (with probability 0.25). Results from these simulations are summarized in Table 4.18. Compared to meta-analyses with full data (Table 2), we observe generally higher standard errors and wider confidence intervals, particularly for the parameters pertaining to event B, as would be expected since less information is used in these analyses. The potential for bias also increased slightly, particularly for the variance and correlation parameters: for example, in analyses using observed within-study correlations, the 5<sup>th</sup> and 95<sup>th</sup> percentiles of the distribution of absolute bias for estimates of  $\rho_D$  were -0.57 and 0.46, respectively, compared to -0.31 and 0.32 with full data. Furthermore, interval estimates of this parameter had lower coverage probabilities (0.83 compared to 0.89).

Analyses based on approximated within-study correlations were similarly affected but did not produce systematically more or less biased or precise estimates than those where the correlations were known. We note slightly higher median biases for correlation estimates in meta-analyses assuming independence or using negative approximations; differences were not large enough, however, to suggest these approaches are any more deficient than the others.

We also examined scenarios where incomplete reporting of outcomes was totally random or underreporting of event B was more likely (with probability 0.75). Once again, we found a greater impact on variance and correlation parameters than on treatment effect parameters. Approximating within-study correlations did not affect the accuracy and precision of estimates noticeably, however.

To determine whether errors in approximation may be more influential in smaller meta-analyses, we simulated scenario 1b assuming only five studies are included (Table 4.19). As would be expected, we observe a substantial reduction in the

precision of the estimates of all parameters. Most notable perhaps, is the case of the variance parameters, where the median width of confidence intervals in meta-analyses with known within-study correlations were 34.2 and 17.1 for events A and B, respectively, compared to 6.7 and 3.3 in simulations of larger meta-analyses (25 studies). This lead to coverage probabilities above the nominal levels, even though the point estimates of these parameters tended to be more biased (median: -16.9% and -16.7%, compared to -0.7%, -1.8%). Similarly, we found greater bias in correlation estimates even when within-study correlations were observed (6.5% (-381.8% - 250.7%) compared to 1.8% (-125.9% - 127.1%) in larger meta-analyses). Interval estimates of this parameter also had poor coverage (around 75%).

The distribution of bias and precision measures from meta-analyses where approximations were used were very similar to those based on observed correlations. We only observed differences in the medians of the distributions for estimates of the correlation parameter. As in previous scenarios, analyses assuming independence or using negative approximations lead to larger median bias in correlation estimates (11.5% and 13.6%, respectively, compared to 6.5% when correlations are known). Furthermore, estimates from meta-analyses underestimating within-study correlations had larger median bias than those using overestimated approximations.

Although we allowed treatment effects to vary across studies, we assumed the between patient covariance matrix ( $S$ ) was constant in all studies. We examined what would happen if this parameter also varied across studies in simulations based on scenario 1b (Table 4.20). Once again, the impact of this change was most apparent for estimates of the correlation parameter. The median bias in meta-analyses where within-study covariances were observed increased from 1.8% in scenario 1b to 11.8%. Meta-analyses in this scenario were less likely to produce underestimates of the correlation, however; the 5<sup>th</sup> percentile of the distribution increased from -125.9% to -80.6%. We observed a similar increase in analyses where within-study covariances were approximated. There was little change in the precision of estimates, however.

		<b>Known</b>	<b>Overestimated Approximation</b>	<b>Underestimated Approximation</b>	<b>Independent</b>	<b>Negative Approximation</b>
		$\rho_S = 0.25$	$\rho_S^* = 0.375$	$\rho_S^* = 0.125$	$\rho_S^* = 0$	$\rho_S^* = -0.25$
<b>Effect on A (True = -10)</b>	Bias	-0.02 (-0.88, 0.75)	-0.02 (-0.88, 0.75)	-0.02 (-0.88, 0.75)	-0.02 (-0.88, 0.75)	-0.02 (-0.88, 0.75)
	% Bias	0.24 (-7.47, 8.78)	0.24 (-7.46, 8.78)	0.24 (-7.47, 8.76)	0.23 (-7.47, 8.76)	0.22 (-7.47, 8.76)
	SE	0.46 (0.35, 0.60)	0.46 (0.35, 0.60)	0.46 (0.35, 0.60)	0.46 (0.35, 0.60)	0.46 (0.35, 0.60)
	Coverage	0.94	0.94	0.94	0.94	0.94
<b>Effect on B (True = -5)</b>	Bias	-0.02 (-0.71, 0.73)	-0.02 (-0.71, 0.73)	-0.02 (-0.71, 0.73)	-0.02 (-0.71, 0.73)	-0.02 (-0.71, 0.74)
	% Bias	0.41 (-14.69, 14.15)	0.42 (-14.66, 14.15)	0.42 (-14.69, 14.14)	0.42 (-14.7, 14.14)	0.42 (-14.71, 14.13)
	SE	0.39 (0.27, 0.53)	0.39 (0.27, 0.53)	0.38 (0.27, 0.53)	0.38 (0.27, 0.53)	0.38 (0.27, 0.53)
	Coverage	0.93	0.93	0.93	0.92	0.92
$D_{11}^2$ (True = 5.0)	Bias	-0.24 (-2.34, 2.87)	-0.24 (-2.33, 2.86)	-0.24 (-2.34, 2.87)	-0.24 (-2.34, 2.87)	-0.24 (-2.34, 2.88)
	% Bias	-4.79 (-46.7, 57.32)	-4.84 (-46.69, 57.14)	-4.81 (-46.77, 57.33)	-4.86 (-46.82, 57.43)	-4.79 (-46.86, 57.53)
	CI Width	7.01 (4.03, 12.41)	7.02 (4.03, 12.41)	7.01 (4.03, 12.41)	7.00 (4.04, 12.41)	7.00 (4.04, 12.42)
	Coverage	0.96	0.96	0.96	0.96	0.96
$D_{22}^2$ (True = 2.5)	Bias	-0.10 (-1.26, 1.75)	-0.10 (-1.27, 1.75)	-0.09 (-1.27, 1.75)	-0.09 (-1.27, 1.75)	-0.09 (-1.26, 1.75)
	% Bias	-3.79 (-50.58, 69.93)	-3.78 (-50.61, 69.91)	-3.72 (-50.61, 69.97)	-3.69 (-50.6, 70.01)	-3.61 (-50.6, 70.04)
	CI Width	4.56 (2.35, 9.09)	4.56 (2.35, 9.09)	4.56 (2.35, 9.10)	4.57 (2.35, 9.10)	4.57 (2.35, 9.09)
	Coverage	0.98	0.98	0.98	0.98	0.98
$\rho_D$ (True = 0.25)	Bias	0.034 (-0.57, 0.46)	0.03 (-0.58, 0.46)	0.037 (-0.56, 0.47)	0.041 (-0.56, 0.47)	0.04 (-0.55, 0.48)
	% Bias	13.45 (-227.87, 184.66)	12.20 (-230.18, 182.52)	14.81 (-225.08, 186.65)	16.21 (-222.53, 188.7)	17.37 (-219.96, 190.75)
	CI Width	1.00 (0.5, 1.53)	1.00 (0.51, 1.53)	0.99 (0.5, 1.52)	0.99 (0.5, 1.52)	0.99 (0.5, 1.52)
	Coverage	0.83	0.83	0.82	0.82	0.81

**Table 4.18(A)** Median (5<sup>th</sup>, 95<sup>th</sup> percentiles) of estimates obtained from 300 simulated meta-analyses based scenario 1b ( $(S_{11}^2, S_{22}^2) = 2(D_{11}^2, D_{22}^2) = (10.0, 5.0)$ ), and  $\rho_S = 2\rho_D = 0.5$ ). Some studies may not report both outcomes. We assume that the probability that a study does not report one of the outcomes is 0.50; given that a study does not report one of the outcomes, the probability that results for event A is not reported is 0.25.

		<b>Known</b>	<b>Overestimated Approximation</b>	<b>Underestimated Approximation</b>	<b>Independent</b>	<b>Negative Approximation</b>
		$\rho_S = 0.25$	$\rho_S^* = 0.375$	$\rho_S^* = 0.125$	$\rho_S^* = 0$	$\rho_S^* = -0.25$
<b>Effect on A (True = -10)</b>	Bias	0.02 (-1.56, 1.62)	0.02 (-1.56, 1.62)	0.018 (-1.56, 1.62)	0.02 (-1.57, 1.62)	0.021 (-1.57, 1.62)
	% Bias	-0.17 (-16.23, 15.64)	-0.17 (-16.22, 15.64)	-0.18 (-16.23, 15.65)	-0.20 (-16.23, 15.65)	-0.21 (-16.23, 15.66)
	SE	0.92 (0.41, 1.50)	0.92 (0.41, 1.50)	0.92 (0.41, 1.50)	0.92 (0.41, 1.50)	0.92 (0.41, 1.50)
	Coverage	0.87	0.87	0.87	0.87	0.87
<b>Effect on B (True = -5)</b>	Bias	0.026 (-1.26, 1.42)	0.027 (-1.26, 1.42)	0.03 (-1.26, 1.42)	0.03 (-1.26, 1.42)	0.03 (-1.26, 1.42)
	% Bias	-0.53 (-28.35, 25.2)	-0.53 (-28.33, 25.19)	-0.53 (-28.37, 25.21)	-0.53 (-28.39, 25.23)	-0.54 (-28.41, 25.24)
	SE	0.65 (0.35, 1.07)	0.65 (0.35, 1.07)	0.65 (0.35, 1.07)	0.65 (0.35, 1.07)	0.65 (0.35, 1.07)
	Coverage	0.88	0.88	0.88	0.88	0.88
$D_{11}^2$ (True = 5.0)	Bias	-0.85 (-4.2, 6.16)	-0.85 (-4.2, 6.16)	-0.85 (-4.2, 6.16)	-0.85 (-4.2, 6.16)	-0.85 (-4.2, 6.16)
	% Bias	-16.93 (-84.06, 123.12)	-16.91 (-84.06, 123.19)	-16.96 (-84.06, 123.17)	-16.98 (-84.07, 123.15)	-17.01 (-84.07, 123.14)
	CI Width	34.24 (7.78, 89.495)	34.25 (7.78, 89.523)	34.23 (7.78, 89.515)	34.18 (7.78, 89.51)	34.17 (7.78, 89.505)
	Coverage	0.98	0.98	0.98	0.98	0.98
$D_{22}^2$ (True = 2.5)	Bias	-0.42 (-1.93, 3.16)	-0.42 (-1.93, 3.16)	-0.42 (-1.93, 3.16)	-0.42 (-1.93, 3.16)	-0.42 (-1.93, 3.17)
	% Bias	-16.68 (-77.05, 126.43)	-16.68 (-77.06, 126.35)	-16.70 (-77.06, 126.49)	-16.72 (-77.05, 126.56)	-16.71 (-77.05, 126.62)
	CI Width	17.11 (5.35, 45.38)	17.11 (5.35, 45.36)	17.11 (5.35, 45.39)	17.11 (5.35, 45.40)	17.11 (5.35, 45.42)
	Coverage	0.97	0.97	0.967	0.967	0.967
$\rho_D$ (True = 0.25)	Bias	0.02 (-0.95, 0.63)	0.01 (-0.96, 0.62)	0.02 (-0.95, 0.63)	0.03 (-0.95, 0.64)	0.03 (-0.95, 0.64)
	% Bias	6.52 (-381.8, 250.66)	3.55 (-382.9, 249.11)	9.32 (-380.64, 252.95)	11.47 (-379.51, 254.53)	13.60 (-378.38, 256.11)
	CI Width	1.62 (0.46, 2.0)	1.62 (0.45, 2.0)	1.62 (0.45, 2.0)	1.62 (0.45, 2.0)	1.62 (0.45, 2.0)
	Coverage	0.75	0.76	0.75	0.75	0.74

**Table 4.19(A)** Median (5<sup>th</sup>, 95<sup>th</sup> percentiles) of estimates obtained from 300 simulated meta-analyses of five studies based on scenario 1b ( $(S_{11}^2, S_{22}^2) = 2(D_{11}^2, D_{22}^2) = (10.0, 5.0)$ ), and  $\rho_S = 2\rho_D = 0.5$ ).

