

University of Southern Queensland

**Statistical Methodology for Ordinal  
Data in Meta-Analysis**

A Dissertation submitted by

**Md. Belal Hossain**

B.Sc.(Hons), M.Sc.

For the degree of

**Doctor of Philosophy**

2011

# Abstract

Meta-analysis combines results from several independent studies. Different methods are available to carry out meta-analyses for binary and continuous outcomes. The effect measures used for binary outcomes are odds ratio (OR), relative risk (RR), risk difference (RD), arcsine difference (AS), hazard ratio (HR) etc. For continuous outcomes mean difference (MD) and standardised mean difference (SMD) are used in meta-analysis. However, there are many medical and health studies in which the outcome variables are measured on an ordinal categorical scale with more than two categories. These categories are non-numerically valued, usually levels. In a typical ordinal categorical data there may be  $L$  categories  $C_1, C_2, \dots, C_L$  ( $C_1$  is the best and  $C_L$  the worst or vice versa) and  $J$  comparison groups  $G_1, G_2, \dots, G_J$ . Hence the count data for such studies are represented by a  $J \times L$  contingency table. As a special case when there are two comparison groups in randomised controlled trials (RCTs), we set a  $2 \times L$  contingency table. As a result, the ordinary OR, log OR or RR can not be used directly without splitting the  $2 \times L$  ( $L > 2$ ) contingency table into  $2 \times 2$  tables.

Among other effect measures for ordinal data there are local and global odds ratios (Dale, 1984), cumulative odds ratios, continuation odds ratio (Agresti, 2010) etc. The local odds ratio measures local association for a specific outcome category not for the whole study. The global odds ratio is

a measure of ratios of the quadrant probabilities ( $J = L \geq 4$ ). Whereas in RCTs there are only two comparison groups ( $J = 2$ ) namely the treatment and control groups. The cumulative odds ratios provide a comparison of pairs of levels of the explanatory variable with respect to their entire conditional distribution of the dependent variable. As a result, these measures are not appropriate in meta-analysis with RCTs.

The data from studies with several ordered categories are analysed by various methods in meta-analysis. Some methods require specific model assumptions while others collapse the  $2 \times L$  ( $L > 2$ ) contingency table into  $2 \times 2$  tables for measuring the effect size. For example, the proportional odds model (Whitehead et al., 2001) requires a proportionality assumption and there is no well defined variance estimate of the pooled estimator for the sample size weight method (Edwardes and Baltzan, 2000) that uses general odds ratio ( $OR_G$ ) as an effect measure.

Therefore we need a method in meta-analysis that can be used for estimating the effect size without any loss of information by merging categories and is not restricted to any model assumptions.

We propose generalised odds ratio (GOR) as an effect measure for ordinal categorical outcomes in meta-analysis (Agresti, 1980). For confidence intervals (CI) of the individual study effects and meta-analysis we employ independent multinomial distribution approach. A general fixed and a random effects models are developed using GOR in meta-analysis for ordinal categorical outcomes.

Heterogeneity is one of the most problematic aspects in many meta-analyses. We have demonstrated a method to remedy the problem of heterogeneity in meta-analysis for ordinal data. Following Saleh (2006) a quasi-empirical Bayes method (QEBM) is developed using predicted generalised

odds ratio (PGOR) for heterogeneous ordinal categorical outcomes. This method identifies the extreme studies and improves the meta-analysis in the presence of heterogeneity. Three different meta-analyses on several studies with different degree of heterogeneity are presented. The first example is of individual patients data (IPD) on tacrine trials with Alzheimer's disease, the second example is of misoprostol trials with insignificant heterogeneity and the third example is from simulation studies with significant heterogeneity. The three examples clearly illustrate detailed implementation process and usefulness of the proposed method.

We apply and compare GOR with OR as an effect measure for binary outcomes in meta-analysis. Three alternative methods for combining results from binary outcomes are presented for meta-analysis. The first method is a sample size weight method (Edwardes and Baltzan, 2000) for binary outcomes using  $OR_G$ . The other two methods employ GOR as an effect measure for binary outcomes in meta-analysis. We present results by analysing six RCTs from meta-analysis of D1 versus D2 gastrectomy for gastric adenocarcinoma (Memon et al., 2011).

This study also proposes GOR as an effect measure and presents method in meta-analysis for latent continuous outcomes. GOR is simple and it has straightforward interpretation. It can be used for more than two treatment groups as well. Hence GOR is a very useful effect measure in meta-analysis not only for multilevel ordinal categorical outcomes but also for binary and latent continuous outcomes.

# Certification of Dissertation

I certify that the ideas, experimental work, results, analyses, software and conclusions reported in this dissertation are entirely my own effort, except otherwise acknowledged. I also certify that the work is original and has not been previously submitted for any other award, except where otherwise acknowledged.

.....

.....

Signature of Candidate

Date

## ENDORSEMENT

.....

.....

Signature of Supervisor

Date

# Acknowledgements

First of all I express my sincere gratitude to Almighty for the completion of this thesis.

I sincerely thank my reverend principal supervisor, Dr Shahjahan Khan, for introducing the research topic, well structured direction, and continuous encouragement to work on this research project. This thesis would not have been completed without his expert guidance, generous support, flexible working time for me and motivation. I thank him for his academic support, arranging financial assistance and providing collaboration opportunities with his research associates in Queensland Health. I am highly benefitted from his excellent professional leadership in enriching my academic experience and collaboration. I also thank his family for kind support from the very beginning in my stay in Australia.

I would like to thank my co-supervisor, Dr Muhammed Ashraf Memon for his invaluable comments and suggestions on this thesis and other projects we worked together.

Special thanks to the University of Southern Queensland and University of Dhaka, Bangladesh for supporting my PhD study, Dept. of Maths & Computing for providing opportunities to work as a tutor/marker, and Australian Centre for Sustainable Catchments (ACSC) for sponsoring my conference travels.

I would like to thank Dr Richard Watson and Dr Stijn Dekeyser, the Heads of Department of Mathematics & Computing, University of Southern Queensland, for offering me teaching related contracts. Those really helped me to enrich my academic career. I also thank Dr Ashley Plank, Ms Christine McDonald and Ms Taryn Swan for welcoming me as a team member of the statistics group. It was a pleasure working with the Data Analysis teaching team and a huge number of students and sharing ideas on social forum. Do you know ‘How many eggs there should be in the hearty breakfast when you finish the exams?’ I do.

My sincere thanks go to the administrative staff of the Department of Maths & Computing, especially Helen Nkansah and Kris Lyon, the Office of the Research and Higher Degrees, the Faculty of Sciences, the Library, the ICT services for quick technical and administrative assistances.

Finally, big applause to my family members who sat me free for three years and took all the responsibilities on themselves to enable me to complete this thesis. Thank you perhaps have no use to them but thank you any way.

Last but not the least, I express my gratitude to my colleagues in University of Dhaka.

# List of Abbreviations

|                 |                                  |
|-----------------|----------------------------------|
| CI              | confidence interval              |
| df              | degrees of freedom               |
| GOR             | generalised odds ratio           |
| LOR             | log odds ratio                   |
| MLE             | maximum likelihood estimator     |
| OR              | odds ratio                       |
| OR <sub>G</sub> | general odds ratio               |
| PGOR            | predicted generalised odds ratio |
| PTE             | preliminary test estimator       |
| QEBM            | quasi-empirical Bayes method     |
| RCTs            | randomised controlled trials     |
| RR              | relative risk                    |

# Contents

|   |             |
|---|-------------|
| <b>Abstract</b>                               | <b>i</b>    |
| <b>Certification of Dissertation</b>          | <b>iv</b>   |
| <b>Acknowledgements</b>                       | <b>v</b>    |
| <b>List of Abbreviations</b>                  | <b>vii</b>  |
| <b>List of Figures</b>                        | <b>xi</b>   |
| <b>List of Tables</b>                         | <b>xiii</b> |
| <b>1 Introduction</b>                         | <b>1</b>    |
| 1.1 Background and Motivation . . . . .       | 2           |
| 1.2 Generalised Odds Ratio (GOR) . . . . .    | 5           |
| 1.3 Contribution . . . . .                    | 7           |
| 1.4 Thesis Outlines . . . . .                 | 8           |
| <b>2 Methods for Ordinal Categorical Data</b> | <b>9</b>    |
| 2.1 Introduction . . . . .                    | 10          |
| 2.2 Review . . . . .                          | 11          |
| 2.2.1 Effect Measures . . . . .               | 11          |
| 2.2.2 Issues of Meta-Analysis . . . . .       | 14          |

|  |           |
|--|-----------|
| <i>CONTENTS</i>  | ix        |
| 2.2.3 Proportional Odds Model . . . . .                                    | 20        |
| 2.2.4 Sample Size Weight Method . . . . .                                  | 21        |
| 2.3 Proposed Method: Meta-Analysis using GOR . . . . .                     | 22        |
| 2.3.1 GOR . . . . .  | 22        |
| 2.3.2 GOR under Fixed Effects Model . . . . .                              | 25        |
| 2.3.3 GOR under Random Effects Model . . . . .                             | 26        |
| 2.4 Application . . . . .  | 27        |
| 2.5 Summary . . . . .  | 31        |
| <b>3 A QEBM for Heterogeneous Ordinal Data</b>                             | <b>33</b> |
| 3.1 Introduction . . . . .   | 34        |
| 3.2 Quasi-Empirical Bayes Method for GOR . . . . .                         | 35        |
| 3.3 Improved Meta-Analysis . . . . .                                       | 39        |
| 3.4 Applications . . . . .   | 40        |
| 3.4.1 Tacrine Trials: Insignificant Heterogeneity . . . . .                | 41        |
| 3.4.2 Misoprostol Trials: Moderately Insignificant Heterogeneity . . . . . | 43        |
| 3.4.3 Simulated Trials: Significant Heterogeneity . . . . .                | 46        |
| 3.5 Conclusion . . . . .   | 50        |
| <b>4 GOR for Binary Outcomes: A Comparison</b>                             | <b>53</b> |
| 4.1 Introduction . . . . .   | 54        |
| 4.2 GOR: As Binary Effect Measure . . . . .                                | 54        |
| 4.3 Review of Methods using OR and $OR_G$ . . . . .                        | 55        |
| 4.3.1 Fixed Effects Model . . . . .  | 57        |
| 4.3.2 Random Effects Model . . . . .                                       | 58        |
| 4.3.3 Sample Size Weight Method . . . . .                                  | 59        |
| 4.4 Proposed Method . . . . .  | 60        |

|  |           |
|--|-----------|
| <i>CONTENTS</i>  | x         |
| 4.5 Results . . . . .                                    | 62        |
| 4.6 Improved Meta-Analysis . . . . .                     | 66        |
| 4.7 Summary . . . . .                                    | 70        |
| <b>5 GOR for Continuous Outcomes</b>                     | <b>72</b> |
| 5.1 Introduction . . . . .                               | 73        |
| 5.2 Review . . . . .                                     | 74        |
| 5.2.1 Method 1: Continuous vs Continuous . . . . .       | 74        |
| 5.2.2 Method 2: Categorical vs Continuous . . . . .      | 76        |
| 5.3 Method 3: Categorical vs Latent Continuous . . . . . | 78        |
| 5.4 Example . . . . .                                    | 80        |
| 5.5 Summary . . . . .                                    | 81        |
| <b>6 Conclusion and Future Work</b>                      | <b>84</b> |
| <b>Bibliography</b>                                      | <b>87</b> |
| <b>Appendices</b>  | <b>94</b> |
| <b>A</b>   | <b>95</b> |
| A.1 Delta Method . . . . .                               | 95        |
| A.2 The Proportional Hazard Model . . . . .              | 96        |
| A.3 Ordered Logistic Regression . . . . .                | 97        |
| A.4 Odds Ratio . . . . .                                 | 97        |
| A.5 Relative Risk . . . . .                              | 98        |
| A.6 Risk Difference . . . . .                            | 99        |
| A.7 Arcsine Difference . . . . .                         | 99        |
| A.8 Continuity Corrections for Zero Cells . . . . .      | 100       |
| A.9 Peto's Method . . . . .                              | 100       |

|   |            |
|---|------------|
| A.10 Mantel-Haenszel Method . . . . .                           | 101        |
| A.11 Estimating Mean and SD from Median, Range and Sample Sizes | 102        |
| A.12 Kendall's tau . . . . .                                    | 103        |
| A.13 Mann-Whitney U Statistic . . . . .                         | 104        |
| <b>B</b>  | <b>105</b> |
| B.1 R Code for Chapter 2 . . . . .                              | 105        |
| B.2 Sample Size Weight Method . . . . .                         | 107        |
| B.3 R Code for Chapter 3 . . . . .                              | 109        |
| B.4 Misoprostol Trials . . . . .                                | 114        |
| B.5 Simulation Studies . . . . .                                | 119        |
| B.6 R Code for Chapter 4 . . . . .                              | 132        |
| B.7 R Code for Chapter 5 . . . . .                              | 145        |
| <b>C</b>  | <b>152</b> |
| C.1 Publications . . . . .                                      | 152        |

# List of Figures

|     |  |    |
|-----|--|----|
| 2.1 | Funnel plot indicating absence of publication bias . . . . .   | 18 |
| 2.2 | Funnel plot showing presence of publication bias . . . . .   | 19 |
| 2.3 | Forest plot of the five RCTs of tacrine using GOR . . . . .  | 30 |
| 2.4 | The funnel plot suggests that there is no publication bias . . . .   | 31 |
| 3.1 | 95% CIs for GOR (solid horizontal lines) and predicted GOR<br>(dashed horizontal lines) for tacrine trials . . . . .   | 42 |
| 3.2 | 95% CIs for GOR (solid horizontal lines) and predicted GOR<br>(dashed horizontal lines) for the randomised trials of misopros-<br>tol by endoscopic classification . . . . . | 45 |
| 3.3 | The horizontal solid lines represent CIs of GOR and dashed-<br>lines CIs of predicted GOR . . . . .  | 48 |
| 3.4 | Display of original GOR and predicted GOR from ten simu-<br>lated heterogeneous trials . . . . .   | 49 |
| 3.5 | Display of GOR and predicted GOR values after deleting Study<br>4 . . . . .  | 50 |
| 3.6 | 95% CIs after deleting Study 4 . . . . .   | 51 |
| 4.1 | Ray plot for GOR of the six studies of gastric carcinoma . . . .   | 67 |
| 4.2 | 95% CI for GOR of the six studies of gastric carcinoma . . . .   | 68 |

|     |  |    |
|-----|--|----|
| 4.3 | Display of GOR and PGOR values of five studies for gastric carcinoma . . . . . | 69 |
| 4.4 | 95% CIs for GORs and PGORs of the five studies of gastric carcinoma . . . . .  | 70 |
| 5.1 | Forest plot from simulated data using GOR . . . . .                            | 82 |

# List of Tables

|     |  |     |
|-----|--|-----|
| 2.1 | Data of a single RCT . . . . .   | 12  |
| 2.2 | Contingency table for the $i$ th study. . . . .                        | 23  |
| 2.3 | Estimate of LOR and LGOR along with SE for tacrine data . . . . .      | 28  |
| 2.4 | Comparison of meta-analysis estimates for the tacrine data . . . . .   | 29  |
| 3.1 | Five randomised trials of anti-cholinesterase drug tacrine . . . . .   | 43  |
| 3.2 | Randomised trials of misoprostol by endoscopic classification. . . . . | 44  |
| 3.3 | Ten simulated heterogeneous studies. . . . .                           | 47  |
| 3.4 | Predicted GOR of nine trials deleting Study 4 . . . . .                | 51  |
| 4.1 | Data of a single RCT . . . . .   | 55  |
| 4.2 | OR for total postoperative complications . . . . .                     | 63  |
| 4.3 | $OR_G$ for total postoperative complication . . . . .                  | 64  |
| 4.4 | GOR for total postoperative complications . . . . .                    | 65  |
| 4.5 | Predicted GOR of six studies of gastric carcinoma . . . . .            | 67  |
| 5.1 | Contingency table for Job Satisfaction by Income . . . . .             | 80  |
| 5.2 | Five simulated studies . . . . .                                       | 83  |
| A.1 | Data of a single RCT . . . . .   | 97  |
| A.2 | The best estimating formulas for mean and sd . . . . .                 | 103 |

# Chapter 1

## Introduction

## 1.1 Background and Motivation

There has been an increasing interest in the development of appropriate measures to ensure that public policy and decision making are based on results of reliable research. The evidence based scientific research has been helping decision makers to determine which interventions are doing good and which are actually harmful, particularly in the health care area. Meta-analysis is a statistical technique which concerns with the analysis of the data extracted from independent studies. It also estimates overall measures of association or effect size and assesses the sensitivity of the results.

History of meta-analysis goes as early as in the beginning of the 20th century when Pearson (1904) developed a statistical technique for making sense of the divergent results from small studies of the effectiveness of inoculation against typhoid fever. Statistical techniques were also used for combining study results in agriculture (Yates and Cochran, 1938) and in the medical area (Beecher, 1995).

Chalmers et al. (1977) conducted one of the first meta-analyses in medicine in the modern era. Then in the mid-1980s meta-analysis started to be used more frequently when Yusuf et al. (1985) published their meta-analysis of beta blockers in myocardial infarction.

Different methods are available to carry out meta-analyses for binary and continuous outcomes. The effect measures used for binary outcomes are odds ratio (OR), relative risk (RR), risk difference (RD), arcsine difference (AS), hazard ratio (HR) etc. For continuous outcomes mean difference (MD) and standardised mean difference (SMD) are widely used in meta-analysis. However, there are many medical and health studies in which the outcome variables are measured on an ordinal categorical scale with more than two categories. These categories are usually levels. For example, study on pain

relief, dementia, head injury, tonsil size etc are measured on ordinal scales with more than two outcomes. In a typical ordinal categorical data there may be  $L$  categories  $C_1, C_2, \dots, C_L$  ( $C_1$  is the best and  $C_L$  the worst or vice versa) and  $J$  comparison groups  $G_1, G_2, \dots, G_J$ . Hence the count data for such studies are represented in a  $J \times L$  contingency table. As a special case in randomised controlled trials (RCTs) when there are two comparison groups (treatment and placebo), a  $2 \times L$  contingency table is used. As a result, the ordinary OR, log OR or RR can not be used directly without splitting the  $2 \times L$  ( $L > 2$ ) contingency table into a number of  $2 \times 2$  tables.

Clayton (1974) introduced some odds ratio statistics for the analysis of ordered categorical data assuming a logit model. Later Clayton (1976) generalised the estimators for the case in which some observations are subject to censorship. McCullagh (1977) used paired comparisons on the ordinal variables.

Dale (1984) introduced the local and global odds ratios for measuring local and global associations for bivariate ordered responses. Unfortunately, neither of these can be used in meta-analysis for ordinal categorical data produced by RCTs, because the local odds ratio measures local relationship for a specific outcome category not for the whole study and the global odds ratio is a measure of ratios of the quadrant probabilities that incur loss of information because of merging quadrant rows and columns.

Among other ordinal measures there are cumulative odds ratios, continuation odds ratio etc (Agresti, 2010). The cumulative odds ratios provide a comparison of pairs of levels of the explanatory variable with respect to their entire conditional distribution of the dependent variable. As a result, these measures are not appropriate in meta-analysis for the data produced by the RCTs.

The data from such studies with several categories are analysed by various methods in meta-analysis. Some methods require a specific model assumption while others collapse the  $2 \times L$  ( $L > 2$ ) contingency table into  $2 \times 2$  tables for measuring the effect size. These methods inherently loss valuable information due to obvious restrictions. Therefore, we need a method that can be used for estimating the effect size without any loss of information and is not restricted to any model assumption. The generalised odds ratio (GOR) is an effect measure free from the above weaknesses. More importantly, ordinal categorical data are naturally in ascending or descending order which makes GOR more suitable effect measure over its competitors.

There are several methods used for combining findings from repeated studies. Hedges and Ingram (1985) demonstrated on various parametric and non-parametric statistical methods for meta-analysis. Sutton et al. (2000) presented various methods and discussed different important issues for meta-analysis in medical research. A more recent comprehensive update of meta-analysis methods and issues are covered in Borenstein et al. (2009). In most cases, they concentrated on the methodologies and issues for binary outcomes in meta-analysis. Now-a-days, a general fixed (Birge, 1932; Cochran, 1937) and standard random effects models (DerSimonian and Laird, 1986) are widely used for binary outcomes in meta-analysis. Some other mixed effects models and Bayesian methods are potential area of research for meta-analysis.

Recently, there has been a sharp increase in the publication of research articles on ordinal categorical data. Many authors have published books on methods for ordinal categorical data analysis. Researchers of diverse disciplines such as sociology, education, public health and wildlife ecology are using these methods frequently. However, meta-analysis with ordinal

categorical outcomes has not been paid much attention so far.

Edwardes and Baltzan (2000) proposed pooling  $\gamma$ 's ( $\gamma$  is also known as Goodman and Kruskal's  $\gamma$ ) instead of general odds ratio ( $OR_G = (1+\gamma)/(1-\gamma)$ ) using sample size weights. This method is useful to calculate the pooled effect when the individual variance estimates are not provided. However, a well defined variance estimate for the pooled effect is unavailable under this method. As a result, we can not find the CI for pooled  $OR_G$  for meta-analysis.

Whitehead et al. (2001) developed a proportional odds model on individual patient data (IPD) using the log odds ratio with a general framework for fixed and random effects models. However, this method needs a 'proportional odds' assumption for computing the effect measures.

## 1.2 Generalised Odds Ratio (GOR)

In this study, we attempt to measure the effect size for the  $2 \times L$  table as a whole without any loss of information or under any model assumption in meta-analysis with RCTs. We propose using GOR (Agresti, 1980) as an effect measure for ordinal categorical outcomes. These outcomes are naturally in an ascending or descending order which makes GOR suitable as an effect measure for these outcomes. We also develop a new meta-analysis method for ordinal categorical outcomes under both the fixed and random effects models. An application is presented using the individual patients data (IPD) of five RCTs of anti-cholinesterase drug tacrine in patients with Alzheimer's disease (Whitehead et al., 2001) and compared with other methods.

Issues of heterogeneity are addressed by computing the value of the chi-square test (Cochran, 1954) and  $I^2$  statistic (Higgins and Thompson, 2002).

Funnel plots are used to check the publication bias.

Heterogeneity is one of the most common issues of meta-analysis. Use of random effects model is not always a cure (cf. Shuster, 2010; Doi SA et al., 2011). To deal with this issue with ordinal categorical outcomes a quasi-empirical Bayes method (QEBM) (cf. Saleh, 2006, p.157) is developed using GOR. This method demonstrates a remedy to the problem of heterogeneity in meta-analysis for ordinal data. This method identifies the extreme studies and improves the meta-analysis in terms of better statistical agreement in the presence of heterogeneity.

Three different meta-analyses on several studies with different degree of heterogeneity are presented. The first example is of IPD of tacrine trials with insignificant heterogeneity, the second example is of misoprostol trials with moderately insignificant heterogeneity and the third example is from simulation studies with significant heterogeneity. The three examples clearly illustrate detailed implementation process and usefulness of the proposed method.

We apply and compare GOR as an effect measure for binary outcomes with OR and  $OR_G$  in meta-analysis. Three alternative methods are also presented for combining estimates from binary outcomes. The first method is the sample size weight method adopted from Edwardes and Baltzan, (2000). The other two methods use GOR as an effect measure and independent binomial distribution approach for estimating the variance for individual study and meta-analysis. We present results by analysing six RCTs from a meta-analysis of D1 versus D2 gastrectomy for gastric adenocarcinoma (Memon et al., 2011).

This study also shows GOR as an effective outcome measure and presents methods in meta-analysis for latent continuous outcomes. The concept of

GOR is straightforward and it has mathematically amenable variance estimates for both individual study and meta-analysis. It can be well used for more than two treatment groups as well. These properties make GOR a useful outcome measure in meta-analysis not only for multilevel ordinal categorical outcomes but also for binary and latent continuous outcomes.

### 1.3 Contribution

Main contributions of this thesis are as follows:

- Propose GOR as an effect measure for ordinal categorical outcomes for data produced by RCTs and apply it in meta-analysis. The GOR has never been used as an effect measure in meta-analysis. Employing GOR as an effect measure for multilevel ordinal categorical outcomes we can estimate the effect size without any loss of information or under the restriction of model assumption or merging multilevel outcome categories into dichotomies.
- Develop general fixed and a random effects models using GOR for ordinal categorical outcomes under independent multinomial distribution.
- Propose a quasi-empirical Bayes method (QEBM) using GOR for heterogeneous ordinal categorical outcomes in meta-analysis.
- Apply and compare GOR as an effect measure with OR for binary outcomes and derive a general fixed and random effects model using GOR for binary outcomes.
- Use GOR for continuous or latent continuous outcomes in meta-analysis.

- Write computer programs (R-codes) for all computations and plotting in this thesis.

## 1.4 Thesis Outlines

The remaining chapters of the thesis are organised as below. Chapter 2 reviews some effect measures and meta-analysis methods used for ordinal categorical outcomes. This covers a new meta-analysis method using GOR for ordered categorical outcomes under both fixed and random effects models.

A QEBM is presented using GOR for heterogeneous ordinal categorical outcomes in Chapter 3. Here the predicted GOR is used for improved meta-analysis in the presence of heterogeneity. We generate ten heterogeneous studies by simulation using multinomial distribution as there are no significant heterogeneous studies available for ordinal categorical outcomes.

Chapter 4 contains a comparison of meta-analysis methods for binary outcomes. This chapter derives a meta-analysis method using GOR for binary outcomes under general fixed and random effects models and then compare with the existing methods currently in use.

Chapter 5 deals with the continuous outcomes in meta-analysis.

In Chapter 6, we rap up by summarising the main concepts proposed and results found in this thesis.

The three appendices contain some related additional topics, R-codes used for computations in the main four chapters and list of publications.

# Chapter 2

## Methods for Ordinal Categorical Data

## 2.1 Introduction

There are many medical and biological studies in which the response variables are measured on an ordinal categorical scale. Often these outcome categories are more than two. As a result, the ordinary odds ratio (OR) or log odds ratio (LOR) can not be used directly without splitting the  $J \times L$  ( $J > 2$ ,  $L > 2$ ) contingency table into  $2 \times 2$  tables. Clayton (1974) introduced some odds ratio statistics for the analysis of ordered categorical data assuming a logit model. Clayton (1976) generalised the estimators for the case in which some observations are subject to censorship. McCullagh (1977) used paired comparisons on the ordinal variables.

Dale (1984) introduced the local and global odds ratios for measuring local and global associations for bivariate ordered responses. However, neither of these can be used in meta-analysis for ordinal categorical data produced by RCTs. Because the local odds ratio measures only the local relationship for a specific outcome category not for the whole study and the global odds ratio incurs loss of information by merging categories.

Edwardes and Baltzan (2000) proposed pooling  $\gamma$ 's (also known as Goodman and Kruskal's  $\gamma$ ) instead of general odds ratio ( $OR_G = (1 + \gamma)/(1 - \gamma)$ ) using sample size weights. This method is useful to calculate individual study effect and associated confidence intervals (CI). However, a well defined variance estimate for the pooled effect is unavailable and hence no CI can be found under this method.

Whitehead et al. (2001) developed a proportional odds model on individual patient data using the LOR with a general framework for fixed and random effects models. However, this method needs a 'proportional odds' assumption for computing the effect measures.

To overcome these issues, we propose GOR (Agresti, 1980) in meta-

analysis as an effect measure for ordinal categorical outcomes. There are several advantages of GOR over other effect measures. The GOR has simple and straightforward interpretation and can be used for continuous outcomes in meta-analysis. The GOR can also be used with multiple treatment groups for ordinal categorical outcomes. For binary outcomes with two comparison groups GOR reduces to ordinary OR. So GOR is a universal effect measure in meta-analysis and is applicable for binary, ordinal and continuous outcomes.

The rest of this chapter is organised as follows. Section 2.2 reviews the effect measures and methods usually used in meta-analysis for ordinal categorical outcomes. Section 2.3 presents the proposed meta-analysis method using GOR for ordinal categorical outcomes. Section 2.4 contains application of an individual patient data set. Finally, a chapter summary is presented in Section 2.5.

## 2.2 Review

In this section we review the effect measures and important methods normally used for ordinal data in meta-analysis.

### 2.2.1 Effect Measures

There are some indices that are used for measuring the effect size in meta-analysis with ordinal categorical outcomes. Some of these measures are discussed below.

#### (a) Odds Ratio (Cornfield, 1951; Edwards, 1963; Mosteller, 1968)

The odds ratio (OR) is one of the most frequently used indices in epidemiology and meta-analysis. The OR can be calculated from the

Table 2.1: Data of a single RCT

| Intervention  | Failure/Dead | Success/Alive |
|---------------|--------------|---------------|
| New treatment | a            | b             |
| Control       | c            | d             |

data (see Table 2.1) as

$$\text{OR} = \frac{a/c}{b/d} = \frac{\text{odds of treated in failure}}{\text{odds of treated in success}} = \frac{a \times d}{b \times c}. \quad (2.1)$$

In a RCT setting for an undesirable outcome  $\text{OR} < 1$  indicates an improvement on the new treatment,  $\text{OR} > 1$  indicates the new treatment is not effective, and  $\text{OR} = 1$  indicates the two comparison groups are very comparable. The converse is true for desirable outcomes. The cell with zero frequency is analysed adding  $\frac{1}{2}$  to each entry before calculation of OR for  $2 \times 2$  tables.

Despite its frequent use with binary outcomes in meta-analysis, it can not be directly used as an effect measure for more than two outcome categories.

**(b)** Local Odds Ratio (Dale, 1984)

The data for RCT with ordinal categorical outcomes can be presented in a  $2 \times L$  contingency table. Then the local odds ratio is defined as

$$\Psi_l = \frac{\pi_{l|1}\pi_{(l+1)|2}}{\pi_{l|2}\pi_{(l+1)|1}}; \quad l = 1, \dots, L - 1. \quad (2.2)$$

Local odds ratio measures the association between the row variable with the preferred category in the column variable in a  $2 \times L$  contingency table.

**(c)** Global Odds Ratio (Dale, 1984)

Global odds ratio is defined as the odds ratio of a  $(2 \times 2)$  contingency

table when the adjacent rows and columns of a  $(J \times L)$  contingency table is collapsed into a  $2 \times 2$  table. That is, a global odds ratio is a function of the quadrant probabilities of the original  $(J \times L)$  contingency table. A  $(J \times L)$  table has  $(J - 1)(L - 1)$  ways of merging rows and columns into  $2 \times 2$  table.

The global cross ratio  $\Psi_{jl}$  may be expressed as an odds ratio of cumulative events:

$$\Psi_{jl} = \text{odds}(Z_1 \leq j | Z_2 \leq l) / \text{odds}(Z_1 \leq j | Z_2 > l) \quad (2.3)$$

$$= \text{odds}(Z_2 \leq l | Z_1 \leq j) / \text{odds}(Z_2 \leq l | Z_1 > j). \quad (2.4)$$

**(d) Cumulative Odds Ratio** (Agresti, 2010, p.18)

The cumulative odds ratios use the sample conditional cumulative distribution functions of  $Y$  given  $x$  as

$$\hat{\theta}_{jl}^C = \frac{\hat{F}_{l|j} / (1 - F_{l|j})}{\hat{F}_{l|j+1} / (1 - \hat{F}_{l|j+1})}. \quad (2.5)$$

These odds ratios are natural when  $x$  is an explanatory variable. They provide a comparison of pairs of levels of  $x$  with respect to their entire conditional distribution on  $Y$ . For a  $2 \times L$  tables, global and cumulative odds ratios are identical. These (local, global and cumulative) odds ratios are the regular odds ratios computed for the  $2 \times 2$  tables obtained by collapsing the row and column classifications into dichotomies. As a result, none of these can be used as an effect measure for the whole table  $2 \times L$  or  $J \times L$  as in meta-analysis.

**(e) General Odds Ratio** (Edwardes and Baltzan, 2000)

Edwardes and Baltzan proposed the general odds ratio ( $\text{OR}_G$ ) for  $J \times L$  tables as

$$\text{OR}_G = \frac{1 + \gamma}{1 - \gamma}, \quad (2.6)$$

where  $\gamma = (P - Q)/(P + Q)$ , is also known as Goodman and Kruskal's  $\gamma$ ,  $P = \sum_{j < j'} \sum_{l < l'} n_{jl} n_{j'l'}$  and  $Q = \sum_{j < j'} \sum_{l > l'} n_{jl} n_{j'l'}$  considering  $n_{jl}$  be the number of observations with exposure  $j$  and severity  $l$ . It is important that the authors showed after few simple algebraic manipulation  $OR_G = P/Q$ , which is exactly the same as Agresti's  $\alpha$  (Agresti, 1980).

## 2.2.2 Issues of Meta-Analysis

Two of the most common issues of meta-analysis are heterogeneity and publication bias. We discuss these issues here briefly.

### Heterogeneity

Heterogeneity refers to the between study effect size variation. Random effects model accounts for heterogeneity to an extent but they do not explain why the study results vary. Subgroup analysis and regression methods can be used to identify the associations between study or patient characteristics and the outcome measures in meta-analysis of RCTs. It has now become common practice reporting  $Q$  statistic (Cochran, 1954) with associated degrees of freedom (df) and  $p$ -value for identifying heterogeneity and  $I^2$  statistic (Higgins and Thompson, 2002) for quantifying heterogeneity.

Graphically heterogeneity can be explored using plot of normalised (Z) scores, forest plot, radial plot (Galbraith, 1988) and L'Abbe plot (L'Abbe et al, 1987). Among these graphical tests, plot of Z scores and forest plot are now widely used for meta-analysis data representation.

- Plot of normalised (Z) scores: The Z-scores or standardised residuals

for each study can be calculated by

$$Z_i = \frac{\Gamma_i - \bar{\Gamma}}{se(\Gamma_i)}, \quad (2.7)$$

where  $\Gamma_i$  is the observed effect size estimate from the  $i$ th study,  $\bar{\Gamma}$  is the weighted average of these effect sizes, that is,  $\bar{\Gamma} = \sum_i^k \omega_i \Gamma_i / \sum_i^k \omega_i$ , where  $\omega_i =$  is the weight for the  $i$ th study and  $se(\Gamma_i) = 1/\omega_i = \sqrt{\text{var}(\Gamma_i)}$ . Under the null hypothesis that all the effect sizes are equal, the distribution of these Z-scores should be normal with mean zero and variance one. Large absolute Z values indicate significant deviations of individual study effect from the pooled or average effect (Greenland, 1987).

- Forest plot: A forest plot is a graphical display of representing the results of a meta-analysis. This plot is usually drawn using the effect estimate on a natural logarithmic scale along with their 95% CIs. A vertical line at the pooled effect estimate representing the meta-analysis estimate and a vertical line representing no effect are also drawn. Sometimes a body of text consisting of study names, publication years, effect estimates and CI values are also included. The forest plot was named after a breast cancer researcher Pat Forrest and hence sometimes spelt ‘forrest plot’ (Lewis and Clarke, 2001).

