UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

EFFECT SIZES OF FIXED AND RANDOM EFFECTS MODEL IN META-ANALYSIS

by

FOO CHUAN HUI

Dissertation submitted in partial fulfillment
of the requirements for the degree
of Master of Science in Statistics

April 2010

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my thesis supervisor, Associate Professor Dr. Low Heng Chin, for her invaluable guidance, relentless contribution of ideas, suggestions, constant advice and patience throughout my study.

I also gratefully acknowledge the Ministry of Higher Education of Malaysia and the Universiti Pendidikan Sultan Idris for providing me the financial assistance during my master study. Further, I would like to express my sincere appreciation to the School of Mathematical Sciences as well as the library of USM for providing various facilities and references.

I would like to convey my heartfelt gratitude to my father, Foo Ah Chye, and my mother, Yew Soo Ngou, for being the pillars of my strength. Their greatest contribution, sacrifice, constant prayers, patience and endless encouragement all this while are certainly the greatest motivation to inspire me to do my very best. Thanks also to all my friends and course mates for lending me their friendship and moral support.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

TABLE OF CONTENTS

PAGE

ACKNOWLEDGEMENTS ii

TABLE OF CONTENTS iii

LIST OF TABLES vi

LIST OF FIGURES viii

ABSTRAK ix

ABSTRACT x

CHAPTER 1 INTRODUCTION

1.1 Introduction to Meta-Analysis 1

1.2 Research Fields in Meta-analysis 4

1.2.1 Impact on Theory Development 5

1.2.2 Impact on Psychology 5

1.2.3 Impact on Medical Field 6

1.2.4 Impact on Other Disciplines 7

1.3 Types of Meta-Analysis 7

1.4 Meta-Analysis Procedures 8

1.5 Introduction to Fixed Effect versus Random Effect Models 9

1.6 Objective 10

1.7 Format of Dissertation 10

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

CHAPTER 4 CASE STUDY

4.1 Introduction 79

4.2 Methodology 80

4.3 Results and Discussion 88

4.4 Conclusion 99

CHAPTER 5 CONCLUSION

5.1 Summary 100

5.2 Contribution 101

5.3 Limitation 102

5.4 Recommendation for Future Research 102

BIBLIOGRAPHY 104

APPENDICES

Appendix A A list of tables from Table A.1 to Table A.4 112

Appendix B A list of data sets from Table B.l to Table B.4 120

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

LIST OF TABLES

		Page
Table 1.1	Comparison between strengths and weaknesses of meta-analysis	4
Table 2.1	The most common meta-analysis terminology	16
Table 2.2	2x2 contingency table	21
Table 3.1	Probability of effect size in vote-count	32
Table 3.2	Common effect size based on distribution of independent and dependent variables	38
Table 3.3	The likelihood function for standardized mean difference	42
Table 3.4	Large sample variance of the estimator of effect size	43
Table 3.5	Summary of data set between student ratings of instructor and student achievement	44
Table 3.6	Log-likelihood function values between student ratings of the instructor and student achievement	45
Table 3.7	Large sample variance of correlation between student ratings of the instructor and student achievement	46
Table 3.8	Meta-analysis on correlation coefficients	49
Table 3.9	Meta-analysis on standardized mean differences	53
Table 3.10	Meta-analysis on differences between proportions	58
Table 3.11	Meta-analysis on quality-weighted approach	63
Table 3.12	Meta-analysis on random effects approach	66
Table 3.13	Meta-analysis on odds ratio	70
Table 3.14	Differences in odds ratio (Mantel-Haenszel)	75
Table 4.1	Data collection for Tamoxifen treatment before 1990	82

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

Table 4.2 Data set of 55 trials for Tamoxifen treatment 83

Table A. 1 Probability of population standardized mean difference 112

Table A.2 Probability of population correlation coefficient 114

Table A.3 Table of standard normal distribution function 116

Table A.4 Table of chi-square distribution 118

Table B. 1 Data set of Open Education on Student Independence and Self-Reliance 120

Table B .2 Data set of Teacher Indirectness and Student Achievement 121

Table B.3 Data set of Open Education on Student Self-Concept 121

Table B.4 Data set of student Ratings of the Instructor with Student Achievement 122

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

LIST OF FIGURES

		Page
Figure 1.1	Systematic reviews versus meta-analysis	3
Figure 1.2	Steps of conducting meta-analysis	8
Figure 2.1	Difference magnitude of effects (treatment versus control)	17
Figure 2.2	Meta-analysis methods for different fields of study	22
Figure 4.1	Recurrences for tamoxifen treatment about 1 year	90
Figure 4.2	Recurrences for tamoxifen treatment about 2 years	91
Figure 4.3	Recurrences for tamoxifen treatment about median of 5 years	92
Figure 4.4	Recurrences for tamoxifen treatment within 55 trials	93
Figure 4.5	Recurrences for tamoxifen treatment among 3 subtotals	93
Figure 4.6	Deaths for tamoxifen treatment about 1 year	95
Figure 4.7	Deaths for tamoxifen treatment about 2 years	96
Figure 4.8	Deaths for tamoxifen treatment about a median of 5 years	97
Figure 4.9	Deaths for tamoxifen treatment within 55 trials	98
Figure 4.10	Deaths for tamoxifen treatment among 3 subtotals	98

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

KESAN SAIZ TERHADAP MODEL KESAN TETAP DAN RAWAK DALAM ANALISIS-META

ABSTRAK

Dalam beberapa dekad yang lalu, analisis-meta telah berkembang dengan pesat dalam pelbagai bidang pengajian untuk meningkatkan kualiti penyelidikan yang sebelumnya menerusi tinjauan kuantitatif. Tinjauan kuantitatif mengukur saiz kesan daripada kajian-kajian secara menyeluruh. Oleh itu, objektif disertasi ini ialah menentukan kaedah penghitungan saiz kesan bagi keadaan yang berbeza dalam analisis meta. Terdapat dua penemuan kunci bagi kajian ini adalah seperti berikut: Pertamanya, kaedah *vote-counting* akan digunakan sekiranya penganggaran saiz kesan tidak didapati. Seterusnya, kaedah kesan tetap dan kesan rawak akan digunakan untuk menganggarkan saiz kesan gabungan sekiranya saiz kesan dapat dihitung. Sebanyak 55 ujikaji rawak berkawal pada rawatan Tamoxifen diberi kepada pesakit-pesakit kanser payudara peringkat awal yang diamati berdasarkan kes-kes kambuh serta kes-kes kematian masing-masing. Kemudian, data tersebut dianalisa dengan menggunakan perisian statistik terkini dan canggih, Comprehensive Meta-Analysis (CMA) versi 2.0. CMA didapati bermanfaat dan berinteraktif dalam proses analisis meta. Pilihan dan penerbitan yang bersifat 'bias' harus dielakkan demi menghasilkan analisis meta yang lebih berkualiti.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

ABSTRACT

Over a few decades ago, meta-analysis has rapidly developed in various fields of studies to improve the quality of previous research by quantitative review. Quantitative review assesses an overall magnitude of the effect from multiple studies. Hence, the objective in this dissertation is to determine the methods of finding effect sizes for different conditions in meta-analysis. Two key findings of this study are as follows: Firstly, vote-counting method will be applied if the effect size estimates are not available. Next, fixed effects and random effects methods will be used in combining the effect size estimates when effect size estimates can be calculated. 55 Randomised Controlled Trials (RCTs) on Tamoxifen treatment for early breast cancer patients are observed according to recurrences and mortality cases, respectively. The data is then analysed by using the latest statistical software, Comprehensive Meta-Analysis (CMA) version 2.0. CMA is found to be beneficial and interactive in the process of meta-analysis. The selection bias as well as the publication bias should be avoided in order to produce better quality meta-analysis.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

CHAPTER 1
INTRODUCTION

1.1 Introduction To Meta-Analysis

Gene V. Glass is the first person to develop the term "meta-analysis" and to demonstrate its use in 1976. Meta-analysis is the integration of data from a large number of research studies through statistical analysis, thus providing conclusions in the simpler forms of relationship to the large collection of studies. It can also be defined as the combination of findings from multiple studies and the deficiencies of information are re-analyzed to give a systematic and plausible conclusion. Meta-analysis can be classified into two different types of systematic review: (1) qualitative review, and (2) quantitative review.