		<b>Known</b>	<b>Overestimated Approximation</b>	<b>Underestimated Approximation</b>	<b>Independent</b>	<b>Negative Approximation</b>
		$\rho_S = 0.25$	$\rho_S^* = 0.375$	$\rho_S^* = 0.125$	$\rho_S^* = 0$	$\rho_S^* = -0.25$
<b>Effect on A (True = -10)</b>	<b>Bias</b>	0.06 (-0.67, 0.7)	0.06 (-0.67, 0.7)	0.07 (-0.67, 0.7)	0.067 (-0.67, 0.7)	0.067 (-0.67, 0.7)
	<b>% Bias</b>	-0.64 (-7.02, 6.71)	-0.62 (-7.01, 6.71)	-0.66 (-7.02, 6.72)	-0.67 (-7.03, 6.71)	-0.67 (-7.04, 6.71)
	<b>SE</b>	0.45 (0.35, 0.57)	0.45 (0.35, 0.57)	0.45 (0.35, 0.57)	0.45 (0.35, 0.57)	0.45 (0.35, 0.57)
	<b>Coverage</b>	0.95	0.95	0.95	0.95	0.95
<b>Effect on B (True = -5)</b>	<b>Bias</b>	-0.01 (-0.52, 0.57)	-0.01 (-0.52, 0.57)	-0.01 (-0.52, 0.57)	-0.01 (-0.52, 0.57)	-0.01 (-0.52, 0.57)
	<b>% Bias</b>	0.11 (-11.43, 10.44)	0.13 (-11.43, 10.43)	0.11 (-11.42, 10.44)	0.12 (-11.41, 10.44)	0.14 (-11.4, 10.44)
	<b>SE</b>	0.31 (0.24, 0.39)	0.31 (0.24, 0.39)	0.31 (0.24, 0.39)	0.31 (0.24, 0.39)	0.31 (0.24, 0.39)
	<b>Coverage</b>	0.94	0.94	0.94	0.94	0.94
$D_{11}^2$ <b>(True = 5.0)</b>	<b>Bias</b>	-0.01 (-2.07, 3.04)	-0.02 (-2.07, 3.04)	-0.02 (-2.08, 3.04)	-0.02 (-2.08, 3.03)	-0.02 (-2.08, 3.03)
	<b>% Bias</b>	-0.25 (-41.48, 60.86)	-0.35 (-41.4, 60.87)	-0.31 (-41.51, 60.73)	-0.33 (-41.56, 60.66)	-0.35 (-41.61, 60.6)
	<b>CI Width</b>	6.74 (4.0, 10.8)	6.74 (4.01, 10.8)	6.74 (4, 10.79)	6.74 (4, 10.79)	6.73 (3.99, 10.78)
	<b>Coverage</b>	0.95	0.95	0.95	0.95	0.95
$D_{22}^2$ <b>(True = 2.5)</b>	<b>Bias</b>	-0.09 (-1.14, 1.3)	-0.089 (-1.14, 1.31)	-0.089 (-1.14, 1.31)	-0.09 (-1.13, 1.3)	-0.09 (-1.13, 1.3)
	<b>% Bias</b>	-3.58 (-45.48, 52.11)	-3.57 (-45.53, 52.29)	-3.55 (-45.44, 52.22)	-3.60 (-45.39, 52.19)	-3.64 (-45.34, 52.17)
	<b>CI Width</b>	3.26 (1.87, 5.11)	3.26 (1.87, 5.11)	3.26 (1.87, 5.11)	3.25 (1.87, 5.11)	3.25 (1.87, 5.11)
	<b>Coverage</b>	0.94	0.94	0.94	0.94	0.94
$\rho_D$ <b>(True = 0.25)</b>	<b>Bias</b>	0.03 (-0.2, 0.28)	0.03 (-0.21, 0.28)	0.03 (-0.2, 0.28)	0.04 (-0.19, 0.29)	0.04 (-0.19, 0.29)
	<b>% Bias</b>	11.80 (-80.55, 112.66)	10.20 (-82.85, 110.52)	13.28 (-79.22, 113.99)	15.38 (-77.41, 115.16)	17.69 (-75.59, 116.32)
	<b>CI Width</b>	0.75 (0.59, 0.81)	0.75 (0.59, 0.81)	0.75 (0.58, 0.81)	0.75 (0.58, 0.812)	0.74 (0.58, 0.81)
	<b>Coverage</b>	0.96	0.96	0.96	0.95	0.95

**Table 4.20(A)** Median (5<sup>th</sup>, 95<sup>th</sup> percentiles) of estimates obtained from 300 simulated meta-analyses based on scenario 1b ( $(S_{11}^2, S_{22}^2) = 2(D_{11}^2, D_{22}^2) = (10.0, 5.0)$ ), and  $\rho_S = 2\rho_D = 0.5$ ), where between-subject variances and correlation are allowed to vary randomly from one subject to another.

## *Chapter 5*

### **(MANUSCRIPT III)**

## **META-ANALYSIS OF LONGITUDINAL STUDIES**

### **PREAMBLE**

This third and last paper of the thesis explored the application of multivariate models for the meta-analysis of longitudinal studies, where treatment effects are typically reported at various occasions. Thus, potentially multiple estimates of effect would be extracted from studies. A separate meta-analysis for effects at each occasion limits the scope of the analyses since changes in trends in effect over time can not be assessed formally. Therefore, a joint analysis of effects from all occasions would be desirable. Since these are repeated measurements from a common unit (i.e., study), however, there are inherent correlations in the data that must be accounted for in the model.

The objective of this paper was to examine alternative approaches that can be used for this purpose using data from a review of the effect of deep brain stimulation (DBS) in patients with Parkinson's disease. Alternative approaches to meta-analyzing such data were explored. More specifically, I considered a naïve approach that ignores the correlated nature of the data, which is equivalent to performing standard time-specific meta-analyses. I also explored methods that capture the total (i.e., within- and between-study) correlations with random-effects models but allow within-study residuals to be independent. The primary purpose was to implement a multivariate model, however, similar to those

employed for the joint meta-analysis of different but related endpoints. These models explicitly capture within- and between-study correlations separately. Thus, both the general benefit from accounting for correlations (standard or naïve vs. other approaches), as well as any advantages to modeling within- and between study correlations (multivariate vs. random-effects) could be ascertained.

This paper addressed the general objective of the thesis to explore new potential applications of multivariate models. Since this was an application of multivariate models in one specific meta-analysis, the scope of the findings is perhaps limited; that is, the advantages we observe may be due to particularities in the data. Therefore, the goal of the paper was mainly to illustrate the implementation of the method and discuss some of the key issues involved in this context.

This article will be submitted to Clinical Trials.

# META-ANALYSIS OF LONGITUDINAL STUDIES

Khajak Ishak<sup>1</sup>, Robert W. Platt<sup>1,2</sup>, Lawrence Joseph<sup>1,3</sup>,  
James A. Hanley<sup>1,4</sup>, J. Jaime Caro<sup>1,5</sup>

<sup>1</sup>Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada;

<sup>2</sup>The Montreal Children's Hospital Research Institute, McGill University, Montreal, Canada; <sup>3</sup>Division of Clinical Epidemiology, Royal Victoria Hospital, Department of Medicine, Montreal, Canada; <sup>4</sup>Division of Clinical Epidemiology, Montreal General Hospital, Department of Medicine, Montreal, Canada; <sup>5</sup>Caro Research Institute, Concord, MA, USA.

---

## SUMMARY

Longitudinal studies typically report estimates of the effect of a treatment or exposure on an endpoint at various times during the course of follow-up. Meta-analyses of these studies must account for the correlation between effect estimates collected from the same study. We describe a linear mixed-effects model that accounts for the correlation between observed estimates with random-effects or in a general multivariate specification that also allows correlated within-study residuals. We use data from a review of studies of the effect of deep-brain stimulation in patients with Parkinson's disease to illustrate the application of these models. Results are contrasted with those obtained from an approach that assumes independence between observations, which is equivalent to meta-analyzing the data at each time point separately. Summary effect estimates from the longitudinal models were more precise and appear to be less affected by apparently outlying observations.

**Keywords:** Longitudinal studies, meta-analysis, multivariate model.

---

## 5.1 INTRODUCTION

Longitudinal studies are commonly used in epidemiological research to track the effect of a treatment or exposure over time. These studies typically involve a series of measurements of the response variable at pre-determined intervals. Treatment effects can then be described by estimates calculated at various times (corresponding to the measurement times in the study). Alternatively, longitudinal data is sometimes analyzed using summary measures<sup>51</sup> (e.g., mean or slope of response values for each participant), in which case the treatment effect is

expressed in terms of (relative or absolute) differences in the summary measure (e.g., difference in mean slope of response among treated and controls).

Meta-analyses of longitudinal studies reporting treatment effects in terms of a single summary measure can be handled with standard approaches<sup>7,18,29</sup>. Special consideration is required, however, when effect estimates are reported at different times or in terms of summary measures involving multiple parameters, (e.g., polynomial functions) since these statistics are inherently correlated. The correlations may be biological (e.g., course of the disease), structural (e.g., effects calculated at different times relative to a common baseline) or statistical (e.g., measurements subject to similar “errors” or derived from same equation – e.g., intercept and slope).

A few examples of meta-analyses of longitudinal data have appeared recently<sup>52-54</sup>. Lopes et al.<sup>52</sup> describe a Bayesian model for the meta-analysis of longitudinal studies with patient-level data using mixtures of multivariate normal distributions. Farlow et al.<sup>53</sup> also present a meta-analysis of longitudinal studies based on patient-level data. Maas et al.<sup>54</sup> describe a mixed model for the meta-analysis longitudinal effect estimates; they handle the correlation between observations by allowing random intercepts and linear time effects.

In this paper, we discuss a general linear mixed-effects model for the meta-analysis of longitudinal effect estimates and discuss alternative ways of specifying the model to account for correlations between the observations. We examine random-study (intercept) and random-time-effects models as well as a more general two-level multivariate model that estimates separate within- and between-study covariance matrices. The latter is a new application of the multiple-outcome models that have been used for the meta-analysis of two or more related outcomes<sup>1,5,6,8-12</sup>. The general linear mixed effects model can also be specified to meta-analyze longitudinal data with an assumption of independence; this is equivalent to conducting separate analyses for the data from each time point.

The unit of analysis in these models is a vector of observed effect estimates at different times, which are modeled assuming a multivariate normal likelihood

(after transformation, if necessary), and allow random-effects to capture heterogeneity across studies. These random-effects are also considered to arise from a multivariate normal distribution to incorporate the possibility that these random-effects may also be correlated. The structure of the relationship between observations is controlled by the way the covariance matrices of the distributions are specified. Although we consider the situation where studies report effect estimates at a set of fixed times, the same method can be applied when longitudinal data are reported in terms of multiple parameters of functions describing the effects of treatment.

We use data from a meta-analysis of the effect of deep-brain stimulation (DBS) on motor skills of patients with Parkinson's disease at 3, 6, 12 months and later after implantation of the stimulator to illustrate the application of the various models considered. We compare results from these approaches to assess the sensitivity of estimates of summary effect and heterogeneity to specifications used in different models. We are particularly interested in differences relative to models that ignore the correlations between observations.

The paper is organized as follows: The next section describes a series of linear mixed-models that may be used to meta-analyze longitudinal effect estimates. The data used to illustrate the models are described in the Section 3; the results from the analyses are summarized in Section 4, after which we contrast findings and propose explanations for the observed differences. We close with a discussion of the key points of the paper.

## **5.2 MODELS FOR LONGITUDINAL META-ANALYTIC DATA**

### **5.2.1 GENERAL LINEAR MIXED MODEL FOR LONGITUDINAL DATA**

We begin with a general linear mixed model that accounts for correlations inherent to longitudinal or repeated measures data. We adapt this model to the meta-analysis context in the following sections. Suppose  $K$  measurements are taken over time on  $N$  units (e.g., usually individuals, or studies in a meta-

analysis); we denote by  $y_i$ , the  $K \times 1$  vector of observed values from the  $i^{\text{th}}$  unit and by  $y_{ij}$  the  $j^{\text{th}}$  observation from this unit. A general linear mixed model that can account for the correlations between the observations is given by<sup>55,56</sup>:

$$y_i = X_i \theta + Z_i \delta_i + \varepsilon_i$$

Here  $X_i$  is a  $K \times p$  matrix of possibly time dependent covariates,  $\theta$  is a  $p \times 1$  vector of fixed effects,  $Z_i$  is a  $K \times q$  matrix of covariates which are usually a subset of those included in  $X_i$ ,  $\delta_i$  is a  $q \times 1$  vector of random-effects and  $\varepsilon_i$  is a  $K \times 1$  vector of residuals. We assume that observations from different studies are independent, so that  $\text{cov}(\varepsilon_{ij}, \varepsilon_{ml}) = 0$  when  $i \neq m$  and for any observations  $j, l = 1 \dots K$ . It is also assumed that the residuals and random effects are independent:  $\text{cov}(\varepsilon_i, \delta_i) = 0$ .

The joint distribution of the random-effects is assumed to be a  $q$ -dimensional multivariate normal (MVN) distribution with mean 0 and  $(q \times q)$  covariance matrix  $D$ :  $\delta_i \sim MVN_q(0, D)$ . The residuals are also assumed to have a joint MVN distribution:  $\varepsilon_i \sim MVN_K(0, S_i)$ , where  $S_i$  is a  $K \times K$  covariance matrix.

The structure of  $D$  and  $S_i$  must be specified in a way that captures the between-and within-unit (study or individual) covariance between observations. This may be done in a number of ways<sup>55</sup>: for instance, one may choose to not include any random-effects and set  $S_i$  to be a general unstructured positive definite matrix that is constant for all units: i.e.,  $S_i = S$ . Another approach is to allow a random intercept (i.e., set  $Z_i$  to be a column of 1s) with (scalar) variance  $D$  and set  $S_i = \sigma^2 I_K$ , where  $I_K$  is the  $K \times K$  identity matrix. This leads to a compound

symmetric correlation structure, whereby  $\text{corr}(y_{ij}, y_{il}) = \frac{D}{D + \sigma^2}$  when  $j \neq l$  and 0 otherwise. When multiple random-effects are involved,  $D$  must be given a structure to describe the relationship between the random-effects within each unit. Some of the most commonly used structures are<sup>61</sup>:

- **Compound Symmetric (Spherical) Correlations:** This structure assumes a constant correlation for all pairs of effects.