### ***Q* statistic**

A very common statistical test, the Cochran’s (1954)  $Q$  test examines the existence of homogeneity (heterogeneity) citing a  $p$ -value. This is defined as

$$Q = \sum_{i=1}^k \omega_i (\hat{\Gamma}_i - \bar{\Gamma})^2, \quad (2.8)$$

where  $\hat{\Gamma}_i$  = Treatment effect estimate in the  $i$ th study,  $k$  = number of studies combined,  $\bar{\Gamma} = \frac{\sum_i \omega_i \hat{\Gamma}_i}{\sum_i \omega_i}$  is the weighted estimator of treatment effect and  $\omega_i$  is the attached weight (inversely proportional to the conditional variance) of the  $i$ th study in the meta-analysis.

The following computation friendly form is available

$$Q = \sum_{i=1}^k \omega_i \hat{\Gamma}_i^2 - \frac{(\sum_{i=1}^k \omega_i \hat{\Gamma}_i)^2}{\sum_{i=1}^k \omega_i} \sim \chi_{k-1}^2 \quad (2.9)$$

under  $H_0 : \Gamma_1 = \Gamma_2 = \dots = \Gamma_k = \Gamma_0$ . For homogeneous studies the null hypothesis is not rejected which does not indicate statistical significance to use the common value whether the treatment is in favour or not. However, Fleiss (1986) recommended using a cut-off significance level of 0.10 rather than the usual 0.05. This test is of low in power with few studies and excessive power with many studies.

### $I^2$ statistic

The  $I^2$  statistic introduced by Higgins and Thompson (2002) is defined as

$$I^2 = \frac{Q - df}{Q} \times 100\%, \quad (2.10)$$

and represents the proportion of total variation in the estimates of treatment effect that is due to heterogeneity between studies. The  $I^2$  statistic can also be expressed in the form

$$I^2 = \frac{\text{Between variance}}{\text{Total variance}} \times 100\% = \frac{\hat{\tau}^2}{s_W^2 + \hat{\tau}^2}, \quad (2.11)$$

that is, the ratio of between variance to total variance across the observed effect estimates. The important aspects of this statistic are that it is easily interpretable and does not depend on the number of studies or type of outcome variable.

### Subgroup Analysis

Subgroup analysis can be carried out in meta-analysis by investigating subsets of studies or patient characteristics. One can also analyse the subsets of patients being pooled. There are three computational (Borenstein et al., 2009) models available such as fixed effect, random effects using separate estimates of  $\tau^2$  and random effects using pooled estimate of  $\tau^2$ . Three methods namely, Z-test, a Q-test based on analysis of variance and a Q-test for heterogeneity can be used for comparing the subgroups.

If it is assumed that the between variation is the same within all subgroups, a pooled estimate can be calculated. If the between variation does differ from one subgroup to the next, then we would estimate  $\tau^2$  within subgroups and use separate estimate of  $\tau^2$  for each subgroup. Although when there are a few studies within subgroups, it may be preferable to pool the estimates.

### Regression models for meta-analysis

Regression models can also be used under fixed effects model (meta-regression model) (Hedges, 1994) and random effects model (mixed effects model) for exploring heterogeneity. However, regression models are mostly useful when the number of studies is large (Raudenbush, 1994).

### Publication Bias

It is a well known fact that research with statistically significant results is more likely to be published or published more quickly than work with insignificant results. This leads to a dominance of false positive results in

the literature. This is known as publication bias. The implications of this for meta-analysis are that combining only the identified published studies.

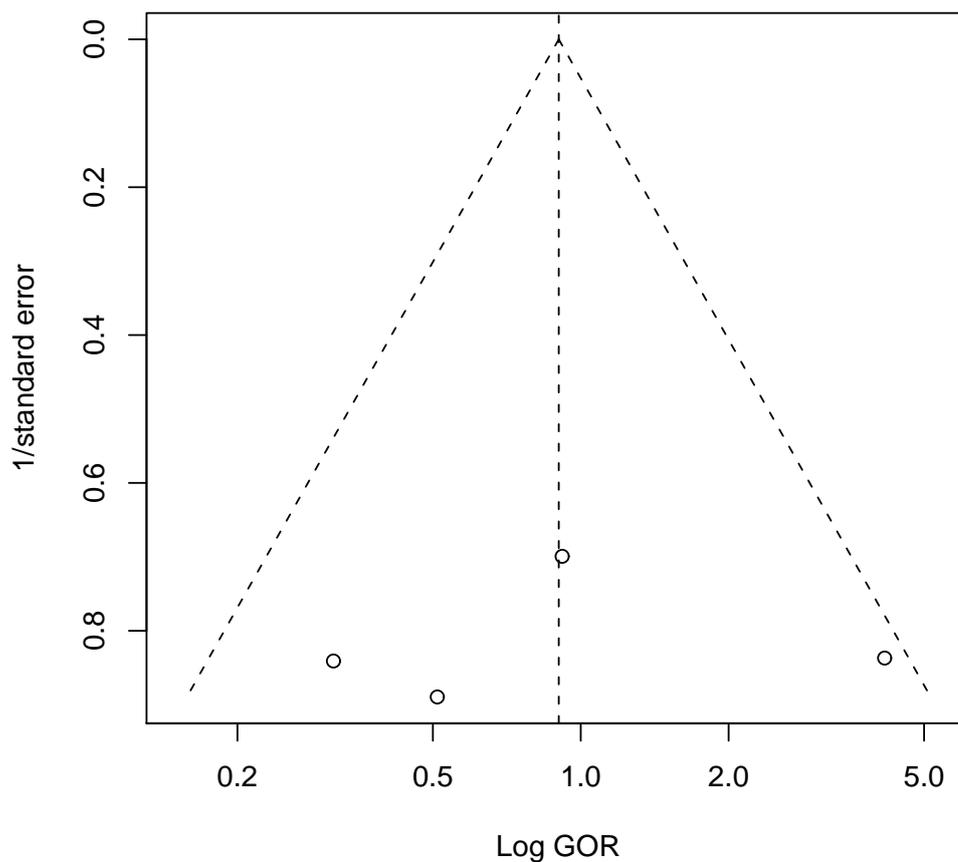


Figure 2.1: Funnel plot indicating absence of publication bias

It is now a common practice to report presence or absence of publication bias and to assess the effects of publication bias in meta-analysis. The following tools can be used for identifying publication bias.

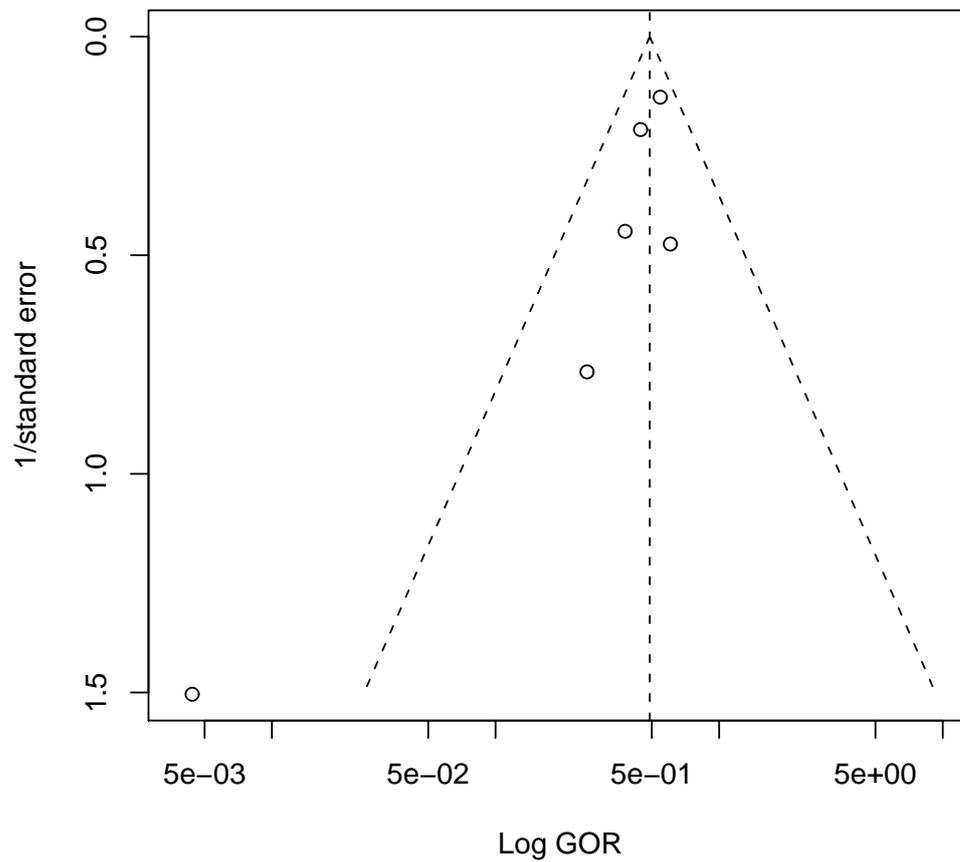


Figure 2.2: Funnel plot showing presence of publication bias

- The funnel plot
- Rank correlation test (Begg and Mazumdar, 1994)
- Linear regression test (Egger et al, 1997)
- Trim and fill method (Duval and Tweedie, 2000)

The funnel plot is now widely used reporting publication bias in meta-

analysis. Therefore, we will only discuss funnel plot here briefly.

### The funnel plot

The treatment effects from individual studies are usually plotted against their standard errors (se) (or inverse of the se) instead of the corresponding sample sizes. A plot of inverse of se versus logarithm of the treatment effect from individual studies in a meta-analysis should thus be shaped like a funnel if there is no publication bias (Light and Pillemar, 1984).

However, plotting treatment effects against standard errors instead of against the inverse of standard errors are not the same. The visual impression may be different in those situations. To detect publication bias by a funnel plot there needs to be a range of studies with varying effect sizes. It is also possible that an asymmetric funnel plot may be caused by factors other than publication bias, such as quality of studies, intervention, differences in underlying risk, poor design of small studies, choice of effect measure, inadequate analysis, chance and number of studies (Egger et al, 1997).

### 2.2.3 Proportional Odds Model

Whitehead et al. (2001) used the proportional odds model in meta-analysis for ordinal outcome categories using individual patient data (IPD). To illustrate this method, let there be  $k$  independent studies each comparing a treatment with a control. Each patient has a response which falls into one of the  $L$  ordered categories  $C_1, \dots, C_L$ ; ( $L > 2$ ) such that  $C_1$  is the best and  $C_L$  the worst. Suppose that there are  $n_i$  patients in the  $i$ th study and  $n = \sum_{i=1}^k n_i$  is the total number of patients in all of the studies combined. Let  $\pi_{ijr}$  be the probability that the  $j$ th subject in the  $i$ th study has a response

in category  $C_r$ . Again consider  $Q_{ijr}$  to be the associated probability of a response in category  $C_r$  or better, so that  $Q_{ijr} = \pi_{ij1} + \pi_{ij2} + \dots + \pi_{ijr}$  and  $Q_{ijL} = 1$ . Then the proportional odds model is defined by

$$\lambda_{ijr} = \log \left( \frac{Q_{ijr}}{1 - Q_{ijr}} \right) = \alpha_r + \beta_{ij}; \quad r = 1, \dots, L - 1, \quad (2.12)$$

where  $\alpha_r$  is the  $r$ th intercept and  $\beta_{ij} = \gamma_1 x_{1ij} + \dots + \gamma_q x_{qij}$  is a linear combination of explanatory variables. This model assumes ‘proportional odds’ that the log odds ratio associated with a unit increase in the explanatory variable does not depend on the intercept. Whitehead et al. (2001) also developed several meta-analysis methods for ordinal outcomes using IPD under fixed and random effects models.

### 2.2.4 Sample Size Weight Method

Edwardes and Baltzan (2000) proposed a meta-analysis method by pooling  $\gamma$  (where  $\gamma = (P - Q)/(P + Q)$ ) is also known as Goodman and Kruskal’s  $\gamma$ ,  $P = \sum_{j < j'} \sum_{l < l'} n_{jl} n_{j'l'}$  and  $Q = \sum_{j < j'} \sum_{l > l'} n_{jl} n_{j'l'}$  considering  $n_{jl}$  be the number of observations with exposure  $j$  and severity  $l$ ) instead of pooling general odds ratio,  $OR_G = (1 + \gamma)/(1 - \gamma)$ , using sample size weight to avoid data change (Agresti, 1980; Goodman and Kruskal, 1972). They proposed weighting each  $\gamma_i$  by

$$T_{Ni} = \left( \sum_{i < j} n_i n_j \right)^2 / \sum_{i < j} (n_i + n_j) n_i n_j. \quad (2.13)$$

As a special case when  $J = 2$ , then  $T_N = n_1 n_2 / (n_1 + n_2)$ . The sample size weighted mean effect is given as

$$\bar{\gamma} = \sum_i T_{Ni} \gamma_i / \sum_i T_{Ni} \quad (2.14)$$

and the pooled  $OR_G$  is calculated as  $\overline{OR}_G = (1 + \bar{\gamma}) / (1 - \bar{\gamma})$ . Confidence intervals (CI) for pooled  $OR_G$  may be formed using

$$SE(\bar{\gamma}) = \left( \sum_i T_{Ni}^2 SE(\gamma_i)^2 \right)^{1/2} / \sum_i T_{Ni}. \quad (2.15)$$

The standard error estimate  $SE(OR_G)$  is derived from Goodman and Kruskal's general formula (Goodman and Kruskal, 1972) as

$$SE(OR_G) = \frac{1}{Q} \left[ \sum_i \sum_j n_{ij} R_{ij}^2 \right]^{1/2}, \quad (2.16)$$

where  $R_{ij} = OR_G(B_{ij} + C_{ij}) - A_{ij} - D_{ij}$ ,  $n_{+j} = \sum_i n_{ij}$ ,  $A_{ij} = \sum_{s<i} \sum_{t<j} n_{st}$ ,  $B_{ij} = \sum_{s<i} \sum_{t>j} n_{st}$ ,  $C_{ij} = \sum_{s>i} \sum_{t<j} n_{st}$ ,  $D_{ij} = \sum_{s>i} \sum_{t>j} n_{st}$ .

An asymptotic standard error of  $\log(OR_G)$  is estimated by

$$SE(\log OR_G) = 2 \times SE(\gamma) / (1 - \gamma)^2 \quad (2.17)$$

and a 95% CI for  $OR_G$  is

$$\{OR_G \exp[-3.92 \times SE(\gamma) / (1 - \gamma^2)], OR_G \exp[3.92 \times SE(\gamma) / (1 - \gamma^2)]\}, \quad (2.18)$$

where  $\exp(\cdot)$  is the exponential function, and 3.92 is  $2 \times 1.96$ . The sample size weight method can be used to estimate the effect size of the studies. Unfortunately, a well defined variance estimate of the estimated pooled effect is unavailable for this method. Hence this method can not be used for constructing CI of meta-analysis.

## 2.3 Proposed Method: Meta-Analysis using GOR

### 2.3.1 GOR

Let  $J$  be the number of comparison groups with  $L$  ordered outcome categories in each group. For RCTs with two comparison groups the  $2 \times L$  contingency

table represents the joint distribution of two ordinal categorical variables:

Table 2.2: Contingency table for the  $i$ th study.

| Groups    | Category 1 | Category 2 | ... | Category $L$ | Sample size |
|-----------|------------|------------|-----|--------------|-------------|
| Treatment | $X_{i11}$  | $X_{i12}$  | ... | $X_{i1L}$    | $n_{i1}$ .  |
| Control   | $X_{i21}$  | $X_{i22}$  | ... | $X_{i2L}$    | $n_{i2}$ .  |

In Table 2.2,  $X_{ijl}$  is the count of the  $l$ th category in the  $j$ th group for the  $i$ th study,  $n_{ij}$  is the total count of  $j$ th group for the  $i$ th study,  $X_{i1L} = n_{i1} - X_{i11} - X_{i12} - \dots - X_{i1(L-1)}$ ,  $X_{i2L} = n_{i2} - X_{i21} - X_{i22} - \dots - X_{i2(L-1)}$ . When  $L = 2$ , the GOR reduces to the OR for a single  $2 \times 2$  contingency table.

The GOR is defined as the ratio of the proportions of concordant and discordant pairs (Agresti, 1980) in a contingency table. A pair is said to be concordant if the subject ranked higher on groups also ranks higher on categories or vice versa. Without loss of generality we assume that the response in category  $l'$  is more severe than the response in category  $l$  where  $l < l'$ . Mathematically, the GOR for the  $i$ th study is defined as

$$\Gamma_i = (\Pi_{di})^{-1} \Pi_{ci}, \tag{2.19}$$

where  $\Pi_{ci} = \sum_{r=1}^{L-1} \sum_{s=r+1}^L \Pi_{ir|1} \Pi_{is|2}$  and  $\Pi_{di} = \sum_{r=2}^L \sum_{s=1}^{r-1} \Pi_{ir|1} \Pi_{is|2}$ . Here,  $\Pi_{ci}$  denotes the probability that the response of a randomly selected subject from group 2 (control) is more severe than the response of a randomly selected subject from group 1 (treatment). Similarly,  $\Pi_{di}$  denotes the probability that the response of a randomly selected subject from group 1 is more severe than the response of a randomly selected subject from group 2. The data with zero cell count is analysed by adding  $\frac{1}{L}$  to each entry before calculation of the GOR. The value of  $\Gamma_i$  may vary from 0 to  $\infty$ .  $\Gamma_i = 1$ , represents identical comparison groups as it is in the OR.

Suppose an independent random sample of size  $n_{ij}$  is taken from group  $j$  ( $j = 1, 2$ ) and  $X_{ijl}$  denote the count falling into category  $l$  of the  $i$ th study. Then the random vector  $(X_{ij1}, X_{ij2}, \dots, X_{ijL})$  follows the multinomial distribution with parameters  $n_{ij}$  and  $\boldsymbol{\pi}'_{ij} = (\pi_{i1|j}, \pi_{i2|j}, \dots, \pi_{iL|j})$ , where  $\pi_{il|j}$  is the probability of a subject to be in the  $l$ th category within the  $j$ th comparison group for the  $i$ th study.

The maximum likelihood estimator (MLE) of  $\pi_{il|j}$  is given by  $\hat{\pi}_{il|j} = X_{ijl}/n_{ij}$  for the  $i$ th study. For large  $n_{ij}$ ,  $\sqrt{n_{ij}}(\hat{\boldsymbol{\pi}}_{ij} - \boldsymbol{\pi}_{ij})$ , where  $\boldsymbol{\pi}_{ij}' = (\pi_{i1|j}, \pi_{i2|j}, \dots, \pi_{iL|j})$ , asymptotically follows a  $L$ -dimensional multivariate normal distribution with mean vector  $\mathbf{0}$  and  $L \times L$  covariance matrix with the diagonal entries  $\pi_{il|j}(1 - \pi_{il|j})$ , and off-diagonal entries  $-\pi_{il|j}\pi_{il'|j}$  for  $l \neq l'$ . The MLE of  $\Gamma_i$ , say  $\hat{\Gamma}_i$ , is defined as

$$\hat{\Gamma}_i = (\hat{\Pi}_{di})^{-1} \hat{\Pi}_{ci}, \quad (2.20)$$

where  $\hat{\Pi}_{ci} = \sum_{r=1}^{L-1} \sum_{s=r+1}^L \hat{\pi}_{ir|1} \hat{\pi}_{is|2}$  and  $\hat{\Pi}_{di} = \sum_{r=2}^L \sum_{s=1}^{r-1} \hat{\pi}_{ir|1} \hat{\pi}_{is|2}$  for the  $i$ th study.

Next, we present a general fixed effects model (FEM) (Birge, 1932) (Cochran, 1937) and standard random effects model (REM) (DerSimonian and Laird, 1986) using GOR for multilevel ordinal outcomes. Consider  $\hat{\Gamma}_i$  to be an estimate of the true effect size  $\Gamma_i$  in the  $i$ th study as  $\hat{\Gamma}_i = \Gamma_i + e_i$ , where  $e_i$  is the error with which  $\hat{\Gamma}_i$  estimates  $\Gamma_i$ . In the FEM,  $\text{var}(\hat{\Gamma}_i) = v_i$ , and in REM  $\text{var}(\hat{\Gamma}_i) = \tau_{\Gamma}^2 + v_i$ , where  $\tau_{\Gamma}^2$  is the random effects variance and  $v_i$  is the variance due to sampling error in the  $i$ th study. If  $\tau_{\Gamma}^2 = 0$ , the above REM would reduce to the FEM.

### 2.3.2 GOR under Fixed Effects Model

The inverse variance weighted method is most widely used amongst the fixed effects models. For  $k$  independent studies if  $\hat{\Gamma}_i$  represents logarithm of GOR (LGOR) and  $v_i$  represents the variance of the effect estimator, then assuming  $\Gamma_1 = \Gamma_2 = \dots, \Gamma_k = \Gamma_0$ , a pooled estimate of the treatment effect is given by

$$\hat{\Gamma}_0 = \frac{\sum_i \omega_i \hat{\Gamma}_i}{\sum_i \omega_i}. \quad (2.21)$$

For an arbitrary number of outcome categories ( $L$ ) in RCTs in which each row is modeled as an independent multinomial distribution, the estimated variance of the  $i$ th study is

$$\hat{\omega}_i^{-1} = \sum_{l=1}^{L-1} \sum_{j=1}^2 \frac{1}{n_{ij} \hat{\pi}_{ijl} (1 - \hat{\pi}_{ijl})}, \quad (2.22)$$

where  $n_{ij}$  is the total count of the  $j$ th group for the  $i$ th study,  $\hat{\pi}_{ijl} = X_{ijl}/n_{ij}$  is the MLE of  $\pi_{ijl}$  and  $X_{ijl}$  is the count of the  $l$ th category in the  $j$ th group for the  $i$ th study.

Assuming  $\hat{\Gamma}_i$ 's are normally distributed, an approximate  $100(1 - \alpha)\%$  CI for the  $i$ th GOR is given by the formula

$$\exp[\hat{\Gamma}_i \pm z_{\alpha/2} \omega_i^{-1/2}], \quad (2.23)$$

where  $z_{\alpha/2}$  is the  $100 \times \alpha/2$  percentage point of a standard normal distribution.

An estimator of the variance of the pooled estimator of  $\Gamma_0$  is given by

$$\hat{\omega}^{-1} = \text{var}(\hat{\Gamma}_0) = 1 / \sum_{i=1}^k \omega_i. \quad (2.24)$$

If  $\hat{\Gamma}_0$  is assumed to be normally distributed, an approximate  $100(1 - \alpha)\%$  CI for the population effect,  $\Gamma_0$ , is given by

$$\exp[\hat{\Gamma}_0 \pm z_{\alpha/2}\omega^{-1/2}] \quad (2.25)$$

for the meta analysis.

### 2.3.3 GOR under Random Effects Model

The standard random effects model (DerSimonian and Laird, 1986) for the GOR can be introduced as follows: Let  $\hat{\tau}_\Gamma^2$  be the estimate of the between study variance. Define  $\bar{\omega}$  and  $s_W^2$  to be the mean and variance of the weights from the  $k$  studies:

$$\bar{\omega} = \sum_{i=1}^k \omega_i/k \text{ and } s_W^2 = \frac{1}{k-1} \left( \sum_{i=1}^k \omega_i^2 - k\bar{\omega}^2 \right). \quad (2.26)$$

Further, define

$$U = (k-1) \left( \bar{\omega} - \frac{s_W^2}{k\bar{\omega}} \right) \text{ and } Q = \sum_{i=1}^k \omega_i (\hat{\Gamma}_i - \bar{\Gamma})^2, \quad (2.27)$$

where  $Q$  is the heterogeneity statistic, also known as Cochran's  $\chi^2$  statistic (Cochran, 1954) for testing the  $H_0 = \Gamma_1 = \Gamma_2 = \dots = \Gamma_k = \Gamma_0$ . The estimated component of variance due to inter-study variation in effect size,  $\hat{\tau}_\Gamma^2$ , is calculated as

$$\hat{\tau}_\Gamma^2 = \begin{cases} 0 & \text{if } Q \leq k-1 \\ (Q - (k-1))/U & \text{if } Q > k-1. \end{cases} \quad (2.28)$$

Then adjusted weights  $\omega_i^*$  for each of the studies can be calculated as

$$\omega_i^* = \frac{1}{[1/\omega_i] + \hat{\tau}_\Gamma^2}. \quad (2.29)$$

A  $100(1 - \alpha)\%$  CI for  $\Gamma_i$  is given by

$$\exp[\hat{\Gamma}_{iR} \pm z_{\alpha/2}/\sqrt{\omega_i^*}] \quad (2.30)$$

under the assumption of normality of  $\hat{\Gamma}_{iR}$ .

The point estimate for the mean treatment effect of all studies,  $\Gamma_0$ , can be computed by

$$\hat{\Gamma}_{0R} = \frac{\sum_{i=1}^k \omega_i^* \hat{\Gamma}_i}{\sum_{i=1}^k \omega_i^*} \text{ with } \text{var}(\hat{\Gamma}_{0R}) = 1 / \sum_{i=1}^k \omega_i^*. \quad (2.31)$$

If normality of  $\hat{\Gamma}_{0R}$  is assumed, a  $100(1 - \alpha)\%$  CI for  $\Gamma_0$  is given by

$$\exp[\hat{\Gamma}_{0R} \pm z_{\alpha/2} / \sqrt{\sum_{i=1}^k \omega_i^*}]. \quad (2.32)$$

## 2.4 Application

We consider the IPD of five RCTs (Table 2.3) of anti-cholinesterase drug tacrine in patients with Alzheimer's disease (Whitehead et al., 2001) to apply the proposed method and compare results from different methods. The categories of this trials are made by the Clinical Global Impression of Change scale (CGIC). The CGIC is based on a seven point scale where 1, 2 and 3 represent 'very much improved', 'much improved' and 'minimally improved' respectively, 4 indicates 'no change', and 5, 6 and 7 represent 'minimally worse', 'much worse' and 'very much worse' respectively. For the analysis purpose they combined categories 1 and 2, and 6 and 7 as there were very few patients in the two extreme categories.

The second last column of Table 2.3 contains the LOR and standard errors (SE) from proportional odds model. The last column contains the log GOR and SE from the proposed method. As  $\text{OR}_{\text{G}}$  values are exactly the same as GORs, those are not included in the table.

Table 2.4 shows the comparison of meta-analysis estimates (in logarithmic scale) and their associated SEs for the three methods used. The last three

Table 2.3: Estimate of LOR and LGOR along with SE for tacrine data

| Study | Groups  | C1 | C2  | C3  | C4 | C5 | T   | LOR(SE)       | LGOR(SE)      |
|-------|---------|----|-----|-----|----|----|-----|---------------|---------------|
| 1     | Tacrine | 4  | 23  | 45  | 22 | 2  | 96  | 0.284 (0.261) | 0.248 (1.033) |
|       | Placebo | 2  | 22  | 54  | 29 | 3  | 110 |               |               |
| 2     | Tacrine | 14 | 119 | 180 | 54 | 6  | 373 | 0.224 (0.242) | 0.197 (1.167) |
|       | Placebo | 1  | 22  | 35  | 11 | 3  | 72  |               |               |
| 3     | Tacrine | 13 | 20  | 24  | 10 | 1  | 68  | 0.360 (0.331) | 0.299 (0.901) |
|       | Placebo | 7  | 16  | 17  | 10 | 3  | 53  |               |               |
| 4     | Tacrine | 21 | 106 | 175 | 62 | 17 | 381 | 0.785 (0.173) | 0.659 (0.569) |
|       | Placebo | 8  | 24  | 73  | 52 | 13 | 170 |               |               |
| 5     | Tacrine | 3  | 14  | 19  | 3  | 0  | 39  | 0.492 (0.422) | 0.434 (1.358) |
|       | Placebo | 2  | 13  | 18  | 7  | 1  | 41  |               |               |

columns of Table 2.4 contain the heterogeneity statistic  $Q$ ,  $p$ -value, and  $I^2$  statistic.

Figure 2.3 represents the forest plot of the five RCTs of tacrine data using GOR. The horizontal lines represent 95% CIs for the five studies and meta-analysis. The pooled estimate at 1.54 is obtained by combining all GORs of the five studies using the proposed method.

In the first meta-analysis of the tacrine trials (Qizilbash et al., 1998) the OR for improvement on the CGIC scale for patients receiving tacrine compared with those receiving placebo was reported 1.58 (SE = 0.270).

Meta-analysis using proportional odds model produced the estimated OR = 1.66 under fixed effects model and OR = 1.62 under random effects model. The heterogeneity statistic  $Q = 4.86$ ,  $df = 4$ ,  $p = 0.30$  and  $I^2 = 17.69\%$  [0%, 99.69%] suggest there exists insignificant heterogeneity (Whitehead et al., 2001; Higgins and Thompson, 2002).

Table 2.4: Comparison of meta-analysis estimates for the tacrine data

| Methods                    | MA    | SE    | $Q$  | $p$ -value | $I^2$        |
|----------------------------|-------|-------|------|------------|--------------|
| <b>Proport. odds model</b> |       |       | 4.86 | 0.300      | 17.69%       |
| FEM                        | 0.504 | 0.112 |      |            | [0%, 99.69%] |
| REM                        | 0.481 | 0.128 |      |            |              |
| <b>Sample size weight</b>  |       |       | 1.60 | 0.809      | 0%           |
| FEM                        | 0.426 | -     |      |            | [0%, 97.7%]  |
| REM                        | 0.426 | -     |      |            |              |
| <b>Proposed method</b>     |       |       | 0.34 | 0.987      | 0%           |
| FEM                        | 0.430 | 0.391 |      |            | [0%, 89%]    |
| REM                        | 0.430 | 0.391 |      |            |              |

Meta-analysis using sample size weight (Edwardes and Baltzan, 2000) gives the estimated pooled  $OR_G = 1.53$ . The authors reported that an advantage to sample size weighting is that the pooled estimate may be based on studies for which variance is not reported. However, a variance estimate is required for the CI of pooled  $OR_G$ . We did not report the estimated variance for the pooled estimate as there is no well defined variance expression for the pooled  $OR_G$ . The authors also did not find the CI for pooled  $OR_G$  in their meta-analysis (Edwardes and Baltzan, 2000). The heterogeneity statistic  $Q = 1.60$ ,  $df = 4$ ,  $p = 0.809$  and  $I^2 = 0\%$  [0%, 97.7%] suggest there exists no heterogeneity in the data set.

The proposed meta-analysis method using GOR results estimated GOR = 1.54. The tests of heterogeneity are statistically insignificant  $Q = 0.341$ ,  $df = 4$ ,  $p = 0.987$ ; and  $I^2 = 0\%$  [0%, 89%].

Although the CI of the pooled effect measure for the proposed method is wider than the proportional odds model, we believe the true scenario has been

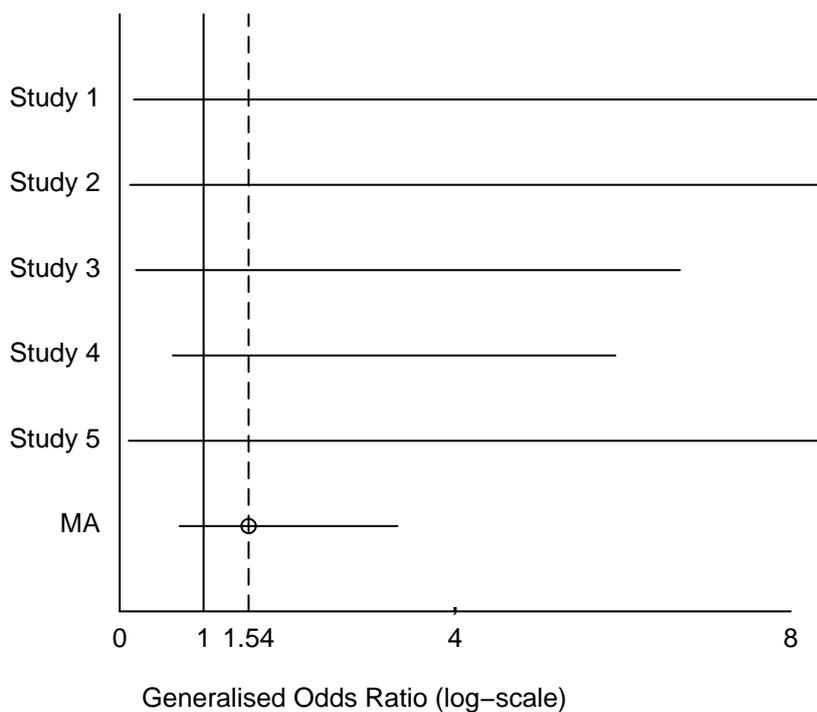


Figure 2.3: Forest plot of the five RCTs of tacrine using GOR

depicted by the GOR. Regarding the heterogeneity statistic  $Q$ , it is observed that as the number of categories increases the value of  $Q$  decreases. For the tacrine data, with five categories  $Q = 0.43$ , with three categories  $Q = 1.83$ , and with two categories  $Q = 4.86$ . As proportional odds model is using LOR, it is unreasonably reducing the SE which entails a higher individual weight. That results the higher  $Q$  value for the proportional odds model.

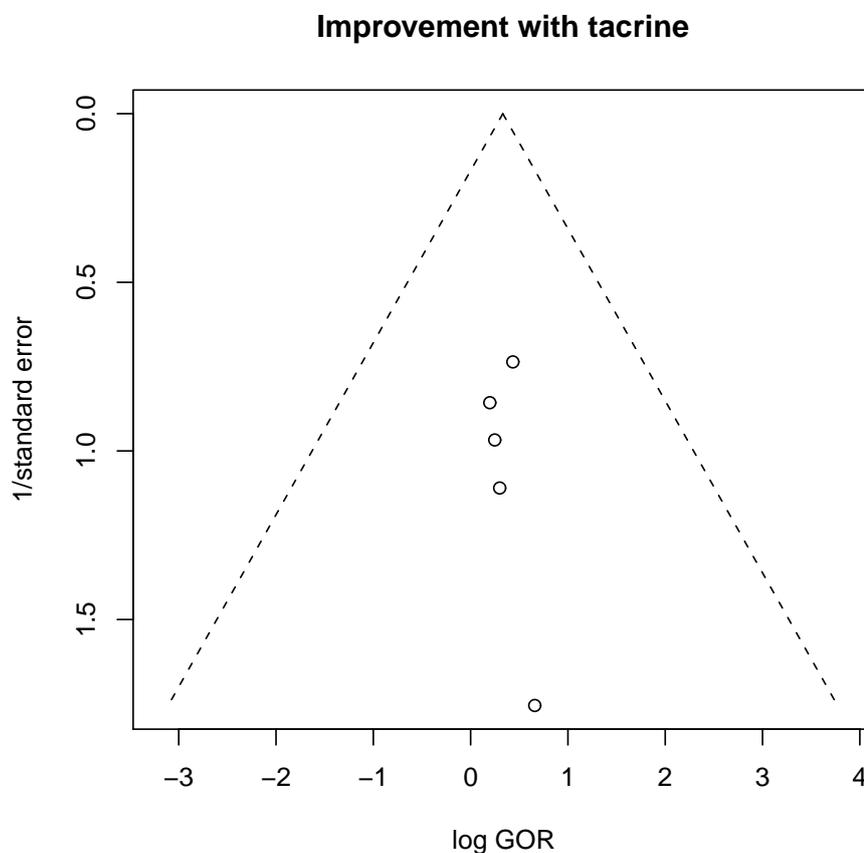


Figure 2.4: The funnel plot suggests that there is no publication bias

## 2.5 Summary

The main contribution of this chapter is that we develop a meta-analysis method using GOR for multi-level ordinal categorical outcomes and compare the results with two existing methods for the tacrine trials data. The currently available proportional odds model is restricted to the proportionality assumption and there is no well defined variance estimate of the pooled estimate for the sample size weight method. Use of the LOR or similar effect measures for multi-level ordinal outcomes by collapsing a  $2 \times L$  table into

$2 \times 2$  tables causes loss of information, inflate the estimate and inappropriately reduce the spread.