Meta-analytic process provides the gathered information from past research findings. Through this, it enables us to discover what are the requirements and even the drawbacks as a guideline for future research. As long as there is a preliminary study, numerous questions can be resolved with the existence of meta-analysis procedures. The study of the relationship can be carried out by using meta-analysis despite the primary studies which did not go through the process. In addition, meta-analysis is able to rectify inaccurate results such as the misrepresented effects of sampling error, inconsistent findings and other measurement inaccuracies. Meta-analytic study is very useful because of its power in the relationship between variables, the correlation between independent variables (variety of treatments) and the efficiency of the interventions (therapy, counselling).

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

Meta-analysis is known to be an exclusive element due to the effect size which enables the degree of findings to be calculated. The three most regularly applied measures of effect sizes include standardized mean differences, the product-moment correlation r, (Rosenthal, 1991) and the odds ratio (Fleiss, 1994; Haddock *et al.*,1998). The purpose of meta-analysis is to comprehend the variation of effect size whenever it diverges greatly from study to study. We then use the one-way analysis of variance (ANOVA) to investigate categorical moderators.

Moderator variable is said to be an interaction either in the form of a qualitative or quantitative variable that changes the direction of the correlation between the independent and dependent variables (Holmbeck, 1997). A moderator can intensify or even weaken treatment effect. If the moderator variable is excluded from the studies, it will be dangerous for those who are undergoing improper treatment and even will reduce the power of statistical testing (Kraemer *et al,* 2001).

Meta-analysis is a way to identifying, synthesizing and accumulating the findings of particular studies, has been widely performed in randomized controlled trials (RCT) which plays an important role in an intervention. Employee training programs, leadership development programs, performance management systems, and organizational development interventions used meta-analysis and found it to be a useful and effective approach.

Systematic reviews may not necessarily include a statistical synthesis called meta-analysis. They take into account the same studies within combination results which are thereby considered a meaningful finding. To be specific, meta-analysis is an optional element (subset) of a systematic review, which uses a specific statistical method to combine the individual studies' findings (see Figure 1.1).

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

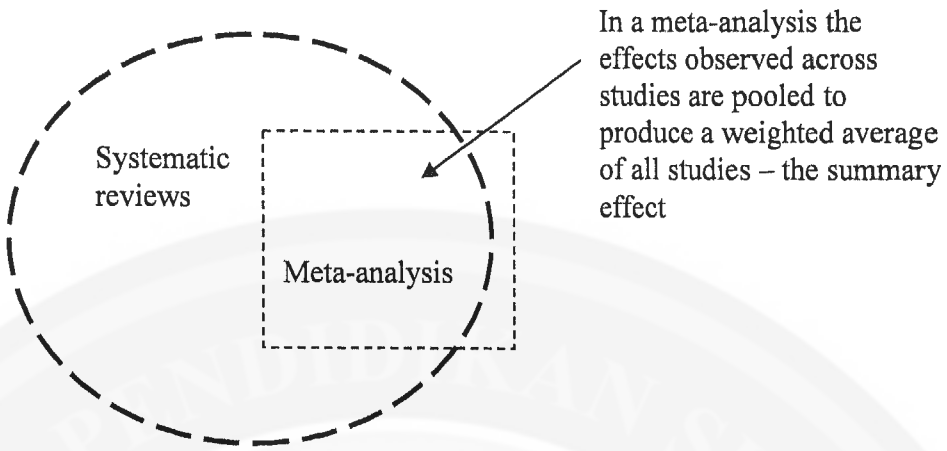


Figure 1.1: Systematic reviews versus Meta-analysis (source: Arvanitis et al., 2004)

The reasons for using meta-analysis are to (i) find out the existence of an effect, (ii) identify the degree of an effect, (iii) ensure the usefulness of moderator variable for an effect, and (iv) overcome the discrepancy in a research. Through meta-analytic studies, the intervention effect can be estimated more thoroughly if further and additional information is provided. Many primary studies vary in characteristics. With meta-analysis, it allows the differences in effect estimates to be examined. The purposes of the meta-analysis can be diverse subject to the need of the investigator. The focus can be on finding the population mean of correlation between variables, some estimating the variance of the population distribution or others that require both the mean and variance for statistical computation.

The importance of meta-analysis is not merely for those who use its procedures or techniques but there is greater impact for researchers who seriously investigate different types of meta-analysis for the sake of research purposes. From a large number of studies, each of them is treated as a data source for meta-analysis. Upon assessment, the impact of various statistical methods will be identified

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

respectively. Therefore, chronological development is considerably crucial for the findings.

Table 1.1 compares the advantages and disadvantages in meta-analysis. In terms of advantages, meta-analysis enables a large number of studies summarize through systematic review. It can detect the relationships whereby the others could not, and solve the limitations of previous study. By quantitative analysis for outcomes in multiple domains, meta-analysis is able to obtain more accurate findings.

For disadvantages, the author requires much effort and expertise to combine a great number of studies for a specific topic. A meta-analysis weak summary produces erratic research or more often it is called "garbage in, garbage out".

Table 1.1 Comparison between strengths and weaknesses of meta-analysis

STRENGTHS	WEAKNESSES
Summarize the research findings on a big-scale	Publication bias - More likely to publish positive significant result rather than negative ones
Curb the troubles faced by traditional narrative reviews - Analyze the impact of moderating variables	Greater effort for the cumulative findings
Able to investigate small or moderate relationships ambiguous in previous studies	Meta-analysis has results with a combination of "poorly" designed studies and "good" studies
Increase the precision of effects summary - discipline and transparency	Only deals with main effects - interaction effects via moderator analyses
Manage to increase power - identify significant effects	Different subjects, variables, and measuring method produce illogical conclusions

1.2 Research Fields in Meta-analysis

Since the 1970s, the applications of meta-analysis in psychological, educational, behavioural, and medical interventions have focused on investigating

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

the effectiveness of each study's treatment respectively. More than three hundred meta-analysis had been reviewed by Lipsey & Wilson (1993) involving the research areas in mental health interventions, counselling and psychoeducational interventions, health related psychological and educational treatment, work or organizational interventions, and educational interventions.

1.2.1 Impact on Theory Development

The development of theory relies on the observations in behaviour and social sciences. In the 1970s, behavioural scientists were seeking for quantitative summaries based on the techniques in meta-analysis in their research literature. Having a good understanding of the empirical relationship among variables is the key to build up theories. The purpose of the studies would not succeed without a strong relationship among variables. Only those with highly correlated variable can proceed to the further interactive or moderator-based theories. From the theory itself, it indicates an explanation prior to the meta-analytic findings. Guzzo, Jackson, and Katzell (1986) criticized this because the theory depends on the strong relationship variable and it is not developed directly from meta-analysis.