- **First-Order Auto-Regression (AR(1)) Correlation:** This structure assumes measurements taken closer to each other in time are more strongly correlated than those further apart.
- **Toeplitz, or Banded Correlation:** This structure also assumes correlations depend on the lag between observations, but does not impose a specific form for the relationship. That is, observations at different lags are allowed to be differently correlated.
- **Unstructured Correlation:** In this specification, no structure is imposed on the correlations between observations, so that any two observations may have a different correlation.

The parameters of the model can be estimated by maximum likelihood (ML), which is built from the marginal distribution of  $y_i$ :  $MVN_K(X_i\theta, V_i)$ , where  $V_i = \text{var}(y_i) = Z'_i D Z_i + S_i$ . Alternatively, restricted or residual maximum likelihood (REML) is sometimes used; in this case, the likelihood is built on the residuals  $y_i - X_i \hat{\theta}$ , and takes into account the loss of degrees of freedom in estimating the fixed effects when estimating the variance parameters of the model<sup>55</sup>.

### 5.2.2 MIXED MODELS FOR META-ANALYSIS OF LONGITUDINAL DATA

The general model described in the previous section can be adapted to meta-analyze longitudinal data with only slight modifications. In fact, the univariate random-effects meta-analysis model that is commonly used in practice<sup>7</sup> is a special case of the above with  $K = 1$  and  $Z_i = 1$  (i.e., random-intercept). The within-study variance which is a scalar in the univariate case, is set to the variance of the observed estimate reported in each study ( $S_{ij}^2$ ) and assumed known without error. This has the effect of weighting each observation by a factor that is inversely proportional to its variance, so that more precise estimates are more influential. The basic model (without covariates) for observations at occasion  $j$  is

given by  $y_{ij} = \theta_j + \delta_i + \varepsilon_{ij}$ , where the index  $j$  is fixed,  $\delta_i \sim N(0, D)$  and  $\varepsilon_{ij} | \delta_i \sim N(0, S_{ij}^2)$ .

In meta-analyses of longitudinal effect estimates, the *units* of observation are studies from which estimates of the treatment effect are collected at  $K$  times of interest. These times may be chosen based on the available data and the objectives of the analysis. Studies are likely to vary with respect to the times at which effect estimates are reported, as the timing of measurements might be different in each study. Thus, not all studies will provide data at all  $K$  times of interest. Observations may also be missing in patient-level data from longitudinal studies; in this case, missing data occurs when patients drop out of studies or fail to return for scheduled visits, which may be related to exposure and prognosis and, hence, can lead to selection bias. In meta-analysis, however, *missingness* is more likely to occur by design, since studies will typically report estimates at all planned measurement occasions; thus, missing observations are prone to occur at random.

Longitudinal models for meta-analytic data also differ from the general case in the way the covariance matrices are specified to account for the correlation between observations, since some of the approaches described above are not appropriate in this context. For instance, random-effects models are increasingly considered to be better suited for meta-analytic data<sup>28,30,31</sup> since they allow heterogeneity in the effect across studies. Thus, omitting random-effects and accounting for correlations through  $S_i$  alone would not be ideal. Furthermore, studies are likely to vary with respect to sample size, measurement methods and other aspects that will cause estimates obtained from one study to be more or less precise than from another. The variance may also vary over time within studies because of loss to follow-up, for instance. Thus, setting the within-study covariance matrix  $S_i$  to  $\sigma^2 I_K$  is not adequate since it assumes a constant variance for all effect estimates in all studies. Therefore,  $S_i$  must be set to be at least a diagonal matrix with components that can have different values. And, to ensure proper weighting of the data, we can proceed as in the univariate case and set the diagonal elements of

$S_i$  to values reported in the studies, or at least use the precision of estimates as weights in the model.

The correlation structure of the data must be controlled through the specification of both random-effects and  $S_i$ . To discuss different ways this may be done, we first specify the fixed-effects, which will guide the choice of random-effects. Since the goal of the analysis is to summarize the effect of treatment over time,  $X_i$  will include covariates that relate the response variable to time. This might consist of some parameterization of time (e.g., linear, quadratic, logarithmic, etc.). In this paper, we consider the case where one wishes to derive summary effect estimates at each of the  $K$  measurement occasions. This may be done by omitting an intercept and including  $K$  indicators corresponding to the measurement times; otherwise,  $K-1$  indicators might be included with an intercept. For example, with  $K = 4$ ,  $X_i$  may be:

$$X_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ or } X_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

In the first case, a summary estimate is obtained for each time while in the second case, the covariates measure differences in summary effect estimates at each time relative to the effect at the first occasion, which is estimated by the intercept. Study-level covariates (e.g., proportion of males, average age, etc) may also be considered for inclusion in  $X_i$ . These are typically considered to have fixed-effects and, therefore, do not impact the specification of the correlation structure.

### 5.2.3 SPECIFICATION OF CORRELATION STRUCTURE

We now describe different ways of modeling the correlated structure in longitudinal meta-analytic data. Unless otherwise stated, we assume  $S_i$  is a diagonal matrix with values set to the variances of estimates reported in the studies (we denote the  $j^{\text{th}}$  diagonal element by  $S_{ij}^2$  for study  $i$ ). To facilitate notation, we assume  $X_i$  only includes time indicators (i.e., no intercept or other

covariates); including other fixed-effects to the models described below would not change the correlation structures.

### *Random Study-Effect*

The simplest way to account for the correlation between observations is to allow a random-effect that is common to all observations from a given study. This can be thought of as a random-intercept model. We set  $Z'_i = [1 \ 1 \ 1 \ 1]$ , so that  $\delta_i$  is a scalar; the model can then be written as:

$$y_{ij} = \sum_{j=1}^K X_{ij} \theta_j + \delta_i + \varepsilon_{ij},$$

where  $\text{var}(\delta_i) = D$ , the covariance of the vector of residuals is  $\text{cov}(\varepsilon_i) = S_i$  and  $\text{cov}(\delta_i, \varepsilon_{ij}) = 0$  for all  $i$  and  $j$ . The variance of the marginal distribution of  $y_{ij}$  is  $D + S_{ij}^2$ . And, the covariance between two observations collected on occasions  $j$  and  $l$  in study  $i$  is given by  $\text{cov}(y_{ij}, y_{il}) = D$ ; observations from different studies are assumed independent. Thus, the correlation between observations is  $\text{corr}(y_{ij}, y_{il}) = \frac{D}{\sqrt{(D + S_{ij}^2)(D + S_{il}^2)}}$ . Since we do not assume a constant within-study variance, this specification allows different correlations between pairs of observations.

### *Random Time-Effects*

The previous model is somewhat restrictive since it assumes that between-study heterogeneity affects all observations in a given study the same way. That is, heterogeneity in effects at each measurement occasion is the same. This may not be true, however. For instance; attrition may occur differently in each study, causing greater heterogeneity in effects estimated at later times. We can extend the random study-effects model by allowing a random-effect at each measurement occasion. That is, we set  $Z_i = X_i$ , so that the model is given by:

$$y_{ij} = \sum_{j=1}^K X_{ij} (\theta_j + \delta_{ij}) + \varepsilon_{ij}, \text{ or}$$

$$y_i = X_i(\theta + \delta_i) + \varepsilon_i.$$

Thus,  $\delta_i$  is now a  $K$ -vector of random-effects with covariance matrix  $D$ . Assuming these are independent (i.e.,  $D$  is diagonal) is equivalent to assuming the observations themselves are independent, since

$$\text{cov}(y_{ij}, y_{il}) = \text{cov}(\delta_{ij} + \varepsilon_{ij}, \delta_{il} + \varepsilon_{il}) = \text{cov}(\delta_{ij}, \delta_{il}) + \text{cov}(\varepsilon_{ij}, \varepsilon_{il}) = 0,$$

because both  $D$  and  $S_i$  are diagonal. This is equivalent to treating each observation as coming from a different study or meta-analyzing the data at each time separately with a univariate model.

A structure must, therefore, be imposed on  $D$  to capture the correlation between observations. Ideally,  $D$  would be left unstructured to allow a different correlation between pairs of random-effects. This adds  $K(K-1)/2$  parameters to the model, however. Considering that data are often scant in meta-analysis, this can hinder the estimation of the model parameters. More parsimonious structures like compound symmetric or AR(1) are more practical, as they involve only one additional parameter. This comes at the cost of more constrained relationships between random-effects, however.

Generally, the variance of the marginal distribution of the observations is given by  $\text{var}(y_i) = Z'_i D Z_i + S_i$ . The correlations between the components of  $y_i$  will depend on the structure chosen for  $D$ . For example, if a compound symmetric covariance matrix is used with  $\text{corr}(\delta_{ij}, \delta_{il}) = \rho$ , we get

$$\text{corr}(y_{ij}, y_{il}) = \frac{\rho \times D_{jj} \times D_{ll}}{\sqrt{(D_{jj} + S_{ij}^2)(D_{ll} + S_{il}^2)}}, \text{ where } D_{jj} \text{ and } D_{ll} \text{ are the diagonal elements}$$

(variances) in  $D$ . With an AR(1) structure,  $\text{corr}(\delta_{ij}, \delta_{il}) = \rho_D^{|j-l|}$  and

$$\text{corr}(y_{ij}, y_{il}) = \frac{\rho_D^{|j-l|} \times D_{jj} \times D_{ll}}{\sqrt{(D_{jj} + S_{ij}^2)(D_{ll} + S_{il}^2)}}.$$

### *Random Within- and Between Study Time Effects*

The models described so far have assumed within-study variation to occur independently at different measurement occasions (diagonal  $S_i$ ) and accounted for correlations between observations through random-effects. Correlations exist at both within- and between-studies, however. For instance, sampling variability and measurement error might affect estimates within a study at different times in a similar way. In general, factors causing variability between studies (e.g., blinding methods, randomization, etc) may cause correlations within-studies as the impact of these factors is shared by all observations. On the other hand, background factors that cause a given study to be prone to overestimate the effect at one occasion might also have a similar impact at other times.

The approaches considered so far do not model these sources of correlation separately, but rather capture the *total* correlation in the marginal distribution of the observations. To do so, we extend the model in the last section by relaxing the assumption that the within-study covariance matrix is diagonal and allowing it to have a structure that reflects the relationship between observations within each study. Two approaches are possible: if the studies report the covariance between estimates, these values may be used to specify the off-diagonal elements of  $S_i$ . As with the variances, the covariances would be assumed known without error; studies rarely report covariances, however, which makes the previous models very appealing.

Alternatively, a hybrid approach may be adopted whereby the diagonal elements of  $S_i$  are held fixed and the off-diagonal elements estimated from the data. This requires specification of a structure for  $S_i$ , however, which introduces more parameters in the model. Since data are usually limited in meta-analyses, some simplifications or approximations that would limit the number of parameters added to the model are in order. One simplification is to assume within-study correlations are constant across studies (i.e.,  $\text{corr}(y_{ij}, y_{il}) = \text{corr}(y_{mj}, y_{ml}) = \rho_{jl}$  for studies  $i$  and  $m$ ), still allowing the variances to differ, however. A further simplification would be to adopt a simple correlation structure, like compound

symmetry ( $\rho_{jl} = \rho$ ) or AR(1) ( $\rho_{jl} = \rho^{|j-l|}$ ), which require only a single parameter.

With either approach, the variance of the marginal distribution of the observations is given by  $V_i = \text{var}(y_i) = Z'_i D Z_i + S_i$ . The structure of  $V_i$  will depend on the structure imposed on  $D$  and assumptions made in specifying the covariances in  $S_i$ .

#### **5.2.4 IMPLEMENTATION OF THE MODEL**

The implementation of linear mixed models for meta-analytic data differs from the general case in that it requires fixing all or some of the components of the within-study covariance matrix to the values observed in the studies. Thus, any software for linear mixed models that provides this option may be used. A commonly used procedure is the SAS/STAT PROC MIXED software. Van Houwelingen et al.<sup>12</sup> describe how the procedure can be set up for general multivariate (multiple-outcome) meta-analyses where within-study covariances are known or set to fixed approximate values. Some manipulation is required, however, if only the within-study variances are fixed and covariance components are to be estimated from the data. Nam et al.<sup>9</sup> implemented similar model in a Bayesian context with BUGS<sup>43</sup>. The observed within-study variances were used to weight the data, but all components of  $S_i$  were treated as unknowns and estimated from the data.

### **5.3 CASE STUDY**

#### **5.3.1 DEEP-BRAIN STIMULATION IN PARKINSON'S DISEASE**

We use data from a meta-analysis of the effect of deep-brain stimulation (DBS) on the relief of symptoms of Parkinson's disease<sup>62</sup> to explore the application of the models described above. Parkinson's disease is a chronic progressive disease characterized by declining motor function and, eventually, severe disability<sup>63</sup>. Although pharmacological treatments are available, medication effects become increasingly unpredictable and short-lived, leaving patients with little or no relief of symptoms for much of the day.<sup>64</sup> To prolong effective ("on") periods,

physicians increase the medication dose, which comes at the cost of increased side-effects<sup>65</sup>. DBS, which is delivered through thin surgically implanted wires in specific areas of the brain and controlled by the patient, is meant to provide relief of this severe disability with lower doses of medication. DBS of the subthalamic nucleus (STN)<sup>66</sup> has been found to be more beneficial than stimulation of other sites in the brain for patients with Parkinson's disease<sup>67</sup>.