Here the treatment effect of tacrine over placebo is evident from the meta-analyses. Of the three meta-analyses the treatment effect produced by the proportional odds model is the highest. The treatment effects produced by the other two methods are very close. This also emphasises the point that use of the proportional odds model may have reduced the SE of the effects by using LOR.

It is revealed that the use of the proportional odds model actually undermines the real variability in the data set of five outcome categories. This is caused by the reduction of the five categories into only two categories which in turn unreasonably reduce the spread. As a result, the SE for LOR is less than what it should be with five categories. Use of the GOR and multinomial distribution can produce the true variance taking into account all the outcome categories in a data set.

It is also observed that the majority of patients from the five tacrine trials fall in the middle three categories. As a result, the effect measure in proportional odds model may be inflated because of the proportionality assumption.

The proposed meta-analysis method using GOR is very simple and has straightforward interpretation. It has simple variance estimate for individual study and meta-analysis. It can also be used for binary and latent continuous outcomes (see Chapter 4 and Chapter 5). Moreover, the ordinal categorical outcomes are naturally in ascending or descending order. Therefore, GOR is a preferable and superior effect measure in meta-analysis for the ordinal categorical data.

## Chapter 3

# A QEBM for Heterogeneous Ordinal Data

### 3.1 Introduction

In recent years, there has been growing interest in meta-analysis in various fields of health science, including analysis of RCTs. The method has impacted significantly on the practice of clinical trials and health care policies. It allows to combine results of independent studies to increase statistical power of the effect size estimate through increased sample size.

The development of the meta-analysis for the binary outcomes is primarily based on the relative risk (RR) and odds ratio (OR). However, in many clinical and epidemiological studies the outcome variables are ordinal and have more than two categories. For example, outcomes of the study on pain relief, dementia, head injury, tonsil size etc are measured on ordinal scales. The commonly used effect size measures for binary outcomes can not be directly used if the data is ordinal with more than two categories. The idea of collapsing the  $2 \times L$ ; ( $L > 2$ ) tables into  $2 \times 2$  tables are often arbitrary and leads to loss of valuable information. The generalised odds ratio (GOR) (Agresti, 1980) is an appropriate measure to estimate the effect size of ordinal categorical outcomes in meta-analysis.

There are some advantages of the GOR over its competitors. Using the GOR in meta-analysis one can overcome the assumption needed in proportional odds model (Whitehead et al., 2001). The GOR can also be used in multi-arm trials. For binary outcomes with two comparison groups the GOR reduces to the ordinary OR. So the GOR can be a universal effect measure in meta-analysis with ordinal data as we can use it for binary, ordinal and continuous outcomes.

One of the main issues in meta-analysis is the heterogeneity of the effect size. The use of random effects model is not always a cure (Shuster, 2010; Doi SA et al., 2011) for the problem. Thompson and Pocock (1987) ar-

gued that meta-analysis provides simple solution of combining several studies with respect to qualitative conclusions but fails to provide conclusive quantitative results for broad treatment policies when there exists heterogeneity among the studies. “Doing meta-analysis is easy”, says Ingram Olkin but “Doing one well is hard” as heterogeneity among studies may lead to incorrect meta-analysis (Mann, 1990). In this chapter, a quasi-empirical Bayes method (QEBM) (Saleh, 2006, p.157) is developed using the predicted GOR to handle the heterogeneity issue in the light of Stein’s shrinkage estimator (Stein, 1956). This QEBM involves a heterogeneity test based on the chi-square statistic (Cochran, 1954) and shrink the estimator towards the common pooled value. This process identifies the extreme studies and finds a statistical agreement to improve the quality of the estimator and trustworthiness of meta-analysis.

This chapter also implements the above method in two real life and one simulated ordinal data sets. The main reason to consider three examples is to illustrate the effectiveness of the method for data sets with varying degree of heterogeneity. If the effect sizes of individual studies are not significantly heterogeneous the meta-analysis produces acceptable results but studies with high degree of heterogeneity require special attention and the method proposed adequately deals with the issue.

## 3.2 Quasi-Empirical Bayes Method for GOR

The use of the empirical Bayes (EB) approaches has received enormous attention in the literature (Van Houwelingen and Stijnen, 1993; Efron, 1996; Raudenbush, 1985; Zhou, 1996; Morris, 1992; Stijnen and Van Houwelingen, 1990 and Saleh, 2006). But EB is rather difficult to define precisely. Some

statisticians refer it to a class of models; others, a style of analysis; still others, a philosophy for screening statistical procedures. However, all the EB methods almost exclusively assumed that the ‘prior distribution’ has been at the second level of the Bayesian analysis and that the hyper-parameters have then been estimated from data. The quasi-empirical Bayes approach results in a similar or an approximation of the EB estimation approach, hence, the name “quasi-empirical Bayes estimation (QEBE)”. Following Saleh et al. (2006) a QEBM has been developed using GOR for heterogeneous ordinal categorical data.

Here the initial objective is to assess the exposure disease association by the common pooled GOR, say  $\Gamma_0$ . However, it is uncertain whether the GORs,  $\Gamma_i$ 's of the  $k$  independent studies are homogeneous or not. That is, first we estimate the vector  $\mathbf{\Gamma} = (\Gamma_1, \Gamma_2, \dots, \Gamma_k)'$  when the hypothesis,

$$H_0 : \Gamma_1 = \Gamma_2 = \dots = \Gamma_k = \Gamma_0 \quad (3.1)$$

is unknown. For  $J = 2$ , the  $i$ th study is modeled as two multinomial distributions with parameters  $(n_{i1}, \boldsymbol{\pi}_{i1})$  and  $(n_{i2}, \boldsymbol{\pi}_{i2})$  for a  $2 \times L$  contingency table in RCTs. Then the GOR of the  $i$ th study is defined as

$$\Gamma_i = (\Pi_{di})^{-1} \Pi_{ci} \quad (3.2)$$

and estimated as  $\mathbf{\Gamma} = (\Gamma_1, \dots, \Gamma_k)'$  by the vector  $\tilde{\mathbf{\Gamma}}_n = (\tilde{\Gamma}_1, \dots, \tilde{\Gamma}_k)'$ , where

$$\hat{\Gamma}_i = (\hat{\Pi}_{di})^{-1} \hat{\Pi}_{ci}, \quad (3.3)$$

in which  $\hat{\Pi}_{ci} = \sum_{r=1}^{L-1} \sum_{s=r+1}^L \hat{\pi}_{ir|1} \hat{\pi}_{is|2}$  and  $\hat{\Pi}_{di} = \sum_{r=2}^L \sum_{s=1}^{r-1} \hat{\pi}_{ir|1} \hat{\pi}_{is|2}$ .

If the null hypothesis is true, there are three choices of efficient estimators of the pooled GOR,  $\Gamma_0$ . These are

(i) weighted arithmetic mean,

$$\hat{\Gamma}_{0n}^{(a)} = \hat{\omega}^{-1} \left( \sum_{i=1}^k \hat{\omega}_i \tilde{\Gamma}_i \right), \quad \text{with } \hat{\omega} = \sum_{i=1}^k \hat{\omega}_i, \quad (3.4)$$

(ii) weighted geometric mean,

$$\hat{\Gamma}_{0n}^{(g)} = \exp \left[ \hat{\omega}^{-1} \left( \sum_{i=1}^k \hat{\omega}_i \ln \tilde{\Gamma}_i \right) \right], \quad (3.5)$$

(iii) weighted harmonic mean,

$$\hat{\Gamma}_{0n}^{(h)} = \hat{\omega} \left( \sum_{i=1}^k \hat{\omega}_i \tilde{\Gamma}_i^{-1} \right)^{-1}, \quad (3.6)$$

where

$$\hat{\omega}_i^{-1} = \sum_{l=1}^{L-1} \sum_{j=1}^2 \frac{1}{n_{ij} \hat{\pi}_{ijl} (1 - \hat{\pi}_{ijl})}. \quad (3.7)$$

We employ the weighted geometric mean, formula (3.5) to obtain  $\hat{\Gamma}_{0n}$ . For large samples the distributions of  $\ln \hat{\Gamma}_i$  and  $\ln \hat{\Gamma}_{0n}$  are approximately normal, so the formula  $\exp[\ln \hat{\Gamma}_i \pm z_{\alpha/2} \omega_i^{-1/2}]$  is appropriate to obtain a  $100(1 - \alpha)\%$  CI for the GOR of the  $i$ th study and  $\exp[\ln \hat{\Gamma}_{0n} \pm z_{\alpha/2} \omega^{-1/2}]$  for the meta-analysis, where  $z_{\alpha/2}$  is the  $(1 - \alpha/2)$  level critical value of the standard normal distribution,  $N(0, 1)$ .

When the  $k$  study effects are homogeneous, i.e.,  $\Gamma_1 = \Gamma_2 = \dots = \Gamma_k = \Gamma_0$ , a vertical line representing the pooled GOR ( $\hat{\Gamma}_{0n}$ ) crosses through the CI for every trial demonstrating the power of meta-analysis to find statistical agreement. When there is suspicion about heterogeneity, then whether meta-analysis is trusted or not depends on the outcome of the test of homogeneity of the GORs,  $H_0 : \Gamma_1 = \Gamma_2 = \dots = \Gamma_k = \Gamma_0$  against the alternative,  $H_A$ : at least one of the pairs  $(\Gamma_i, \Gamma_{i'})$  differ ( $i, i' = 1, \dots, k$ ). We can test this hypothesis using the test statistic,  $\mathcal{L}_n$ , defined by

$$\mathcal{L}_n = (\tilde{\boldsymbol{\psi}}_n - \hat{\psi}_{0n} \mathbf{1}_k)' \widehat{\mathbf{W}}_n (\tilde{\boldsymbol{\psi}}_n - \hat{\psi}_{0n} \mathbf{1}_k), \quad (3.8)$$

where  $\hat{\psi}_{0n} = \ln(\hat{\Gamma}_{0n})$ ,  $\tilde{\psi}_i = \ln(\hat{\Gamma}_i)$ ,  $\tilde{\boldsymbol{\psi}}_n = (\tilde{\psi}_1, \dots, \tilde{\psi}_k)'$ ,  $\mathbf{1}_k = (1, \dots, 1)'$  a  $k$ -tuple of 1's and  $\widehat{\mathbf{W}}_n = \text{diag}(\hat{\omega}_1, \dots, \hat{\omega}_k)$  is a consistent estimator of

$\mathbf{W} = \text{diag}(\omega_1, \dots, \omega_k)$ . Note that  $n = \sum_{i=1}^k (n_{i1} + n_{i2})$ . As  $n \rightarrow \infty$ ,  $\mathcal{L}_n$  follows approximately a central chi-squared distribution under  $H_0$  with  $(k-1)$  degrees of freedom (df). This test statistic measures the departure of the  $k$  GORs from their common GOR. Thus, for the test decision at the  $\alpha$ -level, compare the observed value of  $\mathcal{L}_n$  to the critical value of the chi-square distribution with  $(k-1)$  df. Then  $H_0$  is rejected at the  $\alpha$ -level if the observed  $\mathcal{L}_n > \chi_{k-1}^2(\alpha)$ . Now an estimator of  $\mathbf{\Gamma} = (\Gamma_1, \dots, \Gamma_k)'$  say

$\hat{\mathbf{\Gamma}}_n^{PT} = \exp(\hat{\boldsymbol{\psi}}_n^{PT})$  is defined as

$$\hat{\boldsymbol{\psi}}_n^{PT} = \begin{cases} \hat{\psi}_{0n} \mathbf{1}_k & \text{if } \mathcal{L}_n < \chi_{k-1}^2(\alpha) \\ \tilde{\boldsymbol{\psi}}_n & \text{if } \mathcal{L}_n \geq \chi_{k-1}^2(\alpha), \end{cases} \quad (3.9)$$

which is a choice between the two estimators based on the test outcome. This estimator is known as the preliminary test estimator (PTE) of  $\mathbf{\Gamma}$  (cf. Saleh, 2006, p.55). The PTE is a precursor to the Stein's shrinkage estimator and was proposed and investigated by many authors (Bancroft, 1944; Han and Bancroft, 1968; Khan and Saleh, 1997; Khan and Hoque, 2002; Khan, 2003; and Khan, 2008).

The PTE depends on the choice of  $\alpha$ . For a better estimator of  $\mathbf{\Gamma}$ , the estimator is made dependent on the test statistic,  $\mathcal{L}_n$  instead of the level of significance to produce a Stein-type shrinkage estimator (SE) defined as  $\hat{\mathbf{\Gamma}}_n^S = \exp(\hat{\boldsymbol{\psi}}_n^S)$ , where

$$\hat{\boldsymbol{\psi}}_n^S = \hat{\psi}_{0n} \mathbf{1}_k + (1 - (k-3)\mathcal{L}_n^{-1})(\tilde{\boldsymbol{\psi}}_n - \hat{\psi}_{0n} \mathbf{1}_k), \quad (k \geq 4). \quad (3.10)$$

Details on the derivation of the SE can be found in Stein (1956). If  $\mathcal{L}_n \rightarrow \infty$ , then  $\hat{\boldsymbol{\psi}}_n^S \rightarrow \tilde{\boldsymbol{\psi}}_n$  and if  $(k-3)\mathcal{L}_n^{-1}$  is near 1, then we choose  $\hat{\psi}_{0n} \mathbf{1}_k$ , which is similar to PTE of  $\boldsymbol{\psi}$ . The estimator  $\hat{\mathbf{\Gamma}}_n^S$  is similar to the empirical Bayes type estimator (Saleh, 2006, p.154) and known as quasi-empirical Bayes estimator of  $\mathbf{\Gamma}$ . Another modification to  $\hat{\mathbf{\Gamma}}_n^S$  is given by  $\hat{\mathbf{\Gamma}}_n^{S+} = \exp(\hat{\boldsymbol{\psi}}_n^{S+})$

$$\hat{\psi}_n^{S+} = \hat{\psi}_{0n} \mathbf{1}_k + (1 - (k-3)\mathcal{L}_n^{-1}) I(\cdot)(\mathcal{L}_n > k-3)(\tilde{\psi}_n - \hat{\psi}_0 \mathbf{1}_k), \quad (k \geq 4) \quad (3.11)$$

where  $I(\cdot)$  is an indicator function. This estimator is similar to the positive-rule Stein-type estimator (PRSE). Here  $\hat{\psi}_n^{S+}$  is obtained by adjusting the common estimate  $\hat{\psi}_{0n}$  to an amount  $(1 - (k-3)\mathcal{L}_n^{-1})I(\mathcal{L}_n > k-3)(\tilde{\psi}_n - \hat{\psi}_0 \mathbf{1}_k)$ . Then, we obtain  $k$  predicted values of the GORs which clusters  $\Gamma_{0n}$  more closely than the original GORs. Moreover, the  $100(1 - \alpha)\%$  CI set has at least  $100(1 - \alpha)\%$  as the coverage probability, which may reach close to 1 near the null hypothesis of the equality of the effect measure (Hwang, 1982). If the number of studies are greater than 4, one is advised to use  $\hat{\Gamma}_n^{S+}$  for statistical inference; otherwise, PTE may be used (Casella, 1985) and details about the importance of shrinkage estimators in statistical literature can be found in Efron (1995), Judge and Bock, 1978 and Saleh, 2006, 125 – 211.

### 3.3 Improved Meta-Analysis

Pooling the study effect for homogeneous studies in meta-analysis has an statistical agreement of using a common study effect of the studies, while using a common pooled study effect for the heterogeneous studies lacks such an statistical agreement. To get rid of the problems of interpretation in the case of single weighted averages of heterogeneous GORs, we apply the following improved meta-analysis methodology (Saleh et al., 2006): Firstly, a common value of the GOR is computed as if the hypothesis of homogeneous effect sizes was not rejected. Secondly, a test of significance for homogeneity is conducted using the divergent statistic,  $\mathcal{L}_n$  from equation (3.8) which measures the departure of each of the log GORs from the sample pooled

log GOR. Finally, to produce the predicted GOR (PGOR), the treatment effect is estimated combining the test of homogeneity (heterogeneity) with the usual estimated GORs to achieve the shrinking towards to pooled GOR. The process subsumes the effect of heterogeneity, and is computed by the following formula:

$$Ln(\hat{\Gamma}_{iP}) = \text{pooled } ln(\hat{\Gamma}) + c(\text{observed } ln(\hat{\Gamma}_i) - \text{pooled } ln(\hat{\Gamma})), \quad (3.12)$$

where  $\hat{\Gamma}_{iP} = \text{PGOR}$ ,  $\hat{\Gamma}_i = \text{estimated GOR}$ ,  $c = 1 - \frac{[\text{degree of freedom}] - 2}{\chi^2\text{-value}}$ . The computational formula for  $100(1 - \alpha)\%$  CIs for individual effect size and meta-analysis are given by

$$\exp[ln(\hat{\Gamma}_{iP}) \pm z_{\alpha/2}\omega_i^{-\frac{1}{2}}], \quad (3.13)$$

and

$$\exp[ln(\hat{\Gamma}_P) \pm z_{\alpha/2}\omega^{-\frac{1}{2}}], \quad (3.14)$$

where  $z_{\alpha/2}$  is the  $100 \times \alpha/2$  percentage point of a standard normal distribution.

### 3.4 Applications

In this section, three different meta-analyses on several studies with different degree of heterogeneity are investigated. The first example, tacrine trials of Alzheimer's disease with insignificant heterogeneity and the second example, misoprostol trials with moderately insignificant heterogeneity are presented in subsections 3.4.1 and 3.4.2 respectively. The third example, simulated trials with significant heterogeneity is provided in subsection 3.4.3. The three examples clearly illustrate detailed implementation process and usefulness of the proposed method.

### 3.4.1 Tacrine Trials: Insignificant Heterogeneity

The first example uses data from Whitehead et al. (2001) consisting of 5 RCTs of the anti-cholinesterase drug tacrine in patients with Alzheimer's disease. The categories of this trials are made by the Clinical Global Impression of Change scale (CGIC). The CGIC is based on a seven point scale where 1, 2 and 3 represent 'very much improved', 'much improved' and 'minimally improved', 4 indicates 'no change', and 5, 6 and 7 represent 'minimally worse', 'much worse' and 'very much worse'. To make the example simple and illustrate the computations, we combine categories 1, 2, and 3 representing 'improvement' and categories 5, 6 and 7 representing 'worse' to make three ordered categories (improvement, no change and worse).

The display in Figure 3.1 for the GOR and the associated 95% CIs show that the vertical line at the pooled GOR (1.58) goes through all the CIs of the effects of these trials. Hence the display suggests that these studies are homogeneous. This conclusion is also supported by the divergent statistic in equation (3.8), which is exactly the same as the  $Q$ -statistic. For the given data  $\mathcal{L}_n = 1.83$  with 4 degrees of freedom. So the null hypothesis of equality of the GORs is not rejected at the 5% level of significance ( $p$ -value = 0.7667). The heterogeneity statistic (Higgins and Thompson, 2002)  $I^2 = 0\%$ [0%, 98.03%] also suggests that these trials are homogeneous.

Table 3.1 shows the GORs, their 95% CIs (CI1), and the predicted GORs (PGORs) and their 95% CIs (CI2). Meta-analysis of these trials is found to be 1.58 indicating that there are 1.58 times as many tacrine-placebo pairs in the sample for which tacrine has improved the alzheimer's disease as there are pairs for which placebo has improved the alzheimer's disease. The forest plot, divergent statistic, and  $I^2$  statistic give the same impression that the effects of the tacrine trials to be homogeneous and the values of the predicted

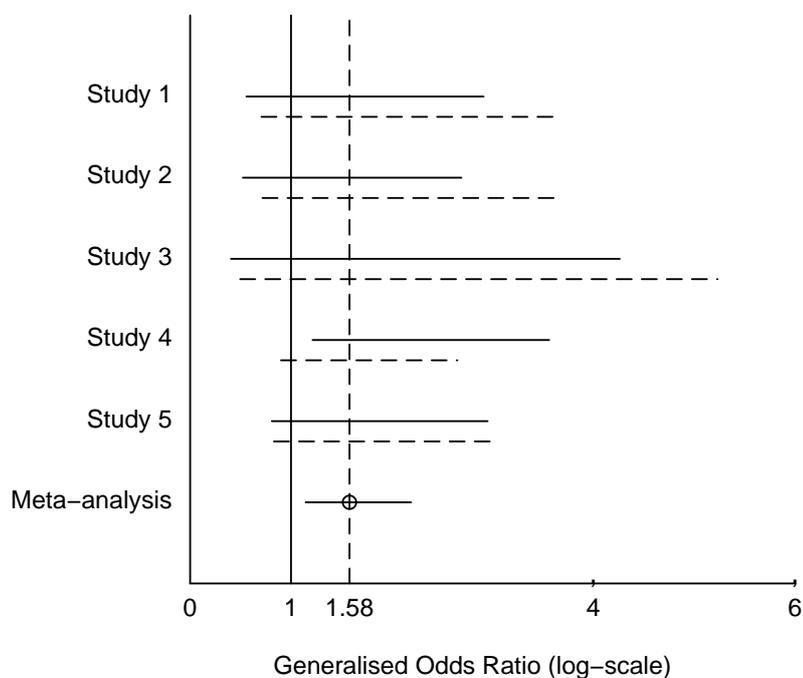


Figure 3.1: 95% CIs for GOR (solid horizontal lines) and predicted GOR (dashed horizontal lines) for tacrine trials

GOR are still concentrating around the common weighted average. Clearly both the GOR and PGOR based CIs are crossing the vertical line at the pooled effect. This is expected for any meta-analysis with homogeneous study effects. Statistical programming in R has been used for computation and plots through out this thesis.

Table 3.1: Five randomised trials of anti-cholinesterase drug tacrine

| Study | Treat.  | W  | NC  | I   | Total | Weight | GOR  | CI1          | PGOR | CI2          |  |
|-------|---------|----|-----|-----|-------|--------|------|--------------|------|--------------|--|
| 1     | Tacrine | 24 | 45  | 27  | 96    | 15.38  | 1.27 | (0.56, 2.91) | 1.62 | (0.71, 3.69) |  |
|       | Placebo | 32 | 54  | 24  | 110   |        |      |              |      |              |  |
| 2     | Tacrine | 60 | 180 | 133 | 373   | 15.64  | 1.18 | (0.52, 2.69) | 1.63 | (0.72, 3.69) |  |
|       | Placebo | 14 | 35  | 23  | 72    |        |      |              |      |              |  |
| 3     | Tacrine | 11 | 24  | 33  | 68    | 7.57   | 1.31 | (0.40, 4.26) | 1.61 | (0.50, 5.23) |  |
|       | Placebo | 13 | 17  | 23  | 53    |        |      |              |      |              |  |
| 4     | Tacrine | 79 | 175 | 127 | 381   | 36.26  | 2.08 | (1.21, 3.56) | 1.54 | (0.90, 2.65) |  |
|       | Placebo | 65 | 73  | 32  | 170   |        |      |              |      |              |  |
| 5     | Tacrine | 3  | 19  | 17  | 39    | 25.14  | 1.55 | (0.81, 2.95) | 1.59 | (0.83, 3.03) |  |
|       | Placebo | 8  | 18  | 15  | 41    |        |      |              |      |              |  |
| Meta  |         |    |     |     |       |        | 1.58 | (1.15, 2.19) |      |              |  |

Here W = Worse, NC = No change and I = Improved are the three rearranged ordered categories from the Clinical Global Impression of Change scale (CGIC), Total = W+NC+I, CI1 = confidence intervals of GORs and CI2 = confidence intervals of PGORs.

### 3.4.2 Misoprostol Trials: Moderately Insignificant Heterogeneity

This subsection uses 10 randomised trials of misoprostol from Whitehead and Jones (1994) investigating the prevention of injury to the gastric mucosa. The subjects had no damage at the initial endoscopy. The results of the final endoscopy were used as the primary response variable. The extent of the damage was recorded in the form of the ordered categorical score for these studies. The data can be fitted into a  $j \times l$  contingency table, where  $j = 2$ , represents the number of comparison groups and  $l = 2, 3, 4$  and  $5$  represent the number of outcome categories. Studies 1 – 6 and 12 used five ordered categories and studies 7 – 8 and 11 used 3 ordered categories. Studies 9 – 10 have only two categories. The second category in studies 9 – 10 are the same as the 5th category of studies 1 – 6 and 12, third category of studies 7 – 8 and 11 and study 13 had no such outcome category representing ‘ulcer of any size’. Pooling for meta-analysis of all these studies may be inappropriate.

Therefore, we have rearranged these categories to three ordered categories (1 = 1, 2 + 3 + 4 = 2 and 5 = 3 for studies 1 – 6 and 12) namely, ‘1 = no erosions’, ‘2 = erosions’, and ‘3 = ulcer’ so that we can find the pooled GOR. We exclude studies 9 – 10 and 13 as they do not meet the inclusion criteria.

Table 3.2: Randomised trials of misoprostol by endoscopic classification.

| Study | Treat.  | NE | E  | U  | T   | W     | GOR   | CI1             | PGOR  | CI2            |
|-------|---------|----|----|----|-----|-------|-------|-----------------|-------|----------------|
| 1     | Misop.  | 21 | 8  | 0  | 29  | 6.94  | 28.53 | (4.35, 187.10)  | 8.65  | (1.32, 56.70)  |
|       | Placebo | 2  | 15 | 13 | 30  |       |       |                 |       |                |
| 2     | Misop.  | 17 | 13 | 0  | 30  | 1.86  | 68.64 | (1.82, 2583.20) | 12.31 | (0.33, 463.20) |
|       | Placebo | 0  | 17 | 13 | 30  |       |       |                 |       |                |
| 3     | Misop.  | 20 | 10 | 0  | 30  | 10.96 | 5.45  | (1.22, 24.30)   | 4.45  | (1.00, 19.90)  |
|       | Placebo | 8  | 17 | 5  | 30  |       |       |                 |       |                |
| 4     | Misop.  | 20 | 8  | 2  | 30  | 1.84  | 23.12 | (0.60, 894.60)  | 7.95  | (0.21, 307.6)  |
|       | Placebo | 0  | 12 | 17 | 29  |       |       |                 |       |                |
| 5     | Misop.  | 1  | 9  | 0  | 10  | 1.27  | 15.71 | (0.19, 1277.50) | 6.81  | (0.08, 553.60) |
|       | Placebo | 0  | 4  | 6  | 10  |       |       |                 |       |                |
| 6     | Misop.  | 93 | 9  | 1  | 103 | 19.99 | 3.13  | (0.99, 9.10)    | 3.49  | (1.15, 10.60)  |
|       | Placebo | 85 | 24 | 5  | 114 |       |       |                 |       |                |
| 7     | Misop.  | 61 | 12 | 0  | 73  | 21.39 | 3.12  | (1.07, 9.10)    | 3.56  | (1.22, 10.40)  |
|       | Placebo | 49 | 28 | 3  | 80  |       |       |                 |       |                |
| 8     | Misop.  | 45 | 1  | 0  | 46  | 3.79  | 3.99  | (0.31, 50.70)   | 3.92  | (0.31, 49.90)  |
|       | Placebo | 65 | 6  | 3  | 74  |       |       |                 |       |                |
| 11    | Misop.  | 30 | 1  | 1  | 32  | 4.51  | 10.72 | (0.81, 85.60)   | 5.27  | (0.51, 54.30)  |
|       | Placebo | 20 | 11 | 7  | 38  |       |       |                 |       |                |
| 12    | Misop.  | 56 | 20 | 0  | 76  | 27.45 | 1.75  | (0.68, 4.50)    | 2.82  | (1.09, 7.30)   |
|       | Placebo | 50 | 32 | 0  | 83  |       |       |                 |       |                |
| Meta  |         |    |    |    |     |       | 3.88  | (2.36, 6.37)    |       |                |

Here, NE = no erosions, E = erosions, U = ulcer, T = total, W = weight, CI1 = confidence intervals of GORs, CI2 = confidence intervals of PGORs.

So the rearranged data represents a  $2 \times 3$  contingency table for each of these studies. Studies with zero cell frequencies are analysed adding  $1/3$  to every cell of each of the  $2 \times 3$  table.

In Table 3.2, we present the GOR and predicted GOR values of these trials along with their CIs. The chi-square value  $Q = 11.71$ ,  $df = 9$ , and  $p$ -value = 0.2304 suggest insignificant heterogeneity.  $I^2 = 23.12\%$  [0%, 62.58%]

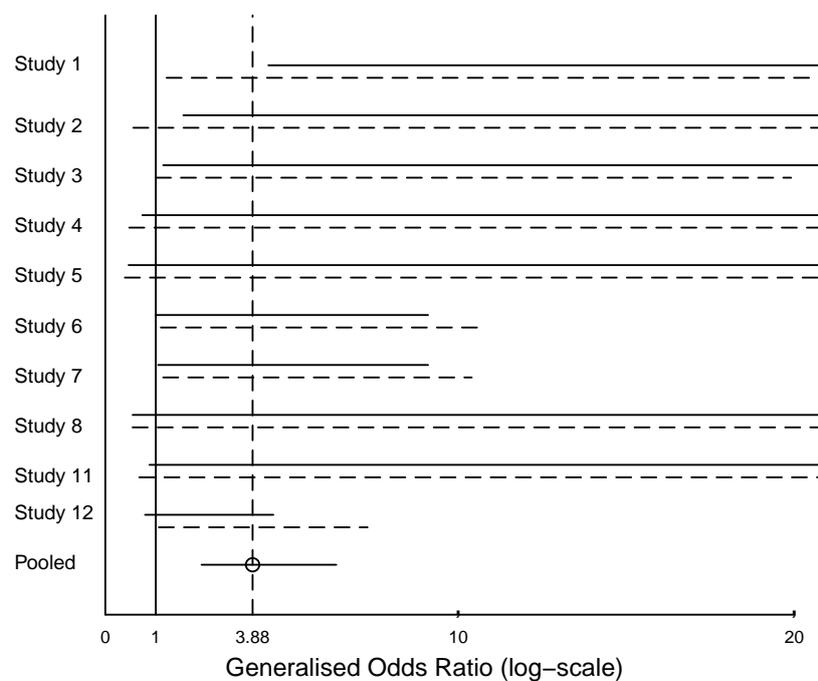


Figure 3.2: 95% CIs for GOR (solid horizontal lines) and predicted GOR (dashed horizontal lines) for the randomised trials of misoprostol by endoscopic classification

suggesting 23.12% of the variability is due to the heterogeneity. Thus, there exists insignificant heterogeneity. Figure 3.2 depicts the CIs of the original GORs in solid horizontal lines and predicted GORs in dashed horizontal lines. It is observed that the vertical line at the common pooled GOR = 3.88 is passing through all the CIs of the original GORs except the first study.

The calculated predicted GORs are more concentrated around the pooled GOR and also the vertical line of the pooled GOR is now crossing all the CIs

of the predicted GORs suggesting a justification of the use of PGORs. Thus, we have shown that the use of predicted GORs are improving the estimates of the GORs with ordinal data in meta-analysis in the presence of insignificant heterogeneity. This is also discussed elaborately in subsection 3.4.3 in the context of significant heterogeneity.

### 3.4.3 Simulated Trials: Significant Heterogeneity

In this section, we generate data for 10 studies using the multinomial sampling procedure. For simulation the cell probabilities are selected in such a way that the outcomes are heterogeneous in order to show the effectiveness of the proposed procedure in heterogeneous data. The simulated data is presented in Table 3.3 along with their weights, GORs, CIs of GORs (CI1), PGORs, CIs of PGORs (CI2) and the common GOR. For the simulated data the heterogeneity statistic  $\mathcal{L}_n = 32.24$  with  $p$ -value = 0.0001 and  $I^2 = 72.09\%$  [47.06%, 85.28%]. The pooled GOR for these trials is 1.48.

In Figure 3.4 the graphical representation shows rays emitting from the pooled GOR to the GORs and the predicted GORs on the vertical lines labeled as the GORs and predicted GORs. The display suggests that the predicted GORs are shrinking towards the pooled GOR. It is observed from Figure 3.3 that the CIs of some studies, for example studies 1, 4, 5, 7, 9 and 10, have moved to the right while that for studies 2, 3, 6 and 8 have moved to the left. The vertical line at the common GOR does not go through all the CIs of these trials. The CI for the study 2 is away to the right while that for the study 4 is away to the left from the common GOR at 1.48.