Science seeks for explanation all the time, and explanation renders the causal. If the data of behavior and social sciences match with one of the methods, either path analysis (Hunter & Gerbing, 1982) or structural equation modelling (SEM) then the method can be utilized to examine causal theories.

1.2.2 Impact on Psychology

The ability of meta-analysis is to establish a deduction once the new knowledge is learnt through different studies based on quantitative review. On top of that, meta-

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

analysis can even come out with new discovery which could hardly be explained from any individual study.

The major review journal in psychology is Psychological Bulletin. Fixed-effect methods are the most common model used in Psychological Bulletin. In fact, random effects model could be more appropriate to be used since it assumes all studies are from different populations while fixed effects model assumes that all studies from a single population. So, fixed effects model is always the first choice because it is easier and its findings are much better compared to traditional narrative review, but the accuracy of the findings would be doubtful.

1.2.3 Impact on Medical Field

According to Hunt (1997), if we make a comparison between behavioural & social sciences and medical research, medical research uses more meta-analysis. This phenomena is noticeably seen in the leading medical journals namely New England Journal of Medicine as well as the Journal of the American Medical Association (approximately 962 to 1411 meta-analyses are recorded in medical literature by 1995; Moher & Olkin, 1995). For medical research, the researchers will not know exactly whether the applicants are receiving the treatments or the placebo in a random allocation. This method is the most frequent mode in medical research, so-called the Randomized Controlled Trial (RCT). Unfortunately, the test that is usually being implemented is based on a small sample size, and so the population effect sizes are always small as well. Despite the same treatment is applied for different RCTs, yet it yields a contradictory outcome in each study.

Thomas Chalmers is the pioneer who suggested the application of meta-analysis in medical research. After doing a survey on meta-analysis and statistical

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

research on different diseases, Chalmers *et al.* (1977) who first used the drug based on different treatments, diseases and various areas of medical practice came to the conclusion that a great number of lives would have been saved as long as the application of meta-analysis is applied and contained chronologically in previous medical research literature. Moreover, the Cochrane Collaboration is another method for Chalmers to be put into practice. It is necessary to update the newest findings in RCT as the development in medical research findings has saved numerous lives.

1.2.4 Impact on other disciplines

Meta-analysis is equally essential for marketing, sociology, finance and even wildlife management research study. Political science is starting to make use of meta-analysis (Pinello, 1999) whereas meta-analysis in economics is now receiving increasing attention, too (Stanley, 2001). In social sciences, a strategy is being developed similar to the method accomplished in the medical field, the Campbell Collaboration (Boruch, 2005).

1.3 Types of Meta-analysis

In qualitative literature review, a researcher combines a set of studies and makes a narrative review. The researcher interprets the findings of the studies in words, rather than mathematical description. Nevertheless, the conclusions of narrative review have always been inconsistent from one literature to the other. By the mid-1990s, quantitative literature review which focuses on effect sizes is used in meta-analysis to avoid the problems of qualitative reviews. The authors review articles by choosing an effect size from the previous research finding. After that

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

meta-analysis is used to describe the overall magnitude (strong or weak) of the effect from a large number of studies based on different situations.

Lately, meta-analytic summaries for primary research papers have been used widely. Meta-analysis is used to provide information about the constancy of an overall relationship among the studies being carried out. Compared to a complete study for quantitative literature review, meta-analytic summaries could be much easier based on a single analysis from different studies.

1.4 Meta-analysis Procedures

The following flow chart shows the basic steps on conducting meta-analysis. Firstly, we need to have the problem statement, and then collect information in a systematic way by reviewing literature from printed research papers or electronic journals (published and unpublished), read through each literature's method and analysis.

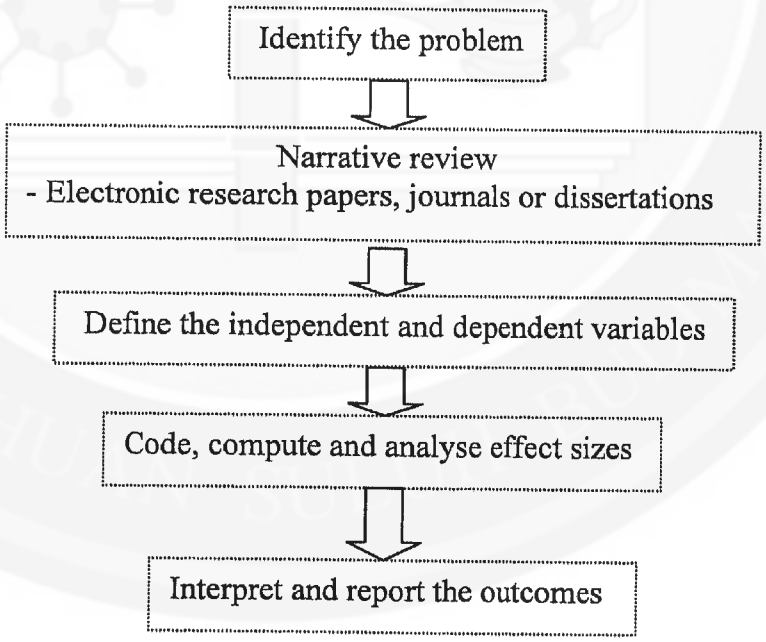


Figure 1.2 Steps of conducting meta-analysis

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Next, we specify the independent and dependent variables of interest (effects of radiotherapy treatment on cancer patients). Once we have identified the objective of the study, we code every data accordingly and analyze the data set to evaluate the magnitude of the effect sizes. Finally, we interpret the results based on the output of data analysis.

Meta-analysis is based on the concept of effect size as defined by Glass (1976). In short, effect size (ES) is the difference between the mean of the experimental group and the mean of the control group divided by the standard deviation of the control group. Meta-analysis combines the effects throughout the studies in order to increase the precision of the overall mean effect. If the difference between the null and alternative means (ES) is greater, the more precise the test is. Those appropriate measures that are applicable in calculating ES are (a) central tendency, (b) pre-post contrast, (c) group contrast, and (d) association between variables.

1.5 Introduction to fixed effect versus random effect models

The major task of many meta-analyses is to forecast how far the effect size varies from each other through the values of moderating variables. The two most ordinary categories of theoretical models in meta-analysis are fixed effect and random effect models depending on the magnitude of heterogeneity in the studies. For the fixed effects methods, we say the population is homogeneous. In this case, the estimated effect sizes are fixed. On the other hand, random effects methods with the heterogeneous population vary within studies and between studies.

The fixed-effect models estimate a common population parameter generated by every effect size from the individual studies. This can be interpreted as a constant

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

effect size will occur throughout the studies, except for the differences from observations due to sampling error.

The meta-analysts are allowed to place the confidence interval around the estimated population mean effect size both in fixed-effect and random-effect analyses. The confidence interval for the random effect model will be wider than the ones for the fixed effect model, given a set of effect sizes (Lipsey & Wilson, 2001). The fixed-effect analysis is selected despite the random effect case is more suitable to be used as its computation is not so complicated. Decision making on either select fixed-effects models or random-effects models is the main concern when drawing a conclusion from the findings (National Research Council, 1992). Lately, there is a suggestion that the assumptions made by random-effects models are more rational, in general, than those made by fixed-effects models (Hunter & Schmidt, 2000).