We reviewed studies published between 1980 and 2004 reporting on the effects of DBS of the STN on changes in medication dose, motor skills, activities of daily living and dyskinesias (side-effect)<sup>62</sup>. We were interested in describing changes in the effect of DBS during the course of the first year of use and in the long term. We included studies that reported estimates of effect for the outcomes of interest with a standard error or a confidence interval.

In this paper, we use data on the effect of DBS on motor skills measured at 3, 6, 12 months and long term after implantation of the stimulator. Motor function was measured in studies with the Unified Parkinson's Disease Rating Scale (UPDRS-part III)<sup>68</sup>. Scores can range between 0 and 108 and lower values are indicative of better motor function. The effect of DBS was measured by comparing scores while the stimulator is active to scores observed at baseline (before implantation of the stimulator) under two conditions: with and without concurrent use of medication. We focus on observations made without medication as these reflect the full potential of DBS.

### **5.3.2 MODELS**

We recorded effect estimates ( $\gamma_{ij}$ ) reported at each of the ( $K=$ ) 4 occasions as well as the variances of the estimates ( $S_{ij}^2$ ). The variances may vary from one measurement occasion to another due to attrition in each study. We also collect the variances of these estimates ( $S_{ijl}^2$ ). The covariances between estimates at different times in each study ( $S_{ijl}$ ) were not reported, however. We recorded the mean age of patients in the study population, the proportion of females, mean duration of disease and the mean baseline UPDRS scores. These were considered

in the analyses as predictors of the effect of DBS over time to explain heterogeneity in effects across studies.

We fitted the following models to meta-analyze these data:

1. **Independence Model:** Linear mixed model with uncorrelated random-effects for time indicators and residuals.
2. **Random Study-Effects Model:** Linear mixed model with random study effect and independent residuals.
3. **Random Time-Effects Model:** Linear mixed model with correlated random-effects for time indicators and independent residuals.
4. **Multivariate Model:** Linear mixed model with correlated random-effects for time indicators and correlated residuals.

The first approach is equivalent to conducting a separate meta-analysis of the data at each occasion since it assumes complete independence between observations, as though each arose from a different study or patient population. We refer to the remaining as longitudinal models as they account for correlations between observations: models 2 and 3 do so with random-effects only, and, therefore, capture the total correlation between the  $y_{ij}$ . The last approach is a two-level linear multivariate model that attempts to separate within- and between-study correlations.

Since the true values of the parameters are not known, it is difficult to evaluate or compare the models with respect to accuracy of estimates or coverage of confidence intervals. Therefore, comparisons we make are meant only to understand the changes we observe relative to the way correlations were handled in each model. For instance, we were interested to know how results from the model that assumes independence (model 1) differs from those that account for correlations. We also wished to determine whether the full multivariate model would yield very different results than models that capture the marginal correlations between observations. We compared the precision of effect estimates, the relative magnitude of between-study variance estimates and

differences in estimated patterns of change in effects over time. We acknowledge, however, that gains in precision are not necessarily a sign of improvement as this may lead to narrower confidence intervals that have coverage probabilities below the nominal values.

We also compare goodness-of-fit using the likelihood-ratio statistics of nested models and Akaike's information criterion (AIC), which penalizes the likelihood by twice the number of parameters in the model. All the models include the same fixed effects but differ in terms of the random-effects and covariance parameters. The random study-effect model is the most parsimonious with only a single random effect. The uncorrelated random-time-effects model includes four random effects; the correlated random-time-effects and multivariate models involve a single correlation parameter (since AR(1) and compound symmetry structures are considered – see below) for the random effects distributions; the multivariate model also requires a correlation parameter for the distribution of residuals.

### **5.3.3 SPECIFICATION OF COVARIANCE MATRICES**

Models 3 and 4 require the specification of covariance structures for random-effects ( $D$ ) and/or residuals ( $S_i$ ). In applications with relatively small data sets, like the current (and likely most) meta-analyses, multi-parameter covariance structures (e.g., unstructured or Toeplitz) can complicate or even inhibit estimation. In fact, we were unable to fit the model to the data in our example with either of these structures; the estimation procedure either diverged or produced non-positive definite covariance matrix estimates with only point estimates.

Our choice was, therefore, limited to the simpler AR(1) and compound symmetric structures. The AR(1) structure seems well-suited for longitudinal data since background factors that induce heterogeneity may vary over time and have a more similar impact on effects measured closer in time than further apart. However, this structure is implemented for data measured at regular (equidistant) time-intervals in SAS/STAT PROC MIXED since the correlations are parameterized

by the order of the observations rather than the actual measurement times ( $(\rho^D)^{|j-k|}$ )<sup>69</sup>. This is not ideal for our example since the lag between observations is not constant. Alternatively, the compound symmetric structure assumes the correlation is constant, and so does not depend on the lag between observations. It too is not ideal since it would fail to capture changes in the correlations over time. Nonetheless, we performed the meta-analysis with both structures and found little difference.

In model 5, we parameterized the covariance components of  $S_i$  in terms of correlation parameters which we assume to be common to all studies. That is, we assumed, for any study  $i$  and observations  $j$  and  $l$ ,  $S_{jl} = \rho_{jl}^S \times \sqrt{S_{jj}^2 \times S_{ll}^2}$ , where  $\rho_{jl}^S$  is the correlation between estimates at times  $j$  and  $l$  in all studies. We further reduced the added burden to the model (in terms of additional parameters) by adopting a spherical ( $\text{corr}(y_{ij}, y_{il}) = \rho^S$ ) or AR(1) ( $\text{corr}(y_{ij}, y_{il}) = (\rho^S)^{|j-k|}$ ) covariance structure. We examined both formulations and, in each case, assumed the same structure for within-study errors and between-study random-effects. Results did not depend on the choice of correlation structures for random-effects and residuals. For brevity, we only report findings based on models using an AR(1) structure for both between- and within-study covariance matrices.

## 5.4 RESULTS

Forty-six studies reporting estimates of the effect of DBS of the STN in the absence of medication in at least one of the time intervals of interest were included in the meta-analysis. Half of these studies reported effects for a single time interval only, 13 reported at two, seven provided data for all but one, and only three studies reported effects in all four times of interest. Eighty-two observations were extracted in all: 24 were measured at three months after implantation of the DBS, 22 at six months, 25 at one year and 11 in the second year or later (long-term).

Figure 5.7(a) and (b) display the observed estimates with corresponding 95% confidence intervals at months 3, 6, 12 and long-term; the point estimates are represented by circles of size proportional to their precision. These figures also show the summary estimates from standard univariate random-effects models (which also correspond to those from the independent random time effects model). Considering the invasive nature of the intervention, many of the studies had small samples, and, thus, yielded estimates with wide confidence intervals that sometimes included the null value. The general pattern of point estimates reveals a considerable improvement in motor skills, however.

Table 5.21 displays the covariance/correlation matrix computed empirically from the observed effect estimates. All correlations were above 0.7, suggesting very strong dependence between effects observed within studies. The correlations tended to decline with increasing lag between measurements. The exception was a strong correlation observed between month three and long-term effects. We note, however, that data from only four studies were available for that calculation. Otherwise, the correlation pattern seems to be generally consistent with an AR(1) structure: effects in consecutive intervals (3-6, 6-12, 12-24) have a correlation of approximately 0.9, while those at a lag of two intervals were 0.71 and 0.77, reasonably close to  $0.9^2=0.81$ . On the other hand, considering the similarity of correlations throughout, a compound symmetric structure would also be acceptable for the covariance matrices in the correlated random time-effects and multivariate models.

Summary estimates from the models we considered in this analysis are presented in Table 5.22. The uncorrelated random time-effects model, which assumes independence between observations, suggests a reduction of 24.9 points (95% CI: -27.3, -22.4) on the UPDRS-Motor scale at month 3. Slightly stronger effects were found at six and 12 months. The long-term effect appears to decline, however, as the summary estimate suggests improvements similar to those observed at month three.

The random study-effects (i.e., random intercepts) and correlated random time-effects models yielded similar point estimates and confidence intervals for summary effects. While the pattern was generally consistent with those from the *independence* model, we note that the estimates from these models suggest a slightly stronger effect at month 3 and long-term. This is evident from both the point estimate and upper-bounds of the confidence intervals, which were narrower than those from the independence model. The long term effect was found to be stronger still in the multivariate model; the upper bound of the confidence interval was -23.8 points compared to -22.6 in the two other models that accounted for correlations and -20.0 points in the independence model.

Table 5.22 also summarizes the estimates and 95% confidence intervals of the between-study variances. These suggest substantial heterogeneity in effects across studies at each of the measurement occasions. While point estimates of between-study variances were similar in all of the models, confidence intervals of these estimates were narrower when the correlations between observations were taken into account, particularly in the long-term interval, which included the fewest observations. The upper bound of the 95% confidence interval was reduced from 186.6 in the independence model to 108.4 (with AR(1) and 112.8 with compound symmetric) in the multivariate model.

Models that accounted for correlations appear to have better fit than the uncorrelated random-time-effects model, which had noticeably higher -2log-likelihood and AIC values. In fact, these statistics were lowered by 39 points in the correlated random-time-effects model which differs from the independence model by a single correlation parameter. The random study effect model had comparable fit to the former with fewer random parameters – in fact it only includes a single random-effect. The multivariate model appears to have the best fit among these models; allowing within-study correlations (with a single parameter) reduced the -2log-likelihood by 11.1 points compared to the correlated random-time-effects model and 14.1 points compared to the random-study-effects model. We note, however, that the last reduction was achieved with five additional (within-study correlation + three additional random-effects + 1

between-study correlation) parameters. In both cases, the differences relative to the multivariate model are statistically significant based on chi-square test of 1 and 5 degrees of freedom at 5% level of significance, respectively.

We attempted to explain some of the heterogeneity between studies by including the mean age of patients, the proportion of females, mean duration of disease and the mean baseline UPDRS scores as predictors of effects over time. The first two did not appear to be predictive of the size of effects and so were dropped from the models. Mean disease duration and mean baseline score were centered at their mean values (14 years and 52 points) and two studies were excluded because the mean duration of disease of the study population was not reported. Table 5.23 summarizes the findings from these analyses. The fits of all models were significantly improved with the inclusion of the covariates; the -2log-likelihood and AICs were reduced by 37.5 points or more. The change was largest for the uncorrelated random-time-effects and multivariate models, for which the reductions exceeded 50 points. Thus, accounting for within-study correlations appears to lead to an even greater improvement in fit relative to the random study or time effects models in this situation as the gap in fit statistics is even greater than in the previous analyses.

Summary effect estimates were very similar to those obtained in the *unadjusted* models (Table 5.22), but confidence intervals were slightly narrower. We attribute this, at least partly, to the substantial reduction in between-study variances by the inclusion of mean disease duration and mean baseline score. Thus, differences in study populations account for an important part of the observed heterogeneity. The effects of mean disease duration and mean baseline score were slightly weaker in the independence model, which also yielded narrower confidence intervals for these factors than the other models.

In addition to effect estimates, the mixed model with correlated random time effects and multivariate models also quantified within- and between-study correlations. In the former case, the covariance matrix of the random-effects was not positive-definite, so that correlation estimates were not returned from the

estimation procedure. In contrast, the within- and between-study covariance matrices of the multivariate model were estimated successfully: with an AR(1) covariance structure, the correlation between effects in successive intervals was estimated to be 0.97 (95% CI: 0.88, 1.06) within-studies ( $\rho_S$ ) and 0.88 (95%CI: 0.79, 0.98) between studies ( $\rho_D$ ) without other covariates\*. We found similar correlations when we included mean disease duration and mean baseline scores: 0.95 (95% CI: 0.88 – 1.03) within and 0.77 (95%CI: 0.52, 1.03) between studies. We attribute these high correlations to the relative stability effects over time†.

## 5.5 COMPARISON OF MODELS

Results were generally consistent between all of the models fitted to the data in our example. A few differences were observed that, while too slight to alter conclusions in our analyses, may be indicative of potential advantages to using models that account for correlations that arise in longitudinal meta-analytic data in general. We acknowledge that since the actual values of parameters are not known, we cannot confirm that these differences are in fact favorable.

Models that captured the correlations between observations appeared to have significantly better fit than the independence model and produced summary effect and between-study variance estimates with narrower confidence intervals. We attribute this to a *borrowing of strength*, or information, across times in these models which does not occur in the uncorrelated random-time-effects model which treats observation as independent, as though each arose from a different study. Furthermore, the multivariate model yielded the narrowest interval estimates for the long-term data and had better fit than the other longitudinal

---

\* Assuming compound symmetric correlations yielded estimates of 0.90 (95% CI: 0.71, 1.08) and 0.88 (95% CI: 0.75, 1.01) for within- and between-study correlations, respectively.

† The estimation procedure yielded upper bounds exceeding the logical boundary of 1.0 for of the confidence intervals for some correlation estimates. We suspect this is due to the relatively high correlations and the lack of precision in estimating correlations which may have pushed the estimation algorithms beyond the logical limit for this parameter.

(non-independence) models. Thus, it seems that modeling within- and between-study correlations separately was beneficial.