This indicates that study 2 with GOR = 6.28 and study 4 with GOR = 0.22 may be with extreme effect sizes. We compute the  $\mathcal{L}_n$  statistic by excluding one extreme study at a time. After excluding study 2 we find

Table 3.3: Ten simulated heterogeneous studies.

| Study | Groups    | C1 | C2 | C3 | T   | W     | GOR  | CI1            | PGOR | CI2          |
|-------|-----------|----|----|----|-----|-------|------|----------------|------|--------------|
| 1     | Treatment | 2  | 24 | 3  | 29  | 3.91  | 1.18 | (0.12, 11.24)  | 1.24 | (0.13,11.81) |
|       | Placebo   | 3  | 22 | 5  | 30  |       |      |                |      |              |
| 2     | Treatment | 24 | 5  | 1  | 30  | 11.51 | 9.38 | (2.52, 34.83)  | 6.28 | (1.69,23.34) |
|       | Placebo   | 7  | 19 | 4  | 30  |       |      |                |      |              |
| 3     | Treatment | 26 | 1  | 3  | 30  | 3.02  | 8.62 | (0.67, 111.55) | 5.88 | (0.45,76.12) |
|       | Placebo   | 4  | 21 | 5  | 30  |       |      |                |      |              |
| 4     | Treatment | 2  | 24 | 4  | 30  | 6.93  | 0.13 | (0.02, 0.70)   | 0.22 | (0.04,1.19)  |
|       | Placebo   | 18 | 9  | 2  | 29  |       |      |                |      |              |
| 5     | Treatment | 1  | 8  | 1  | 10  | 1.30  | 0.01 | (0.00, 0.60)   | 0.03 | (0.00,1.71)  |
|       | Placebo   | 9  | 1  | 0  | 10  |       |      |                |      |              |
| 6     | Treatment | 8  | 84 | 11 | 103 | 13.74 | 4.39 | (1.32, 14.60)  | 3.47 | (1.04,11.53) |
|       | Placebo   | 18 | 10 | 86 | 114 |       |      |                |      |              |
| 7     | Treatment | 16 | 50 | 7  | 73  | 35.95 | 1.46 | (0.69, 3.07)   | 1.46 | (0.70,3.08)  |
|       | Placebo   | 14 | 53 | 13 | 80  |       |      |                |      |              |
| 8     | Treatment | 37 | 1  | 8  | 46  | 3.91  | 3.93 | (0.41, 37.34)  | 3.18 | (0.33,30.22) |
|       | Placebo   | 14 | 55 | 5  | 74  |       |      |                |      |              |
| 9     | Treatment | 7  | 19 | 6  | 32  | 8.16  | 0.29 | (0.06, 1.37)   | 0.41 | (0.09,1.96)  |
|       | Placebo   | 26 | 6  | 6  | 38  |       |      |                |      |              |
| 10    | Treatment | 3  | 64 | 9  | 76  | 11.57 | 0.78 | (0.21, 2.91)   | 0.90 | (0.24,3.34)  |
|       | Placebo   | 11 | 59 | 13 | 83  |       |      |                |      |              |
|       | Meta      |    |    |    |     |       | 1.48 | (0.95, 2.31)   |      |              |

Here C1=category 1, C2=category 2, C3=category 3, T=total, W=weight, CI1=confidence intervals of GOR, CI2=confidence intervals of PGOR.

$\mathcal{L}_n = 23.67$  with  $p$ -value = 0.0048 and  $I^2 = 66.20\%$  [31.46%, 83.33%]. This implies rejection of the null hypothesis of the equality of the GORs with a reduction of 8.67 from the chi-square value of original 10 studies. The pooled GOR excluding study 2 is 1.17. Dropping study 4 yields the same values for  $\mathcal{L}_n$  and  $I^2$ . This also suggests rejection of the null hypothesis. The pooled GOR of the remaining 9 study effects is 1.78. This directs us to exclude study 4 from the meta-analysis as study 2 contains more information than study 4 in terms of relative weight. Dixon's test (Dixon, 1950) for outliers also produces the same result. Thus, we recompute the predicted GORs excluding study 4 and present the recalculated results in Table 3.4 along

with the 95% CIs.

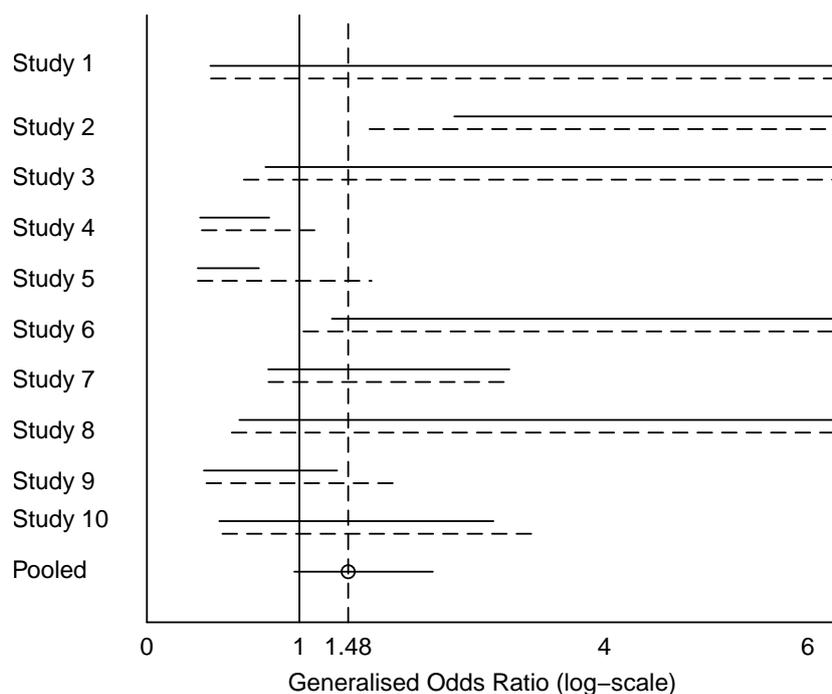


Figure 3.3: The horizontal solid lines represent CIs of GOR and dashed-lines CIs of predicted GOR

The graphical display in Figure 3.5 shows that the predicted GOR values are now more concentrated around the pooled common GOR (= 1.78) and the vertical line at the new common GOR at 1.78 goes through all the CIs of the predicted GORs. This suggests a statistical agreement about the power of the new meta-analysis which was initially rejecting the null hypothesis of equality at the 5% level of significance. Therefore, the common GOR value of 1.78 implies that there are 1.78 times as many treatment-placebo pairs in

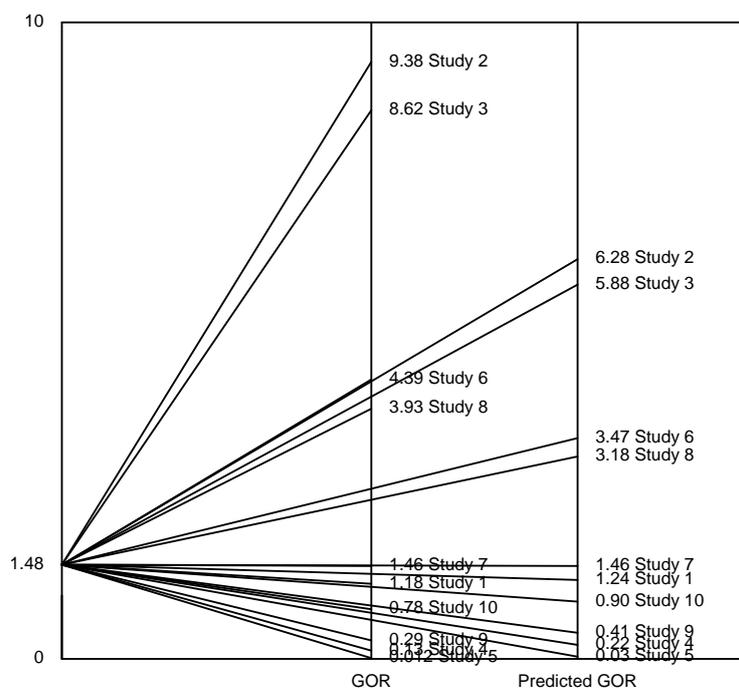


Figure 3.4: Display of original GOR and predicted GOR from ten simulated heterogeneous trials

the sample for which the treatment improve the disease as there are pairs for which placebo improve the disease. It is evident from this analysis that the PGORs provides statistically valid meta-analysis for ordinal data from heterogeneous trails. Figure 3.6 depicts the improvement in the homogeneity of the predicted GORs after deletion of the extreme trial. In this study, we have shown that the traditional meta-analysis has been improved using the quasi-empirical Bayes estimation and the Dixon's test of outliers. These improved estimates of the GORs along with their CIs make easier to redress

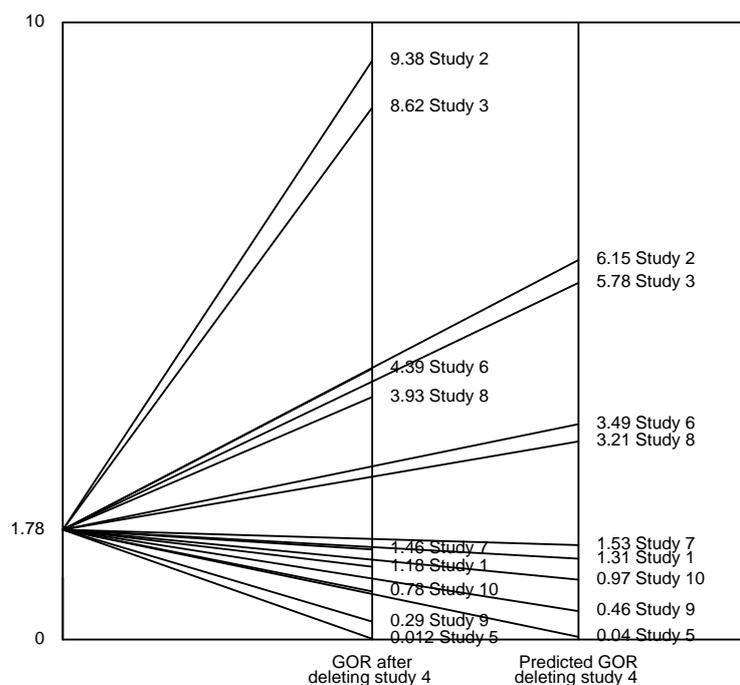


Figure 3.5: Display of GOR and predicted GOR values after deleting Study 4

the issue of heterogeneity among diverse studies for ordinal data in order to ensure the validity and trustworthy meta-analysis.

### 3.5 Conclusion

Heterogeneity is one of the most problematic aspects in many meta-analyses. We have demonstrated a method to remedy the problem of heterogeneity in meta-analysis for ordinal categorical data. To improve the pooled estimator

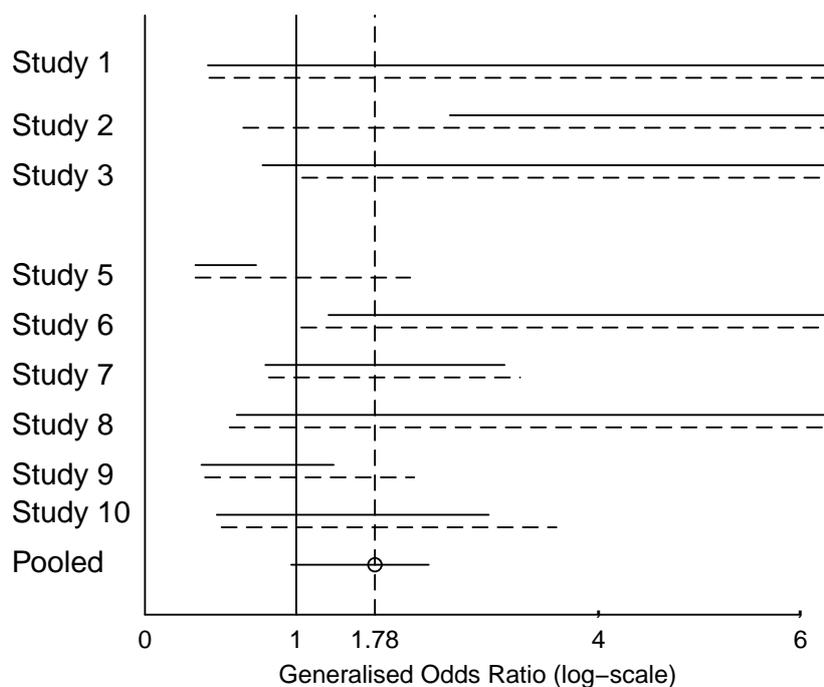


Figure 3.6: 95% CIs after deleting Study 4

Table 3.4: Predicted GOR of nine trials deleting Study 4

| Study | Study 1      | Study 2      | Study 3       | Study 4 | Study 5     |
|-------|--------------|--------------|---------------|---------|-------------|
| GOR   | 1.18         | 9.38         | 8.62          | deleted | 0.012       |
| 95%CI | (0.12,11.25) | (2.52,34.84) | (0.67,111.55) | deleted | (0.00,0.60) |
| PGOR  | 1.31         | 6.15         | 5.78          | deleted | 0.043       |
| 95%CI | (0.14,12.47) | (0.48,79.61) | (1.06,31.33)  | deleted | (0.00,2.13) |

| Study | Study 6      | Study 7     | Study 8      | Study 9     | Study 10     |
|-------|--------------|-------------|--------------|-------------|--------------|
| GOR   | 4.39         | 1.46        | 3.93         | 0.29        | 0.78         |
| 95%CI | (1.32,14.60) | (0.69,3.07) | (0.41,37.34) | (0.06,1.37) | (0.212,2.91) |
| PGOR  | 3.49         | 1.53        | 3.21         | 0.46        | 0.97         |
| 95%CI | (1.05,11.61) | (0.73,3.22) | (0.34,30.54) | (0.10,2.17) | (0.26,3.58)  |

a quasi-empirical Bayes method based on the preliminary test approach is presented. The use of the predicted GOR improves the meta-analysis significantly in terms of shrinking the CIs of the study effects towards the pooled effect. Three examples are discussed with different degree of heterogeneity. The standard meta-analysis for homogeneous studies are simple and straightforward. We have shown that our method improves the meta-analysis in the presence of insignificant and significant heterogeneity. With the use of the predicted GOR the vertical line at the pooled GOR is passing through all the CIs of the effect measures suggesting an statistical agreement of using the new pooled effect measure. For insignificant heterogeneity we do not need to exclude any study whereas for significant heterogeneity the method suggests to exclude the extreme study. After excluding the extreme study, the statistical agreement in terms of shrinking the CIs of the study effects towards the pooled effect is achieved.

Dropping extreme studies without further investigation of the clinical reasons is unwise and unwarranted. Normally clinicians would search for any reasonable factors that may have caused it, and explain its reasons and impact. However, standard meta-analysis leads to invalid and misleading interpretation and conclusion when the studies are heterogeneous (Saleh et al., 2006). The proposed method provides a procedure to conduct valid meta-analysis of ordinal data in the presence of heterogeneity. The graphical analysis (ray plot) is an integral part of the method that is very useful in understanding the mechanism that improves the meta-analysis by implementing the idea of shrinking and using the test of homogeneity.

## Chapter 4

# GOR for Binary Outcomes: A Comparison

## 4.1 Introduction

For binary outcomes relative risk (RR) and odds ratio (OR) are widely used as effect measures to analyse data produced by RCTs. In this chapter, we employ GOR to measure effect size of binary outcomes in meta-analysis of RCTs. The GOR can be a very useful effect measure in meta-analysis not only for binary outcomes but also for more than two outcome categories in RCTs (Hossain and Khan, 2011). It is simple to compute and its interpretation is very straightforward. It can be used for the continuous ordinal outcomes in RCTs. It can also be used with multi arm trials (more than two comparison groups).

In this chapter, we cover the fixed effects model (FEM) (Birge, 1932; Cochran, 1937) and random effects model (REM) (DerSimonian and Laird, 1986) using ordinary OR. Three alternative methods for combining results from binary outcomes are also presented for meta-analysis. The first alternative method is based on the sample size weight using  $OR_G$ . This is a FEM devised from Edwardes and Baltzan (2000). The other alternative methods use weights from the estimated variance under independent binomial distribution using GOR as the effect measure. Of these, the first method is a FEM approach using GOR and the second method is a REM approach using GOR. The results from a meta-analysis of six RCTs (Memon et al., 2011) evaluating the efficacy and drawbacks of limited (D1) versus extended lymphadenectomy (D2) for proven gastric adenocarcinoma are compared.

## 4.2 GOR: As Binary Effect Measure

Consider a RCT setting for binary outcomes as in Table 4.1.

The generalised odds ratio (GOR) can be used for such binary outcomes.

Table 4.1: Data of a single RCT

| Groups    | Failure/Dead | Success/Alive | Size        |
|-----------|--------------|---------------|-------------|
| Treatment | a = $X_{11}$ | b = $X_{12}$  | $n_1 = a+b$ |
| Placebo   | c = $X_{21}$ | d = $X_{22}$  | $n_2 = c+d$ |

In fact, GOR can be used as an effect measure for a number of situations in meta-analysis, such as for binary outcomes, more than two ordinal category outcomes and latent continuous outcomes. It has been widely used in psychology and education disciplines for over 30 years. The GOR for the  $i$ th study is defined as

$$\Gamma_i = (\Pi_d)^{-1}\Pi_c, \tag{4.1}$$

where  $\Pi_c = \sum_{r=1}^{L-1} \sum_{s=r+1}^L \Pi_{r|1}\Pi_{s|2}$  and  $\Pi_d = \sum_{r=2}^L \sum_{s=1}^{r-1} \Pi_{r|1}\Pi_{s|2}$ . For binary outcomes  $L = 2$ , so  $\Gamma_i$  is estimated by its maximum likelihood estimator, say  $\hat{\Gamma}_i$ , as

$$\hat{\Gamma}_i = (\hat{\Pi}_d)^{-1}\hat{\Pi}_c = (\hat{\Pi}_{2|1}\hat{\Pi}_{1|2})^{-1}\hat{\Pi}_{1|1}\hat{\Pi}_{2|2} = \frac{ad}{bc}, \tag{4.2}$$

where  $\Pi_c$  denotes the probability that the response of a randomly selected subject from group 2 is severer than the response of a randomly selected subject from group 1, and  $\Pi_d$  denotes the probability that the response of a randomly selected subject from group 1 is severer than the response of a randomly selected subject from group 2.

### 4.3 Review of Methods using OR and OR<sub>G</sub>

In this section we discuss the widely used general FEM (Birge, 1932), (Cochran, 1937) and standard REM (DerSimonian and Laird, 1937) for binary outcomes using OR in meta-analysis. A great debate is going on over which model is

theoretically and/or practically superior. Although REMs have been criticised for its unrealistic distribution assumptions (Peto, 1987), it has been argued that they are consistent with the standard specific aims of generalisation (Raudenbush, 1994). Greenland (1994) showed that REMs are more sensitive to publication bias, as those offer greater relative weight to smaller studies. Now-a-days, many meta-analyses employ both FEM and REM for the same set of studies as neither FEM nor REM can be considered as ideal (Thompson, 1993). Although when there is no evidence of heterogeneity, the REM produces the same pooled estimate as of FEM. However, when there is evidence of heterogeneity, FEMs can not take into account this extra variation, whereas REMs do. That indicates a REM may still be worth using as it can not be assumed that true homogeneity exists (Thompson and Pocock, 1991).

However, Shuster (2010) put into serious question the validity of empirically based weighting in random effects meta-analysis and proposed two alternative methods. The first method estimates the arithmetic mean of the population of study effect sizes according to the classical model for random effects meta-analysis and the second method estimates a patient level effect size. In response to the criticism of Waksman, (2010), Thompson and Higgins, (2010), Laird et al, (2010), and Rucker et al (2010), Shuster et al (2010) wrote “we continue to recommend that past meta-analyses which have influenced public policy or clinical paradigms be reanalyzed by unweighted methods”.

Now to discuss the FEM and REM for OR let us consider  $\hat{\Gamma}_i$  is an estimate of the true effect size  $\theta_i$  in the  $i$ th study as  $\hat{\Gamma}_i = \theta_i + e_i$ , where  $e_i$  is the error with which  $\hat{\Gamma}_i$  estimates  $\theta_i$ . For the FEM,  $\text{var}(\hat{\Gamma}_i) = v_i$  and for REM  $\text{var}(\hat{\Gamma}_i) = \tau_\theta^2 + v_i$ , where  $\tau_\theta^2$  is the between study variance and  $v_i$  is the

variance due to sampling error for the  $i$ th study. If  $\tau_\theta^2 = 0$ , the above REM would reduce to the FEM.

### 4.3.1 Fixed Effects Model

The inverse variance-weighted method is one of the most widely used method among the FEMs in meta-analysis. To illustrate this, consider  $T_i$  is the observed estimate for  $i$ th study effect with variance  $v_i$  and  $\theta_i$  is the underlying population effect size, where  $i = 1, \dots, k$ .

Assuming  $T_i$  to be normally distributed, an approximate  $100(1 - \alpha)\%$  confidence interval (CI) for the  $i$ th study effect,  $\theta_i$ , is given by

$$T_i - z_{\alpha/2} \sqrt{1/\omega_i} \leq \theta_i \leq T_i + z_{\alpha/2} \sqrt{1/\omega_i}, \quad (4.3)$$

where  $z_{\alpha/2}$  is the  $\alpha/2$  percentage point of a standard normal distribution and

$$\omega_i = \frac{1}{v_i}. \quad (4.4)$$

Then assuming  $\theta_1 = \theta_2 = \dots, \theta_k = \theta$ , a pooled estimate of the treatment effect is given by

$$\bar{T}. = \frac{\sum_i \omega_i T_i}{\sum_i \omega_i}. \quad (4.5)$$

An estimator of the variance of the pooled estimate  $\bar{T}.$  is given by

$$\text{var}(\bar{T}.) = 1 / \sum_{i=1}^k \omega_i. \quad (4.6)$$

If  $\bar{T}.$  is assumed to be normally distributed, an approximate  $100(1 - \alpha)\%$  CI for the population effect,  $\theta$ , is given by:

$$\bar{T}. - z_{\alpha/2} \sqrt{1 / \sum_{i=1}^k \omega_i} \leq \theta \leq \bar{T}. + z_{\alpha/2} \sqrt{1 / \sum_{i=1}^k \omega_i}, \quad (4.7)$$

where  $z_{\alpha/2}$  is the  $\alpha/2$  percentage point of a standard normal distribution.

### 4.3.2 Random Effects Model

Consider  $\hat{\tau}_\theta^2$  denote the estimate of the between study variance of the study effect sizes  $\tau_\theta^2$ . Define  $\bar{\omega}$  and  $s_W^2$  to be the mean and variance of the weights from the  $k$  studies:

$$\bar{\omega} = \sum_{i=1}^k \omega_i/k \quad (4.8)$$

and

$$s_W^2 = \frac{1}{k-1} \left( \sum_{i=1}^k \omega_i^2 - k\bar{\omega}^2 \right). \quad (4.9)$$

Further, define

$$U = (k-1) \left( \bar{\omega} - \frac{s_W^2}{k\bar{\omega}} \right) \quad (4.10)$$

and

$$Q = \sum_{i=1}^k \omega_i (T_i - \bar{T})^2, \quad (4.11)$$

where  $Q$  is the heterogeneity test statistic, (also known as Cochran's  $\chi^2$  statistic) for testing the  $H_0 = \theta_1 = \theta_2 = \dots, \theta_k = \theta$ . The estimated component of variance due to inter-study variation in effect size  $\hat{\tau}_\theta^2$ , is calculated as

$$\hat{\tau}_\theta^2 = \begin{cases} 0 & \text{if } Q \leq k-1 \\ (Q - (k-1))/U & \text{if } Q > k-1. \end{cases} \quad (4.12)$$

Then adjusted weights  $\omega_i^*$  for each of the studies can be calculated as

$$\omega_i^* = \frac{1}{[1/\omega_i] + \hat{\tau}_\theta^2}. \quad (4.13)$$

A  $100(1 - \alpha)\%$  CI for  $\theta_i$  is given by

$$T_{iR} - z_{\alpha/2}/\sqrt{\omega_i^*} \leq \theta_i \leq T_{iR} + z_{\alpha/2}/\sqrt{\omega_i^*} \quad (4.14)$$

under the assumption of normality of  $T_{iR}$ , where  $T_{iR}$  represents the observed estimate of the  $i$ th study effect size under REM. The point estimate for the mean treatment of all studies,  $\theta$ , can be computed by

$$\bar{T}_{.R} = \sum_{i=1}^k \omega_i^* T_i / \sum_{i=1}^k \omega_i^*. \quad (4.15)$$

The variance of this estimate is

$$\text{var}(\bar{T}_{\cdot R}) = 1 / \sum_{i=1}^k \omega_i^*, \quad (4.16)$$

and if normality of  $\bar{T}_{\cdot R}$  is assumed, a  $100(1 - \alpha)\%$  CI for  $\theta$  is given by

$$\bar{T}_{\cdot R} - z_{\alpha/2} / \sqrt{\sum_{i=1}^k \omega_i^*} \leq \theta \leq \bar{T}_{\cdot R} + z_{\alpha/2} / \sqrt{\sum_{i=1}^k \omega_i^*}. \quad (4.17)$$

### 4.3.3 Sample Size Weight Method

Here we adopt the sample size weight method (Edwardes and Baltzan, 2000) for binary data pooling in meta-analysis. This method does pooling Goodman and Kruskal's  $\gamma = (ad - bc)/(ad + bc)$  instead of pooling  $\text{OR}_{\text{G}} = (1 + \gamma)/(1 - \gamma)$  using the sample size weights  $T_N = n_1 n_2 / (n_1 + n_2)$  as

$$\bar{\gamma} = \sum_i T_{Ni} \gamma_i / \sum_i T_{Ni} \quad (4.18)$$

taking  $\gamma_i, T_{Ni}$  as  $\gamma, T_N$  for the  $i$ th study. Data can be combined for meta-analysis with  $\overline{\text{OR}}_{\text{G}} = (1 + \bar{\gamma}) / (1 - \bar{\gamma})$ .

An asymptotic standard error of  $\log(\text{OR}_{\text{G}})$  is estimated by

$$\text{SE}(\log \text{OR}_{\text{G}}) = 2 \times \text{SE}(\gamma) / (1 - \gamma)^2 \quad (4.19)$$

and a 95% CI for  $\text{OR}_{\text{G}}$  is

$$\{\text{OR}_{\text{G}} \exp[-3.92 \times \text{SE}(\gamma) / (1 - \gamma^2)], \text{OR}_{\text{G}} \exp[3.92 \times \text{SE}(\gamma) / (1 - \gamma^2)]\}, \quad (4.20)$$

where  $\exp(\cdot)$  is the exponential function. This model is a FEM approach. Although the formula for the variance estimate of  $\bar{\gamma}$  is available, there is no well defined variance estimate for the pooled  $\text{OR}_{\text{G}}$ , and hence the CI for pooled  $\text{OR}_{\text{G}}$  is unavailable under this method.

## 4.4 Proposed Method

The proposed method is to use GOR as an effect measure for binary outcomes under both FEM and REM. For illustration consider  $\hat{\Gamma}_i$  representing logarithm of GOR, the observed effect size with variance  $v_i$ , and  $\theta_i$  is the underlying population effect size. Then assuming  $\theta_1 = \theta_2 = \dots = \theta_k = \theta$ , a pooled estimate of the treatment effect can be obtained by

$$\hat{\Gamma}_0 = \frac{\sum_i \omega_i \hat{\Gamma}_i}{\sum_i \omega_i}. \quad (4.21)$$

For an arbitrary number of ordered outcome categories ( $L$ ) in RCTs in which each row is modeled as an independent multinomial distribution, the estimated variance,  $v_i = \hat{\omega}_i^{-1}$  of the  $i$ th study effect is given by

$$\hat{\omega}_i^{-1} = \sum_{l=1}^{L-1} \sum_{j=1}^2 \frac{1}{n_{ij} \hat{\pi}_{ijl} (1 - \hat{\pi}_{ijl})}, \quad (4.22)$$

where  $n_{ij}$  is the total count of  $j$ th group for the  $i$ th study,  $\hat{\pi}_{ijl} = X_{ijl}/n_{ij}$  is the maximum likelihood estimator (MLE) of  $\pi_{ijl}$ ,  $X_{ijl}$  is the count of the  $l$ th category in the  $j$ th group for the  $i$ th study. For binary outcomes the estimated variance expression reduces to

$$\hat{\omega}_i^{-1} = \sum_{j=1}^2 \frac{1}{n_{ij} \hat{\pi}_{ij1} (1 - \hat{\pi}_{ij1})}. \quad (4.23)$$

Assuming  $\hat{\Gamma}_i$ 's are normally distributed, an approximate  $100(1 - \alpha)\%$  CI for the GORs is given by the formula

$$\exp[\hat{\Gamma}_i \pm z_{\alpha/2} \omega_i^{-1/2}]. \quad (4.24)$$

An estimator of the variance of the pooled estimate  $\hat{\Gamma}_0$  is given by

$$\text{var}(\hat{\Gamma}_0) = 1 / \sum_{i=1}^k \omega_i. \quad (4.25)$$

If  $\hat{\Gamma}_0$  is assumed to be normally distributed, an approximate  $100(1 - \alpha)\%$  CI for the population effect,  $\theta$ , is given by

$$\exp[\hat{\Gamma}_0 \pm z_{\alpha/2}\omega^{-1/2}], \quad (4.26)$$

where  $z_{\alpha/2}$  is the  $\alpha/2$  percentage point of a standard normal distribution.

Furthermore, an asymptotic variance of  $\hat{\Gamma}_i$  (Agresti, 1980; Lui, 2004) for binary outcomes can also be found as

$$\text{var}(\hat{\Gamma}_i) = \frac{\pi_{i22}^2\pi_{i11} + (\Gamma_i\pi_{i21})^2\pi_{12}}{n_{i1}\Pi_d^2} + \frac{\pi_{i11}^2\pi_{i22} + (\Gamma_i\pi_{i12})^2\pi_{21}}{n_{i2}\Pi_d^2}. \quad (4.27)$$

Then using the logarithmic transformation a  $100(1 - \alpha)\%$  percent asymptotic confidence interval for  $\Gamma_i$  can be obtained as

$$\left[ \hat{\Gamma}_i \exp(-z_{\alpha/2}\sqrt{\text{var}(\hat{\Gamma}_i)/\hat{\Gamma}_i}), \hat{\Gamma}_i \exp(z_{\alpha/2}\sqrt{\text{var}(\hat{\Gamma}_i)/\hat{\Gamma}_i}) \right]. \quad (4.28)$$

The above procedure is a FEM approach. For REM approach, consider  $\hat{\tau}_\theta^2$  to be the estimate of the between study variance. Define  $\bar{\omega}$  and  $s_W^2$  to be the mean and variance of the weights from the  $k$  studies as

$$\bar{\omega} = \sum_{i=1}^k \omega_i/k \text{ and } s_W^2 = \frac{1}{k-1} \left( \sum_{i=1}^k \omega_i^2 - k\bar{\omega}^2 \right). \quad (4.29)$$

Further define

$$U = (k-1) \left( \bar{\omega} - \frac{s_W^2}{k\bar{\omega}} \right) \text{ and } Q = \sum_{i=1}^k \omega_i (\hat{\Gamma}_i - \bar{\Gamma})^2, \quad (4.30)$$

where  $Q$  is the heterogeneity statistic. The estimated component of variance due to inter-study variation in effect size,  $\hat{\tau}_\theta^2$ , is calculated as

$$\hat{\tau}_\theta^2 = \begin{cases} 0 & \text{if } Q \leq k-1 \\ (Q - (k-1))/U & \text{if } Q > k-1. \end{cases} \quad (4.31)$$

Then adjusted weights  $\omega_i^*$  for each of the studies can be calculated as

$$\omega_i^* = \frac{1}{[1/\omega_i] + \hat{\tau}_\theta^2}. \quad (4.32)$$

A  $100(1 - \alpha)\%$  CI for  $\theta_i$  is given by

$$\exp[\hat{\Gamma}_{iR} \pm z_{\alpha/2}/\sqrt{\omega_i^*}] \quad (4.33)$$

under the assumption of normality of  $\hat{\Gamma}_{iR}$ .

The point estimate for the mean treatment effect of all studies,  $\theta$ , can be computed by

$$\hat{\Gamma}_{0R} = \frac{\sum_{i=1}^k \omega_i^* \hat{\Gamma}_i}{\sum_{i=1}^k \omega_i^*} \text{ with } \text{var}(\hat{\Gamma}_{0R}) = 1/\sum_{i=1}^k \omega_i^*. \quad (4.34)$$

If normality of  $\hat{\Gamma}_{0R}$  is assumed, a  $100(1 - \alpha)\%$  CI for  $\theta$  is given by

$$\exp[\hat{\Gamma}_{0R} \pm z_{\alpha/2}/\sqrt{\sum_{i=1}^k \omega_i^*}]. \quad (4.35)$$

## 4.5 Results

This example uses data from meta-analysis of RCTs evaluating the efficacy and drawbacks of limited (D1) versus extended lymphadenectomy (D2) for proven gastric adenocarcinoma (Memon et al, 2011). Six trials totaling 1876 patients (D1 = 946, D2 = 930) were analyzed. The meta-analysis was prepared in accordance with the Preferred Reporting Items for Systematic reviews and Meta-analyses statement. The six outcome variables analysed included length of hospital stay; overall complication rate; anastomotic leak rate; re-operation rate; 30-day mortality rate and 5-year survival rate. The outcome postoperative complication rate is chosen out of the six outcomes studied because of its significant heterogeneity.

We apply the above five models for estimation and illustrate using these six studies. The first two models are a FEM and a REM using OR, the third model is a FEM using  $OR_G$ , the proposed last two models are a FEM and a REM using GOR.

Table 4.2: OR for total postoperative complications

| Study | Name    | e.t | n.t | e.c | n.c | OR    | CI         | W(f)  | W(r)  |
|-------|---------|-----|-----|-----|-----|-------|------------|-------|-------|
| 1     | Dent    | 3   | 22  | 8   | 21  | 0.26  | 0.06, 1.15 | 1.98  | 7.23  |
| 2     | Robert. | 0   | 25  | 24  | 29  | 0.004 | 0.00, 0.08 | 0.52  | 2.20  |
| 3     | Bonenk. | 128 | 513 | 183 | 483 | 0.54  | 0.41, 0.71 | 60.67 | 32.62 |
| 4     | Cusch.  | 55  | 200 | 92  | 200 | 0.44  | 0.29, 0.67 | 25.78 | 28.10 |
| 5     | Degiuli | 8   | 76  | 14  | 86  | 0.60  | 0.24, 1.53 | 5.18  | 14.36 |
| 6     | Wu      | 8   | 110 | 19  | 111 | 0.38  | 0.16, 0.91 | 5.88  | 15.49 |

Here, e.t = events in the treatment group, n.t = total number of patients in the treatment group, e.c = events in the control group, and n.c = total number of patients in the control group. W(f) = weights from FEM and W(r) = weights from REM

FEM: OR = 0.4894, 95% CI = (0.396, 0.60),  $z = -6.62$ ,  $p$ -value < 0.0001.

REM: OR = 0.4210, CI = (0.2683, 0.6606),  $z = -3.76$ ,  $p$ -value = 0.0002.

The homogeneity statistic,  $Q$  (Cochran, 1954) using OR for these data is  $Q = 11.85$  with 5 df, giving  $p$ -value = 0.0369. This indicates there exists significant heterogeneity in the data set. Using REM of DerSimonian and Laird (1986), we obtain  $\hat{\tau}^2 = 0.1428$ . Heterogeneity statistic  $I^2 = 57.8\%$  [0%, 82.9%] indicates about 58% variation is due to the between study variation.

Table 4.2 shows the data for the six studies, estimated study specific OR, 95% CI, weights for fixed and random effects models. Using fixed effects model, there are 51% reduction of relative odds of developing postoperative complications (OR = 0.49, CI = (0.40, 0.60),  $p$ -value < 0.0001). Under random effects model, there are 58% reduction in relative odds of developing postoperative complications (OR = 0.42, CI = (0.27, 0.66),  $p$ -value = 0.0002).