The random-effect model assumes that there is variation in population effect size rather than it is a constant. It assumes a random sampling from a multiple population. In fact the random sampling is assumed from a single population for a fixed-effect model. As a result, some of the parameter estimates will vary from the others.

1.6 Objective

To determine methods of finding effect sizes in meta-analysis.

1.7 Format of dissertation

This thesis is divided into five chapters. Chapter 1 is gives the introduction on meta-analysis, a brief discussion about the contemporary development of meta-analysis in different research of studies, objective as well as the format of the

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

dissertation. Chapter 2 focuses on the literature review that is related to meta-analysis. It commences with an introduction to meta-analysis. The origin of meta-analysis since 1904 until the current development of meta-analysis in various fields of studies will be included. Moreover, a brief discussion on effect size estimates, and meta-analysis methods for different research of studies are reviewed.

Chapter 3 presents the methodology. The methods of finding effect size estimate in different circumstances are introduced. A brief discussion on the measurement of clinical intervention is given.

Chapter 4 includes a case study with a set of secondary data set from clinical trials. Comprehensive Meta-Analysis (CMA) 2.0 statistical software is used to analyze the data. This chapter will emphasize on the usage of CMA in conducting meta-analysis. The outcomes of the data will be evaluated and interpreted.

Conclusion and further research relevant to the meta-analysis will be discussed in Chapter 5. The main contributions as well as the limitation of the findings are included.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

Meta-analysis has been developed for more than a century since 1904. The impact of meta-analysis is apparent in different fields of study. Through the research literature in the past, numerous methods of meta-analysis can be observed, and the strengths and weaknesses of the meta-analytic studies by the researchers from their findings will be discussed in this chapter. A number of meta-analysis methods and the related studies are included in this chapter. The application of meta-analysis in the past and the current development will be discussed in detail in the following parts.

2.2 Genesis of meta-analysis

Karl Pearson performed the first meta-analysis in 1904 in a study on the effectiveness of vaccines on typhoid fever. He realised that certain errors exist which did not allow him to produce a reliable conclusion due to the small sample sizes in the corresponding studies. One of the earliest fields that used meta-analysis was agriculture. The British statistician Ronald Fisher developed a method similar to Pearson's approach which uses a product of p-values to obtain an overall p-value on agriculture experimentation. According to Loughin (2004), an industrial statistician, Tippett (1931) modified Fisher's method by using a minimum p-value to obtain the overall p-value, and William Cochran (1937) discussed the methods and problems in combining results after getting the inspiration from agricultural experiments.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

In 1940, psychologists analyzed the research results collected for 60-year long by quantitative synthesis approach. Similar studies using the quantitative synthesis methods started to expand and were applied by other American social scientists and statisticians. The only difference is the research results were separated this time. One of them, Gene Glass (1976), created and referred the term 'meta-analysis' as 'the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings'. In 1976, Glass and Smith's article entitled "Primary, Secondary, and Meta-analysis of Research" is believed to be the main source for modern meta-analysis. The article discussed the advantages of psychotherapy in 1977 which compared 375 psychotherapy studies.

In the 1980s, several medical questions about treatment of heart disease and cancer had been answered by meta-analyses. For instance, the falling rate of mortality after a myocardial infarction for the long-term beta blockade patients from the coronary care unit (Yusuf *et ah,* 1985), and also a decreasmg rate in mortality for women above 50 with early breast cancer according to the research of Early Breast Cancer Trialists' Collaborative Group (1988). According to Peter Elwood (2006), Archie Cochrane and their colleagues first conducted a randomized trial in 1971 to investigate if aspirin reduced recurrences of heart attack. Archie Cochrane or known as the father of systematic reviews and meta-analysis had his book published in 1972. It was the main idea and the major contribution for the Cochrane Collaboration regarding healthcare interventions. Subsequently, DerSimonian and Laird (1986) proceeded to the latter approach and promoted to medical researchers concerning simple approximate formulas for Cochran's formal random effects model. Meta-analyses were well-known by the 1990s, a statement saying that 'it is obvious that

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

the new scientific discipline of meta-analysis is here to stay' (Chalmers and Lau, 1993).

Meta-analysis seems to be a usual practice in medical and psychological research by providing answers to clinical questions and thus is deemed as "a tool to save lives". Rosenthal (1961) was one of the first behavioural scientists who collected the outcomes accordingly from various studies. The process of the combination mostly used by Rosenthal or others is the ideas upon the previous literature back in the beginning of the 20th century (Olkin, 1990). For example, by using antibiotics before colon surgery for 26 clinical trials, effects combined indicated antibiotic reduced infection and reduced death rates from 11.2% to 4.5% (Baum *et at,* 1981). Through this approach, medical researchers would be able to conclude which treatment to be used as an effective way to the patients. Finally, meta-analysis is used to test the effects of watching violent TV versus antisocial behaviour from more than 200 studies. It proved antisocial behaviour is due to violence in TV program (Paik & Comstock, 1994).

Since the early 1980s, meta-analysis had been used in economics which began in environmental economics and further used in labor economics, industrial organization, and marketing except in 'macroeconomics' which is limited for its application in meta-analysis. Meanwhile, several lengthy explanations regarding the concepts, techniques, and statistical theory of different versions of meta-analysis had been published (Hedges & Olkin, 1985; Wolf, 1986).

In the 1980s, these meta-analytic studies continued to grow and became well-known techniques in behavioural, social sciences, and medical research. During the same period, the variance in effect sizes plays an important role in meta-analysis (Hunter & Schmidt, 1990). Different approaches dealing with variability results. A

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

number of researchers decided to apply fixed-effects models in meta-analysis (e.g., Hunter, Schmidt, & Jackson, 1982; Rosenthal, 1991). If the effect sizes from different studies are randomly sampled from a population of effect sizes, then the random-effects models are to be selected (DerSimonian & Laird, 1986).

Despite meta-analysis had been applied substantially during twentieth century, however, there are certain drawbacks as well as dilemmas in clinical treatments for the cumulative research (Hunt, 1997). By the early 1990s, as the terminology was becoming more puzzling, Chalmers and Altaian (1995) recommended that the term 'meta-analysis' should be focused on the combination of statistical analyses. Meta-analyses have terminologies which are similar to the usual terms in other statistical methods but with a few additional unique terms (see Table 2.1).

2.3 Magnitude of Effect Measures

The purpose of doing a research should be concentrating on the magnitude of effect rather than the previous measurement of "statistical significance" based on p-value (Cohen, 1990). One of the main aims of a magnitude of effect is to determine the power of the independent variable(s) in controlling and estimating the dependent variable (Rosnow and Rosenthal, 1989). If the analysis has a comparatively small p-value, it does not mean there is a strong relationship between independent and dependent variables of interest in a study (Snyder and Lawson, 1993).

Effect size in meta-analysis is the significance value that is combined from different studies into a single analysis. By combining effect size estimates we can get an overall estimate of the average effect size (Choi *et al*, 2003). However, many researchers did not include effect size in the discussion section of their findings. It has caused the estimation to become inaccurate and even worse the investigators

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

could hardly apply meta-analyses and make any comparison from primary literature in the future. The use of effect size measures remains inconsistent (Osborne, 2008).