Although we found generally similar effect estimates from all approaches, models that accounted for correlations yielded slightly stronger long-term summary estimates. We hypothesized that this may be due to an apparently outlying effect estimate (identified by an arrow in Figure 5.8) in this time interval. This observation came from a study that also reported relatively low effects at other times. In fact, patients in this study had better motor skills (lower UPDRS score) at baseline compared to other studies, as a result there was lower potential for improvement. We removed the long-term observation from this interval and refitted the models. Figure 5.9 illustrates the new results (dashed lines) and the original (solid lines) effect estimates for the uncorrelated random-effects and multivariate models, which had previously yielded the weakest and strongest estimates, respectively. A greater change can be seen in results from the independence model, while those from the multivariate (and other two longitudinal models – not shown in Figure 5.9) model is essentially unaffected. This is because the outlying point is modeled as an independent observation in the uncorrelated random-time-effects model, even though other effects were observed from the same study at previous times. Models that account for correlations consider the entire set of observations from this study as the unit of analysis. Furthermore, the relatively constant pattern in the effects over time in this study implies strong within-study correlations; thus, the *effective* information provided by the observed effects is less than it would be had they been observed in different studies (i.e., if they were truly independent). This is taken into account in models that capture the correlations in the data; the extreme observation in the long term interval is, therefore, less influential in these models because very similar effects from the same study are also included in the analyses. Consequently, the outlying observation exerts far less pull on estimates from these models compared to the naïve model.

To explore this further, we also examined the impact of removing an apparently outlying observation at month six - the only observation from that particular study

- and found similar changes in summary estimates from all models. Therefore, we infer that accounting for the correlated nature of the data reduced the potential influence of the extreme observation. We must consider, however, that the correlations between observations were very strong in our data since there was little fluctuation in effects over time within studies; it is not clear whether the extreme observation would have been more influential if the correlations were weaker (i.e., effects were not relatively constant).

## 5.6 DISCUSSION

In this paper, we examined three approaches to meta-analyzing longitudinal effect estimates: one approach would be to summarize the data at each time independently; alternatively, random-effects may be used to account for the correlations between observations by allowing either random-study effects or correlated random time-effects. A more general multivariate model can also be specified by allowing both within-study residuals and random-time-effects to be correlated. Thus, within- and between-study correlations between effects were modeled separately, whereas the random-effects models capture the total or marginal correlation in the observations. By comparing these approaches in a case study, we were able to assess the potential impact of ignoring the correlations in the data (i.e., independence model vs. others), and whether there may be any benefit to modeling within- and between-study correlations separately (multivariate vs. random-effects).

Models that accounted for correlations had better fit and produced slightly more precise summary effect and heterogeneity estimates, particularly in intervals where data were limited. In fact, confidence intervals from the multivariate model were slightly narrower than those from random study- and time-effects models, which also had poorer fit. Furthermore, we found that the multivariate model (and to some extent the random-effects models) was less affected by an apparently *outlying* estimate in the last interval, since this was strongly correlated with effects observed at earlier times in this study. In fact, we also fitted the models after removing all of the observations from this study; summary estimates

at earlier times were only slightly affected in all models, although longitudinal models were less affected. We suspect the differences we observed between standard and longitudinal approaches in our example, particularly in the long term, were due to the relatively strong correlations ( $> 0.75$ ) in the data. This allowed a borrowing of strength across times, particularly for the long-term effects, for instance, which were reported relatively infrequently.

The slight gain in precision from the models that accounted for correlations seems surprising, since ignoring the correlation between observations would be expected to lead to underestimated variances<sup>70</sup>. However, the models we implemented stratified on time, so that only one observation per study is used at each occasion. Thus, the variances in the naïve model would not necessarily be underestimated. Rather, it is possible that, in the context of multivariate meta-analysis, allowing independent random-effects for each outcome perhaps exaggerates the heterogeneity between studies since the factors that induce this variability likely affect the outcomes (measurements at different times, in our example) in a similar way. Therefore, incorporating a correlation structure on the random-effects may account for the overlap in heterogeneity between time-intervals. In our example, this was apparent only when we accounted for factors causing systematic variability between studies (mean baseline score, mean disease duration); in this case, the multivariate model suggested lower heterogeneity than the independence model. In the unadjusted models, we suspect the greater precision in the multivariate model was in large part due to the reduced impact of the outlying observation noted above. This is evidenced by the fact that lower bounds of the confidence intervals were almost identical, while upper bounds (and point-estimates) changed.

Although the random-time-effects and multivariate models generally allow more flexibility in the specification of covariance structures, their estimation may be hampered by the added parameters due to the relatively small size of meta-analyses. In our example, we restricted the covariance matrices to have AR(1) or compound symmetry structures since estimation algorithms did not converge with other covariance structures. Despite these simplifications, the covariance matrix

estimate of the random-effects in the correlated random-time-effects model was not positive-definite and the correlation parameter was not estimated. The situation is potentially more difficult in the multivariate case since this model includes two covariance matrices. If within-study covariances are reported by studies or approximated<sup>1,8</sup>, the values may be treated as known in the covariance matrix of the likelihood (as with the variances). This can possibly avoid the simplifications made in our analysis (i.e., assuming constant correlation across studies) and allow more complex structures to be considered for the random-effects distribution.

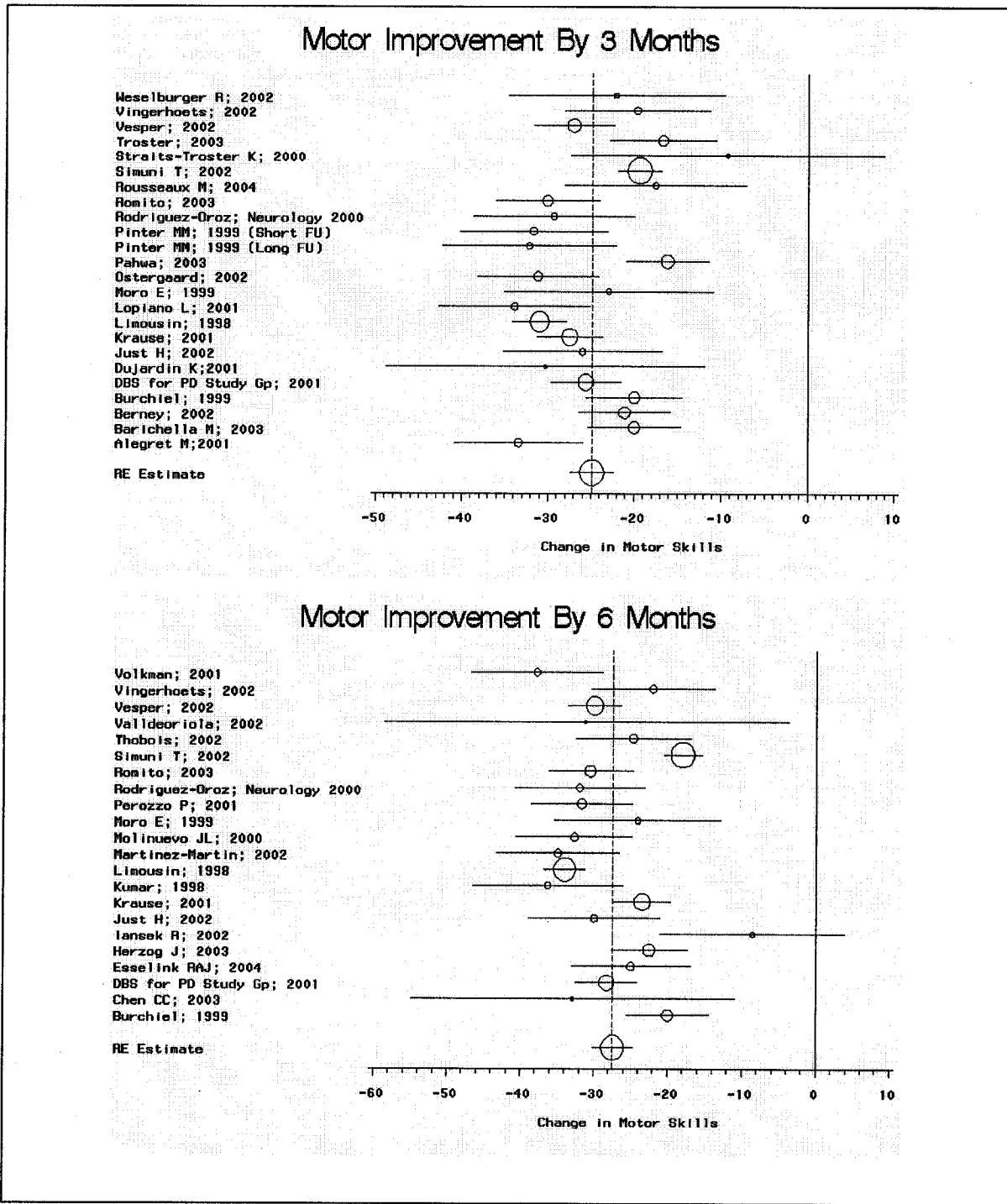
Half of the studies in our meta-analysis reported only one effect estimate (i.e., at one time). Data from other intervals for these studies are treated as missing values in the analysis. In fact, these are implicitly assumed to be missing completely at random. Results may be biased, however, if there effects at certain times were selectively omitted from publications. In the longitudinal context, however, it is plausible to expect that studies likely report results at all planned measurement occasions, so that the risk of selective reporting may be less than for reporting of different endpoints, for instance.

The models presented in this paper can be extended in a number of ways. They may be used for the meta-analysis of different outcome measures, as long as these have (at least approximately) normal sampling distributions. For example, they may be used to meta-analyze log-odds-ratio or risk difference estimates calculated at various times in studies of chronic diseases. Furthermore, although we used categorical time intervals, the analysis can also be performed with continuous formulations of the time variable, assuming the shape of the relationship is known or can be specified reasonably well. In this case, studies included in the meta-analysis need not be restricted to those publishing specific time points of interest.

Multivariate meta-analytic models have appeared in the recent literature<sup>1,5,6,8-12</sup> as an approach to meta-analyze two or more related outcomes. These applications have typically shown little difference in findings from multivariate and standard univariate (outcome-specific) meta-analyses. The example presented in this paper

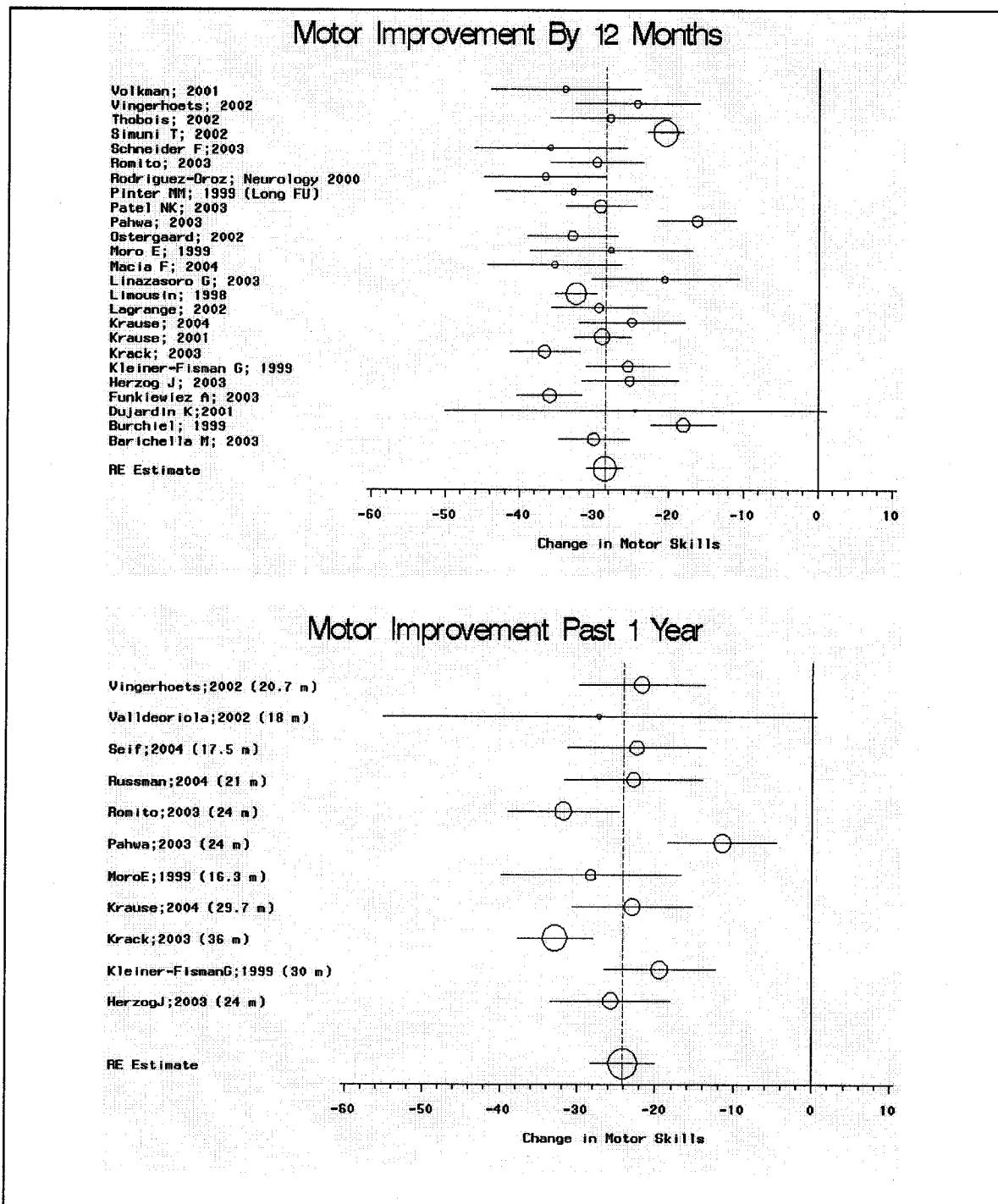
suggests there may be a potential for gaining precision with a multivariate analysis; since the true values of the parameters of the model were not known, however, we can not confirm that the observed differences were favorable. Further simulation studies are required to verify our findings.