Table 4.3:  $OR_G$  for total postoperative complication

| Study | Name          | $OR_G$ | Wf(S) | CI          | $Q/I^2$                               |
|-------|---------------|--------|-------|-------------|---------------------------------------|
| 1     | Dent          | 0.256  | 2.21  | 0.06 1.15   | $Q = 302.08, df = 5,$<br>$p < 0.0001$ |
| 2     | Robertson     | 0.004  | 3.59  | 0.00 22.91  |                                       |
| 3     | Bonenkamp     | 0.545  | 57.64 | 0.415 0.715 | $I^2 = 98.34\%$<br>[97.60%, 98.86%]   |
| 4     | Cuschieri     | 0.445  | 25.05 | 0.293 0.675 |                                       |
| 5     | Degiuli       | 0.605  | 5.12  | 0.238 1.533 |                                       |
| 6     | Wu            | 0.379  | 6.39  | 0.158 0.909 |                                       |
|       | Meta-analysis |        | 0.475 | -           |                                       |

Table 4.3 contains study names, estimated  $OR_G$ , sample size weight, CI of individual  $OR_G$  and values of heterogeneity statistics  $Q$  and  $I^2$ . The heterogeneity statistic,  $Q = 302.08$ ,  $df = 5$ ,  $p$ -value  $< 0.0001$  using sample size weights. The  $I^2$  statistic = 98.34% [97.60%, 98.86%] (Higgins and Thompson, 2002). Both the values of the heterogeneity statistics for sample size weight method are significantly higher than other methods. Under sample size weight method the pooled estimate  $OR_G = 0.475$ . As there is no well defined variance estimate for the pooled  $OR_G$  (Edwardes and Baltzan, 2000), we can not find the CI for the pooled  $OR_G$ .

The proposed method conducts meta-analysis both under FEM and REM using GOR. The heterogeneity statistic  $Q = 11.85$ ,  $df = 5$ ,  $p$ -value = 0.0369. This suggests presence of significant heterogeneity among these studies. The statistic  $I^2 = 57.80\%$  [0%, 82.93%]. This indicates about 58% of the total variation is accounted for heterogeneity. Meta-analysis under FEM using GOR = 0.49, CI = (0.40, 0.60), which means 51% reduction in relative odds of developing postoperative complications. Under REM the meta-analysis estimate is GOR = 0.42, CI = (0.27, 0.66). This implies there are 58%

Table 4.4: GOR for total postoperative complications

| Study | Name            | GOR   | Wf(M)        | Wr(M)        | CI             |
|-------|-----------------|-------|--------------|--------------|----------------|
| 1     | Dent et al      | 0.256 | 1.98         | 7.23         | (0.06, 1.15)   |
| 2     | Robertson et al | 0.004 | 0.51         | 2.18         | (0.00, 0.08)   |
| 3     | Bonenkamp et al | 0.545 | 60.67        | 32.62        | (0.415, 0.715) |
| 4     | Cuschieri et al | 0.445 | 25.78        | 28.10        | (0.293, 0.675) |
| 5     | Degiuli et al   | 0.605 | 5.18         | 14.36        | (0.238, 1.533) |
| 6     | Chew-Wun et al  | 0.379 | 5.87         | 15.49        | (0.158, 0.909) |
|       | Meta-analysis   |       | 0.49         | 0.42         |                |
|       | 95% CI          |       | (0.40, 0.60) | (0.27, 0.66) |                |

Heterogeneity test:  $Q = 11.85$ ,  $df = 5$ ,  $p$ -value = 0.0369,

$I^2 = 57.80\%$  [0%, 82.93%]

reduction of relative odds of developing postoperative complications.

It is interesting that the meta-analysis estimates under both FEM and REM using GOR are exactly the same using ordinary OR for both FEM and REMs. This tells us that use of GOR for binary outcomes in meta-analysis is as efficient as OR. The sample size weight method produces a lower pooled estimate than the rests under FEM. This method produces extraordinarily higher  $Q$  value and too sensitive to zero cell counts (Edwardes and Baltzan, 2000).

Standard meta-analysis works fine with homogeneous studies but how trustworthy is meta-analysis that use a single common effect measure for all the studies when there exists significant heterogeneity? Because of the issues (Shuster, 2010; Doi SA et al., 2011) of the REM, when used in a poorly designed studies, it still results in biased estimates even though there is statistical adjustment for effect size heterogeneity (Senn, 2007). Now how

representative is the common meta-analysis value 0.49 under FEM or 0.42 under REM for study Robertson et al?

## 4.6 Improved Meta-Analysis

To overcome the problem we can apply the improved methodology for meta-analysis. The first step in the traditional meta-analysis produce the common effect size = 0.49. In the second step we find the divergent statistic  $\mathcal{L}_n = 11.85$ ,  $df = 5$ ,  $p$ -value = 0.0369. It shows significant evidence that the treatment effects are different in the six trials. Thus, the interpretation of the common weighted average of these trials raise question; to which population of patients does the overall GOR apply? Even in meta-analysis with evidence of homogeneity, the same difficulty arises because of the existence of buried heterogeneity in clinical trials.

Finally, we estimate the predicted GORs (PGORs), the treatment effects combining the test of heterogeneity with the usual estimated GORs. This process subsumes the effect of heterogeneity using the following formula.

$$Ln(\hat{\Gamma}_{iP}) = \text{pooled } ln(\Gamma) + c(\text{observed } ln(\hat{\Gamma}_i) - \text{pooled } ln(\Gamma)), \quad (4.36)$$

where  $\hat{\Gamma}_{iP}$  = PGORs,  $\hat{\Gamma}_i$  = estimated GORs,  $c = 1 - \frac{[\text{degree of freedom}] - 2}{\chi^2\text{-value}}$ .

The computational formula for the  $100(1 - \alpha)\%$  CIs is given by

$$\Gamma_L = \exp(ln(\hat{\Gamma}_{iP}) - z_{\alpha/2}\omega_i^{-\frac{1}{2}}) \leq \theta \leq \exp(ln(\hat{\Gamma}_{iP}) + z_{\alpha/2}\omega_i^{-\frac{1}{2}}) = \Gamma_U. \quad (4.37)$$

For these six studies  $c = 0.7468$ . The six PGOR values and their 95% CI are presented in Table 4.5. The first two rows of the table show the original GOR and their 95% CIs. The graphical display in Figure 4.1 shows the rays emitting from the estimated common GOR toward the sample GORs and the PGORs on the vertical line labeled as the predicted GOR. It is evident

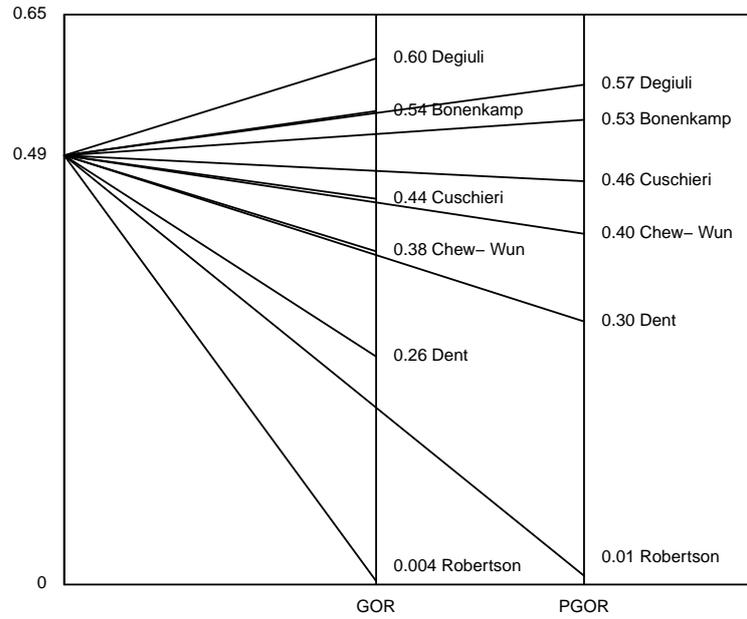


Figure 4.1: Ray plot for GOR of the six studies of gastric carcinoma

Table 4.5: Predicted GOR of six studies of gastric carcinoma

| Study | Dent       | Robert.   | Bonenk.   | Cusch.    | Degiuli   | Wu        |
|-------|------------|-----------|-----------|-----------|-----------|-----------|
| GOR   | 0.26       | 0.00      | 0.54      | 0.44      | 0.60      | 0.38      |
| 95%CI | 0.06,1.15  | 0.00,0.08 | 0.41,0.71 | 0.29,0.67 | 0.24,1.53 | 0.16,0.91 |
| PGOR  | 0.30       | 0.01      | 0.53      | 0.46      | 0.57      | 0.40      |
| 95%CI | 0.07, 1.36 | 0.00,0.28 | 0.40,0.69 | 0.30,0.69 | 0.23,1.45 | 0.17,0.97 |

from the display that the PGORs are more concentrated around the common GOR at 0.49. Figure 4.2 shows the CIs for GOR (solid horizontal lines) and CIs for PGOR (dashed horizontal lines) of the six studies. The vertical line at common pooled value (GOR = 0.49) is passing through the CIs of all the studies except the Robertson et al. The estimated PGOR values are

more concentrated around the common GOR = 0.49 which corresponds to an estimated 51% reduction of the odds of overall complications attributed to D1. From Figure 4.2 we see that the study Robertson is away from

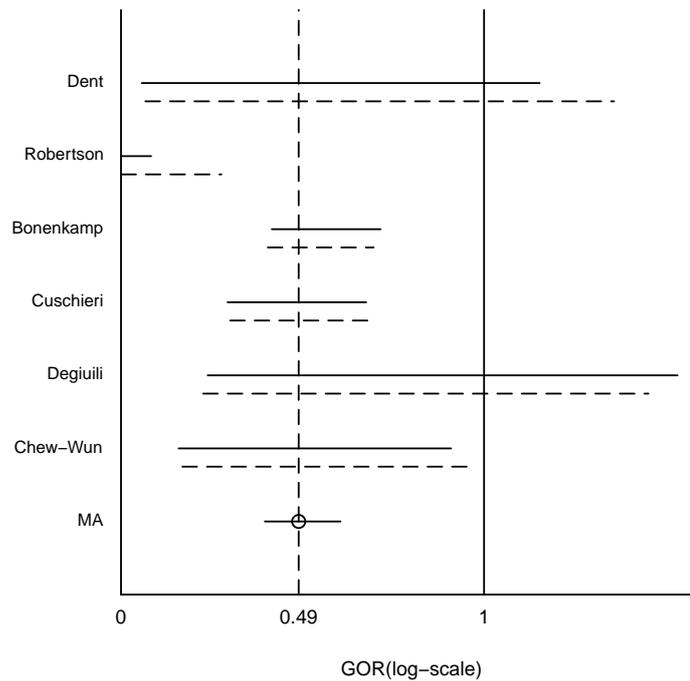


Figure 4.2: 95% CI for GOR of the six studies of gastric carcinoma

the common GOR, therefore may be considered as an outlier although it is difficult to identify the clinical reason initially. Now we compute the  $\mathcal{L}_n$  with the five studies excluding Robertson. In this case  $\mathcal{L}_n = 1.98$ ,  $df = 4$  with  $p\text{-value} = 0.7389$ . When we apply the Dixon's test (Dixon, 1950) for outlier, we also obtain the same result. So the null hypothesis of equality of the GORs is accepted. The recomputed weighted average becomes 0.50. Thus using the formula  $\ln\text{PGOR} = \ln(0.50) + c(\ln\text{GOR} - \ln(0.50))$ , where  $c = -0.0088$ , we find

the PGOR values. Figure 4.3 shows the ray plot and Figure 4.4 shows the

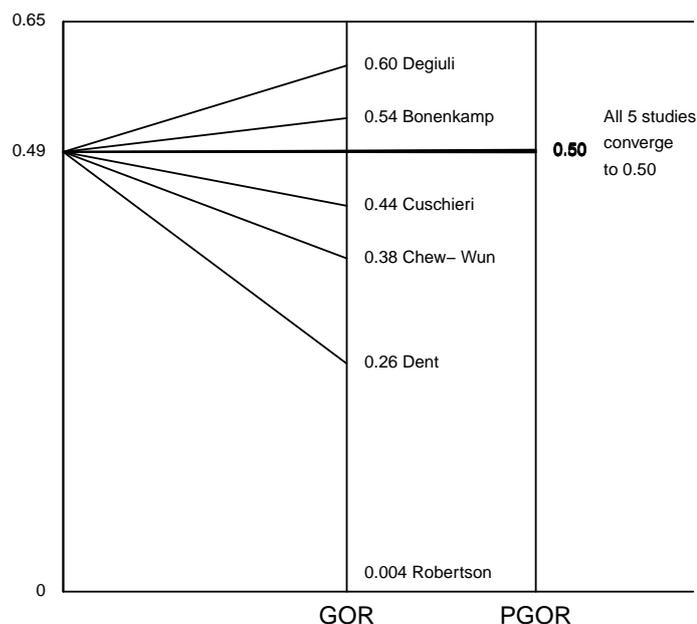


Figure 4.3: Display of GOR and PGOR values of five studies for gastric carcinoma

CI plot of the five studies excluding Robertson. The display of the PGORs shows that these predicted values are more concentrated toward estimated common value 0.50. The vertical dashed line at the common value 0.50 goes through the CIs of every study, demonstrating the power of the new meta-analysis. It is therefore statistically valid and trustworthy to combine the GORs for a common value 0.50 for these five studies. Figures 4.3 and Figure 4.4 show sharp improvement in the homogeneity of the PGORs after deleting the outlier study.

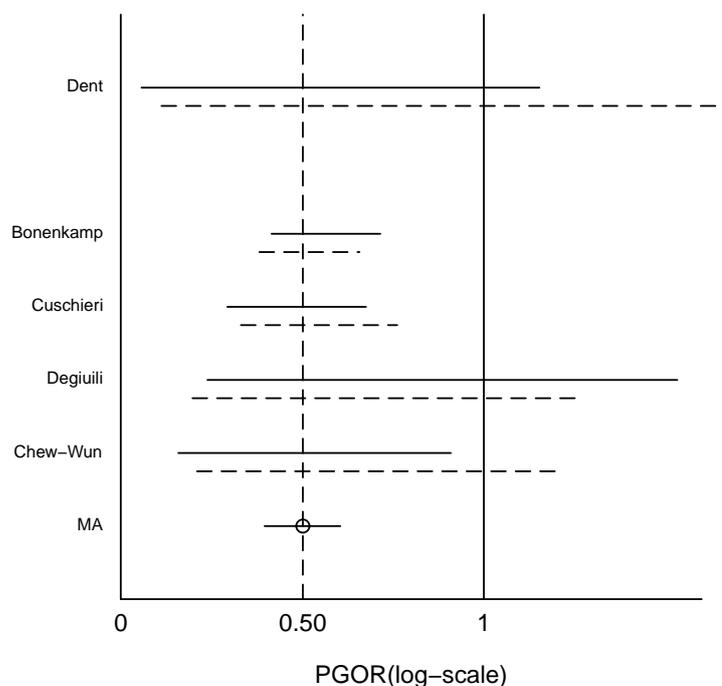


Figure 4.4: 95% CIs for GORs and PGORs of the five studies of gastric carcinoma

## 4.7 Summary

In this chapter we have compared results from five meta-analysis methods for binary outcomes. Meta-analyses using GOR for binary outcomes are as efficient as OR for both FEM and REMs. This study suggests that GOR can be used as an effect measure for binary outcomes in meta-analysis.

The sample size weighted method for binary outcomes entails lower meta-analysis estimate as compared with the other FEMs using OR and GOR. However, unavailability of a well defined variance estimate for the pooled estimate, inflated  $Q$ -value, sensitivity to the zero cell frequency are important

issues of concern for this method.

Use of GOR in meta-analysis for binary outcomes along with the improved methodology can facilitate achieving more valid and trustworthy meta-analysis.

## Chapter 5

### GOR for Continuous Outcomes

## 5.1 Introduction

In Chapter 4, we showed GOR as an effective binary outcome measure in meta-analysis. In this chapter, we discuss methods in meta-analysis for continuous outcomes as well as mixture of one continuous and one categorical outcomes.

The weighted mean difference (WMD) is used as an effect measure for continuous outcomes for a long time. Sometimes standardised weighted mean difference (SWMD) is also used to pool continuous outcomes in a meta-analysis combining parallel and cross-over trial designs (Curtin et al, 2002).

In this chapter, three methods for multilevel continuous outcomes are considered for meta-analysis. The first method is applicable when both the variables are continuous in multilevel outcomes and comparison groups. The second method is useful only when the multilevel continuous outcomes are observed in RCTs with two comparison groups. The third method is for multilevel continuous outcomes when those are made ordinal from two or more comparison groups. While the second method uses mean difference (MD) as an effect measure, the first and third method employ GOR (Agresti, 1980) as an effect measure. Although the first method is moderately complex, the other two methods are simple and straightforward.

Due to the unavailability of any real data for the multi-level continuous outcome we present results for the third method (Table 5.2) from simulated data sets.

## 5.2 Review

### 5.2.1 Method 1: Continuous vs Continuous

Consider RCTs where both the treatment and outcome variables are continuous. The treatment variable is represented along the rows and the outcome variable in the columns. Let us consider two pairs of variables  $(Y_1, Y'_1)$  and  $(Y_2, Y'_2)$  both distributed according to a bivariate distribution. When both the variables are continuous,  $\pi_c + \pi_d = 1$ , where  $\pi_c$  is the probability that  $(Y_2, Y'_2)$  is concordant with  $(Y_1, Y'_1)$  and  $\pi_d$  is the probability that  $(Y_2, Y'_2)$  is discordant with  $(Y_1, Y'_1)$ . Then the GOR (Agresti, 1980) can be written as

$$\Gamma = \frac{(1 + \tau)}{(1 - \tau)}, \quad (5.1)$$

where  $\tau$  is Kendall's tau (Kendall and Gibbons, 1990; p.3) and  $\hat{\tau}$  is asymptotically normally distributed. Then

$$\frac{(\hat{\Gamma}_i - \Gamma_i)\sqrt{n}}{\sqrt{\{\lim \text{var}(\hat{\Gamma}_i\sqrt{n})\}}} \sim N(0, 1), \text{ where } 0 < \Gamma_i < \infty, \quad (5.2)$$

where  $\hat{\Gamma}_i$  represents the estimated GOR for the  $i$ th study. The variance expression from Noether (1967, p.74) gives

$$\begin{aligned} \text{var} \{n(n-1)\hat{\tau}/2\} &= 2N(N-1) \pi_c(1 - \pi_c) \\ &+ 4N(N-1)(N-2)(\pi_{cc} - \pi_c^2), \end{aligned} \quad (5.3)$$

where  $\pi_{cc}$  is the probability that the first and the second member are both concordant with the third for a random sample of three members.

Using the above expression, we get

$$\begin{aligned} \lim_{n \rightarrow \infty} \text{var}(\hat{\Gamma}\sqrt{n}) &= \lim_{n \rightarrow \infty} \frac{\text{var}(\hat{\tau}\sqrt{n})}{4(1 - \pi_c)^4} \\ &= 4 \frac{(\pi_{cc} - \pi_c^2)}{(1 - \pi_c)^4}. \end{aligned} \quad (5.4)$$

For bivariate normal distribution

$$\Gamma_i = \frac{(\pi + 2\sin^{-1}\rho)}{(\pi - 2\sin^{-1}\rho)} \quad (5.5)$$

and it follows from Esscher (1924) that

$$\text{var}(\tau) = \frac{2}{n(n-1)} \left\{ 1 - \left( \frac{2}{\pi} \sin^{-1}\rho \right)^2 + 2(n-2) \left[ \frac{1}{9} - \left( \frac{2}{\pi} \sin^{-1}\frac{1}{2}\rho \right)^2 \right] \right\}. \quad (5.6)$$

Then,

$$\lim_{n \rightarrow \infty} \text{var}(\hat{\Gamma}_i \sqrt{n}) = \left[ \frac{1}{9} - \left( \frac{2}{\pi} \sin^{-1}\frac{1}{2}\rho \right)^2 \right] / \left( \frac{\pi + 2\sin^{-1}\rho}{2\pi} \right)^4. \quad (5.7)$$

Suppose  $Y_1$  and  $Y_2$  are two independent continuous random variables, then the GOR,

$$\Gamma_i = \frac{P(Y_2 > Y_1)}{\{1 - P(Y_2 > Y_1)\}} \quad (5.8)$$

is used to compare the two distributions. Let us consider  $Y_1$  and  $Y'_1$  are independent random variables from the treatment distribution and further  $Y_2$  and  $Y'_2$  from the outcome distribution. Again, let  $P_{21} = P(Y_2 > Y_1)$ ,  $P_{221} = P(Y_1 < Y_2, Y_1 < Y'_2)$ ,  $P_{211} = P(Y_1 < Y_2, Y'_1 < Y_2)$ , and suppose that random samples of sizes  $n_1 = w_1 n$  and  $n_2 = w_2 n$  are selected from the two distribution so that  $w_1 + w_2 = 1$ . Then using the asymptotic normality of the Mann-Whitney statistic  $U =$  number of pairs for which  $Y_2 > Y_1$ ,  $\hat{\Gamma}_i$  is asymptotically normally distributed with

$$\lim_{n \rightarrow \infty} \text{var}(\hat{\Gamma}_i \sqrt{n}) = \sigma_{21}^2 / (1 - P_{21})^4, \quad (5.9)$$

where  $\sigma_{21}^2 = \lim_{n \rightarrow \infty} \text{var}(\hat{P}_{21} \sqrt{n})$  in which  $\hat{P}_{21} = U/n_1 n_2$ . From Lehmann (1975, p. 70) we have

$$\text{var}(U) = n_1 n_2 P_{21} (1 - P_{21}) + n_1 n_2 (n_2 - 1) (P_{221} - P_{21}^2) + n_1 n_2 (n_1 - 1) (P_{211} - P_{21}^2). \quad (5.10)$$

Therefore,

$$\sigma_{21}^2 = (P_{221} - P_{21}^2)/w_1 + (P_{211} - P_{21}^2)/w_2. \quad (5.11)$$

The pooled estimate of GOR can be calculated as  $\hat{\Gamma}_0 = \frac{\sum_i \omega_i \hat{\Gamma}_i}{\sum_i \omega_i}$ , where  $\omega_i = 1/\text{var}(\hat{\Gamma}_i)$ . The CIs for individual studies under FEM can be computed using the estimate of  $\text{var}(\hat{\Gamma}_i)$  as  $\exp[\ln \hat{\Gamma}_i \pm z_{\alpha/2} \sqrt{\text{var}(\hat{\Gamma}_i)}]$  and for meta-analysis using formula  $\exp[\ln \hat{\Gamma}_0 \pm z_{\alpha/2} / \sqrt{\sum_{i=1}^k \text{var}(\hat{\Gamma}_i)}]$ .

The pooled estimate and the CI can also be constructed under REM in a similar fashion as in Chapter two incorporating the between study variance.

## 5.2.2 Method 2: Categorical vs Continuous

When data from RCTs with multiple outcome categories are continuous, Mean difference (MD) or standardised mean difference (SMD) can be used as an effect measure for multilevel continuous outcome categories for two comparison groups. Using MD one can estimate the effect measure in the original unit of measurement while SMD has the advantage of being a unit free measurement. The treatment effect ( $T_i$ ) for the  $i$ th study is calculated by

$$T_i = \bar{X}_{ti} - \bar{X}_{ci}, \quad (5.12)$$

where  $\bar{X}_{ti}$  and  $\bar{X}_{ci}$  are the sample mean responses in the treatment and control groups respectively. Now the standard error of this treatment difference ( $T_i$ ) can be found as

$$se(T_i) = \sqrt{\text{var}(T_i)} = \sqrt{s_p^2 (1/n_t + 1/n_c)}, \quad (5.13)$$

where  $s_p^2 = \sqrt{\frac{(n_t-1)s_t^2 + (n_c-1)s_c^2}{n_t+n_c-2}}$  is the estimate of the assumed common variance (Hedges and Olkin, 1985),  $\sigma_t^2 = \sigma_c^2 = \sigma^2$  for the two population variances and  $n_t$  and  $n_c$  are the within study sample sizes in the treatment and control groups respectively. If the population variances are assumed different

$se(T_i) = \sqrt{\text{var}(T_i)} = \sqrt{s_t^2/n_t + s_c^2/n_c}$ . An appropriate 95% CI for individual study can be calculated then by  $T_i \pm 1.96 * se(T_i)$ .

The weighted mean difference (WMD) in meta-analysis under FEM can be found as

$$\text{WMD} = \bar{T} = \frac{\sum_{i=1}^k \omega_i T_i}{\sum_{i=1}^k \omega_i}, \quad (5.14)$$

where  $\omega_i = 1/\text{var}(T_i)$  is the weight and  $T_i$  is the effect size for the  $i$ th study. A  $100(1-\alpha)\%$  CI for pooled estimate can be calculated as  $\bar{T} \pm z_{\alpha/2}/\sqrt{\sum_{i=1}^k \omega_i}$ .

Under REM the  $\omega_i$ 's are calculated incorporating the between study variance  $\hat{\tau}^2$  as

$$\omega_i^* = \frac{1}{[(1/\omega_i) + \hat{\tau}^2]}, \quad (5.15)$$

where  $\hat{\tau}^2 = (Q - (k - 1))/U$  if  $Q > k - 1$ , 0 otherwise,  $U = (k - 1) \left( \bar{\omega} - \frac{s_W^2}{k\bar{\omega}} \right)$ ,  $s_W^2 = \frac{1}{k-1} \left( \sum_{i=1}^k \omega_i^2 - k\bar{\omega}^2 \right)$ ,  $\bar{\omega} = \sum_{i=1}^k \omega_i/k$ ,  $Q$  is the heterogeneity statistic and  $k$  is the number of studies. Under normality assumption a  $100(1 - \alpha)\%$  CIs for individual effect size estimate and pooled estimate can be calculated by  $T_i \pm 1.96 * se(T_i)$  and  $\bar{T} \pm z_{\alpha/2}/\sqrt{\sum_{i=1}^k \omega_i^*}$  respectively.

### Standardised Mean Difference

The standardised mean difference (SMD) is estimated by

$$d_i = \frac{\bar{X}_{ti} - \bar{X}_{ci}}{s^*}, \quad (5.16)$$

where  $\bar{X}_{ti}$  and  $\bar{X}_{ci}$  are the sample mean responses in the treatment and control groups respectively, and  $s^*$  is the pooled estimate of the standard deviations as before. The estimate  $d_i$  has small sample bias and the bias can be removed using a simple correction (Hedges, 1981) that produce an unbiased estimate of the population SMD. The variance of  $d_i$  is difficult to

compute exactly. However, a very good approximation of the variance of  $d_i$  is given by

$$V_{di} = \frac{n_{ti} + n_{ci}}{n_{ti}n_{ci}} + \frac{d_i^2}{2(n_{ti} + n_{ci})}. \quad (5.17)$$

More simpler variance approximation (Fleiss, 1993) is also available if the  $n_{ti}$  and  $n_{ci}$  are large and the population variances are equal as

$$V_{di} = \frac{n_{ti} + n_{ci}}{n_{ti}n_{ci}}. \quad (5.18)$$

The use of SMD has been criticised by saying that studies with identical results may vaguely appear to yield different results. The transformation can even make a study whose original estimate was smaller in magnitude than another study appear greater and vice versa (Greenland, 1987). There are other continuous outcome measures rarely used in medicine, such as correlation coefficient etc (Rosenthal, 1994).

### 5.3 Method 3: Categorical vs Latent Continuous

When the continuous outcomes are made ordinal in RCTs or the outcomes are available in continuous range, GOR is an effective outcome measure. These data are in counts in multilevel categories. These categories are either in ascending or descending order. As a result, we can set them as concordant or discordant pairs with two or more comparison groups as needed. Even when there are more than two treatment groups (multi-arm trials) with multilevel outcomes, GOR is still a meaningful effect measure. The effect size can be estimated using GOR as

$$\hat{\Gamma}_i = \hat{\Pi}_{ci} / \hat{\Pi}_{di}, \quad (5.19)$$

where  $\hat{\Pi}_{ci} = \sum_{r=1}^{L-1} \sum_{s=r+1}^L \hat{\pi}_{ir|1} \hat{\pi}_{is|2}$  and  $\hat{\Pi}_{di} = \sum_{r=2}^L \sum_{s=1}^{r-1} \hat{\pi}_{ir|1} \hat{\pi}_{is|2}$  for the  $i$ th study.

For  $i = 1, \dots, k$  independent studies if  $\ln \hat{\Gamma}_i$  representing logarithm of GOR be the observed effect size with variance  $v_i$  and  $\Gamma_i$  the underlying population effect size. Then assuming  $\Gamma_1 = \Gamma_2 = \dots = \Gamma_k = \Gamma_0$ , a pooled estimate of the treatment effect is given by

$$\hat{\Gamma}_0 = \frac{\sum_i \omega_i \ln \hat{\Gamma}_i}{\sum_i \omega_i}. \quad (5.20)$$

Then the estimated variance of  $\ln \hat{\Gamma}_i$  with  $J$  different treatment and  $L$ -level outcome categories can be found as

$$\hat{\omega}_i^{-1} = \sum_{l=1}^{L-1} \sum_{j=1}^J \frac{1}{n_{ij} \hat{\pi}_{ijl} (1 - \hat{\pi}_{ijl})}, \quad (5.21)$$

where  $n_{ij}$  is the total count of  $j$ th group for the  $i$ th study,  $\hat{\pi}_{ijl} = X_{ijl}/n_{ij}$  is the maximum likelihood estimator (MLE) of  $\pi_{ijl}$ ,  $X_{ijl}$  is the count of the  $l$ th category in the  $j$ th group for the  $i$ th study. Here each row of the  $J \times L$  table is modeled as an independent multinomial distribution.

For RCTs with two comparison groups the estimated variance,  $\hat{\omega}_i^{-1}$  of the  $i$ th study, becomes

$$\hat{\omega}_i^{-1} = \sum_{l=1}^{L-1} \sum_{j=1}^2 \frac{1}{n_{ij} \hat{\pi}_{ijl} (1 - \hat{\pi}_{ijl})}. \quad (5.22)$$

An approximate  $100(1 - \alpha)\%$  CI for the GORs is given by the following formula under normality assumption

$$\exp[\ln \hat{\Gamma}_i \pm z_{\alpha/2} \omega_i^{-1/2}], \quad (5.23)$$

where  $z_{\alpha/2}$  is the  $100(1 - \alpha)\%$  percentage point of a standard normal distribution.

An estimator of the variance of  $\ln \hat{\Gamma}_0$  is given by

$$\text{var}(\ln\hat{\Gamma}_0) = 1/\sum_{i=1}^k \omega_i. \quad (5.24)$$

Under normality assumption, an approximate  $100(1 - \alpha)\%$  CI for the population effect,  $\Gamma_0$ , is given by

$$\exp \left[ \ln\hat{\Gamma}_0 \pm z_{\alpha/2} \sqrt{1/\sum_{i=1}^k \omega_i} \right] \quad (5.25)$$

for the meta analysis, where  $z_{\alpha/2}$  is the  $100(1 - \alpha)\%$  percentage point of a standard normal distribution.

## 5.4 Example

Here we consider the data (Table 5.1) for two variables income and job satisfaction (JS), measured for the black males in U.S.A. The income variable has four category levels in range and job satisfaction has four categories as well, such as very dissatisfied (VD), little dissatisfied (LD), moderately satisfied (MS), and very satisfied (VS).

Table 5.1: Contingency table for Job Satisfaction by Income

| Income<br>(dollars) | Job satisfaction     |                        |                         |                   | Total |
|---------------------|----------------------|------------------------|-------------------------|-------------------|-------|
|                     | Very<br>Dissatisfied | Little<br>Dissatisfied | Moderately<br>satisfied | Very<br>satisfied |       |
| <15,000             | 1                    | 3                      | 10                      | 6                 | 20    |
| 15,000-25,000       | 2                    | 3                      | 10                      | 7                 | 22    |
| 25,000-40,000       | 1                    | 6                      | 14                      | 12                | 33    |
| > 40,000            | 0                    | 1                      | 9                       | 11                | 21    |

Source: 1996 General Social Survey, National Opinion Research Center, USA.

As the level of Income increases, responses on job satisfaction tend to increase towards higher levels. Here, the GOR can be used as an effect measure considering these two variables as concordant and discordant pairs.

Following this example, we generate data for five studies using simulation techniques under independent multinomial distribution for meta-analysis. The data is presented in Table 5.2. This data has four groups in the row variable and four categories in the column variable for each study. The last two columns of the table contain the estimated GORs and weights for the five studies. Meta-analysis using these GORs is found to be 1.30 with CI (0.16, 10.91).

Figure 1 shows the forest plot for the five simulated studies. The solid horizontal lines represent the 95% CI for the five studies and meta-analysis. The dashed vertical line represents the pooled estimate from these five studies. The heterogeneity statistic  $Q = 0.004$ ,  $df = 4$ ,  $p$ -value = 0.99 suggests that these studies are homogeneous. The  $I^2$  statistic = 0%[0%, 0%] also indicates the homogeneity of these simulated studies.

## 5.5 Summary

The main contribution of this chapter is that we presented three different meta-analysis methods for continuous outcomes. The first method is applicable when both the variables in a  $J \times L$  contingency table are continuous. That is, the outcome categories and the treatment categories are both continuous. The second method is suitable when the multilevel outcome categories are continuous and there are two comparison groups. The third method is applicable when there are multiple outcome categories observed for a variable that is inherently continuous with two or more comparison groups.

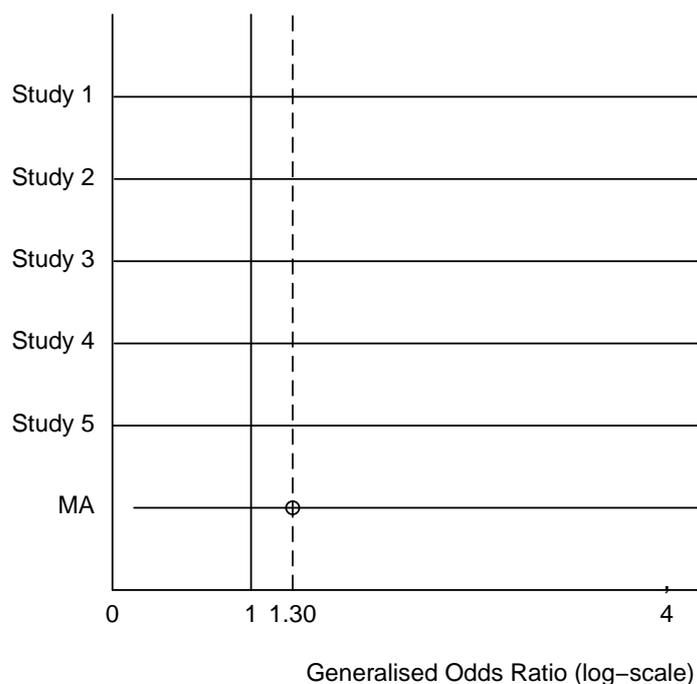


Figure 5.1: Forest plot from simulated data using GOR

The first and the third method used GOR as an effect measure while the second method employed MD as an effect measure. Although the first situation is not very common in meta-analysis of RCTs, GOR is being used for long time in such situation in psychology, education and ophthalmology. However, the mathematical formulation is moderately complex. The other two methods are simple and straightforward. Using MD and GOR, we can estimate the effect sizes in two different scenarios. We trust these methods will facilitate conducting more meta-analyses with continuous outcomes.