Table 2.1 The most common meta-analysis terminology

Term	Definition
Bias	Systematic error introduced by limiting the scope of studies included in meta-analyses, thereby causing an incorrect estimate of the effect; can be caused by language, publication, or other types of bias.
Cohen's d	Common effect size estimate; expressed as units of standard deviation.
Cumulative meta-analysis	A meta-analysis whereby new studies related to the studied topic are added as they appear in the literature.
Effect size	The standardized estimate of the magnitude of an intervention's outcome in a study; the effect sizes calculated from the included studies are used to compare the magnitude and direction of an intervention in an individual study or a meta-analysis.
Eligibility criteria	Conditions that must be met to include a study in a meta-analysis; also known as Inclusion Criteria.
Error type I	Occurs when a correct null hypothesis is rejected as false; also known as a error.
Error, type 11	Occurs when an incorrect null hypothesis is accepted as true; also known as (3 error.
Fixed-effects model	Assumes that the studies analyze the same, exact effect and variability occurs only due to random chance.
Heterogeneity	Variation and differences between studies due to differences in the populations, study design, or methodology; variation not explained by random chance.
Homogeneity	The similarity and absence of variation between studies.
Ineligibility criteria	Conditions that, once met, will exclude a study from an analysis; also known as Exclusion Criteria.
Random-effects model	Assumes the effect differs between studies and must be represented by a range or distribution of values versus a single value.

(Source: Livingston et al., 2008)

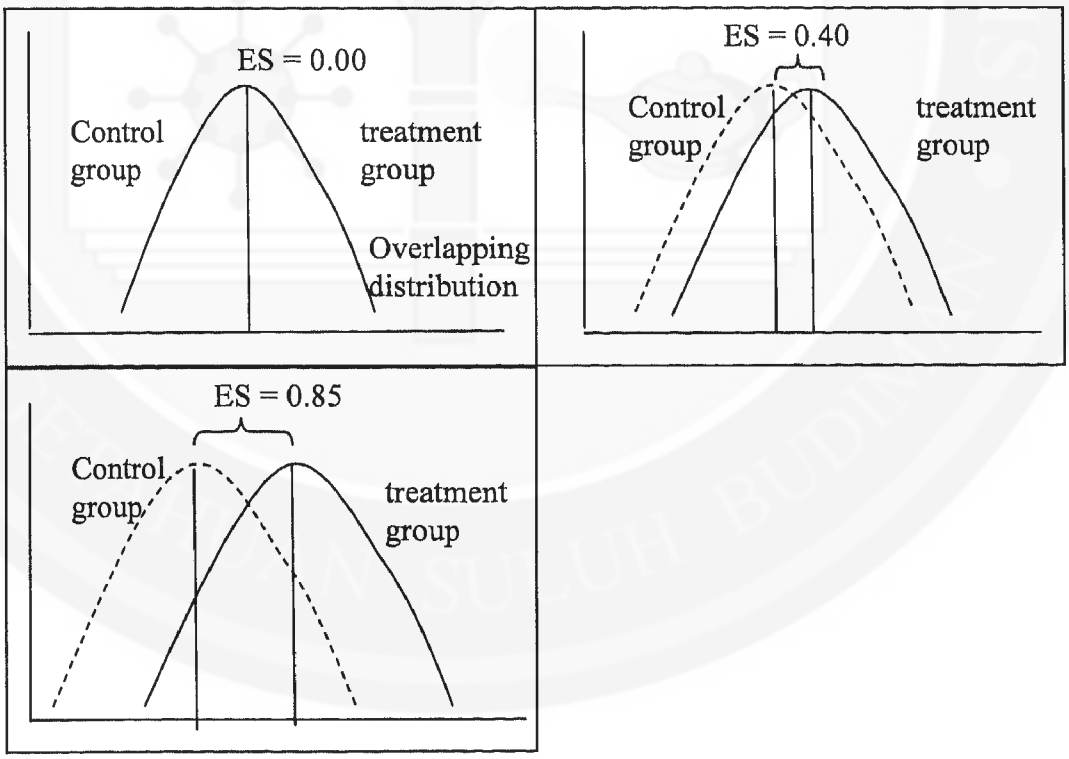
UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Besides estimating the magnitude of effects, the effect sizes ought to be explained and the results interpreted (Thomas, Salazar, and Landers, 1991). When the effect size is ≤ 0.20, this is interpreted as small effect size, 0.50 as moderate effect size and ≥ 0.80 as large effect size for standardized differences (Cohen, 1998). Figure 2.1 illustrates the effect sizes in different magnitudes.

If the homogeneity of effect size is found, we say the entire studies are sharing the same effect size in the meta-analysis. If the effect size is heterogeneous, the 'true' effect size remains unknown.

Generally, effect size (ES) can be thought of as falling into four general categories (Vacha-Haase & Thompson, 2004). The four categories are (1) Group difference indices (e.g. Cohen's *d*), (2) Strength of association (e.g. Pearson r), (3) Corrected estimates (e.g. adjusted R^2), and (4) Risk estimates (e.g. relative risk, RR and odds ratio, OR).



(Source: Bernard & Abrami., 2005)

Figure 2.1 Different magnitudes of effects between control and treatment groups

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

Each effect size value is not similar in an analysis due to the reliability of the information in the study. For example, an effect size of 500 subjects (sample) is much better than the one with 5 subjects. When effect sizes from both studies combine together, the estimation is worse than taking only the larger sample size. This results in a lack of precision in conducting a meta-analysis.

The inaccuracy of combined effect sizes can be resolved with weighing each effect size value based on its standard error. The value of effect size becomes imprecise when the standard error is increased. Thus, the inverse of the square standard error value is applied as the actual weights or simply called the inverse variance weight.

In meta-analysis, the research findings of one-variable relationship are rarely to be analyzed. The general effect size measures are proportions and arithmetic means. Two-variable relationships are most commonly used in meta-analysis such as (a) pre-post contrast where comparison between two variables due to difference in time measurement, (b) group contrast where a dependent variable is compared by a dichotomous independent variable, (c) association between two different variables are compared in a single sample.

Pre-post contrasts

There are two effect size indices for pre-post contrasts: (a) unstandardized mean gain for the studies from exact same constructs and (b) standardized mean gain.

For unstandardized mean gain, the statistics of effect size are,

$$ES_{ug} = \overline{X}_2 - \overline{X}_1 = \overline{G} \qquad (2.1)$$

where $\overline{X}_1$ is the mean at Time 1, $\overline{X}_2$ is the mean at Time 2, $\overline{G}$ is the mean gain difference between Time 1 and Time 2.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

For the standardized mean gain, the effect size statistics are developed by Becker (1992) is given as

$$ES_{sg} = \frac{\overline{X}_2 - \overline{X}_1}{s_p} = \frac{\overline{G}}{s_g / \sqrt{2(1-r)}} \qquad (2.2)$$

where $\overline{X}_1$ is the mean at Time 1, $\overline{X}_2$ is the mean at Time 2, $\overline{G}$ is the mean gain score, s_p is the pooled standard deviation of the Time 1 and Time 2 gain scores such that $s_p = \sqrt{(s_1^2 + s_2^2)/2}$, s_g is the standard deviation of the gain score, and r is the correlation between Time 1 and Time 2 scores.