## 5.7 FIGURES AND TABLES

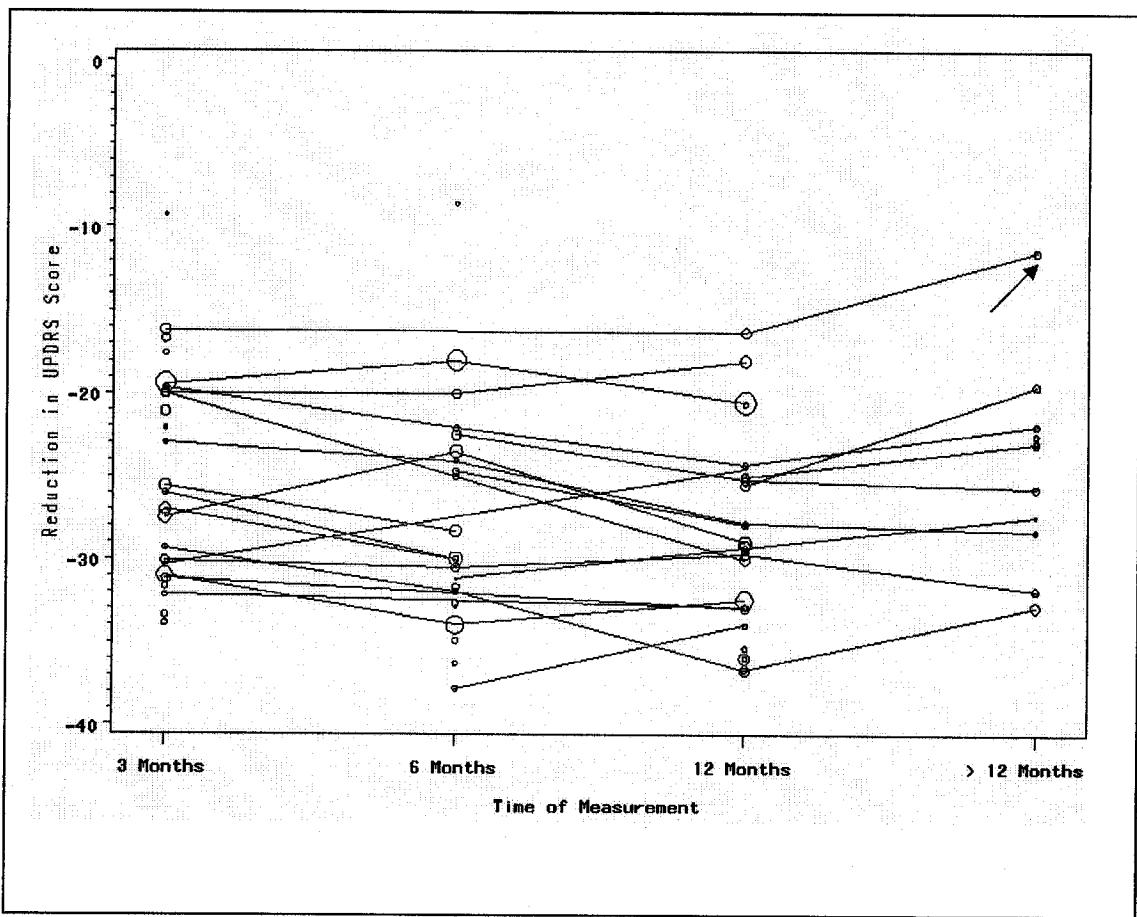


**Figure 5.7(a)** Effect estimates by months 3 and 6 with corresponding 95% confidence intervals, and as summary estimates from univariate random-effects models. The size of circles reflects the precision of estimates.

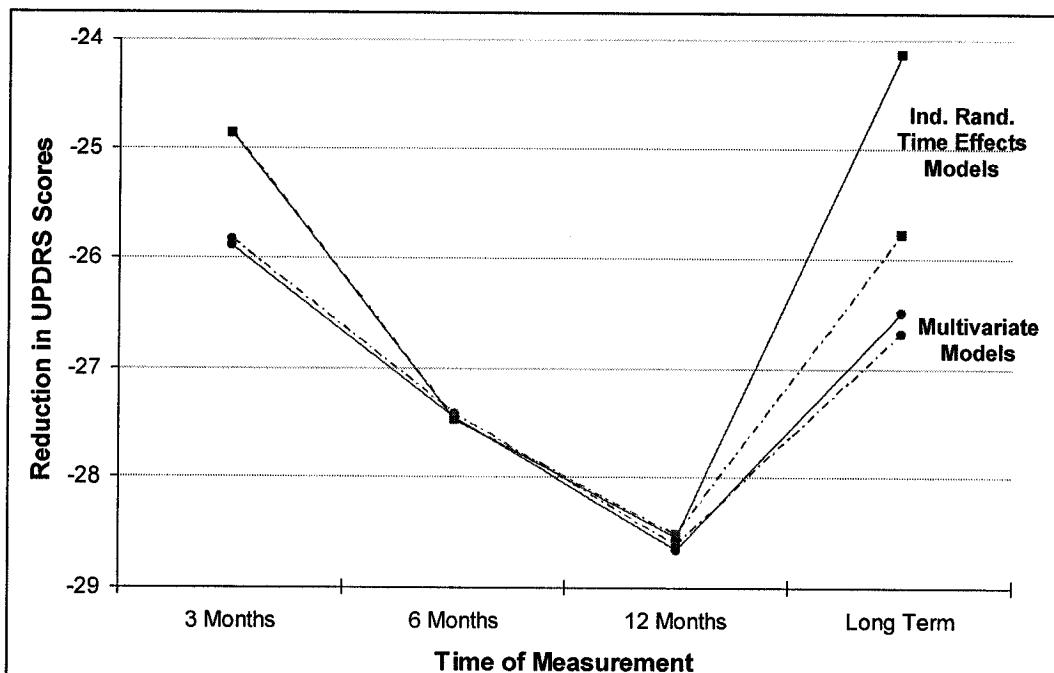
**NOTE:** The two studies by Pinter MM in month 3 based on two different populations.



**Figure 5.7(b)** Effect estimates by month 12 and long term with corresponding 95% confidence intervals, and summary estimates from univariate random-effects models. The size of circles reflects the precision of estimates.



**Figure 5.8** Observed effects of DBS of the STN by months 3, 6, 12 and long term and summary effect estimates. Connected points arise from the same study; unconnected points are from studies reporting an effect estimate at a single time. The size of circles reflects the precision of estimates. The arrow points to an apparently extreme observation, which was removed to assess its impact on summary estimates (Figure 5.9).



**Figure 5.9** Summary estimates of effect with all observations (solid lines) and when an apparently *outlying* observation in the long-term interval is removed (dashed line).

	<b>3 Months</b>	<b>6 Months</b>	<b>12 Months</b>	<b>Long Term</b>
<b>3 Months</b>	<b>43.0</b>	0.90	0.77	0.92
<b>6 Months</b>	20.1	<b>48.0</b>	<b>0.88</b>	0.71
<b>12 Months</b>	27.6	30.9	<b>34.8</b>	0.92
<b>Long-Term</b>	48.7	11.5	36.9	<b>35.9</b>

**Table 5.21** Correlation/covariance matrix of observed estimates of effect.

Diagonal elements are the variances, the upper diagonal shows the correlations and the lower diagonal shows the covariances between the observed estimates.

	<b>Uncorrelated Random Time Effects</b>	<b>Random Study Effect</b>	<b>Correlated Random Time Effects<sup>#</sup></b>	<b>Multivariate Model<sup>##</sup></b>
<b><i>Summary Effect Estimates</i></b>				
<b>3 Months</b>	-24.9 (-27.3, -22.4)	-26.2 (-28.3, -24.1)	-26.0 (-27.9, -23.9)	-25.9 (-27.9, -23.9)
<b>6 Months</b>	-27.5 (-30.2, -24.7)	-27.2 (-29.3, -25.1)	-27.5 (-29.7, -24.9)	-27.5 (-29.7, -25.2)
<b>12 Months</b>	-28.5 (-31.0, -26.0)	-28.5 (-30.6, -26.5)	-28.6 (-30.6, -26.5)	-28.7 (-30.7, -26.6)
<b>Long- Term</b>	-24.1 (-28.3, -20.0)	-25.6 (-28.9, -22.5)	-25.8 (-29.0, -22.6)	-26.5 (-29.2, -23.8)
<b><i>Estimates of Between-Study Variances</i></b>				
<b>3 Months</b>	23.1 (11.2, 71.6)	26.7 (16.1, 52.4)*	20.4 (11.0, 49.7)	22.6 (11.6, 61.3)
<b>6 Months</b>	27.8 (13.4, 88.9)		36.0 (18.7, 85.1)	33.7 (18.7, 78.4)
<b>12 Months</b>	27.7 (14.7, 69.6)		26.4 (15.1, 58.0)	26.1 (14.7, 58.7)
<b>Long- Term</b>	29.9 (11.5, 186.6)		30.1 (13.1, 124.7)	31.1 (14.5, 108.4)
<b><i>Model Fit</i></b>				
<b>-2LogL</b>	524.7	488.7	485.7	474.6
<b>AIC</b>	532.7	490.7	493.7	486.6

**Table 5.22** Summary effect and between-study variance estimates and 95% confidence intervals from various meta-analysis models.

\* Estimate and 95% confidence interval of variance of the random intercept.

# Implemented with AR(1) covariance structure for the joint distribution of random-effects.

## Implemented with an AR(1) covariance structure for within- and between-study covariance matrices.

	Uncorrelated Random Time Effects	Random Study Effect	Correlated Random Time Effects <sup>#</sup>	Multivariate Model <sup>##</sup>
<i><b>Summary Effect Estimates</b></i>				
<b>3 Months</b>	-25.0 (-26.4, -23.7)	-26.2 (-28.0, -24.4)	-25.7 (-27.2, -24.3)	-25.7 (-26.9, -24.5)
<b>6 Months</b>	-27.0 (-29.2, -24.8)	-27.1 (-28.9, -25.3)	-27.5 (-29.6, -25.4)	-27.7 (-29.4, -25.9)
<b>12 Months</b>	-28.1 (-30.3, -26.0)	-28.4 (-30.2, -26.7)	-28.5 (-30.3, -26.7)	-28.8 (-30.4, -27.1)
<b>Long- Term</b>	-24.0 (-27.1, -20.0)	-25.1 (-28.0, -22.2)	-25.2 (-28.1, -22.3)	-26.1 (-28.1, -24.1)
<b>Disease Duration</b>	-0.58 (-1.18, 0.02)	-0.81 (-1.67, 0.04)	-0.78 (-1.53, -0.04)	-0.63 (-1.29, 0.02)
<b>Baseline Score</b>	-0.60 (-0.74, -0.46)	-0.54 (-0.76, -0.33)	-0.57 (-0.76, -0.38)	-0.69 (-0.86, -0.52)
<i><b>Estimates of Between-Study Variances</b></i>				
<b>3 Months</b>	1.8 (0.35, 3491.2)	13.3 (7.4, 30.4)*	5.6 (2.1, 38.1)	1.8 (0.44, 152.0)
<b>6 Months</b>	14.1 (6.1, 60.2)		20.9 (10.3, 62.3)	12.7 (5.9, 43.4)
<b>12 Months</b>	16.3 (7.6, 55.6)		14.8 (7.6, 40.8)	12.0 (6.1, 33.0)
<b>Long- Term</b>	9.9 (2.7, 407.5)		12.9 (4.4, 137.5)	6.9 (2.0, 161.3)
<i><b>Model Fit</b></i>				
<b>-2LogL</b>	472.0	451.2	445.6	423.2
<b>AIC</b>	480.0	453.2	453.6	435.2

**Table 5.23** Summary effect and between-study variance and 95% confidence intervals of between-study variance parameters from meta-regressions including mean disease duration and mean baseline UPDRS-motor scores of patients in each study.

\* Estimate and 95% confidence interval of variance of the random intercept.

# Implemented with AR(1) covariance structure for the joint distribution of random-effects. Final estimate of covariance matrix was not positive-definite.

## Implemented with an AR(1) covariance structure for within- and between-study covariance matrices.

## *Chapter 6*

### **DISCUSSION AND CONCLUSION**

Multivariate meta-analysis has been proposed and used as an approach to summarize evidence regarding the effect of a treatment (or exposure) on multiple, possibly correlated outcomes reported in clinical trials (or epidemiological studies) using a single joint model. The broad objective of this thesis was to evaluate the usefulness and reliability of this method and to investigate new potential applications. More specifically, I explored the *meaning* of the correlation between outcomes estimated in multivariate models and whether these can be used reliably to deduce the underlying true relationship between treatment effects; I also examined the sensitivity of estimates from the model to poor approximations of within-study covariance matrices, which are assumed known in the analyses but rarely reported. These analyses revealed that correlations between outcomes do not reflect the intended relationships, and their estimates are usually prone to bias and lack precision. Finally, I implemented a multivariate meta-analysis of longitudinal data from a review of the effect of DBS in patients with Parkinson's disease and examined potential differences relative to models that capture marginal correlations and univariate time-specific meta-analyses.