Table 5.2: Five simulated studies

| JS      |                     |    |    |    |    |       |      |        |
|---------|---------------------|----|----|----|----|-------|------|--------|
| Study   | Income<br>(dollars) | VD | LD | MS | VS | Total | GOR  | Weight |
| Study 1 | <15,000             | 1  | 6  | 5  | 8  | 20    | 1.34 | 0.17   |
|         | 15,000-25,000       | 2  | 4  | 12 | 4  | 22    |      |        |
|         | 25,000-40,000       | 1  | 8  | 8  | 16 | 33    |      |        |
|         | > 40,000            | 1  | 2  | 7  | 11 | 21    |      |        |
| Study 2 | <15,000             | 2  | 6  | 4  | 8  | 20    | 1.36 | 0.19   |
|         | 15,000-25,000       | 2  | 5  | 10 | 5  | 22    |      |        |
|         | 25,000-40,000       | 1  | 9  | 8  | 15 | 33    |      |        |
|         | > 40,000            | 1  | 3  | 6  | 11 | 21    |      |        |
| Study 3 | <15,000             | 1  | 7  | 4  | 8  | 20    | 1.24 | 0.18   |
|         | 15,000-25,000       | 2  | 3  | 12 | 5  | 22    |      |        |
|         | 25,000-40,000       | 1  | 9  | 9  | 14 | 33    |      |        |
|         | > 40,000            | 1  | 4  | 5  | 11 | 21    |      |        |
| Study 4 | <15,000             | 1  | 8  | 8  | 3  | 20    | 1.39 | 0.18   |
|         | 15,000-25,000       | 2  | 4  | 11 | 5  | 22    |      |        |
|         | 25,000-40,000       | 1  | 11 | 9  | 12 | 33    |      |        |
|         | > 40,000            | 1  | 4  | 7  | 9  | 21    |      |        |
| Study 5 | <15,000             | 1  | 6  | 8  | 5  | 20    | 1.12 | 0.12   |
|         | 15,000-25,000       | 3  | 5  | 5  | 9  | 22    |      |        |
|         | 25,000-40,000       | 1  | 9  | 7  | 16 | 33    |      |        |
|         | > 40,000            | 0  | 5  | 6  | 10 | 21    |      |        |

Meta-analysis=1.30 (0.16, 10.91)

Heterogeneity tests:

$Q = 0.004$ ,  $df = 4$ ,  $p = 0.99$

$I^2 = 0\%$ [0%, 0%].

# Chapter 6

## Conclusion and Future Work

In this study, we considered the data from RCTs with multilevel ordinal categorical outcomes in meta-analysis. We aimed at estimating the effect sizes without any loss of information by merging or splitting the  $J \times L$  contingency table into  $2 \times 2$  tables. We employed GOR as an effect measure for meta-analysis with ordinal outcome categories. GOR has the advantages of estimating the effect sizes not only with RCTs of two comparison groups but also any number of comparison groups and outcome categories.

### **Meta-analysis for ordinal outcomes**

We developed a meta-analysis method for multilevel ordinal categorical outcomes under independent multinomial distribution approach. The proposed meta-analysis method is much easier than other methods and provide more realistic results. More importantly, the proposed method estimates the effect sizes without merging categories and is free from any model assumption. These properties lead the proposed method preferable and more realistic than the existing methods for multilevel ordinal categorical outcomes.

### **Heterogeneous ordinal outcomes**

Heterogeneity is one of the most problematic aspects in meta-analysis. Statistical agreement of results is under serious question in the presence of heterogeneity. In this study, we have developed a QEBM using GOR for heterogeneous ordinal categorical outcomes. This method provides a remedy to the issue of heterogeneity for ordinal categorical data in meta-analysis. The QEBM identifies the extreme study along with the Dixon's test of outliers. This method also improves the meta-analysis using the predicted GOR taking into account the heterogeneity in the data set. The use of the predicted GOR improves the meta-analysis sharply in terms of shrinking the CIs of the study effects towards the pooled study effect. The proposed method demonstrates a methodology for trustworthy meta-analysis with heterogeneous ordinal categorical outcomes.

### **GOR for binary outcomes**

We have presented a meta-analysis method for binary outcomes using GOR under both the fixed and random effects models. A sample size weight method for binary outcomes using  $OR_G$  is also covered. We have applied GOR,  $OR_G$  for binary outcomes and compared results obtained with the methods using ordinary OR. It is observed that the meta-analysis estimates and their associated CIs under both FEM and REMs using GOR are exactly the same as that of using OR under FEM and REMs respectively. It is observed that the weights computed from the variance estimate of the individual GORs are the same with that of the ORs. Therefore, this study

suggests that GOR can also be used for binary outcomes as efficiently as OR in meta-analysis. The sample size weight method have produced a lower meta-analysis estimate as compared to other methods. However, a well defined variance estimate for pooled  $OR_G$  is not available.

### **Methods for continuous outcomes**

We have also presented three methods for multilevel continuous outcomes in meta-analysis in this thesis. The first method is applicable when both the variables are continuous in multilevel outcomes and comparison groups. The second method is useful only when the multilevel continuous outcomes are observed in RCTs with two comparison groups. The third method is for multilevel continuous outcomes when those are made ordinal from two or more comparison groups. While the second method uses MD as an effect measure, the first and third method employ GOR as an effect measure. Although the first method is moderately complex, the other two methods are simple and straightforward. We trust these methods will contribute widely conducting more meta-analyses with continuous outcomes.

### **Challenges and future work**

This study analyses outcome data in meta-analysis that occur in multilevel ordinal categories. As in most cases the ordinal outcomes are naturally in ascending or descending orders, it makes our study table setting easier for the computation of the GOR as an effect measure. Because GOR is defined as the proportion of concordant and discordant pairs, we need to set the study table in such a way that as the level of outcome variable in the columns

increases, so does the comparison groups in the rows in terms of severity or vice versa.

Meta-analysis with more than two outcomes has not been carried out as frequently as with binary outcomes. As a result, we have found insufficient data in the literature. In particular, we did not find a set of studies with significant heterogeneity in these situations. Therefore, we used simulated data along with the real data sets available.

Future research can be carried out using GOR under quality effects model (Doi SA et al., 2011) for heterogeneous ordinal outcomes. GOR can also be used under Bayesian meta-analysis of ordinal categorical data. However, the challenges remain regarding the availability of the quality scores from studies using GOR and development of softwares for computational purpose. We will endeavour to resolve these challenges in our future work.

- Agresti, A. (1980), ‘Generalised odds ratios for ordinal data’, *Biometrics*, **36**, 59 – 67.
- Agresti, A. (2002), *Categorical Data Analysis*, Second Edition, Hoboken, New Jersey. John Wiley & Sons, Inc.
- Agresti, A. (2010), *Analysis of Ordinal Categorical Data*, 2nd Edition, Hoboken, New Jersey. John Wiley & Sons, Inc.
- Bancroft, T. A. (1944), ‘On biases in estimation due to the use of preliminary test of significance’, *Ann. Math. Stat.*, **15**, 190 – 194.
- Beecher, H. K. (1995), ‘The powerful placebo’, *J. Am. Medical Assoc.*, **159**, 1602 – 1606.
- Begg, C. B., and Mazumdar, M. (1994), ‘Operating characteristics of a rank correlation test for publication bias’, *Biometrics*, **50**, 1088 – 1101.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T. and Rothstein, H. R. (2009), *Introduction to Meta-Analysis*, First Edition, United Kingdom, John Wiley

& Sons Ltd.

- Casella, G. (1985), ‘An introduction to empirical Bayes data analysis’, *J. Am. Stat. Assoc.*, **39**, 83 – 87.
- Chalmers, T. C., Matta, .J., Smith, H. Jr., and Kunzler, A. M. (1977), ‘Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction’, *New Engl. J. Med.*, **297**, 1091 – 6.
- Cornfield, J. (1951), ‘A method of estimating comparative rates from clinical data; applications to cancer of the lung, breast, and cervix’, *J Natl Cancer Inst*, **11(6)**, 1269 – 75.
- Curtin, F., Altman, G. D. and Elbourne, D. (2002), ‘Meta-analysis combining parallel and cross-over clinical trials. I: Continuous outcomes’, *Statistics in Medicine*, **21**, 2131 – 2144.
- Dale, J. R. (1984), ‘Local vs global association for bivariate ordered responses’, *Biometrika*, **71**, 507 – 514.
- DerSimonian, R. and Laird, N. (1986), ‘Meta-analysis in clinical trials’, *Control Clin Trials*, **7**, 177 – 88.
- Dickersin, K. and Berlin, J. A. (1992), ‘Meta-analysis: state-of-the-science’, *Epidemil. Rev.*, **14**, 154 – 76.
- Dixon, W. J. (1950), ‘Analysis of Extreme values’, *Ann. Math. Stat.*, **21**, 488 – 506.
- Doi, S. A., Barendregt, J. J., and Mozurkewich, E. L. (2011), ‘Meta-analysis of heterogeneous clinical trials: An empirical example’, *Contemporary Clinical Trials*, **32(2)**, 288 – 98.
- Duval, S. and Tweedie, R. (2000), ‘Trim and fill: A simple funnel plot based method of testing and adjusting for publication bias in meta-analysis’, *Biometrics*, **56(2)**, 455 – 463.
- Edwardes, M. D. and Baltzan, M. (2000), ‘The generalization of the odds ratio,

- risk ratio and risk difference to  $r \times k$  tables', *Statistics in Medicine*, **19**, 1901 – 1914.
- Edwards, A. W. F. (1963), 'The measure of association in a  $2 \times 2$  table', *Journal of the Royal Statistical Society*, **126**(1), 109 – 114.
- Efron, B. (1995), 'A statistical century', *R. Stat. Soc. News*, **22**, 1 – 2.
- Efron, B. (1996), 'Empirical Bayes methods for combining likelihoods', *J. Am. Stat. Assoc.*, **91**, 538 – 50.
- Efron, B. and Morris, C. N. (1973), 'Stein's estimation rule and its competitors—an empirical Bayes approach', *Sci. Am.*, **236**, 119 – 127.
- Efron, B. and Morris, C. N. (1977), 'Stein's paradox in Statistics', *J. Am. Stat. Assoc.*, **68**, 117 – 130.
- Emerson, J. D. (1994), 'Combining estimates of the odds-ratio: the state of the art', *Stat. Methods Med. Res.*, **3**, 157 – 178.
- Fleiss, J. L. (1993), 'The statistical basis of meta-analysis', *Stat. Methods Med.*, **2**, 121 – 145.
- Fleiss, J. L. (1994), 'Measures of effect size for categorical data. In: Cooper, H., Hedges, L. V., (editors)', *The Handbook of Research Synthesis*, New York: Russel Sage Foundation; 245 – 60.
- Galbraith, R. F. (1988), 'A note on graphical presentation of estimated odds ratios from several clinical trials'. *Statistics in Medicine*, **7**, 889 – 894.
- Greenland, S. (1987), 'Quantitative methods in the review of epidemiologic literature', *Epidemio. Rev.*, **9**, 1 – 30.
- Gupta, A. K. and Saleh, A. K. Md. E. (1997), 'Estimating odds-ratio: Homogeneous constraints', *J. Ital. Stat. Soc.*, **6**, 67 – 81.
- Han, C. P., Bancroft, T. A. (1968), 'On pooling means when variance is known', *J. Am. Stat. Assoc.*, **63**, 1333 – 1342.
- Hedges, L. V. and Olkin, I. (1985), *Statistical Methods for Meta-Analysis*, San

- Diego: Academic Press Inc.
- Higgins, J. P. T and Thompson S. G. (2002), ‘Quantifying heterogeneity in a meta-analysis’, *Statistics in Medicine*, **21**, 1539 – 1558.
- Hossain, M. B. and Khan, S. (2011), ‘Meta-analysis for ordinal categorical outcomes’, *Submitted to Statistics in Medicine*.
- Hozo, S. P., Djulbegovic, B. and Hozo, I. (2005), ‘Estimating the mean and variance from the median, range, and the size of a sample’, *BMC Med Res Methodolo.*, **5(13)**. doi: 10.1186/1471 – 2288 – 5 – 13.
- Hwang, J. T. and Casella, G. (1982), ‘Minimax confidence sets for the mean of a multivariate normal distribution’, *Ann. Stat.*, **10**, 868 – 881.
- Judge, G. G., and Bock, M. E. (1978), *The Statistical Implications of Pre-test and Stein-rule Estimators in Econometrics*, North-Holland, Amsterdam.
- Kendall, M. and Gibbons, J. D. (1990), *Rank Correlation Methods*. First Edition, New York: Oxford University Press.
- Khan, S. and Hoque, Z. (2002), ‘James-Stein estimators for the mean vector of a multivariate normal population based on independent samples from two normal populations with common covariance structure’, *Pakistan Journal of Statistics*, Special Issue, **18(3)**, 359 – 381.
- Khan, S. (2003), ‘Estimation of the Parameters of two Parallel Regression Lines Under Uncertain Prior Information’, *Biometrical Journal*, **44**, 73 – 90.
- Khan, S. (2008), ‘Shrinkage Estimators of Intercept Parameters of Two Simple Regression Models with Suspected Equal Slopes’, *Communications in Statistics - Theory and Methods*, **37**, 247 – 260.
- L’Abbe, K. A., Detsky, A.S., and O’Rourke, K. (1987), ‘Meta-analysis in clinical research’, *Annals of Internal Medicine*, **107**, 224 – 233.
- Laird, N., Fitzmaurice, G. and Ding, X. (2010), Comments on ‘Empirical vs natural weighting in random effects meta-analysis’, *Statistics in Medicine* ,

- 29(12)**, 1266 – 1267.
- Lehmann, E. L. (1975), *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day.
- Lewis, S. and Clarke, M. (2001), ‘Forest plots: trying to see the wood and the trees’, *BMJ*, **322**, 1479 – 1480.
- Light, R. J. and Pillemer, D. B. (1984), *Summing Up: The Science of Reviewing Research*, Cambridge, MA: Harvard University Press.
- Mann, C. (1990), ‘Meta-analysis in breech’, *Science*, **249**, 476 – 490.
- Mantel, N. and Haenszel, W. (1959), ‘Statistical aspects of the analysis of data from retrospective studies of disease’, *J. Nat. Cancer Inst.*, **22**, 719 – 748.
- Mantel, N. (1963), ‘Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure’, *J. Am. Statist. Assoc.*, **58**, 690 – 700.
- McCullagh, P. (1980), ‘Regression models for ordinal data’, *J. R. Statist. Soc. B*, **42(2)**, 109 – 142.
- Memon, M. A., Subramanya, M. S., Khan, S., Hossain, M. B., Osland, E., and Memon, B. (2011), ‘Meta-analysis of D1 versus D2 gastrectomy for gastric adenocarcinoma’, *Annals of Surgery*, **253(5)**, 900 – 911.
- Morris, C. N. (1983), *Parameter Empirical Bayes Confidence Intervals, Scientific Inference, Data Analysis and Robustness*. New York: Academic Press Inc., pp. 22 – 50.
- Morris, C. N. (1992), ‘Hierarchical models for combining information and for meta-analysis’, *Bayesian Statistics*, **4**, 321 – 44.
- Mosteller, F. (1968), ‘Association and estimation in contingency tables’, *Journal of the American Statistical Association*, **63(321)**, 1 – 28.
- Noether, G. E. (1967), *Elements of Non-parametric Statistics*. New York, London, Sydney: John Wiley and Sons, Inc. pp. 70 – 83.
- Pearson, K. (1904), Report on certain enteric fever inoculation statistics. *B.*

- Med J.*, **3**, 1243 – 6.
- Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R. and Howard, S. V. (1977), ‘Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. analysis and examples’, *Br. J. Cancer*, **35**, 1 – 39.
- Peto, R. (1987), ‘Why do we need systematic overview of randomized trials?’, *Statistics in Medicine*, **6**, 233 – 240.
- Qizilbash N. Whitehead. A, Higgins. J, Wilcock. G, Schneider. L and Farlow, M. (1998), ‘Cholinesterase Inhibition for Alzheimer Disease: A Meta-analysis of the Tacrine Trials’, *Journal of the American Medical Association*, **280**, 1777-1782.
- Raudenbush, S. W. and Bryk, A. S. (1985), ‘Empirical Bayes meta-analysis’, *J. Educ. Statist*, **10**, 75 – 98.
- Raudenbush, S. W. (1994), Random effects models. In: Cooper, H., Hedges, L.V., (editors). *The Handbook of Research Synthesis*. New York: Russell Sage Foundation. 301 – 22.
- Rosenthal, R. (1994), *The Handbook of Research Synthesis*, New York: Russell Sage Foundation, pp. 231 – 244.
- Rucker, G., Schwarzer, G., Carpenter, J. and Schumacher, M. (2010), Comments on ‘Empirical vs natural weighting in random effects meta-analysis’, **29**, 2963 – 2966.
- Saleh, A. K. Md. E., Hassanein, K. M., Hassanein, R. S., and Kim, H. M. (2006), ‘Quasi-empirical Bayes methodology for improving meta-analysis’, *Journal of Biopharmaceutical Statistics*, **16**, 77 – 90.
- Saleh, A. K. Md. E. (2006), *Theory of Preliminary Test and Stein-type Estimation with Applications*, Hoboken, New Jersey. John Wiley & Sons, Inc.
- Schlesselman, J. and Stolley, P. (1982), *Case-control Studies. Design, Conduct,*

*Analysis*, New York: Oxford University Press.

- Schork, M. A. and Remington, R. D. (2000), *Statistics with Applications to the Biological and Health Sciences*, Upper Saddle River, New Jersey: Prentice Hall.
- Senn, S. (2007), ‘Trying to be precise about vagueness’, *Statistics in Medicine*, **26**(7), 1417 – 30.
- Shuster, J. J. (2010), ‘Empirical vs natural weighting in random effects meta-analysis’, *Statistics in Medicine*, **29**, 1259 – 1265.
- Spector, T. D. and Thompson, S. G. (1991), ‘Research methods in epidemiology 5. The potential and limitation of meta-analysis’, *J. Epidemiol. Comm. Hlth.*, **45**, 89 – 92.
- Stein, C. (1956), *Inadmissibility of the usual estimator of mean of a multivariate normal distribution*, Proceedings 3rd Berkeley Symposium on Mathematical Statistics and Probability, pp. 197 – 206.
- Stijnen, T. and Van Houwelingen, J. C. (1990), ‘Empirical Bayes methods in clinical trials meta-analysis’, *Biometrical J.*, **32**, 335 – 46.
- Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A. and Song, F. (2000), *Methods for Meta-Analysis in Medical Research*, West Sussex, PO191UD, England: John Wiley & Sons, Ltd.
- Thompson, S. G. and Higgins, J. P. T. (2010), Comments on ‘Empirical vs natural weighting in random effects meta-analysis’, *Statistics in Medicine*, **29**(12), 1270 – 1271.
- Thompson, S. G. and Pocock, S. J. (1987), ‘Can Meta-analysis be trusted?’, *Lancet*, **338**, 1127 – 1130.
- Van Houwelingen, H. C. and Stijnen, T. (1993), ‘Monotone empirical Bayes estimators based on more informative samples’, *J. Am. Statist. Assoc.*, **88**, 1438 – 43.

- Waksman, J. A. (2010), Comments on ‘Empirical vs natural weighting in random effects meta-analysis’, *Statistics in Medicine*, **29**(12), 1268 – 1269.
- Whitehead, A. and Jones N. M. B. (1994), ‘A meta-analysis of clinical trials involving different classifications of response into ordered categories’, *Statistics in Medicine*, **13**, 2503 – 2515.
- Whitehead, A., Omar, R. Z., Higgins, J. P. T., Savaluny, E., Turner, R. M. and Thompson, S. G. (2001), ‘Meta-analysis of ordinal outcomes using individual patient data’, *Statistics in Medicine*, **20**, 2243 – 2260.
- Woolf, B. (1955), ‘On estimating the relation between blood group and disease’, *Ann. Hum. Genet.*, **19**, 251 – 253.
- Yates, F. and Cochran, W.G. (1938). The analysis of groups of experiments. *J. Agricultural Sci.*, **28**, 55 – 80.
- Yusuf, S., Peto, R., Lewis, J., Collins, R. and Sleight, P. (1985). Beta blockade during and after myocardial infarction: an overview of the randomised trials. *Progress in Cardiovascular Diseases*, **27**, 335 – 71.
- Zhou, X. H. (1996), ‘Empirical Bayes combination of estimated areas under ROC curves using estimating equations’, *Med. Decision Making*, **16**, 24-8.

# Appendix A

## A.1 Delta Method

Suppose that  $\hat{\boldsymbol{\mu}}' = (\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_S)$  asymptotically follows the multivariate normal distribution with mean  $\boldsymbol{\mu}' = (\mu_1, \mu_2, \dots, \mu_S)$  and covariance matrix  $\boldsymbol{\Sigma}/n$  and further that  $g(\mathbf{x})$  has a continuous non-zero differential  $\delta g(\mathbf{x})/\delta x_i$  at  $\mathbf{x} = \boldsymbol{\mu}$ . Define  $\delta g/\delta \mathbf{x}|_{\mathbf{x}=\boldsymbol{\mu}}$  as the vector  $(\delta g(\mathbf{x})/\delta x_1, \delta g(\mathbf{x})/\delta x_2, \dots, \delta g(\mathbf{x})/\delta x_S)'$  evaluated at  $\mathbf{x} = \boldsymbol{\mu}$ . Then,  $\sqrt{n}(g(\hat{\boldsymbol{\mu}})-g(\boldsymbol{\mu}))$  asymptotically follows the multivariate normal distribution with mean  $\mathbf{0}$  and variance-covariance matrix  $(\delta g/\delta \mathbf{x})'|_{\mathbf{x}=\boldsymbol{\mu}} \boldsymbol{\Sigma} (\delta g/\delta \mathbf{x})|_{\mathbf{x}=\boldsymbol{\mu}}$ . For example, if a random vector  $(Y_1, Y_2, \dots, Y_S)'$  follows a multinomial distribution with parameters  $n$  and  $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_S)'$  then, by the CLT, for large  $n$  the random vector  $(\hat{\pi}_1, \hat{\pi}_2, \dots, \hat{\pi}_S)'$  asymptotically has the multivariate normal distribution with mean  $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_S)'$  and co-variance matrix  $[\text{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}']/n$ , where  $\hat{\pi}_i = Y_i/n$ , and  $\text{diag}(\boldsymbol{\pi})$  is a diagonal matrix with diagonal elements equal to  $\pi_i$ . Thus using the delta method, we may claim that  $\sqrt{n}(g(\hat{\boldsymbol{\mu}})-g(\boldsymbol{\mu}))$  asymptotically follows the multivariate normal distribution with mean  $\mathbf{0}$  and variance  $\sum_i \pi_i (\delta g/\delta x_i)^2|_{\mathbf{x}=\boldsymbol{\pi}} - (\sum_i \pi_i (\delta g/\delta x_i)|_{\mathbf{x}=\boldsymbol{\pi}})^2$ .

## A.2 The Proportional Hazard Model

The hazard function  $\lambda(t; \mathbf{x})$  is defined to be the instantaneous failure probability at time  $t$  conditional on survival up to time  $t$ . The proportional hazard model for an individual with covariate  $\mathbf{x}$  is given by

$$\lambda(t; \mathbf{x}) = \lambda_0(t) \exp(-\boldsymbol{\beta}^T \mathbf{x}), \quad (\text{A.1})$$

where  $\lambda_0(t)$  is the hazard function at  $\mathbf{x}=\mathbf{0}$  and  $\boldsymbol{\beta}$  is a vector of unknown parameters. This model can be used in survival analysis (Cox, 1972). Here the survival function  $S(t; \mathbf{x})$ , the probability of surviving beyond time  $t$  given the co-variate  $\mathbf{x}$ , satisfies

$$-\log \{S(t; \mathbf{x})\} = \Lambda_0(t) \exp(-\boldsymbol{\beta}^T \mathbf{x}), \quad (\text{A.2})$$

where  $\Lambda_0(t) = \int_0^t \lambda(s) ds$ . So for any two individuals with covariates  $\mathbf{x}_1$  and  $\mathbf{x}_2$  respectively the survival function satisfy

$$\log S(t; \mathbf{x}_1) / \log \{S(t; \mathbf{x}_2)\} = \exp\{\boldsymbol{\beta}^T (\mathbf{x}_2 - \mathbf{x}_1)\}. \quad (\text{A.3})$$

That is, the ratio of log survival functions, like the ratio of the hazard functions depends only on the difference between the covariate values  $\mathbf{x}_2 - \mathbf{x}_1$  and is constant for all  $t$ . For discrete data, the proportional hazards model (A.2) becomes (McCullagh, 1980)

$$-\log \{1 - \gamma_j(\mathbf{x})\} = \exp(\theta_j - \boldsymbol{\beta}^T \mathbf{x}), \quad (\text{A.4})$$

where  $1 - \gamma_j(\mathbf{x})$  is the probability of survival beyond category  $j$  given covariate values  $\mathbf{x}$ . Logarithm of the above equation gives us a more appropriate linear structure analogous to the linear logistic model as follows

$$\log [-\log \{1 - \gamma_j(\mathbf{x})\}] = \theta_j - \boldsymbol{\beta}^T \mathbf{x}, \quad (\text{A.5})$$

the transformation to linearity being called the complementary log-log transform.

### A.3 Ordered Logistic Regression

The ordered logistic regression model is an extension of the binary model. The model is frequently employed with three or more ordered levels. The foremost ordered logistic model is the proportional odds model. The proportional odds model assumes that the model coefficients for each level of response are equal or slopes can be parameterized for any of the binomial links e.g., logit or logistic, the probit, complementary loglog, loglog and Cauchit as found in generalised linear models.

### A.4 Odds Ratio

The odds ratio (OR) is defined as the ratio of two odds of interest and is calculated from Table A.1 as

Table A.1: Data of a single RCT

| Intervention  | Success/Alive | Failure/Dead |
|---------------|---------------|--------------|
| New treatment | a             | b            |
| Control       | c             | d            |

$$\text{OR} = \frac{(a/n)/(b/n)}{(c/n)/(d/n)} = \frac{ad}{bc}, \quad (\text{A.6})$$

where  $a$ ,  $b$ ,  $c$ , and  $d$  are the cell frequencies of the four cells in a RCT setting for a  $2 \times 2$  table. For desirable outcomes OR greater than one indicates improvement by the new treatment while an OR less than one means the new treatment is less effective. For undesirable outcomes the converse is true. The following large sample variance (Fleiss, 1993) of the log OR is commonly

used for the purpose of pooling in meta-analysis and for computing the CI.

$$\text{var}(\ln \text{ OR}) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}. \quad (\text{A.7})$$

For zero cell count problem, it is recommended adding 0.5 to each cell of all the studies of  $2 \times 2$  tables. This also reduces the bias caused by one or more small cells in the study table. Under the normality assumption of the  $\ln \text{ OR}$ , a 95% CI for  $\ln \text{ OR}$  can be found as

$$\exp[\ln \text{ OR} \pm 1.96 \times \sqrt{\text{var}(\ln \text{ OR})}]. \quad (\text{A.8})$$

## A.5 Relative Risk

The relative risk or risk ratio (RR) is defined as the probability of an event in the treatment group divided by the probability of an event in the control group (Table A.1) as

$$\text{RR} = \frac{a/(a+b)}{c/(c+d)}. \quad (\text{A.9})$$

Useful variance expression of the  $\log \text{ RR}$  can be found as

$$\text{var}(\ln \text{ RR}) = \frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}. \quad (\text{A.10})$$

A 95% CI for  $\ln \text{ RR}$  can be calculated under the assumption of normality as

$$\exp[\ln \text{ RR} \pm 1.96 \times \sqrt{\text{var}(\ln \text{ RR})}]. \quad (\text{A.11})$$

There are debates on the choice of binary effect measures between OR and RR. Some researchers prefer OR as an effect measure than RR because OR can be estimated and interpreted for RCTs and case-control study reasonably for rare outcomes. Moreover, some study designs select subjects on the basis of outcome rather than the treatment type (case-control study). RR can not be used for this study design.

## A.6 Risk Difference

Apart from the OR and RR that measure the association between treatment and outcome, risk difference (RD) provides an indication of the impact of the treatment or exposure. The RD is defined for a  $2 \times 2$  table simply as risk in the experimental group minus risk in the control group and is calculated as

$$\text{RD} = a/(a + b) - c/(c + d). \quad (\text{A.12})$$

The variance estimate of RD can be found (Fleiss, 1993) as

$$\text{var}(\text{RD}) = \frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}, \quad (\text{A.13})$$

where  $p_1 = a/(a + b)$ ,  $p_2 = c/(c + d)$ ,  $n_1 = a + b$  and  $n_2 = c + d$ . A 95% CI for RD can be computed as

$$\text{RD} \pm \times \sqrt{\text{var}(\text{RD})}. \quad (\text{A.14})$$

## A.7 Arcsine Difference

The arcsine difference (AS) is defined as

$$\text{AS} = \arcsin\sqrt{\frac{a}{a + b}} - \arcsin\sqrt{\frac{c}{c + d}}. \quad (\text{A.15})$$

The AS is a measure rarely used in medical science which handles zero frequencies naturally. The asymptotic variance of AS which does not depend on the event of probability is given as

$$\text{var}(\text{AS}) = \frac{1}{4(a + b)} + \frac{1}{4(c + d)}. \quad (\text{A.16})$$

## A.8 Continuity Corrections for Zero Cells

Consider the cells of binary data for the  $i$ th study in the following table as:

| Study $i$ | Event | No Event | Total       |
|-----------|-------|----------|-------------|
| Treatment | $a_i$ | $b_i$    | $a_i + b_i$ |
| Control   | $c_i$ | $d_i$    | $c_i + d_i$ |

When any of the four cells is zero, we add 0.5 to all cells of the contingency table. This is otherwise called the Woolf-Haldane correction for the odds ratio (Schlesselman, 1982). Alternatively, statistical software Rmeta adds 0.5 to all the cells of all the studies involved in the meta-analysis. As a result, there is an increase of 1 subject in each row involved in the analysis. In our analysis with  $J \times L$  contingency table, we add  $1/L$  to each cell with zero frequency. For example, in a  $2 \times 3$  contingency table we add  $1/3$  to each cell and for a  $2 \times 4$  contingency table, we add  $1/4$  and so on. Therefore, there is an increase of one subject in each group involved in the analysis as in the case of OR.

## A.9 Peto's Method

This method was first introduced by Peto et al. (1977) and later more elaborately by Yusuf et al. (1987). This method is considered as a break through in the field of meta-analysis. Although this method may produce serious under estimates (Fleiss, 1994) for large treatment or exposure effects or in the meta-analysis of epidemiology (Spector and Thompson, 1991), it is unlikely to be a problem in clinical trials. For the working formula, let us define the following:

$n_i$  = number of subjects in the  $i$ th trial

$n_{ti}$  = number of subjects in the new treatment group

$d_i$  = total number of events from both treatment and control groups

$O_i$  = observed number of events in the treatment group

$E_i = (n_{ti} \times d_i)/n_i$  = expected number of subjects in the treatment group.

For  $k$  studies,  $i = 1, 2, \dots, k$ , the pooled estimate of the OR under this method is calculated as

$$\bar{T}_{Peto(OR)} = \exp \left[ \frac{\sum_{i=1}^k (O_i - E_i)}{\sum_{i=1}^k v_i} \right], \quad (\text{A.17})$$

where  $v_i = E_i[(n_i - n_{ti})/n_i][(n_i - d_i)/(n_i - 1)]$ . An approximate variance estimate is given by

$$\text{var}(\ln \bar{T}_{Peto(OR)}) = \left( \sum_{i=1}^k v_i \right). \quad (\text{A.18})$$

An asymmetric  $100(1 - \alpha)\%$  CI is thus found as

$$\exp \left( \frac{\sum_{i=1}^k (O_i - E_i) \pm z_{\alpha/2} \sqrt{\sum_{i=1}^k v_i}}{\sum_{i=1}^k v_i} \right), \quad (\text{A.19})$$

where  $z_{\alpha/2}$  is the  $(1 - \alpha/2)100$  percentage point of a standard normal distribution.

## A.10 Mantel-Haenszel Method

The Mantel-Haenszel method also known as M-H method was first described by Mantel and Haenszel (1959) for combining ORs for stratified case-control studies. This method is a fixed effects model approach and could be used for a variety of laboratory experiments (Mantel, 1963). In meta-analysis it is

widely used considering each study as an individual stratum (Dickersin and Belin, 1992). For the  $2 \times 2$  table with  $a_i$ ,  $b_i$ ,  $c_i$  and  $d_i$ , the four cell frequencies and  $n_i$ , the total number of subjects in the  $i$ th study,  $i = 1, \dots, k$ , the pooled estimate is calculated as

$$\bar{T}_{MH(OR)} = \frac{\sum_{i=1}^k a_i d_i / n_i}{\sum_{i=1}^k b_i c_i / n_i}. \quad (\text{A.20})$$

A variance estimate (Robin et al., 1986) for the  $\log \bar{T}_{MH(OR)}$  can be computed as

$$v_{MH(\ln OR)} = \frac{\sum_{i=1}^k P_i R_i}{2 \left( \sum_{i=1}^k R_i \right)} + \frac{\sum_{i=1}^k (P_i S_i + Q_i R_i)}{2 \left( \sum_{i=1}^k R_i \right) \left( \sum_{i=1}^k S_i \right)} + \frac{\sum_{i=1}^k (Q_i S_i)}{2 \left( \sum_{i=1}^k S_i \right)}, \quad (\text{A.21})$$

where  $P_i = (a_i + d_i)/n_i$ ,  $Q_i = (b_i + c_i)/n_i$ ,  $R_i = (a_i d_i)/n_i$ , and  $S_i = (b_i c_i)/n_i$ . A  $100(1 - \alpha)\%$  CI for the pooled OR,  $\theta$  can be found as

$$\exp \left[ \ln \bar{T}_{MH(OR)} \pm z_{\alpha/2} (v_{MH(\ln OR)})^{1/2} \right], \quad (\text{A.22})$$

where  $z_{\alpha/2}$  is the  $(1 - \alpha/2)100$  percentage point of a standard normal distribution.