Group Contrasts

In group contrasts, two or more groups of respondents measure a variable and make a comparison among groups. Various groups are normally investigated with statistical tests like t-test, analysis of variance (ANOVA), chi-square, Mann-Whitney U-test, and the like. Four effect size statistics are discussed as follows:

i) The unstandardized mean difference

An unstandardized effect size statistic is just the difference between two group means,

$$ES_{um} = \overline{X}_{g1} - \overline{X}_{g2} \qquad (2.3)$$

where $\overline{X}_{g1}$ is the mean from Group 1, $\overline{X}_{g2}$ is the mean from Group 2.

ii) The standardized mean difference

Standardized mean difference is mostly used in testing treatment effectiveness by comparing between experimental and control groups in the research. There are numerous literature and a great number of research studies showing this effect size statistic measure in meta-analysis (Cooper & Hedges, 1993; Hedges & Olkin, 1985).

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

The effect size for the standardized mean difference

$$ES_{sm} = \frac{\overline{X}_{g1} - \overline{X}_{g2}}{s_p} \quad (2.4)$$

where pooled standard deviation, $s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)}}$, s_1 is the standard deviation for Group 1, s_2 is the standard deviation for Group 2, n_1 is the number of subjects in Group 1 and n_2 is the number of subjects in Group 2. However, $\frac{(\mu_{t2}-\mu_{t1})}{\sigma} - \frac{(\mu_{c2}-\mu_{c1})}{\sigma}$ this effect size seems to be biased if the sample size is small especially samples less than 20 is being used (Hedges, 1981). Hedges corrected the bias and came out with an unbiased effect size estimate as:

$$ES'_{sm} = [\, 1 - \frac{3}{4n-9} \,]\, ES_{sm} \quad (2.5)$$

where $N = n_1+n_2$ is the total sample size, ES_{sm} is the biased standardized mean difference as in Equation (2.4), n_1 is the number of subjects in Group 1, and n_2 is the number of subjects in Group 2. When the treatment and control groups are contrasted, a positive sign of effect size shows that treatment works "better" than control group; otherwise, the negative sign is assigned to the effect size.

iii) The proportion difference

Proportion difference is applicable for group contrast findings. The statistics for effect size are given as

$$ES_{pd} = p_1 - p_2 \quad (2.6)$$

where p_1 is the proportion for Group 1, and p_2 is the proportion for Group 2.

iv) The Odds Ratio

Odds ratio is a dichotomous construct based on 2x2 contingency table where its effect size compares two groups of relative odds of event such as death, successful

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

outcome, gender, treatment received, illness and so forth (DerSimonian & Laird, 1986; Fleiss, 1994).

Table 2.2 2x2 contingency table

	Frequencies	
	Success	Failure
Treatment Group	a	b
Control Group	c	d

The effect size for odds ratio is given by

$$ES_{OR} = ad / bc \quad (2.7)$$

If a, b,c or d equal to zero, we add 0.5 to all cells (see also Haddock et al., 1998 p. 343, or Fleiss, 1994 p. 252).

Association between variables

The most widely used measure of correlation is the Pearson Product-Moment Correlation, particularly noted as "r". This method is developed by Karl Pearson for agricultural research purpose. The correlation coefficient for the relationship between two variables, say x and y, is

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

where n is the number of subjects.

2.4 Meta-analysis methods and related studies

There are many methods which can be applied in meta-analysis based on the areas of study that are being carried out. According to Shadish, Cook, and Campbell (2002), meta-analysis is "one of the most important social science developments in

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

the second half of the 20th century". There were both *Meta-Analysis in Social Research* (Glass, McGaw, and Smith 1981) and *Statistical Methods for Meta-Analysis* (Hedges and Olkin, 1985) published in the early 1980s. For example: (1) the Pretest and Post-test for organizational interventions, (2) Randomized Controlled Trial and Cochrane Collaboration are commonly used in medical research, (3) Cooperative Learning methods are most widespread in education, and (4) Meta-regression analysis is suitable for economics study (see Figure 2.2).

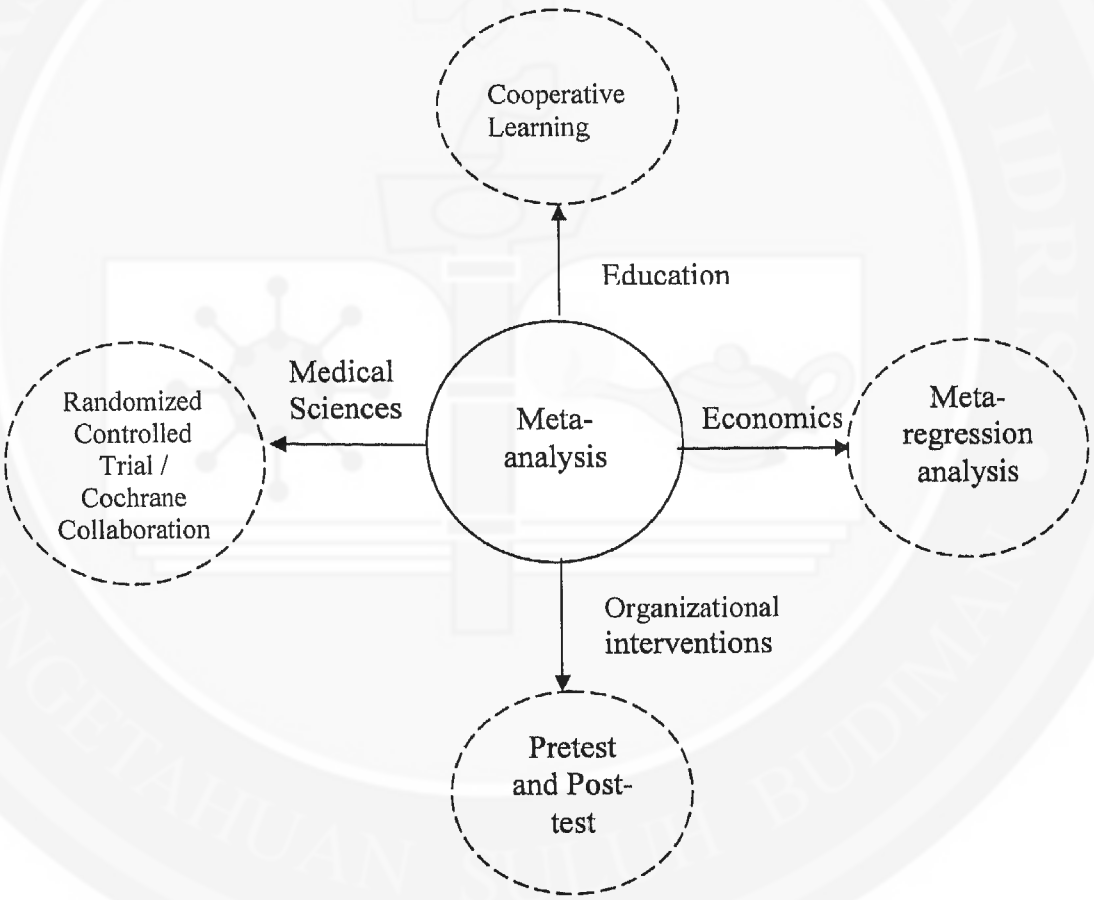


Figure 2.2 Meta-analysis methods applied to relevant fields of study

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

2.4.1 Randomized Controlled Trial (RCT) and Cochrane Collaboration

Commonly, RCTs can be seen from the publications focusing on health care, heart disease, diabetes mellitus, cancer, asthma, and anesthetics. Heart disease consists of hypertension, coronary artery operations, congestive heart failure, and blood lipids. The most frequent use of RCT in cancer treatment is anticancer drug combinations and breast cancer. As for the health literature, RCT is normally applied in depressive disorder, exercise, and quality of life. Overall, drug therapy and surgeries are mostly applied by RCT.