#### **6.1 RELIABILITY OF MULTIVARIATE META-ANALYSIS**

My research provided some evidence to support the reliability of the multivariate approach but also highlighted some areas where the method might have some limitations. In frequentist applications of multivariate meta-analysis models (e.g., references 1, 5, 8), the within-study covariance matrix of observed estimates is assumed known in order to properly weight the observations; this is analogous to treating as fixed the variances in univariate meta-analyses. While variances are

usually reported or may be derived from the available information, the covariances are rarely available and must, therefore, be approximated<sup>1,8</sup>. I assessed the impact of errors made in these approximations on summary effect estimates as well as other parameters of the multivariate model. Summary effects and, to a slightly lesser extent, between-study variances were generally estimated accurately and their confidence intervals had the expected coverage probabilities, regardless of how poorly within-study covariances were approximated. Thus, these findings suggest that the model is relatively robust to erroneous approximations of within-study covariances for the estimation of effect and heterogeneity parameters. In fact, assuming independence between observations from the same study (i.e., independent within-study residuals) building in a correlation between the random-effects yielded very similar findings to those from analyses where observed within-study covariances were used.

Estimates of the correlation between outcomes, which is an important advantage of the multivariate approach as it can provide added insight about the treatment being studied<sup>6</sup>, were not estimated as reliably. Although the estimates of the correlation were unbiased on average, there was substantial variability over replications within each scenario of the simulations, sometimes producing relative biases in excess of 200%. Furthermore, confidence intervals consistently had coverage below the nominal values. In fact, these limitations were apparent even when within-study covariances were known. Thus, correlations between outcomes might generally not be estimated reliably in multivariate meta-analyses. This may have much to do with the fact that meta-analytic data are usually of limited size; this is further complicated in the multivariate case, where even if relatively large number of studies are included in the review, only part of these may report all of the outcomes of interest. Thus, the data may carry very limited information about the relationship between outcomes.

Not only are correlations between treatment effects on different outcomes difficult to estimate reliably, the relationships captured by this measure do not necessarily capture the dependencies of interest. This was examined in simulations reported

in the first manuscript (Chapter 3) which aimed to determine whether the correlation between the effects of a treatment on two (or more) outcomes measured across studies would accurately reflect the true underlying dependencies in treatment effects. I formulated a conceptual model consisting of modules representing the disease, treatment and study processes and described how each of these involve dependencies (between endpoints, treatment effects and *shared* errors across outcomes, respectively) that can affect the correlations measured in meta-analyses. The simulations showed that the correlation between marginal log-ORs measured across studies often did not reflect the *biological* relationship between treatment effects. Correlated random-effects and strong dependence between the events tended to overwhelm the dependence between treatment effects, regardless of the actual strength of the relationship.

The broader implication from these simulations is that marginal effect estimates, which are typically the units of meta-analysis, may not be sufficient to study the relationship between treatment effects on different outcomes. Richer data may help in isolating the various sources of correlations. For instance, having patient level data or, at least, knowing the joint distribution of event counts (A but not B, not A but B and both A and B) in treated and control groups would allow more sensitive approaches.

## 6.2 APPLICATIONS OF MULTIVARIATE META-ANALYSIS

I explored two potential applications for multivariate models in meta-analysis. The first was to measure the correlation between treatment effects for different outcomes to gain insight into the underlying biological relationship. This is a potentially interesting application as it may be used, for example, to measure the association between efficacy and safety of new drugs. This may be difficult to assess in any single study since adverse events are typically rare. Although the performance of multivariate models were not formally tested in this capacity, I simulated the treatment, disease and study processes to understand how the relationship between effects measured across studies are affected by dependencies that exist at each of these levels. That is, the simulations produced the true values

of the correlation between effects that would be estimated in multivariate meta-analyses. The accuracy and precision of estimates in meta-analyses would further depend on other factors such as the number of available studies and observations, sample sizes of studies, incomplete observations etc. As discussed in the previous section, the correlation between marginal effect-sizes observed across studies often did not accurately reflect the underlying relationships and was distorted by dependencies from other sources. Thus, even if multivariate meta-analyses models were capable of estimating these correlations accurately (which is also doubtful based on findings reported in Chapter 4), the estimated values would not be representative of the intended measure. Therefore, the usefulness of multivariate models in this area seems to be considerably limited (at least when only marginal effect estimates are available).

The second application I examined was for the meta-analysis of longitudinal studies where effect estimates are often reported at several occasions and are, therefore, inherently correlated. Meta-regression (with time as a covariate) can be used to analyze data from all times jointly, but random-effects must be specified carefully to build in a correlated structure between observations. I implemented random study-effects and random time-effects models that captured the marginal correlation between observations, as well as a *full* multivariate model with separate within- and between-study correlations. Although summary effect estimates were similar from all models in the meta-analysis of the effect of DBS, estimates from the univariate models had slightly wider confidence intervals, notably in the interval where the fewest observations were reported, and appeared to be more affected by an apparently outlying observation. The random study-effects and random time-effects models performed better than the standard approach, but not as good as the full multivariate model in terms of both fit and precision. Whether these differences are improvements over the standard approach (lower bias and better coverage) can not be determined, however. Further simulation or theoretical work is needed necessary to establish this more definitively. Nonetheless, this application showed that the method can be

implemented successfully and shows some potential benefits over other approaches.

### **6.3 ADVANTAGES OF A MULTIVARIATE APPROACH**

Conducting a multivariate meta-analysis when the effect of a treatment on a number of outcomes is of interest can have several advantages. A single model can produce summary estimates for all outcomes in one step, instead of performing separate meta-analyses for each. More importantly, multivariate hypotheses about the outcomes (e.g., comparability of effect sizes) can be examined<sup>5</sup>. Although the multivariate model was initially expected to lead to potential improvements in accuracy or precision of summary estimates<sup>1,5</sup>, applications that have appeared so far have shown little difference compared to the univariate approach.

This was not the case, however, in the meta-analysis of the effect of DBS, where the multivariate meta-analysis produced slightly more precise summary estimates, particularly in the last interval which had the fewest observations. This model was also less influenced by an apparently outlying observation in this same interval because of which a standard meta-analysis of the data in the last interval was producing weaker summary effect estimates. In fact, I suspect that this is what led to the gain in precision from the multivariate model, since the lower bounds of the confidence intervals were similar but differed significantly at the upper bound. The impact of the outlying observation is limited in the longitudinal models as these take into account the strong correlation between the outlying effect and others from the same study included in the meta-analysis. Since correlation implies less information, the outlying effect is less influential in models that reflect the repeated nature of the data, particularly in the DBS example where the within-study correlations were very strong.

Accounting for correlations between random-effects is another mechanism through which the multivariate model may possibly produce more precise estimates. Standard time-specific meta-analyses allow a separate independent

random-effect for each time interval; it seems plausible, however, that underlying factors that cause heterogeneity across studies act similarly for measurements at different times within each study. Therefore, imposing a correlation between the random-effects, as in the multivariate and random-time-effects models, might account for this and provide a more accurate measure of heterogeneity. This occurred in the DBS example only when controlling for mean disease duration and mean baseline scores, and while this occurred in all models, the reduction was greatest in the multivariate model. Furthermore, the confidence intervals of summary estimates were slightly narrower than those from the multivariate model without covariates. Thus, controlling for systematic sources of heterogeneity reduced the variances of the random-effects, which in turn improved the precision of summary estimates.

The advantages described for the multivariate model were also observed for the random study-effects and random time-effects models, which account for the marginal or total correlations in the data. The multivariate model performed slightly better than these in terms of fit and precision in the last interval. The differences were not substantial in the case study I explored, however, so that the additional benefit of modeling within- and between-study correlations separately is not clear. Assuming independence within-study and imposing a correlation on the random-effects (which is the random time-effects model) yielded comparable results, with the exception of slightly higher between-study variances when controlling for study characteristics. This is consistent with findings in Chapter 4, which showed that assuming within-study covariances were 0 did not alter results substantially compared to analyses using the observed covariances.

Although I could not confirm that the multivariate estimates in the longitudinal example were more accurate and confidence intervals had better coverage probabilities, both of the *apparent* advantages suggest a potential borrowing of strength across data from different outcomes (or times in the longitudinal case). This would clearly not be possible with separate time- or outcome-specific analyses, since observations collected from a given study are not *linked* in this

approach as is done through the correlation structures imposed in the multivariate model.

## 6.4 LIMITATIONS OF ANALYSES

Some limitations of the analyses presented in this dissertation warrant some discussion to delineate the scope of the conclusions. Two of the manuscripts included in this thesis were based on simulations; therefore, the generalizability of the findings depends to some extent on the way the data were generated and situations that were examined.

In the first manuscript, conceptual model incorporating dependencies between events, treatment effects and between-study random-effects was formulated to examine the correlations measured between outcomes across studies. This conceptual model is likely a simplification of how events and effects are related in practice but served as a useful framework to parameterize the data generating procedure. In light of the findings, this simplification is likely conservative, as distortions in the correlations are prone to be more severe in a more complex conceptual model. I focused specifically on dichotomous endpoints and used conditional specifications of the event rates and treatment effects (parameterized as odds-ratios) to build in dependencies. I did not explore whether the results depended on the choice of treatment effect measure - using conditional risk differences instead of conditional odds-ratios, for instance; or, whether correlations between treatment effects would be better measured when the endpoints were continuous.

The simulations to assess the impact of errors in specifying within-study covariances (second manuscript) were based solely on continuous endpoints. The motivation for this was primarily to simplify the data generating procedure. Within-study covariances of effect estimates on continuous endpoints can be derived by generating patient-level data and calculating empirically from the covariance between responses within patients. With two dichotomous endpoints, on the other hand, the covariance between odds-ratios for the two endpoints can

not be derived from patient-level data directly. In fact, a joint analysis (repeated-measures logistic regression) of each simulated trial data would have to be conducted to obtain the covariance matrix of the effect estimates. This does not completely limit the findings to meta-analyses of continuous endpoints, however, since meta-analyses of dichotomous (or time-to-event endpoints) are typically conducted using the same model as the one evaluated in the simulation (the model is fitted to log-OR or log-hazard-ratio estimates). Thus, findings pertaining to the performance of the model should apply to both situations.

The scenarios examined in this manuscript were targeted to situations where the potential impact of errors in specifying within-study covariances would likely be most apparent. Thus, some potentially influential parameters may have been left out unknowingly. Considering the consistency of findings across all the scenarios examined, it is plausible that the conclusions from these simulations will hold in other situations.

Although the application of multivariate models in the meta-analysis of longitudinal effect estimates confirmed the feasibility of using the method in this context, because the data were real rather than simulated, and hence the truth unknown, I could not verify that the differences observed compared to the standard approach were in fact beneficial - that is, whether the stronger summary estimate in the last interval measured from the multivariate model is actually more accurate; or, whether the narrower confidence interval has at least as good or better coverage probabilities as that from the standard model.

Due to the limited size of the available data in the DBS example, choices of correlation structures for within- and between-study covariance matrices were limited. First-order auto-regressive or compound symmetry covariance matrices were used as they involve only a single correlation parameter. The estimates from the model were very similar for these two approaches; it not clear, however, whether results would not be different if an unrestricted (positive-definite) matrix or a Toeplitz correlation structure could be implemented, for instance. The

estimation algorithm did not converge when the models were fitted with these structures.

## 6.5 SUGGESTIONS FOR FUTURE RESEARCH

I have identified both advantages and shortcomings of multivariate meta-analyses that should be investigated further. The meta-analysis of the effect of DBS suggested that the multivariate approach can potentially provide more precise estimates by reducing the impact of extreme observations and possibly through other mechanisms. Whether these differences lead to more accurate estimates with maintained or improved coverage probabilities could not be confirmed, however, since the true values of the parameters of the model were not known. Simulations studies would be helpful to determine if the findings in my case study were due to particularities of the data or a general property of the model.

I also found that multivariate meta-analyses do not quantify the relationship between treatment effects reliably; the correlations that are measured are not necessarily representative of the underlying association. This may be because meta-analyses are usually based on effect estimates found in study reports; these can be thought of as ecological measures, and hence some of the underlying associations may be distorted or hidden. It would, therefore, be interesting to investigate whether one could do better if actual patient-level data were available. This would allow more complex models to be fitted in which patient-level and study-level sources of correlations could be better isolated. Obtaining patient-level data is currently possible but difficult, but may become increasingly more feasible through trial registries (e.g., [www.clinicaltrials.gov](http://www.clinicaltrials.gov)) and policies aiming to ensure reproducible research.