## A.11 Estimating Mean and SD from Median, Range and Sample Sizes

Hozo et al. (2005) produced an excellent work on estimating the mean and standard deviation from the median, range and sample size. The sample mean,  $\bar{x}$  can be used to estimate the population mean as

$$\bar{x} = \frac{a + 2m + b}{4}, \quad (\text{A.23})$$

where  $m$  is the median and  $a$  and  $b$  are the low and high end of the range respectively. The variance can be estimated using the formula as

$$S^2 \approx \frac{1}{12} \left( \frac{(a - 2m + b)^2}{4} + (b - a)^2 \right). \quad (\text{A.24})$$

In addition, the well estimators of standard deviation Range/4 for a normal distribution and Range/6 for any random distribution also useful in meta-analysis. The best estimating formulae for mean and standard deviation (sd) depending on the sample sizes for any unknown distribution are listed below:

Table A.2: The best estimating formulas for mean and sd

| Sample size   | $n \leq 15$  | $15 < n \leq 25$ | $25 < n \leq 70$ | $70 < n$ |
|---------------|--------------|------------------|------------------|----------|
| Estimate mean | Formula A.23 | Formula A.23     | Median           | Median   |
| Estimate sd   | Formula A.24 | Range/4          | Range/4          | Range/6  |

## A.12 Kendall's tau

Suppose we take two members  $x_i$  and  $x_j$  at random from a continuous population. Then we may consider the probability  $x_i < x_j$  and  $x_i > x_j$ . Again let us take two members  $x_i, y_i$  and  $x_j, y_j$  from a bivariate continuous population. Then we may consider the probabilities of concordance of type 1 as the conditional probability that  $y_i < y_j$  given that  $x_i < x_j$ . That is,

$$\pi_c = \text{Prob}(y_i < y_j | x_i < x_j). \quad (\text{A.25})$$

and

$$1 - \pi_c = \text{Prob}(y_i > y_j | x_i < x_j). \quad (\text{A.26})$$

The probability  $\pi_c$  represents a property of the population. Let us draw a sample of  $n$  values at random from the bivariate population and arrange the  $x$  values in ascending order of magnitude. Of the  $1/2n(n - 1)$  pairs of  $x$

values which we may choose for comparison, some have the corresponding  $y$  values in ascending order and some do not. The number of those which do divided by  $1/2n(n-1)$  is clearly an estimator (unbiased) of  $\pi_c$ . If  $p_c$  is this proportion and  $q_c=1-p_c$ , then the Kendall's tau coefficient ( $\tau$ ) for this example is

$$\tau = p_c - q_c = 2p_c - 1. \quad (\text{A.27})$$

Thus we could therefore define  $\tau$  in terms of concordance and arrive at a coefficient which has an analogue in the continuous case.

### A.13 Mann-Whitney U Statistic

The Mann-Whitney statistic  $U$  is obtained by ordering two samples ( $n_1 + n_2$ ) observations according to their magnitude and counting the number of observations in sample II that precede each observation in sample I. For comparing two distributions of independent continuous random variables  $U$  can be written as

$$U = \hat{P}_{21}/n_1n_2, \quad (\text{A.28})$$

where  $\hat{P}_{21}$  is the sample version of  $P_{21} = P(Y_2 > Y_1)$ . The GOR ( $\Gamma$ ) for two independent continuous random variables is written as

$$\Gamma = P(Y_2 > Y_1)/\{1 - P(Y_2 > Y_1)\}. \quad (\text{A.29})$$

Therefore, the relationship between  $\Gamma$  and  $U$  can be written as

$$\begin{aligned} \Gamma &= \frac{U/n_1n_2}{(n_1n_2 - U)/n_1n_2} \\ &= \frac{U}{n_1n_2 - U}. \end{aligned} \quad (\text{A.30})$$

As a non-parametric approach the  $U$  statistic has the advantage that the two sample under consideration may not necessarily have the same number of observations.

# Appendix B

## B.1 R Code for Chapter 2

Listing B.1: R code for the GOR functions, weights, CIs, PGORs and CIs

---

```
# GOR functions for 2 x 5 contingency table
tr <- -c(f11, f12, f13, f14, f15, n11) # f11 = count in cell (1, 1), f12 = count
in cell (1, 2) and so on
cl <- -c(f21, f22, f23, f24, f25, n21)
e1 <- -cbind(tr, cl)
e <- -t(e1)
n11 <- -e[1,6]
n21 <- -e[2,6]
p11 <- -e[1,1]/e[1,6]
p12 <- -e[1,2]/e[1,6]
p13 <- -e[1,3]/e[1,6]
p14 <- -e[1,4]/e[1,6]
p15 <- -e[1,5]/e[1,6]
p21 <- -e[2,1]/e[2,6]
p22 <- -e[2,2]/e[2,6]
```

```

p23 < -e[2,3]/e[2,6]
p24 < -e[2,4]/e[2,6]
p25 < -e[2,5]/e[2,6]
pc < -((p11p22)+(p11p23)+(p11p24)+(p11p25)+(p12p23)+(p12p24)+
(p12p25) +(p13p24)+(p13p25)+(p14p25))
pd < -((p12p21)+(p13p21)+(p13p22)+(p14p21)+(p14p22)+(p14p23)+
(p15p21) +(p15p22)+(p15p23)+(p15p24))
gi < -pc/pd
#Weight for the ith study
tr < -c(f11, f12, f13, f14, f15, n1)
cl < -c(f21, f22, f23, f24, f25, n2)
e1 < -cbind(tr, cl)
e < -t(e1)
n1 < -e[1,6]
n2 < -e[2,6]
p11 < -e[1,1]/e[1,6]
q11 < -1-p11
p12 < -e[1,2]/e[1,6]
q12 < -1-p12
p13 < -e[1,3]/e[1,6]
q13 < -1-p13
p14 < -e[1,4]/e[1,6]
q14 < -1-p14
p21 < -e[2,1]/e[2,6]
q21 < -1-p21
p22 < -e[2,2]/e[2,6]
q22 < -1-p22

```

```

p23 <- -e[2,3]/e[2,6]
q23 <- -1-p23
p24 <- -e[2,4]/e[2,6]
q24 <- -1-p24
wli <- -(1/(n1p11q11)+1/(n1p12q12)+1/(n1p13q13)+ 1/(n1p14q14)
+ 1/(n2p21q21)+1/(n2p22q22)+1/(n2p23q23)+1/(n2p24q24))
wi <- -1/wli; i=1, 2, ..., k=5. # number of studies.
# pooling GOR
wi <- -c(w1,w2,w3,w4,w5)
sei <- -sqrt(1/wi)
g1 <- -log(gi)
t <- -sum(g1*wi)/sum(wi).
#calculation of ci for individual GOR
ll <- -format(exp(log(g) -1.96wi-1/2),digits=4)
ul <- -format(exp(log(g) +1.96wi-1/2),digits=4)
ci <- -data.frame(ll,ul)
#CI for pooled GOR
w <- -sum(wi)
ll <- -format(exp(log(t) -1.96w-1/2),digits=4)
ul <- -format(exp(log(t) +1.96w-1/2),digits=4)
ci <- -data.frame(ll,ul)

```

---

## B.2 Sample Size Weight Method

Listing B.2: R code for sample size weights

---

```

# CI for ORG
selg1 <- -(2*seg1)/((1 - ga1)2)

```

```

# selg < -SE(logORG)
ci1 < -c(or1*exp((-3.92*seg1)/(1-(ga1)2),or1*exp((3.92*seg1)/(1-(ga1)2)))
# weights
n1 < -1st col total
n2 < -2nd col total
n3 < -3rd col total
n4 < -4th col total
n5 < -5th col total
ga < -gammas
seg < -standard error of gammas
se2 < -seg2
tni < -n1*n2/(n1+n2)+ n1*n3/(n1+n3)+ n3*n2/(n3+n2)
+ n1*n4/(n1+n4)+n1*n5/(n1+n5)+n4*n2/(n4+n2)+ n5*n2/(n5+n2)
+n3*n4/(n3+n4)+n3*n5/(n3+n5)+n4*n5/(n4+n5)
# Pooling by Sample size weighting
ga < - gammas for the studies # easily obtainable from crosstab command
of SPSS
seg < - standard errors of gammas # easily obtainable from crosstab com-
mand of SPSS
se2 < -seg2
tn2 < -tni2
gb < -sum(tni*ga)/sum(tni)
ORb < -(1+gb)/(1-gb)
segb < -sqrt(sum(tn2*se2))/sum(tni)
cip < -c(ORb*exp((-3.92*segb)/(1-(gb)2),ORb*exp((3.92*segb)/(1-(gb)2)))

```

---

### B.3 R Code for Chapter 3

Listing B.3: R code for GORs, weights for three categories of tacrine trials

---

```
# GORs for 2 x 3 contingency tables
tr <- -c(f11, f12, f13, Total)
cl <- -c(f21, f22, f23, Total)
e1 <- -cbind(tr, cl)
e <- -t(e1)
n11 <- -e[1,4]
n21 <- -e[2,4]
p11 <- -e[1,1]/e[1,4]
p12 <- -e[2,1]/e[2,4]
p13 <- -e[1,3]/e[1,4]
p21 <- -e[1,2]/e[1,4]
p22 <- -e[2,2]/e[2,4]
p23 <- -e[2,3]/e[2,4]
pc <- -(p11*p22)+(p11*p23)+(p21*p23)
pd <- -(p12*p21)+(p13*p12)+(p13*p22)
gi <- -pc/pd
# Weights
p <- -c(32,54,24,110)
t <- -c(24,45,27,96)
e1 <- -cbind(p,t)
e <- -t(e1)
n1 <- -e[1,4]
n2 <- -e[2,4]
p11 <- -e[1,1]/e[1,4]
```

```

q11 <- -1-p11
p12 <- -e[1,2]/e[1,4]
q12 <- -1-p12
p21 <- -e[2,1]/e[2,4]
q21 <- -1-p21
p22 <- -e[2,2]/e[2,4]
q22 <- -1-p22
wli <- -1/(n1*p11*q11)+1/(n1*p12*q12)+1/(n2*p21*q21)+1/(n2*p22*q22)
wi <- -1/wli
# Calculation of ci for individual GOR for tacrine trials
g <- gor for the studies
wi <- weights
ll <- format(exp(log(g) -1.96*wi(-1/2)),digits=4)
ul <- format(exp(log(g) +1.96*wi(-1/2)),digits=4)
ci <- data.frame(ll,ul)
# pooling GOR
g2 <- c(1.274390, 1.185958, 1.312500, 2.079957, 1.546875 )
g <- log(g2)
wi <- c(5.623291, 5.718636, 2.767237, 13.253025, 9.190350)
t <- -sum(g*wi)/sum(wi)
pt <- exp(t)
# CI for pooled GOR
w <- sum(wi)
ll <- format(exp(log(1.583732) -1.96*w(-1/2)),digits=4)
ul <- format(exp(log(1.583732) +1.96*w(-1/2)),digits=4)
ci <- data.frame(ll,ul)
# calculation of PGOR

```

```

g<-c(1.274390, 1.185958, 1.312500, 2.079957, 1.546875 )
df<-length(g)-1
w<-c(5.623291, 5.718636, 2.767237, 13.253025, 9.190350)
q2<-1.831287# value of the chi-square statistic
k<-1-((df-2)/q2)
pg<-log(1.583732)+k*(log(g)-log(1.583732))
g1<-exp(pg)
#CI for PGOR
wi<-c(5.623291, 5.718636, 2.767237, 13.253025, 9.190350)
ll<-format(exp(pg-1.96*wi(-1/2)),digits=4)
ul<-format(exp(pg+1.96*wi(-1/2)),digits=4)
cip<-data.frame(ll,ul)
#Standard chisquare test with logGOR
w<-c(5.623291, 5.718636, 2.767237, 13.253025, 9.190350)
t<-c(1.274390, 1.185958, 1.312500, 2.079957, 1.546875 )
T<-log(t)
T2<-T*T
T3<-w%*%T
T4<-T3/sum(w)
T5<-T32
T6<-sum(wT2)
T7<-sum(w)
q<-T6-(T5/T7)
q2<-sum(w*((T-T4)**2))
Q1<-sum(w*T2)-(T5/T7)
# I2 for tacrine
Q<-1.831287

```

```

df <- -4
I2 <- -(Q-df)/Q * 100
if Q > df+1 then
d1 <- -2*df-1
q1 <- -2 * Q
B <- -0.5 * (log(Q)-log(df))/(sqrt(q1)-sqrt(d1))
or
if Q < -df+1
p1 <- -3 * (df - 1)2
p2 <- -1/2*(df-1)*(1-(1/p1))
B <- -sqrt(p2)
L <- -exp(0.5*log(Q/df)-1.96*B)
U <- -exp(0.5*log(Q/df)+1.96*B)
LL <- -(L2 - 1)/L2 * 100
UL <- -(U2 - 1)/U2 * 100
#95% CI for GOR tacrine trials
x <- -1.58
y <- -1
plot(x,y,xlim=c(-1.5,6),ylim=c(-1,7),xaxt="n",
yaxt="n",frame.plot = NULL,cex=1)
segments(0, 0, 0, 7, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(0, 0, 6, 0, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(1, 0, 1, 7, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(1.58, 0, 1.58, 7, col = "black", lty = 5, lwd = par("lwd"))
x5 <- -c(0.558, 2.91)
y5 <- -c(6,6)
lines(x5,y5,ty="l")

```

```
x6 < -c(0.523, 2.69)
y6 < -c(5,5)
lines(x6,y6,ty="l")
x7 < -c(0.404, 4.26)
y7 < -c(4,4)
lines(x7,y7,ty="l")
x8 < -c(1.214, 3.56)
y8 < -c(3,3)
lines(x8,y8,ty="l")
x9 < -c(0.81, 2.95)
y9 < -c(2,2)
lines(x9,y9,ty="l")
x10 < -c(1.145,2.19)
y10 < -c(1,1)
lines(x10,y10,ty="l")
text(0, 6, labels="Study 1", pos=2,adj = c(0,0), cex = .8)
text(0, 5, labels="Study 2", pos=2,adj = c(0,0), cex = .8)
text(0, 4, labels="Study 3", pos=2,adj = c(0,0), cex = .8)
text(0, 3, labels="Study 4", pos=2,adj = c(0,0), cex = .8)
text(0, 2, labels="Study 5", pos=2,adj = c(0,0), cex = .8)
text(0, 1, labels="Meta-analysis",adj = c(0,0), pos=2, cex =.8)
#PGOR
x5 < -c(0.707, 3.69)
y5 < -c(5.75,5.75)
lines(x5,y5,ty="l",col="black",lty=5)
x6 < -c(0.717, 3.69)
y6 < -c(4.75,4.75)
```

```

lines(x6,y6,ty="l",col="black",lty=5)
x7< -c(0.496, 5.23)
y7< -c(3.75,3.75)
lines(x7,y7,ty="l",col="black",lty=5)
x8< -c(0.901, 2.65)
y8< -c(2.75,2.75)
lines(x8,y8,ty="l",col="black",lty=5)
x9< -c(0.831, 3.03)
y9< -c(1.75,1.75)
lines(x9,y9,ty="l",col="black",lty=5)
text(1.58, 0, labels="1.58", pos=1, adj = c(0,0),cex = .8)
text(0, 0, labels="0", pos=1, adj = c(0,0),cex = .8)
text(1, 0, labels="1", pos=1, adj = c(0,0),cex = .8)
#text(2, 0, labels="2", pos=1, adj = c(0,0),cex = .8)
text(4, 0, labels="4", pos=1, adj = c(0,0),cex = .8)
text(6, 0, labels="6", pos=1, adj = c(0,0),cex = .8)
text(2.8, -.71, labels="Generalised Odds Ratio (log-scale)", pos=1, adj =
c(2.50,-1),cex = .8)
points(6,0,pch="",cex =1)
points(4,0,pch="",cex =1)

```

---

## B.4 Misoprostol Trials

Listing B.4: R code for GORs, weights, CI, PGORs,  $L_n$  and  $I^2$

---

```

# GORs for 2 x 3 contingency table
tr< -c(f11, f12, f13, n11)
cl< -c(f21, f22, f23, n21)
e1< -cbind(nc,ca)

```

```
e < -t(e1)
n11 < -e[1,4]
n21 < -e[2,4]
p11 < -e[1,1]/e[1,4]
p12 < -e[2,1]/e[2,4]
p13 < -e[1,3]/e[1,4]
p21 < -e[1,2]/e[1,4]
p22 < -e[2,2]/e[2,4]
p23 < -e[2,3]/e[2,4]
pc < -(p11*p22)+(p11*p23)+(p21*p23)
pd < -(p12*p21)+(p13*p12)+(p13*p22)
gi < -pc/pd, i = 1, 2, ..., 10.
# weights
x11 < -f11
x12 < -f12
x13 < -f13
x21 < -f21
x22 < -f22
x23 < -f23
n1 < -1st column total
n2 < -2nd column total
p11 < -x11/n1
q11 < -1-p11
p12 < -x12/n1
q12 < -1-p12
p21 < -x21/n2
q21 < -1-p21
```

```

p22 < -x22/n2
q22 < -1-p22
wli < -1/(n1*p11*q11)+1/(n1*p12*q12)+1/(n2*p21*q21)+
1/(n2*p22*q22)
w1 < -1/wli, i = 1, 2, ..., 10.
# Pooling GOR
g < -c(28.526316, 68.645161, 5.450704, 23.123711, 15.714286, 2.990099, 3.126642
, 3.987988, 8.307692, 1.757374)
sn < -log(g)
w < -c(1.0858756, 0.2918908, 1.7150532, 0.2874851, 0.1986054, 3.1277129,
3.3480683, 0.5934123, 0.7058769, 4.2951792)
t < -sum(sn*w)/sum(w)
pt < -exp(t)
# Calculation of the  $L_n$  statistic
sn < -c(28.526316, 68.645161, 5.450704, 23.123711, 15.714286, 2.990099,
3.126642 , 3.987988, 8.307692, 1.757374)
w < -c(1.0858756, 0.2918908, 1.7150532, 0.2874851, 0.1986054, 3.1277129,
3.3480683, 0.5934123, 0.7058769, 4.2951792)
wd < -diag(w)
pi < -log(sn)
s0 < -rep(log(3.878293),10)
x < -(pi-s0)
p2 < -t(x)% * %wd
ln < -p2% * %x
p3 < -(t(x)%*%wd%*%x)
# CI for individual GOR for misoprostol trials
g < -c(28.526316, 68.645161, 5.450704, 23.123711, 15.714286, 2.990099, 3.126642

```

```

, 3.987988, 8.307692, 1.757374)
wi< -c(1.0858756, 0.2918908, 1.7150532, 0.2874851, 0.1986054, 3.1277129,
3.3480683, 0.5934123, 0.7058769, 4.2951792)
ll< -format(exp(log(g) -1.96*wi(-1/2)),digits=4)
ul< -format(exp(log(g) +1.96*wi(-1/2)),digits=4)
ci< -data.frame(ll,ul)
# ci for pooled GOR
w< -sum(wi)
ll< -format(exp(log(3.878293) -1.96*w(-1/2)),digits=3)
ul< -format(exp(log(3.878293) +1.96*w(-1/2)),digits=3)
cip< -data.frame(ll,ul)
# Calculation of PGOR
g< -c(28.526316, 68.645161, 5.450704, 23.123711, 15.714286, 2.990099, 3.126642
, 3.987988, 8.307692, 1.757374)
df< -length(g)-1
q2< - 11.70625
k< - 1-((df-2)/q2)
pg< -log(3.878293)+k*(log(g)-log(3.878293))
g1< -format(exp(pg),digits=3)
g1< -c(8.65,12.31, 4.45, 7.95,6.81, 3.49, 3.56, 3.92,5.27,2.82)
wi< -c(1.0858756, 0.2918908, 1.7150532, 0.2874851, 0.1986054, 3.1277129,
3.3480683, 0.5934123, 0.7058769, 4.2951792)
ll< -format(exp(log(g1) -1.96*wi(-1/2)),digits=3) ul< -format(exp(log(g1)
+1.96*wi(-1/2)),digits=3) pci< -data.frame(ll,ul)
# Standard chisquare test
w< -c(1.0858756, 0.2918908, 1.7150532, 0.2874851, 0.1986054, 3.1277129,
3.3480683, 0.5934123, 0.7058769, 4.2951792)

```

```

t <- -c(28.526316, 68.645161, 5.450704, 23.123711, 15.714286, 2.990099, 3.126642
, 3.987988, 8.307692, 1.757374)
T <- -log(t)
T2 <- -T*T
T3 <- -w%%*%T
T4 <- -T3/sum(w)
T5 <- -T3^2
T6 <- -sum(w*T2)
T7 <- -sum(w)
q <- -T6-(T5/T7)
q2 <- -sum(w*((T-T4)**2))
Q1 <- -sum(w*T2)-(T5/T7)
p <- -1-pchisq(11.70625, df= 9)
# I^2 for misoprostol Q <- -11.70625
df <- -9
I2 <- -(Q-df)/Q*100
if Q > df+1 then
d1 <- -2*df-1
q1 <- -2 * Q
B <- -0.5*(log(Q)-log(df))/(sqrt(q1)-sqrt(d1))
or
if Q < -df+1 then
p1 <- -3 * (df - 1)^2
p2 <- -1/2*(df-1)*(1-(1/p1))
B <- -sqrt(p2)
L <- -exp(0.5*log(Q/df)-1.96*B)
U <- -exp(0.5*log(Q/df)+1.96*B)

```

$$LL < -(L^2 - 1)/L^2 * 100$$

$$UL < -(U^2 - 1)/U^2 * 100$$

---

## B.5 Simulation Studies

Listing B.5: R code for GORs and weights from simulated studies

---

```
set.seed(1000)
# Study 1
x1 <- rmultinom(1, size = 29, prob=c(0.1099476,0.8089005,0.08115183))
x2 <- rmultinom(1, size = 30, prob=c(0.1797753,0.6910112,0.1292135))
x11 <- data.frame(x1,x2)
x0 <- t(x11)
s <- c(29,30)
e1 <- cbind(x0,s)
n11 <- e1[1,4]
n21 <- e1[2,4]
p11 <- e1[1,1]/e1[1,4]
p12 <- e1[2,1]/e1[2,4]
p13 <- e1[1,3]/e1[1,4]
p21 <- e1[1,2]/e1[1,4]
p22 <- e1[2,2]/e1[2,4]
p23 <- e1[2,3]/e1[2,4]
pc <- -(p11*p22)+(p11*p23)+(p21*p23)
pd <- -(p12*p21)+(p13*p12)+(p13*p22)
g1 <- pc/pd
# Weight
n1 <- e1[1,4]
```

```

n2 <- -e1[2,4]
p11 <- -e1[1,1]/e1[1,4]
q11 <- -1-p11
p12 <- -e1[1,2]/e1[1,4]
q12 <- -1-p12
p21 <- -e1[2,1]/e1[2,4]
q21 <- -1-p21
p22 <- -e1[2,2]/e1[2,4]
q22 <- -1-p22
w11 <- -1/(n1*p11*q11)+1/(n1*p12*q12)+1/(n2*p21*q21)+1/(n2*p22*q22)
w1 <- -1/w11
# Study 2
x1 <- -rmultinom(1, size = 30, prob=c(0.8089005,0.1099476,0.08115183))
x2 <- -rmultinom(1, size = 30, prob=c(0.1797753,0.6910112,0.1292135))
# Study 3
x1 <- -rmultinom(1, size = 30, prob=c(0.8089005,0.08115183,0.1099476))
x2 <- -rmultinom(1, size = 30, prob=c(0.1797753,0.6910112,0.1292135))
# Study 4
x1 <- -rmultinom(1, size = 30, prob=c(0.1099476,0.8089005,0.08115183))
x2 <- -rmultinom(1, size = 29, prob=c(0.6910112,0.1797753,0.1292135))
# Study 5
x1 <- -rmultinom(1, size = 10, prob=c(0.1099476,0.8089005,0.08115183))
x2 <- -rmultinom(1, size = 10, prob=c(0.6910112,0.1292135,0.1797753))
# Study 6
x1 <- -rmultinom(1, size = 103, prob=c(0.1099476,0.8089005,0.08115183))
x2 <- -rmultinom(1, size = 114, prob=c(0.1797753,0.1292135,0.6910112))
x11 <- -data.frame(x1,x2)

```

```
x6< -t(x11)
s< -c(103,114)
e6< -cbind(x6,s)
# Study 7
x1< -rmultinom(1, size = 73, prob=c(0.2099476,0.7089005,0.08115183))
x2< -rmultinom(1, size = 80, prob=c(0.1797753,0.6910112,0.1292135))
x11< -data.frame(x1,x2)
x7< -t(x11)
s< -c(73,80)
e7< -cbind(x7,s)
# Study 8
x1< -rmultinom(1, size = 46, prob=c(0.7089005,0.08115183,0.2099476))
x2< -rmultinom(1, size = 74, prob=c(0.1797753,0.6910112,0.1292135))
x11< -data.frame(x1,x2)
x8< -t(x11)
s< -c(46,74)
e8< -cbind(x8,s)
# Study 9
x1< -rmultinom(1, size = 32, prob=c(0.2099476,0.7089005,0.08115183))
x2< -rmultinom(1, size = 38, prob=c(0.6910112,0.1292135,0.1797753))
x11< -data.frame(x1,x2)
x9< -t(x11)
s< -c(32,38)
e9< -cbind(x9,s)
# Study 10
x1< -rmultinom(1, size = 76, prob=c(0.1099476,0.8089005,0.08115183))
x2< -rmultinom(1, size = 83, prob=c(0.1292135,0.6910112,0.1797753))
```

```

x11<-data.frame(x1,x2)
x10<-t(x11)
s<-c(76,83)
e10<-cbind(x10,s)
# Q Statistic
w<-c(0.7578464, 2.2304833, 0.5860116, 1.3440000, 0.2526316, 2.6620506,
6.9675671, 0.7579955, 1.5819698, 2.2422012)
g<-c(1.18367347, 9.37704918, 8.62025316, 0.12962963, 0.01219512, 4.39120879,
1.45936698, 3.93109541, 0.28862974, 0.78560720)
T<-log(g)
T2<-T*T
T3<-w%%T
T4<-T3/sum(w)
T5<-T3*T3
T6<-sum(w*T2)
T7<-sum(w)
q<-T6-(T5/T7)
q2<-sum(w*((T-T4)**2))
Q1<-sum(w*T2)-(T5/T7)
p<-1-pchisq(32.24718, df= 9)
#I2 statistic
Q<-32.24718
df<-9
I2<-(Q - df)/Q * 100
if Q > df + 1 then
d1<-2 * df - 1
q1<-2 * Q

```

```

B <- -0.5*(log(Q)-log(df))/(sqrt(q1)-sqrt(d1))
or
if Q <= df + 1
p1 <- -3 * (df - 1)^2
p2 <- -1/2 * (df - 1) * (1 - (1/p1))
B <- -sqrt(p2)
L <- -exp(0.5*log(Q/df)-1.96*B)
U <- -exp(0.5*log(Q/df)+1.96*B)
LL <- -(L^2 - 1)/L^2 * 100
UL <- -(U^2 - 1)/U^2 * 100
# Pooling GOR
w <- -c(0.7578464, 2.2304833, 0.5860116, 1.3440000, 0.2526316, 2.6620506,
6.9675671, 0.7579955, 1.5819698, 2.2422012)
g <- -c(1.18367347, 9.37704918, 8.62025316, 0.12962963, 0.01219512, 4.39120879,
1.45936698, 3.93109541, 0.28862974, 0.78560720)
sn <- -log(g)
t <- -sum(sn*w)/sum(w)
pt <- -exp(t)
# Calculation of the divergent statistic  $L_n$ 
w <- -c(0.7578464, 2.2304833, 0.5860116, 1.3440000, 0.2526316, 2.6620506,
6.9675671, 0.7579955, 1.5819698, 2.2422012)
sn <- -c(1.18367347, 9.37704918, 8.62025316, 0.12962963, 0.01219512, 4.39120879,
1.45936698, 3.93109541, 0.28862974, 0.78560720)
wd <- -diag(w)
pi <- -log(sn)
s0 <- -rep(log( 1.482130),10)
x <- -(pi-s0)

```

```

p2< -t(x)%*%wd
ln< -p2%*%x
p3< -(t(x)%*%wd%*%x)
# Calculation of CI for individual study
wi< -c(0.7578464, 2.2304833, 0.5860116, 1.3440000, 0.2526316, 2.6620506,
6.9675671, 0.7579955, 1.5819698, 2.2422012)
g< -c(1.18367347, 9.37704918, 8.62025316, 0.12962963, 0.01219512, 4.39120879,
1.45936698, 3.93109541, 0.28862974, 0.78560720)
ll< -format(exp(log(g) -1.96*wi(-1/2)),digits=3)
ul< -format(exp(log(g) +1.96*wi(-1/2)),digits=3)
ci< -data.frame(ll,ul)
# CI for pooled GOR
w< -sum(wi)
ll< -format(exp(log(1.482130) -1.96*w(-1/2)),digits=4)
ul< -format(exp(log(1.482130) +1.96*w(-1/2)),digits=4)
cip< -data.frame(ll,ul)
# Calculation of predicted GOR
g< -c(1.18367347, 9.37704918, 8.62025316, 0.12962963, 0.01219512, 4.39120879,
1.45936698, 3.93109541, 0.28862974, 0.78560720)
df< -length(g)-1
q2< -32.24718
k< -1-((df-2)/q2)
pg< -log(1.482130)+k*(log(g)-log(1.482130))
g1< -format(exp(pg),digits=3)
g1< -c(1.2429,6.2828,5.8822,0.2200,0.0346,3.4689, 1.4643,3.1809 ,0.4117,0.9017)
wi< -c(0.7578464, 2.2304833, 0.5860116, 1.3440000, 0.2526316, 2.6620506,
6.9675671, 0.7579955, 1.5819698, 2.2422012)

```

```

ll< -format(exp(pg -1.96*wi(-1/2)),digits=3)
ul< -format(exp(pg +1.96*wi(-1/2)),digits=3)
pci< -data.frame(ll,ul)
# Q statistic deleting 4th study
wi< -c(0.7578464, 2.2304833, 0.5860116, 0.2526316, 2.6620506, 6.9675671,
0.7579955, 1.5819698, 2.2422012)
ti< -c(1.18367347, 9.37704918, 8.62025316, 0.01219512, 4.39120879, 1.45936698,
3.93109541, 0.28862974, 0.78560720)
lt< -log(ti)
tb< -sum(lt*wi)/sum(wi)
t< -exp(tb)
q< -(wi*(lt-tb)2)
Q< -sum(wi*(lt - tb)2)
# Pooling GOR deleting 4th study
w< -c(0.7578464, 2.2304833, 0.5860116, 0.2526316, 2.6620506, 6.9675671,
0.7579955, 1.5819698, 2.2422012)
g< -c(1.18367347, 9.37704918, 8.62025316, 0.01219512, 4.39120879, 1.45936698,
3.93109541, 0.28862974, 0.78560720)
sn< -log(g)
t< -sum(sn*w)/sum(w)
pt< -exp(t)
# CI of pooled
wi< -c(0.7578464, 2.2304833, 0.5860116, 0.2526316, 2.6620506, 6.9675671,
0.7579955, 1.5819698, 2.2422012)
w< -sum(wi)
ll< -format(exp(log(1.777164) -1.96*w(-1/2)),digits=3)
ul< -format(exp(log(1.777164) +1.96*w(-1/2)),digits=4)

```

```

cip<-data.frame(ll,ul)
# Calculation of the  $L_n$  statistic deleting 4th study
w<-c(0.7578464, 2.2304833, 0.5860116, 0.2526316, 2.6620506, 6.9675671,
0.7579955, 1.5819698, 2.2422012)
sn<-c(1.18367347, 9.37704918, 8.62025316, 0.01219512, 4.39120879, 1.45936698,
3.93109541, 0.28862974, 0.78560720)
wd<-diag(w)
pi<-log(sn)
s0<-rep(log(1.777164),9)
x<-(pi-s0)
p2<-t(x)%*%wd
ln<-p2%*%x
p3<-(t(x)%*%wd%*%x)
# Calculation of predicted GOR deleting study 4
g<-c(1.18367347, 9.37704918, 8.62025316, 0.01219512, 4.39120879, 1.45936698,
3.93109541, 0.28862974, 0.78560720)
df<-length(g)-1
q2<-23.67364
k<-1-((df-2)/q2)
pg<-log(1.777164)+k*(log(g)-log(1.777164))
g1<-format(exp(pg),digits=3)
g1<-c(1.3121,6.1517,5.7771,0.0431,3.4915, 1.5341,3.2146,0.4575,0.9662)
w<-c(0.7578464, 2.2304833, 0.5860116, 0.2526316, 2.6620506, 6.9675671,
0.7579955, 1.5819698, 2.2422012)
ll<-format(exp(pg -1.96*wi(-1/2)),digits=3)
ul<-format(exp(pg +1.96*wi(-1/2)),digits=3)
pci<-data.frame(ll,ul)