Since Bradford Hill first used the RCTs in 1948, the use of RCTs has developed exponentially in the later year as they play an important role in evidence-based medicine. Procedures of the RCT have five key items:

(i) a population of eligible patients

(ii) a group of patients sampled from the population

(iii) at least two treatment groups exist

(iv) randomization of the treatment groups

(v) comparison of results for both treatment groups

The Cochrane Collaboration improves the research works of controlled trials by cooperating with the U.S. National Library of Medicine. It is aimed to facilitate the health care decision makers worldwide about the effectiveness of health care interventions. For instance, antioxidant supplements (beta-carotene, vitamin A, and vitamin E) may increase mortality while no evidence shows that Vitamm C prevents mortality (Bjelakovic *et al*,, 2008)

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

2.4.2 Pretest and Post-test

Current research concentrates on the pretest-posttest-control (PPC) design which is useful for evaluating organizational interventions. PPC is a useful way to estimate the treatment effect. The participants are either allocated to treatment or control group where each of them is observed twice before and after the treatment has been carried out.

Variance, σ^2, is normally distributed in the PPC scores in each population. Let μ_{t1} be the mean for the treatment population pretest whilst μ_{t2} for the treatment population post-test, μ_{c1} for the control group pretest, and μ_{c2} for the control group post-test. The standardized mean change in each population is defined as the mean difference between PPC scores, divided by the common standard deviation.

Treatment group: Standardized mean change,

$$\partial_t = \frac{\mu_{t2} - \mu_{t1}}{\sigma}$$

Control group: Standardized mean change,

$$\partial_c = \frac{\mu_{c2} - \mu_{c1}}{\sigma}$$

The effect size for the PPC design is given by

$$\Delta = \partial_t - \partial_c = \frac{(\mu_{t2} - \mu_{t1}) - (\mu_{c2} - \mu_{c1})}{\sigma}$$

where Δ is the difference between the standardized mean of the treatment and control groups.

2.4.3 Cooperative Learning

Cooperative learning is widely used by social psychology and it is now a standard educational practice in most of the elementary as well as secondary schools.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Cooperative learning is whereby a class of students work in small groups as joint learning groups (Johnson, Johnson, & Holubec, 1992, 1993). By using such approach, students made greater progress in their knowledge of studies than the students taught using a traditional method (Armstrong *et al.*, 2007). A great number of issues such as achievement, ethnic integration, and prevention of drug abuse can be found in this cooperative learning method.

Through cooperative learning, students are more attentive of their knowledge (meta-cognition), better social skills as well as attitudes towards the course. On top of that, it helps to build up students' self-esteem and to improve the relationships among themselves (Johnson *et al.,* 1998). Recently, Springer *et al.* (1999) found out that by using meta-analysis, group study has positive effects on achievement, persistence, and attitude for cooperative learning in science, technology, engineering, and mathematics (STEM) courses rather than traditional teaching approach in college and university institution.

2.4.4 Meta-regression Analysis

Meta-regression analysis (MRA) is now being used by economic studies but is rarely used in medical and social sciences. MRA plays a vital role in correcting the excess variation found in recent economic research due to the publication selection biases.

A conventional regression model in applied economics is as follows:

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \qquad (2.8)$$

where **Y** is the nx1 dependent variable vector or the effect size, **X** are the explanatory variables with different characteristics, β is the magnitude of regression coefficient, and ε is the random error vector or the between study variation. $X\beta$ is the explanatory

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

HAK MILIK PERPUSTAKAAN
UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID

model in the original literature, independent and identically distributed. In order to distinguish from a number of meta-regression models, we say Equation (2.8) is the fundamental model.

When an empirical economics is explanatory, the corresponding regression coefficient, β (one of the parameters from kx1 vector β of equation (2.8)) is the main consideration. The following meta-regression model explains the significance of a particular β from different areas of studies in economics.

$$b_j = \beta + \sum_{k=1}^{K} \alpha_k Z_{jk} + e_j, \qquad j = 1,2,\ldots,q$$

where b_j is the estimate of β of the jth study which is made up of q studies in a research, β is the 'true' value of the parameter of interest, Z_{jk} is the meta-independent variables from particular jth study which measures the corresponding characteristics and shows its systematic variation from other results in the literature, α_ks are the meta-regression coefficients treated as the bias effect of the corresponding characteristics, and e_j is the meta-regression disturbance term (Stanley and Jarrell, 1989, p. 165).

In economics study, the practice of the application of MRA for more than twenty years suggests that additional information for moderator variables, Z_{jk} should include:

i) Standard error or degree of freedom of the estimate.

ii) Characteristics of the author (gender, income and the like).

iii) Selected model characteristics of Equation (2.8).

iv) Datasets characteristics

Whenever the biases exist, the 'true' effect, β, will vary for the estimates. In the absence of 'true' effect throughout the statistical significance in the studies, the model below can be used for correction of the biases (Stanley, 2005, 2006).

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

$$b_j = \beta + \sum_{k=1}^{K} \alpha_k Z_{jk} + \beta_0 Se_j + e_j , \qquad j = 1,2,\ldots,q \qquad (2.9)$$

where β is the 'true' value of the parameter of interest or the estimated overall effect size, Z_{jk} is the meta-independent variables from particular jth study, α_ks are the meta-regression coefficients treated as the bias effect of the corresponding characteristics, e_j is the meta-regression disturbance term, and Se_j is the standard error for *j*th study. However, this model (2.9) is rarely applied for a particular reason. As we know, small samples have a greater tendency to get larger standard errors. The publication selection is more likely to choose those researches with significant result. Even with small-samples studies, researchers will attempt to estimate until a larger value (significant result) is obtained in the hope that their research would be published.

Model (2.9) is rarely used because the variance of the dependent variable varies across the data. The following model (t-test) is the weighted least squares which divides equation (2.9) by Se_j :

$$t_j = \beta_0 + \beta\left(\frac{1}{Se_j}\right) + \sum_{k=1}^{K} \frac{\alpha_k Z_{jk}}{Se_j} + v_j \qquad (2.10)$$

Here β_0 is the intercept, Z_{jk} is the meta-independent variables from particular jth study, α_ks are the meta-regression coefficients, Se_j is the standard error for *j*th study, β is the slope coefficients, and v_j is the variance of *j*th study. $(1/Se_j)$ is the independent variable for this model. The intercept, β_0, is the test for publication selection. β in equation (2.10) would be the genuine effect corrected for publication selection.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

2.5 Conclusion

Meta-analysis is used frequently in medical, psychological sciences and educational even though it is not much used in economics. However, many other areas of study now attempt to use meta-analysis to investigate their research findings. A recent meta analysis in technology has examined topics such as the effectiveness of interactive distance education (Cavanaugh, 2001), computer-assisted instruction in science education (Bayraktar, 2001-2002), and computer-based instructional simulation (Lee, 1999). Effect size measurement is an important tool in assisting investigators to develop further insights into the true meaning of their research outcomes. It is crucial to know that there is no effect size measure which is perfect for all situations.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

CHAPTER 3
METHODOLOGY

3.1 Introduction

Two particular methods will be discussed throughout this chapter which are vote count method and combining estimates of effect size. Vote-counting procedures can be used to overcome some of the limitations of meta-analysis. The advantage of vote-counting estimator is to replace the calculation of parametric estimates when they are not available to be used. When a large group of studies where the effect size estimates are not available, vote count method takes this into account. However, this method has affected statistical power as it does not consider the sample size. Since it does not give an estimate for the size of the effect, this method is said to have its drawback as well (Hedges & Olkin, 1980). The second method is to combine the effect size estimates as long as there is the presence of effect size in each individual study. The data sets throughout the studies in this chapter are mainly obtained from both reference books of Hedges & Olkin (1985), and Williamson *et al.* (2002).