Multivariate meta-analysis models can have interesting applications in situations not yet considered in this thesis or the literature to date. One such application may be for the analysis of composite endpoints to capture the *overall* effect of a treatment. Typically, the meta-analyses of such outcomes would be based on effect estimates reported for a specifically defined composition of the endpoint

(e.g., headache relief within 4 hours without adverse events). As a result, studies that only reported the components or used similar but not identical definitions of the composite cannot be included. A multivariate meta-analysis based on the components could provide added insight and allow a more inclusive review. It may be necessary to extend the usual normal distribution based approach, however; for example, a multinomial model for the joint distribution of the components of the composite of dichotomous endpoints may be better suited as it would allow summarizing the effect of treatment on various combinations of the two components including the composite definition. For instance, if studies report headache relief, adverse events and the combined endpoint headache relief without adverse events, the event counts corresponding to all possible pairs of the two outcomes (no relief/no adverse event, relief without adverse event, etc) could be modeled as a multinomial outcome. This would allow joint inferences about the effect of the treatment on the components as well as the composite of the endpoints.

One particularity of the use of multivariate models in meta-analysis is that these methods are applied to perform joint meta-analyses of outcomes that will have usually been analyzed independently (e.g., verbal and math scores on SATs<sup>1</sup>) in the studies included in the review. It would be interesting to examine whether multivariate models of correlated outcomes have advantages in the analysis of patient data in a given study, and whether the limitations observed in the estimation of correlations between treatment effects apply in this case as well. I suspect that they do, as the conceptual model used in Chapter 3 can be interpreted in the context of patient level data. Further study would be helpful to verify this hypothesis.

## 6.6 SUMMARY OF CONCLUSIONS

Multivariate meta-analyses models, at least in theory, can be a useful tool as they potentially allow inferences about the relationships between the outcomes being studied. This appears to be the primary advantage of this method; previous work has failed to show benefits in terms of added accuracy or precision. Indeed,

considering that only the shared or common “information” about the outcomes can potentially influence the results, a joint analysis does not incorporate new information in deriving summary estimates for each outcome. Therefore, it is not surprising that univariate and multivariate meta-analyses typically produce very similar results. Using multivariate models for longitudinal effect estimates in the DBS example suggested that a potential for gain in precision may exist in some specific situations (in the presence of an extreme observation and high within-study correlations).

The robustness of multivariate meta-analyses to poor approximations of within-study covariances when these are not reported (as is usually the case) is also encouraging. In fact, assuming independence within-studies and accounting for correlations through random-effects appears to have no impact on the accuracy and precision of summary estimates. The correlations between outcomes were not as reliably estimated, however. Furthermore, the correlations measured across studies may not always be representative of the underlying relationship between treatment effects. Thus, multivariate meta-analyses should not be undertaken with the aim of quantifying the correlation between the effects of a treatment on different outcomes.

## REFERENCES

---

- <sup>1</sup> Raudenbush SW, Becker BJ, Kalaian H. Modeling multivariate effect sizes. *Psychological Bulletin*, **103**(1):111-20 (1988).
- <sup>2</sup> Staessen JA, Wang J, Thijs L. Cardiovascular prevention and blood pressure reduction: a quantitative overview updated until 1 March 2003. *Journal of Hypertension*, **21**(6):1055–1076 (2003).
- <sup>3</sup> Pocock SJ. Clinical trials with multiple outcomes: a statistical perspective on their design, analysis, and interpretation. *Controlled Clinical Trials*, **18**:530-545 (1997).
- <sup>4</sup> Pocock SJ, Geller NL, Tsiatis AA. The analysis of multiple endpoints in clinical trials. *Biometrics*, **43**:487- 498 (1987).
- <sup>5</sup> Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F, Colditz GA. Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine*, **17**:2537-50 (1998).
- <sup>6</sup> Arends LR, Voko Z, Stijnen T. Combining multiple outcome measures in a meta-analysis: an application. *Statistics in Medicine*, **22**:1335-53 (2003).
- <sup>7</sup> DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials*, **7**, 177-88 (1986).
- <sup>8</sup> Berkey CS, Anderson JJ, Hoaglin DC. Multiple outcome meta-analysis of clinical trials. *Statistics in Medicine*, **15**:537-57 (1996).
- <sup>9</sup> Nam I, Mengersen K, Garthwaite P. Multivariate meta-analysis. *Statistics in Medicine*, **22**:2309-33 (2003).
- <sup>10</sup> Dear KBG. Iterative generalized least squares for meta-analysis of survival data at multiple times. *Biometrics*, **50**:989-1002 (1994).

- 
- <sup>11</sup> Van Houwelingen HC, Zwinderman K, Stijnen T. A bivariate approach to meta-analysis. *Statistics in Medicine*, **12**:2272-84 (1993).
- <sup>12</sup> Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine*, **21**:589-624 (2002).
- <sup>13</sup> McIntosh MW. The population risk as an explanatory variable in research synthesis of clinical trials. *Statistics in Medicine*, **15**:1713-28 (1996).
- <sup>14</sup> Whitehead A. Meta-analysis of Controlled Clinical Trials. New York: John Wiley & Sons, Ltd. (2002).
- <sup>15</sup> Glass GV. Primary, secondary and meta-analysis of research. *Education Researcher*, **5**, p.3–8 (1976).
- <sup>16</sup> Cornell J, Mulrow C. Meta-Analysis. In *Research Methodology in the Social, Behavioural & Life Sciences*, Herman J. Adér and Gideon J. Mellenbergh editors. Great Britain: Sage Publications, 1999.
- <sup>17</sup> Eggar M, Davey Smith G, Altman DG (editors). *Systematic Reviews in Health Care: Meta-analysis in Context*. London: BMJ Books (2001).
- <sup>18</sup> Normand ST. Tutorial in Biostatistics: Meta-analysis: Formulating, evaluating, combining and reporting. *Statistics in Medicine*, **18**:321-59 (1999).
- <sup>19</sup> Hedges LV, Olkin I. *Statistical Methods for Meta-Analysis*. Orlando, Fl. Academic, 1985.
- <sup>20</sup> Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *British Medical Journal*; **327**:57-560 (2003).
- <sup>21</sup> Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine*, **21**:1575-1600 (2002).
- <sup>22</sup> Schmid CH, Lau J, McIntosh MW, Cappelleri JC. An empirical study of the effect of control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Statistics in Medicine*, **17**:1923042 (1998).

- 
- <sup>23</sup> Warn DE, Thompson SG, Spiegelhalter DJ. Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales. *Statistics in Medicine*; **21**:1601-23 (2002).
- <sup>24</sup> Cochran WG. The combination of estimates from different experiments. *Biometrics*, **10**:101-129 (1954).
- <sup>25</sup> Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine*; **17**:841- 856 (1998).
- <sup>26</sup> Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, **21**:1539-58 (2002).
- <sup>27</sup> Baujat B, Mahe C, Pignon J, Hill C. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. *Statistics in Medicine*, **21**:2641-52 (2002).
- <sup>28</sup> Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine*, **18**:2693-2708 (1999).
- <sup>29</sup> Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. *Statistics in Medicine*, **15**:619-629 (1996).
- <sup>30</sup> Mosteller F, Colditz GA. Understanding research synthesis (meta-analysis). *Annual Review of Public Health*, **17**:1-23 (1996).
- <sup>31</sup> Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Statistics in Medicine*, **20**:825-40 (2001).
- <sup>32</sup> Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Statistics in Medicine*, **14**:2685-99 (1995).
- <sup>33</sup> Greenland S. Invited commentary: A critical look at some popular meta-analytic methods. *American Journal of Epidemiology*, **140**(3) :290-6 (1994).
- <sup>34</sup> Thompson SG. Systematic Review: Why sources of heterogeneity in meta-analysis should be investigated. *British Medical Journal*, **309**:1351-55 (1994).
- <sup>35</sup> Glasziou PP, Sanders SL. Investigating causes of heterogeneity in systematic reviews. *Statistics in Medicine*, **21**:1503-11 (2002).

- 
- <sup>36</sup> Thompson SG, Higgins JPR. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, **21**:1559-73 (2002).
- <sup>37</sup> Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in Medicine*, **10**:1665-77 (1991).
- <sup>38</sup> Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. *Statistics in Medicine*, **14**:395-411 (1995).
- <sup>39</sup> Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random-effects. *Statistics in Medicine*, **15**:619-29 (1996).
- <sup>40</sup> Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine*, **19**:3417-32 (2000).
- <sup>41</sup> Carlin JB. Meta-analysis for  $2 \times 2$  tables: a Bayesian approach. *Statistics in Medicine*, **11**:141-58 (1992).
- <sup>42</sup> Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis (Second Edition)*. Chapman & Hall/CRC, Boca Raton, 2004.
- <sup>43</sup> Spiegelhalter D, Thomas A, Best N, Gilks W, Lunn D. BUGS: Bayesian inference using Gibbs sampling. MRC Biostatistics unit, Cambridge, England. [www.mrc-bsu.cam.ac.uk/bugs/](http://www.mrc-bsu.cam.ac.uk/bugs/)
- <sup>44</sup> Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, **22**:719-48 (1959).
- <sup>45</sup> Platt RW, Leroux BG, Breslow N. Generalized linear mixed models for meta-analysis. *Statistics in Medicine*, **18**:643-54 (1999).
- <sup>46</sup> Breslow N. Regression analysis of the log odds ratio: a method for retrospective studies. *Biometrics*, **32(2)**:409-16 (1976).
- <sup>47</sup> Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88(421)**:9-25 (1993).

- 
- 48 McIntosh MW. The population risk as an explanatory variable in research synthesis of clinical trials. *Statistics in Medicine*, **15**:1713-28 (1996).
- 49 Korolija D, Tadic S, Simic D. Extent of oncological resection in laparoscopic vs. open colorectal surgery: meta-analysis. *Langenbecks Archives of Surgery*, **387**:366-71 (2003).
- 50 Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine*, **23**:3105-24 (2004).
- 51 Senn S, Stevens L, Chaturvedi N. Repeated measures in clinical trials: simple strategies for analysis using summary measures. *Statistics in Medicine*, **19**:861-77 (2000).
- 52 Lopes HF, Muller P, Rosner GL. Bayesian meta-analysis for longitudinal data models using multivariate mixture priors. *Biometrics*, **59**:66-75 (2003).
- 53 Farlow MR, Small GW, Quarg P, Krause A. Efficacy of Rivastigmine in Alzheimer's disease patients with rapid disease progression: Results of a meta-analysis. *Dement Geriatr Cogn Disord* **20**:192-7 (2005).
- 54 Maas CJM, Hox JJ, Lensveld-Mulders GJLM. Longitudinal meta-analysis. *Quality & Quantity*, **38**:381-9 (2004)
- 55 Demidenko E. *Mixed Models. Theory and Applications*. Hoboken, NJ: Wiley Series in Probability and Statistics, 2004
- 56 Cnaan A, Laird NM, Slasor P. Using the general linear mixed model to analyze unbalanced repeated measures and longitudinal data. *Statistics in Medicine*, **16**:2349-80 (1997).
- 57 Hutton JL, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. *Journal of the Royal Statistical Society (Series C)*, **49(3)**: 359-70 (2000).
- 58 Rothman KJ, Greenland S. *Modern Epidemiology* (Second Edition), Lippincott Williams & Wilkins, Philadelphia PA, 1998.
- 59 Greenland S, Brumback B. An overview of relations among causal modeling methods. *International Journal of Epidemiology* 2002; **31**:1030-7.

- 
- 60 Kannel W. Risk stratification in hypertension: new insights from the Framingham study. *American Journal of Hypertension*, **13**:3S-10S (2000).
- 61 Littell RC, Pendergast J, Natarajan R. Modeling covariance structures in the analysis of repeated measures data. *Statistics in Medicine*, **19**:1793-1819 (2000).
- 62 Ishak KJ, Proskorovsky I, Caro I, Caro JJ. A meta-analysis of the effect of deep brain stimulation of the subthalamic nucleus in Parkinson's disease. *Submitted for publication.*
- 63 Dodel R, Berger K, Oertel W. Health-related quality of life and healthcare utilization in patients with Parkinson's disease. Impact of motor fluctuations and dyskinesias. *Pharmacoeconomics*, **19**:1013-1038 (2001).
- 64 Dewey RB. Management of motor complications in Parkinson's disease. *J Neurology*, **62**(Suppl 4):S3-7 (2004).
- 65 Krack P, Pollak P, Limousin P, Benazzouz A, Deuschl G, Benabid AL. From off-period dystonia to peak-dose chorea. The clinical spectrum of varying subthalamic nucleus activity. *Brain*, **122**:1133-1146 (1999).
- 66 Benabid A, Koudsié A, Benazzouz A, et al. Subthalamic stimulation for Parkinson's disease. *Archives of Medical Research*, **31**:282-289 (2000).
- 67 Kumar R, Lozano AM, Montgomery E, Lang AE. Pallidotomy and deep brain stimulation of the pallidum and subthalamic nucleus in advanced Parkinson's disease. *Movement Disorders*, **13** Suppl 1:73-82 (1998).
- 68 Fahn S, Elton RL, the UPDRS Development Committee. Unified Parkinson's Disease Rating Scale. In: Fahn S , Marsden CD , Calne D , Goldstein M , editors. *Recent developments in Parkinson's disease*. Vol **2**. Florham Park, NJ: Macmillan Healthcare Information; 1987. p 153-163.
- 69 SAS Institute Inc. SAS OnlineDoc®, Version 8, February 2000, Copyright ©2000, SAS Institute Inc.
- 70 Liang KY, Zeger SL. Regression analysis for correlated data. *Annual Review of Public Health*, **14**:43-68 (1993).