```

---

```
# R code for forest plot of simulation studies
```

---

```
x < -1.48
y < -1
plot(x,y,xlim=c(-1.5,6),ylim=c(-1,12),xaxt="n",
yaxt="n",frame.plot = NULL,cex=1)
segments(-0.5, 0, -0.5, 12, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(-0.5, 0, 7, 0, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(1, 0, 1, 12, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(1.48, 0, 1.48, 12, col = "black", lty = 5, lwd = par("lwd"))
# GOR
x1 < -c(0.124,11.247)
y1 < -c(11,11)
lines(x1,y1,ty="l")
x2 < -c(2.524,34.835)
y2 < -c(10,10)
lines(x2,y2,ty="l")
x3 < -c(0.666,111.55)
y3 < -c(9,9)
lines(x3,y3,ty="l")
x4 < -c(0.023,0.703)
y4 < -c(8,8)
lines(x4,y4,ty="l")
x5 < -c(0.0002,0.602)
y5 < -c(7,7)
lines(x5,y5,ty="l")
x6 < -c(1.32,14.598)
y6 < -c(6,6)
```

```
lines(x6,y6,ty="l")
x7 < -c(0.694,3.066)
y7 < -c(5,5)
lines(x7,y7,ty="l")
x8 < -c(0.41,37.344)
y8 < -c(4,4)
lines(x8,y8,ty="l")
x9 < -c(0.06,1.371)
y9 < -c(3,3)
lines(x9,y9,ty="l")
x10 < -c(0.212,2.909)
y10 < -c(2,2)
lines(x10,y10,ty="l")
x11 < -c(0.95,2.313)
y11 < -c(1,1)
lines(x11,y11,ty="l")
# PGOR
x1 < -c(0.13,11.81)
y1 < -c(10.75,10.75)
lines(x1,y1,ty="l",col="black", lty=5)
x2 < -c(1.69,23.34)
y2 < -c(9.75,9.75)
lines(x2,y2,ty="l",col="black", lty=5)
x3 < -c(0.454,76.12)
y3 < -c(8.75,8.75)
lines(x3,y3,ty="l",col="black", lty=5)
x4 < -c(0.04,1.19)
```

```
y4 <- -c(7.75,7.75)
lines(x4,y4,ty="l",col="black", lty=5)
x5 <- -c(0.0007,1.71)
y5 <- -c(6.75,6.75)
lines(x5,y5,ty="l",col="black", lty=5)
x6 <- -c(1.04,11.53)
y6 <- -c(5.75,5.75)
lines(x6,y6,ty="l",col="black", lty=5)
x7 <- -c(0.696,3.08)
y7 <- -c(4.75,4.75)
lines(x7,y7,ty="l",col="black", lty=5)
x8 <- -c(0.3349,30.22)
y8 <- -c(3.75,3.75)
lines(x8,y8,ty="l",col="black", lty=5)
x9 <- -c(0.086,1.96)
y9 <- -c(2.75,2.75)
lines(x9,y9,ty="l",col="black", lty=5)
x10 <- -c(0.243,3.34)
y10 <- -c(1.75,1.75)
lines(x10,y10,ty="l",col="black", lty=5)
text(-2, 11, labels="Study 1", pos=4,adj = c(0,0), cex = 1)
text(-2, 9.75, labels="Study 2", pos=4,adj = c(0,0), cex = 1)
text(-2, 8.75, labels="Study 3", pos=4,adj = c(0,0), cex = 1)
text(-2, 7.75, labels="Study 4", pos=4,adj = c(0,0), cex = 1) text(-2, 6.75,
labels="Study 5", pos=4,adj = c(0,0), cex = 1)
text(-2,5.75, labels="Study 6", pos=4,adj = c(0,0), cex = 1) text(-2,4.75,
labels="Study 7", pos=4,adj = c(0,0), cex = 1)
```

```

text(-2, 3.75, labels="Study 8", pos=4,adj = c(0,0), cex = 1)
text(-2,2.75, labels="Study 9", pos=4,adj = c(0,0), cex = 1)
text(-2, 2, labels="Study 10", pos=4,adj = c(0,0), cex = 1)
text(-2, 1, labels="Pooled",adj = c(0,0), pos=4, cex =1)
text(1.48, 0, labels="1.48", pos=1, adj = c(0,0),cex = 1)
text(-0.5, 0, labels="0", pos=1, adj = c(0,0),cex = 1)
text(1, 0, labels="1", pos=1, adj = c(0,0),cex = 1)
text(4, 0, labels="4", pos=1, adj = c(0,0),cex = 1)
text(6, 0, labels="6", pos=1, adj = c(0,0),cex = 1)
text(2.8, -.71, labels="Generalised Odds Ratio (log-scale)", pos=1, adj =
c(2.50,-1),cex = 1)
points(6,0,pch="'",cex =1)
points(4,0,pch="'",cex =1)

```

---

```
# R code for rayplot for 10 simulated studies
```

---

```

x1 <- -c(0,0,0,0,0,0,0,0,0,0)
y1 <- -c(1.48,1.48,1.48,1.48,1.48,1.48,1.48,1.48,1.48,1.48)
plot(x,y,xlim=c(-.1,1),ylim=c(0,10), type = "n",xaxt="n",
yaxt="n",frame.plot = NULL,cex=1)
segments(0, 0, 1, 0, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(0, 0, 0, 10, col = "black", lty = par("lty"), lwd = par("lwd"))
text(0,1.48, "1.48",adj = c(0,0), pos=2,cex=.6)
segments(0, 1.48, .45, 1.18, col = "black", lty = 1, lwd = par("lwd"))
segments(0, 1.48, .45, 9.38, col = "black", lty = 1, lwd = par("lwd"))
segments(0, 1.48, 0.45, 8.62, col = "black", lty = 1, lwd = par("lwd"))
segments(0, 1.48, 0.45, 0.13, col = "black", lty = 1, lwd = par("lwd"))
segments(0, 1.48, .45, 0.012, col = "black", lty = 1, lwd = par("lwd"))

```

```

segments(0, 1.48, 0.45, 4.39, col = "black", lty = 1, lwd = par("lwd"))
segments(0, 1.48, .45, 1.46, col = "black", lty = 1, lwd = par("lwd"))
segments(0, 1.48, .45, 3.93, col = "black", lty = 1, lwd = par("lwd"))
segments(0, 1.48, .45, 0.29, col = "black", lty = 1, lwd = par("lwd"))
segments(0, 1.48, .45, 0.78, col = "black", lty = 1, lwd = par("lwd"))
segments(.45, 0, .45, 10, col = "black", lty = par("lty"), lwd = par("lwd"))
# Predicted OR
segments(0, 1.48, .75, 1.24, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(0, 1.48, .75, 6.28, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(0, 1.48, .75, 5.88, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(0, 1.48, .75, .22, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(0, 1.48, .75, 0.0346, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(0, 1.48, .75, 3.4689, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(0, 1.48, .75, 1.46, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(0, 1.48, .75, 3.18, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(0, 1.48, .75, 0.41, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(0, 1.48, .75, 0.90, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(.75, 0, .75, 10, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(0, 10, 1, 10, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(0, 0, 0, 1, col = "black", lty = par("lty"), lwd = par("lwd"))
text(.45,1.18, "1.18 Study 1",adj = c(0,0), pos=4,cex=.6)
text(.45,9.38, "9.38 Study 2",adj = c(0,0),pos=4, cex=.6)
text(.45,8.62, "8.62 Study 3",adj = c(0,0), pos=4,cex=.6)
text(.45,0.13, "0.13 Study 4",adj = c(0,0),pos=4, cex=.6)
text(.45,0.012, "0.012 Study 5",adj = c(0,0), pos=4,cex=.6)
text(.45, 4.39, "4.39 Study 6",adj = c(0,0),pos=4, cex=.6)
text(.45,1.46, "1.46 Study 7",adj = c(0,0), pos=4,cex=.6)

```

```

text(.45,3.93, "3.93 Study 8",adj = c(0,0), pos=4,cex=.6)
text(.45,0.29, "0.29 Study 9",adj = c(0,0), pos=4,cex=.6)
text(.45, 0.78, "0.78 Study 10",adj = c(0,0),pos=4, cex=.6)
text(.75,1.24, "1.24 Study 1",adj = c(0,0),pos=4, cex=.6)
text(.75,6.28, "6.28 Study 2",adj = c(0,0), pos=4,cex=.6)
text(.75,5.88, "5.88 Study 3",adj = c(0,0),pos=4, cex=.6)
text(.75,0.22, "0.22 Study 4",adj = c(0,0),pos=4, cex=.6)
text(.75,0.0346, "0.03 Study 5",adj = c(0,0), pos=4,cex=.6)
text(.75, 3.4689, "3.47 Study 6",adj = c(0,0),pos=4, cex=.6)
text(.75,1.46, "1.46 Study 7",adj = c(0,0),pos=4, cex=.6)
text(.75,3.18, "3.18 Study 8",adj = c(0,0),pos=4, cex=.6)
text(.75, 0.41, "0.41 Study 9",adj = c(0,0), pos=4,cex=.6)
text(.75, 0.90, "0.90 Study 10",adj = c(0,0), pos=4,cex=.6)
text(.45,0, "GOR", pos=1, cex = .6)
text(.75,0, "Predicted GOR", pos=1, cex = .6)
text(0,0, "0", pos=2, cex = .6)
text(0,10, "10", pos=2, cex = .6)

```

---

## B.6 R Code for Chapter 4

Listing B.6: R code for GORs, weights, CIs for Gastric carcinoma

---

```

# Dent et al
nc < -c(3,19,22)
ca < -c(8,13,21)
e1 < -cbind(nc,ca)
e < -t(e1)
n1 < -e[1,3]

```

```

n2 < -e[2,3]
p11 < -e[1,1]/e[1,3]
p12 < -e[2,1]/e[2,3]
p21 < -e[1,2]/e[1,3]
p22 < -e[2,2]/e[2,3]
pc < -(p11*p22)
pd < -(p12*p21)
g1 < -pc/pd
v1 < -(p11*((p22)^2))+(p21*((-g1*p12)^2))
v2 < -((-g1*(p21))^2)*(p12)+((p11)^2)*(p22)
vg1 < -v1/((n1)*(pd^2))+v2/((n2)*(pd^2))
sevg1 < -sqrt(vg1)
ll2 < -g1*exp(-1.96*sevg1/g1)
ul2 < -g1*exp(1.96*sevg1/g1)
ci21 < -c(ll2,ul2)
# Robertson et al
nc < -c(0.5,25.5,26)
ca < -c(24.5,5.5,30)
# Bonenkamp et al
nc < -c(128,385,513)
ca < -c(183,300,483)
# Cuschieri et al
nc < -c(55,145,200)
ca < -c(92,108,200)
# Degiuli et al
nc < -c(8,68,76)
ca < -c(14,72,86)

```

```
# Chew et al
nc < -c(8,102,110)
ca < -c(19,92,111)
# R code for weights
# Dent et al
x11 < -f11
x12 < -f12
x21 < -f21
x22 < -f22
n1 < -first row total
n2 < -second row total
p11 < -x11/n1
q11 < -1-p11
p12 < -x12/n1
q12 < -1-p12
p21 < -x21/n2
q21 < -1-p21
p22 < -x22/n2
q22 < -1-p22
wli < -1/(n1*p11*q11)+1/(n2*p21*q21)
wi < -1/wli; i= 1, 2, ..., 6.
# Pooling logGOR
g2 < -c(0.256578947, 0.004401761, 0.545028742, 0.445277361, 0.605042017,
0.379772962)
g < -log(g2)
wi < -c(1.7010043, 0.4421159, 52.0624086, 22.1203730, 4.4439834, 5.0427355)
t < -sum(g*wi)/sum(wi)
```

```

pt< -exp(t)
# Calculation of ci for individual GOR for gastric carcinoma
wi< -c(1.7010043, 0.4421159, 52.0624086, 22.1203730, 4.4439834, 5.0427355)
g< -c(0.256578947, 0.004401761, 0.545028742, 0.445277361, 0.605042017,
0.379772962)
ll< -format(exp(log(g) -1.96*wia(-1/2)),digits=2)
ul< -format(exp(log(g) +1.96*wi(-1/2)),digits=2)
ci< -data.frame(ll,ul)
# CI for pooled
w< -sum(wi) ll< -format(exp(log(0.4893854) -1.96*w(-1/2)),digits=4)
ul< -format(exp(log(0.4893854) +1.96*w(-1/2)),digits=4)
ci< -data.frame(ll,ul)
# Ln Statistic
sn< -c(0.2566, 0.0044, 0.545, 0.4453, 0.6050, 0.3798)
w< -c(1.7010043, 0.4421159, 52.0624086, 22.1203730, 4.4439834, 5.0427355)
wd< -diag(w)
pi< -log(sn)
s0< -rep(log(0.4893854),6)
x< -(pi-s0)
p2< -t(x)% * %wd
ln< -p2% * %x
# Calculation of PGOR g< -c(0.256578947, 0.004401761, 0.545028742,
0.445277361, 0.605042017, 0.379772962)
df< -length(g)-1
w< -c(1.7010043, 0.4421159, 52.0624086, 22.1203730, 4.4439834, 5.0427355)
q2< -11.85
k< -1-((df-2)/q2) pg< -log(0.4893854) + k * (log(g) - log(0.4893854))

```

```

g1<-exp(pg) # CI for PGOR
wi<-c(1.7010043, 0.4421159, 52.0624086, 22.1203730, 4.4439834, 5.0427355)
ll<-format(exp(pg-1.96*wi(-1/2)),digits=4)
ul<-format(exp(pg+1.96*wi(-1/2)),digits=4)
cip<-data.frame(ll,ul)

# ray plot of gastric carcinoma with 6 studies x1<-c(0,0,0,0,0,0)
y1<-c(0.4893854,0.4893854,0.4893854,0.4893854,0.4893854,0.4893854)
plot(x,y,xlim=c(-.1,1),ylim=c(-.10,1), type = "n",xaxt="n",
yaxt="n",
frame.plot = NULL,cex=1)
segments(0, 0, 1, 0, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(0, 0, 0, 1, col = "black", lty = par("lty"), lwd = par("lwd"))
text(0,0.4893854, "0.49",adj = c(0,0), pos=2,cex=.6)
segments(0, 0.4893854, .45, 0.26, col = "black", lty = 1, lwd = par("lwd"))
segments(0, 0.4893854, .45, 0.004, col = "black", lty = 1, lwd = par("lwd"))
segments(0, 0.4893854, 0.45, 0.54, col = "black", lty = 1, lwd = par("lwd"))
segments(0, 0.4893854, 0.45, 0.44, col = "black", lty = 1, lwd = par("lwd"))
segments(0, 0.4893854, .45, 0.60, col = "black", lty = 1, lwd = par("lwd"))
segments(0, 0.4893854, 0.45, 0.38, col = "black", lty = 1, lwd = par("lwd"))
segments(.45, 0, .45, 1, col = "black", lty = par("lty"), lwd = par("lwd"))
# PGOR
segments(0, 0.4893854, .75, 0.30, col = "black", lty = par("lty"), lwd =
par("lwd"))
segments(0, 0.4893854, .75, 0.01, col = "black", lty = par("lty"), lwd =
par("lwd"))
segments(0, 0.4893854, .75, 0.53, col = "black", lty = par("lty"), lwd =
par("lwd"))

```

```

segments(0, 0.4893854, .75, 0.46, col = "black", lty = par("lty"), lwd =
par("lwd"))
segments(0, 0.4893854, .75, 0.57, col = "black", lty = par("lty"), lwd =
par("lwd"))
segments(0,0.4893854, .75, 0.40, col = "black", lty = par("lty"), lwd =
par("lwd"))
segments(.75, 0, .75, 1, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(0, 1, 1, 1, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(0, 0, 0, 1, col = "black", lty = par("lty"), lwd = par("lwd"))
text(.45,0.26, "0.26 Dent",adj = c(0,0), pos=4,cex=.6)
text(.45,0.02, "0.004 Robertson",adj = c(0,0),pos=4, cex=.6)
text(.45, 0.54, "0.54 Bonenkamp",adj = c(0,0), pos=4,cex=.6)
text(.45,0.44, "0.44 Cuschieri",adj = c(0,0),pos=4, cex=.6)
text(.45,0.60,"0.60 Degiuli",adj = c(0,0), pos=4,cex=.6)
text(.45, 0.38, "0.38 Chew- Wun",adj = c(0,0),pos=4, cex=.6)
text(.75,0.30, "0.30 Dent",adj = c(0,0),pos=4, cex=.6)
text(.75,0.03, "0.01 Robertson",adj = c(0,0), pos=4,cex=.6)
text(.75,0.53, "0.53 Bonenkamp",adj = c(0,0),pos=4, cex=.6)
text(.75,0.46, "0.46 Cuschieri",adj = c(0,0),pos=4, cex=.6)
text(.75,0.57, "0.57 Degiuli",adj = c(0,0), pos=4,cex=.6)
text(.75, 0.40, "0.40 Chew- Wun",adj = c(0,0),pos=4, cex=.6)
text(.45,0, "GOR", pos=1, cex = .6)
text(.75,0, "PGOR", pos=1, cex = .6)
text(0,0, "0", pos=2, cex = .6)
text(0,1, "1", pos=2, cex = .6)
# Forest plot gastric carcinoma
x< -0.4893854

```

```
y < -1
plot(x,y,xlim=c(-.5,1.6),ylim=c(-1,8),xaxt="n",
yaxt="n",frame.plot = NULL,cex=1)
segments(0, 0, 0, 8, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(0, 0, 1.6, 0, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(1, 0, 1, 8, col = "black", lty = par("lty"), lwd = par("lwd"))
segments(0.4893854, 0, 0.4893854, 8, col = "black", lty = 5, lwd = par("lwd"))
# GOR
x5 < -c(0.057,1.1532)
y5 < -c(7,7)
lines(x5,y5,ty="l")
x6 < -c(0.00023,0.083)
y6 < -c(6,6)
lines(x6,y6,ty="l")
x7 < -c(0.4153,0.7151)
y7 < -c(5,5)
lines(x7,y7,ty="l")
x8 < -c(0.2935,0.6755)
y8 < -c(4,4)
lines(x8,y8,ty="l")
x9 < -c(0.2387,1.533)
y9 < -c(3,3)
lines(x9,y9,ty="l")
x10 < -c(0.1586,0.9091)
y10j-c(2,2)
lines(x10,y10,ty="l")
x11 < -c(0.3961, 0.6047)
```

```
y11 < -c(1,1)
lines(x11,y11,ty="l")
text(0, 7, labels="Dent", pos=2,adj = c(0,0), cex = .6)
text(0, 6, labels="Robertson", pos=2,adj = c(0,0), cex = .6)
text(0, 5, labels="Bonenkamp", pos=2,adj = c(0,0), cex = .6)
text(0, 4, labels="Cuschieri", pos=2,adj = c(0,0), cex = .6)
text(0, 3, labels="Degiuli", pos=2,adj = c(0,0), cex = .6)
text(0, 2, labels="Chew-Wun", pos=2,adj = c(0,0), cex = .6)
text(0, 1, labels="MA",adj = c(0,0), pos=2, cex =.6)
# PGOR
x5 < -c(0.067229, 1.3579)
y5 < -c(6.75,6.75)
lines(x5,y5,ty="l",col="black",lty=5)
x6 < -c(0.000761, 0.2766)
y6 < -c(5.75,5.75)
lines(x6,y6,ty="l",col="black",lty=5)
x7 < -c(0.404210, 0.6959)
y7 < -c(4.75,4.75)
lines(x7,y7,ty="l",col="black",lty=5)
x8 < -c(0.300628, 0.6918)
y8 < -c(3.75,3.75)
lines(x8,y8,ty="l",col="black",lty=5)
x9 < -c(0.226293, 1.4529)
y9 < -c(2.75,2.75)
lines(x9,y9,ty="l",col="black",lty=5)
x10 < -c(0.169178, 0.9693)
y10 < -c(1.75,1.75)
```

```

lines(x10,y10,ty="l",col="black",lty=5)
text(0.4893854, 0, labels="0.49", pos=1, adj = c(0,0),cex = .7)
text(0, 0, labels="0", pos=1, adj = c(0,0),cex = .7)
text(1, 0, labels="1", pos=1, adj = c(0,0),cex = .7)
text(6, 0, labels="6", pos=1, adj = c(0,0),cex = .7)
text(.8, -.71, labels="GOR(log-scale)", pos=1, adj = c(2.50,-1),cex = .7)
text(-0.25, 8, labels="(b)",adj = c(0,0), pos=2, cex =.6)

# Sample size weighting for gastric carcinoma
# Dent et al
cl< -c(8,13,21)
tr< -c(3,19,22)
e1< -cbind(tr,cl)
e< -t(e1)
n11< -e[1,1]
n12< -e[1,2]
n21< -e[2,1]
n22< -e[2,2]
p1< -n11*n22
q1< -n12*(n21)
or1< -p1/q1
ga1< -(p1-q1)/(p1+q1)
seg1< -0.249#From SPSS
#seg=se(gamma)
selg1< -(2*seg1)/((1-ga1)2)
# selg< -SE(log ORG)
ci1< -c(or1*exp((-3.92*seg1)/(1-(ga1)2)),or1*exp((3.92*seg1)/(1-(ga1)2)))

```

```
# Robertson et al
cl <- c(24.5, 5.5, 30)
tr <- c(0.5, 25.5, 26)
e1 <- cbind(tr, cl)
e <- -t(e1)
n11 <- -e[1,1]
n12 <- -e[1,2]
n21 <- -e[2,1]
n22 <- -e[2,2]
p1 <- -n11*n22
q1 <- -n12*(n21)
or2 <- -p1/q1
ga2 <- -(p1-q1)/(p1+q1)
seg2 <- -0.020 # From SPSS
selg2 <- -(2*seg2)/((1-ga2)2)
ci2 <- -c(or2*exp((-3.92*seg2)/(1-(ga2)2)), or2*exp((3.92*seg2)/(1-(ga2)2)))

# Bonenkamp et al
tr <- c(128, 385, 513)
cl <- c(183, 300, 483)
e1 <- cbind(tr, cl)
e <- -t(e1)
n11 <- -e[1,1]
n12 <- -e[1,2]
n21 <- -e[2,1]
n22 <- -e[2,2]
p1 <- -n11*n22
q1 <- -n12*(n21)
```

```

or3 < -p1/q1
ga3 < -(p1-q1)/(p1+q1)
seg3 < - 0.063
selg3 < -(2*seg3)/((1-ga3)2)
ci3 < -c(or3*exp((-3.92*seg3)/(1-(ga3)2)),or3*exp((3.92*seg3)/(1-(ga3)2)))
# Cuschiei et al
cl < -c(92,108,200)
tr < -c(55,145,200)
e1 < -cbind(tr,cl)
e < -t(e1)
n11 < -e[1,1]
n12 < -e[1,2]
n21 < -e[2,1]
n22 < -e[2,2]
p1 < -n11*n22
q1 < -n12*(n21)
or4 < -p1/q1
ga4 < -(p1-q1)/(p1+q1)
seg4 < - 0.091
selg4 < -(2*seg4)/((1-ga4)2)
ci4 < -c(or4*exp((-3.92*seg4)/(1-(ga4)2)),or4*exp((3.92*seg4)/(1-(ga4)2)))
# Degiuli et al
cl < -c(14,72,86)
tr < -c(8,68,76)
e1 < -cbind(tr,cl)
e < -t(e1)
n11 < -e[1,1]

```

```

n12 < -e[1,2]
n21 < -e[2,1]
n22 < -e[2,2]
p1 < -n11*n22
q1 < -n12*(n21)
or5 < -p1/q1
ga5 < -(p1-q1)/(p1+q1) seg5 < - 0.223 selg5 < -(2*seg5)/((1-ga5)2)
ci5 < -c(or5*exp((-3.92*seg5)/(1-(ga5)2)),or5*exp((3.92*seg5)/(1-(ga5)2)))
# Chew et al
cl < -c(19,92,111)
tr < -c(8,102,110)
e1 < -cbind(tr,cl)
e < -t(e1)
n11 < -e[1,1]
n12 < -e[1,2]
n21 < -e[2,1]
n22 < -e[2,2]
p1 < -n11*n22
q1 < -n12*(n21)
or6 < -p1/q1
ga6 < -(p1-q1)/(p1+q1)
seg6 < - 0.178
selg6 < -(2*seg6)/((1-ga6)2)
ci6 < -c(or6*exp((-3.92*seg6)/(1-(ga6)2)),or6*exp((3.92*seg6)/(1-(ga6)2)))
# Sample size weights
#n1 < -n11+n21
#n2 < -n12+n22

```

```
#n3 < -n13+n23
# Dent et al
n1 < -11
n2 < -32
ga < --0.592
seg < -0.249
se2 < -seg2
tn1 < -n1*n2/(n1+n2)
# Robertson et al
n1 < -24
n2 < -30
tn2 < -n1*n2/(n1+n2)
# Bonenkamp et al
n1 < -311
n2 < -685
tn3 < -n1*n2/(n1+n2)
# Cuschieri et al
n1 < -147
n2 < -253
tn4 < -n1*n2/(n1+n2)
# Degiuli et al
n1 < -22
n2 < -140
tn5 < -n1*n2/(n1+n2)
# Chew et al
n1 < -27
n2 < -194
```

```

tn6 <- -n1*n2/(n1+n2)
tn <- -c(tn1,tn2,tn3,tn4,tn5,tn6)
tn <- -c(8.186047, 13.333333, 213.890562, 92.977500, 19.012346, 23.701357)
# Pooling by Edwardes
ga <- -c(-0.591623, -0.9912351, -0.2944743, -0.3838174, -0.246073, -0.449513)
seg <- -c(0.249, 0.020, 0.063, 0.091, 0.223, 0.178)
se2 <- -seg2
tn <- -c(8.186047, 13.333333, 213.890562, 92.977500, 19.012346, 23.701357)
tn2 <- -tn2
gb <- -sum(tn*ga)/sum(tn)
ORb <- -(1+gb)/(1-gb)
segb <- -sqrt(sum(tn2*se2))/sum(tn)
cip <- -c(ORb*exp((-3.92*segb)/(1-(gb)2)), ORb*exp((3.92*segb)/(1-(gb)2)))

```

---

## B.7 R Code for Chapter 5

```

# R code for continuous outcomes from simulated data

```

---

```

set.seed(12100)
x1 <- -rmultinom(1, size = 20, prob=c(0.1,0.15,0.45,0.3))
x2 <- -rmultinom(1, size = 22, prob=c(0.0909091, 0.1363636, 0.4545455,
0.3181818))
x3 <- -rmultinom(1, size = 33, prob=c(0.03030303, 0.1818182, 0.4242424
,0.3636364))
x4 <- -rmultinom(1, size = 21, prob=c(0.03, 0.04761905, 0.3285714 , 0.5938095))
x11 <- -data.frame(x1,x2,x3,x4)
x0 <- -t(x11)
s <- -c(20,22,33,21)

```

```

e<-cbind(x0,s)
p11<-e[1,1]/e[1,5]
p12<-e[1,2]/e[1,5]
p13<-e[1,3]/e[1,5]
p14<-e[1,4]/e[1,5]
p21<-e[2,1]/e[2,5]
p22<-e[2,2]/e[2,5]
p23<-e[2,3]/e[2,5]
p24<-e[2,4]/e[2,5]
p31<-e[3,1]/e[3,5]
p32<-e[3,2]/e[3,5]
p33<-e[3,3]/e[3,5]
p34<-e[3,4]/e[3,5]
p41<-e[4,1]/e[4,5]
p42<-e[4,2]/e[4,5]
p43<-e[4,3]/e[4,5]
p44<-e[4,4]/e[4,5]
pc<-p11*(p22+p23+p24+p32+p33+p34+p42+p43+p44)+
p12*(p23+p24+p33+p34+p43+p44)+p21*(p32+p33+p34+p42+p43+p44)+
p22*(p33+p34+p43+p44)+p23*(p34+p44)+p31*(p42+p43+p44)+
p32*(p43+p44)+p33*p44
pd<-p12*(p21+p31+p41)+p13*(p21+p22+p31+p32+p41+p42)+
p14*(p21+p22+p23+p31+p32+p33+p41+p42+p43)+
p22*(p31+p41)+p23*(p31+p32+p41+p42)+p24*(p31+p32+p33+p41+p42+p43)+
p32*p41+p33*(p41+p42)+p34*(p41+p42+p43)
g1<-pc/pd
# weight 1

```

$$n_1 < -e[1,5]$$

$$n_2 < -e[2,5]$$

$$n_3 < -e[3,5]$$

$$n_4 < -e[4,5]$$

$$p_{11} < -e[1,1]/e[1,5]$$

$$q_{11} < -1-p_{11}$$

$$p_{12} < -e[1,2]/e[1,5]$$

$$q_{12} < -1-p_{12}$$

$$p_{13} < -e[1,3]/e[1,5]$$

$$q_{13} < -1-p_{13}$$

$$p_{21} < -e[2,1]/e[2,5]$$

$$q_{21} < -1-p_{21}$$

$$p_{22} < -e[2,2]/e[2,5]$$

$$q_{22} < -1-p_{22}$$

$$p_{23} < -e[2,3]/e[2,5]$$

$$q_{23} < -1-p_{23}$$

$$p_{31} < -e[3,1]/e[3,5]$$

$$q_{31} < -1-p_{31}$$

$$p_{32} < -e[3,2]/e[3,5]$$

$$q_{32} < -1-p_{32}$$

$$p_{33} < -e[3,3]/e[3,5]$$

$$q_{33} < -1-p_{33}$$

$$p_{41} < -e[4,1]/e[4,5]$$

$$q_{41} < -1-p_{41}$$

$$p_{42} < -e[4,2]/e[4,5]$$

$$q_{42} < -1-p_{42}$$

$$p_{43} < -e[4,3]/e[4,5]$$

```

q43 <- -1-p43
w11 <- -(1/(n1*p11*q11)+1/(n2*p21*q21)+1/(n3*p31*q31)
+1/(n4*p41*q41)
+1/(n1*p12*q12)+1/(n2*p22*q22)+1/(n3*p32*q32)+1/(n4*p42*q42)
+1/(n1*p13*q13)+1/(n2*p23*q23)+1/(n3*p33*q33)+1/(n4*p43*q43))
w1 <- -1/w11
# Study 2
set.seed(12100)
x1 <- -rmultinom(1, size = 20, prob=c(0.15,0.15,0.45,0.3))
x2 <- -rmultinom(1, size = 22, prob=c(0.0909091, 0.1863636, 0.4045455,
0.3181818))
x3 <- -rmultinom(1, size = 33, prob=c(0.03030303, 0.21818182, 0.4242424
,0.336364))
x4 <- -rmultinom(1, size = 21, prob=c(0.02, 0.09761905, 0.285714, 0.6038095))
x11 <- -data.frame(x1,x2,x3,x4)
x0 <- -t(x11)
s <- -c(20,22,33,21)
e <- -cbind(x0,s)
# Study 3
set.seed(12100)
x1 <- -rmultinom(1, size = 20, prob=c(0.1,0.20,0.4,0.3))
x2 <- -rmultinom(1, size = 22, prob=c(0.0909091, 0.1063636, 0.4845455,
0.3181818))
x3 <- -rmultinom(1, size = 33, prob=c(0.03030303, 0.21818182, 0.4542424
,0.306364))
x4 <- -rmultinom(1, size = 21, prob=c(0.01, 0.14761905, 0.245714, 0.6038095))
x11 <- -data.frame(x1,x2,x3,x4)

```

```
x0<-t(x11)
s<-c(20,22,33,21)
e<-cbind(x0,s)
# Study 4
set.seed(12100)
x1<-rmultinom(1, size = 20, prob=c(0.07,0.25,0.33,0.35))
x2<-rmultinom(1, size = 22, prob=c(0.0909091, 0.1463636, 0.4045455,
0.3581818))
x3<-rmultinom(1, size = 33, prob=c(0.05030303, 0.26818182, 0.4342424
,0.256364))
x4<-rmultinom(1, size = 21, prob=c(0.01, 0.1761905, 0.355714 , 0.5838095))
x11<-data.frame(x1,x2,x3,x4)
x0<-t(x11)
s<-c(20,22,33,21)
e<-cbind(x0,s)
# Study 5
set.seed(12100)
x1<-rmultinom(1, size = 20, prob=c(0.10,0.15,0.3,0.45))
x2<-rmultinom(1, size = 22, prob=c(0.1109091, 0.1863636, 0.3445455,
0.3581818))
x3<-rmultinom(1, size = 33, prob=c(0.03030303, 0.20818182, 0.3942424
,0.376364))
x4<-rmultinom(1, size = 21, prob=c(0.00, 0.1761905, 0.285714 , 0.6838095))
x11<-data.frame(x1,x2,x3,x4)
x0<-t(x11)
s<-c(20,22,33,21)
e<-cbind(x0,s)
```

```

g< -c(g1,g2,g3,g4,g5)
w< -c(w1,w2,w3,w4,w5)
# Pooling log GOR
g< -c(1.344747, 1.355527, 1.240727, 1.391849, 1.118055)
g1< -log(g)
w< -c(0.1731767, 0.1953294, 0.1766055, 0.1851793, 0.1206612)
t< -sum(g1*w)/sum(w)
T< -exp(t)
# calculation of ci for individual GOR
g< -c(1.344747, 1.355527, 1.240727, 1.391849, 1.118055)
wi< -c(0.1731767, 0.1953294, 0.1766055, 0.1851793, 0.1206612)
ll< -format(exp(log(g) -1.96*wi(- 1/2)),digits=5)
ul< -format(exp(log(g) +1.96*wi(- 1/2)),digits=5)
ci< -data.frame(ll,ul)
# CI for pooled GOR
w< -sum(wi)
ll< -format(exp(log(1.303739) -1.96*w(- 1/2)),digits=4)
ul< -format(exp(log(1.303739) +1.96*w(- 1/2)),digits=4)
cip< -data.frame(ll,ul)
# Q value
g< -c(1.344747, 1.355527, 1.240727, 1.391849, 1.118055)
w< -c(0.1731767, 0.1953294, 0.1766055, 0.1851793, 0.1206612)
T< -log(g)
T2< -T*T
T3< -w%*%T
T4< -T3/sum(w)
T5< -T32

```

```
T6 < -sum(w*T2)
T7 < -sum(w)
q < -T6 - (T5/T7)
q2 < -sum(w*((T-T4)**2))
Q1 < -sum(w*T2)-(T5/T7)
p < -1-pchisq(0.004530595, df= 4)
# I2
Q < -0.004530595
df < -4
I2 < -(Q-df)/Q*100
#if Q>df+1 then
# d1 < -2*df-1
#q1 < -2*Q
#B < -0.5*(log(Q)-log(df))/(sqrt(q1)-sqrt(d1))
#or if Q<=df+1, compute
p1 < -3*(df-1)2
p2 < -1/2*(df-1)*(1-(1/p1))
B < -sqrt(p2)
L < -exp(0.5*log(Q/df)-1.96*B)
U < -exp(0.5*log(Q/df)+1.96*B)
LL < -(L2-1)/L2 * 100
UL < -(U2-1)/U2 * 100
```

# Appendix C

## C.1 Publications

### Refereed Journal Articles and Conference Proceedings

- (1) Meta-Analysis of D1 Versus D2 Gastrectomy for Gastric Carcinoma. *Annals of Surgery*, 2011; 253: 900-911. Jointly with M A Memon, M S Subramanya, S Khan, E Osland, and B Memon.
- (2) Meta-Analysis for Ordinal Categorical Outcomes. Submitted to *Statistics in Medicine*. Jointly with S Khan.
- (3) Statistical Methodology for Heterogeneous Ordinal Data in Meta-Analysis. Submitted to *Journal of Biopharmaceutical Statistics*. Jointly with S Khan.
- (4) Statistical Methodology for Ordinal Data in Meta-Analysis. Presented in Australian Statistical Conference, Dec 6-10, 2010. Jointly with S Khan.
- (5) Effect of Immune-enhanced Enteral Formulas on Gastrointestinal Surgical Postoperative Outcomes: A Meta-analysis. Presented in The Australian Society for Parenteral and Enteral Nutrition (AuSPEN) annual

scientific meeting Oct 21-23, 2010. Jointly with E Osland, S Khan, and M A Memon.

- (6) Meta-analysis of Laparoscopic Posterior and Anterior Fundoplication for Gastro-Oesophageal Reflux Disease. Presented in the Australasian Research Society meeting at Adelaide. Nov 2009. Jointly with M S Subramanya, B Memon, S Khan, and M A Memon.
- (7) Effect of Immune-enhanced Supplement Formulas in Perioperative outcomes: A Meta-Analysis. Presented in Royal Australian College of Surgeons Annual Scientific Congress, Brisbane, 2009. Jointly with E Osland, S Khan, and M A Memon.
- (8) Preliminary Results of a Meta-Analysis Evaluating the Effect of Immunonutrition on Outcomes of Elective Gastrointestinal Surgery. Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I, The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 132-139. Jointly with E Osland, S Khan, and M A Memon.
- (9) Meta-Analysis of Laparoscopic Posterior and Anterior Fundoplication for Gastro-oesophageal Reflux Disease. Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (ICCS-X), Volume I, The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, (2010): 148-156. Jointly with M S Subramanya, S Khan, B Memon, and M A Memon.