3.2 Vote-Counting Procedures

From research studies, at least one of the three types of data has to be used, either (1) data for calculating effect size estimate (standard deviations and means), (2) data on statistically significant relationship for hypothesis testing, or (3) data about the significance of the results. Vote-counting methods are suitable for the second and

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

third type of data. Vote-counting can be utilised in case there is insufficient information to calculate the effect size estimates for the multiple studies.

Although the primary studies do not provide complete data for effect size estimates, the reports usually contain conclusion in the form of significance test or the magnitude of the effects (a positive mean difference or a positive correlation). There are two ways to determine if the test statistic is more than the critical value, C_a. The first is a conventional critical value at a given significance level, such as a=0.05 whereby the researchers use it to calculate how many are there statistically significant positive results. The other is an unconventional critical value at significance level a=0.5 is used to calculate for the number of positive outcomes.

3.2.1 Conventional vote-counting methodology

Conventional vote-counting also known as box-score method summarizes the treatment effect magnitude by combining all the significant test results from numerous studies. Vote-counting procedures is initially described by Light and Smith (1971).

According to the methodology, a comparison is performed between a group under experimental treatment and one without any treatment among a collection of studies. There are three possible outcome in terms of the relationship between the independent and dependent variables: significantly positive, significantly negative or non-significant results. When the majority falls into one of the three categories, the others fall into the remaining two categories, respectively. Thus the modal categorization can be assumed the best to estimate the real relationship between the explanatory and response variables (Light & Smith, 1971, p.433).

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI

Example:

In k studies for two groups, the standardized mean difference correspond to Student's t or k studies for correlation studies, the population correlation coefficient correspond to statistic T as sample correlation coefficient. We make an assumption that the magnitude of population effect (either the standardized mean difference or the population correlation coefficient) is the same for all studies. Thus, the probability of any study producing a positive significant result is based on the sample size, n, and the population effect magnitude, δ. Generally,

the probability of a positive significant result is

$$p \equiv \text{Prob}\{\text{significant result} \mid \delta, n\} = \int_{C\alpha}^{\infty} f(t; \delta, n)\, dt,$$

whereby $f(t; \delta, n)$ is the probability density function of statistic T in sample size, n, and magnitude effect size, δ, and the critical value, C_α.

The probability, p, or simply denoted as the power of the statistical test is treated equally likely for all studies. Each integration is a trial of success (a positive significant result) or a failure (a nonsignificant or negative significant result). Through this, the number of successes from those trials has a binomial distribution with parameter, p.

As Hedges and Olkin (1980) noted, the vote-counting approach has very low statistical power for the range of effect sizes and sample sizes especially in social sciences. Table 3.1 shows the probability that a vote count fails to detect an effect through different sample sizes. When effect sizes become smaller, mostly, the conventional vote-counting method fails to detect the effects. Besides that, there is almost loss of power for conventional vote-counting when the k studies increase. An effect size, δ, is detected if the proportion of positive significant result is more than one-third. Let X be the frequency of successes,

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

$$\text{Prob}\{\text{proportion of successes} > C_o\} = \text{Prob}\left\{\frac{X}{k} > C_o\right\}$$

where C_o is a preset measure in the range $0 \leq C_o \leq 1$, usually is set to be one-third in order to detect an effect size.

The effect size is

$$\delta = (\mu_E - \mu_C) / \sigma$$

where μ_E is the population mean for experimental groups while μ_C is the population mean for control groups, and σ is the population standard deviation for each group.

Table 3.1 Probability that a Vote Count with $C_o = {}^1/_3$ fails to detect an effect for different sample sizes per group, n, effect sizes δ, and total number of studies, k

k	n	Effect size $\delta \equiv (\mu_E - \mu_C) / \sigma$							
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
10	10	1.000	0.999	0.998	0.994	0.985	0.968	0.935	0.880
10	20	1.000	0.998	0.990	0.966	0.906	0.987	0.606	0.395
10	30	0.999	0.995	0.975	0.906	0.947	0.502	0.252	0.089
10	40	0.999	0.991	0.950	0.813	0.547	0.254	0.073	0.012
10	50	0.999	0.986	0.914	0.694	0.358	0.105	0.016	0.001
20	10	1.000	1.000	1.000	0.999	0.997	0.989	0.966	0.914
20	20	1.000	1.000	0.998	0.988	0.941	0.800	0.545	0.265
20	30	1.000	1.000	0.993	0.941	0.747	0.400	0.118	0.016
20	40	1.000	0.999	0.978	0.834	0.463	0.119	0.011	0.000
20	50	1.000	0.997	0.948	0.672	0.222	0.023	0.001	0.000
50	10	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.986
50	20	1.000	1.000	1.000	1.000	0.994	0.915	0.589	0.174
50	30	1.000	1.000	1.000	0.994	0.862	0.363	0.035	0.000
50	40	1.000	1.000	0.999	0.942	0.461	0.036	0.000	0.000
50	50	1.000	1.000	0.995	0.773	0.124	0.001	0.000	0.000

(Source: Hedges & Olkin, 1985)

3.2.2 The Sign Test

The sign test is the oldest nonparametric test that all the non-significant effect sizes from k independent studies. Sign test is just the binomial test with probability $\pi = 0.5$ (Conover 1980, p.122). We say the probability to obtain a positive outcome is 0.5 if all the treatment effects are not significant (zero). Otherwise, the probability

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

of having positive result is larger than 0.5 when an effect size is detected in the studies.

The hypothesis testing is as follows:

$$H_0: \pi = 0.5$$
$$H_1: \pi > 0.5$$

where π denote the population proportion of positive results. The π estimator, $p = \frac{U}{k}$ where $U = \sum_{i=1}^{k} x_i$ be the frequency of positive results in k studies.

Example

Case I:

Assume the researcher found 14 positive results in 20 studies. The π estimator, $p = \frac{14}{20} \approx 0.7$. When $\pi = 0.5$, according to the Binomial distribution table, the probability is 0.0577 (i.e., 0.0370 + 0.0148 + 0.0046 + 0.0011+ 0.0002 + 0.0000 + 0.0000). So, we fail to reject H_0 since the p-value is greater than $\alpha = 0.05$. We conclude that there is no significant effect size in the studies at the 5% significance level.

Case II:

When there are 15 positive results in 20 studies, then the probability from the Binomial distribution table is now 0.0207 (i.e., 0.0148 + 0.0046 + 0.0011 + 0.0002 + 0.0000 + 0.0000). Now we can reject H_0 as 0.0207 is smaller than 0.05. We make a deduction that an effect size exists at 5% significance level in 20 independent studies. Similar to the conventional vote-counting procedures, the sign test does not take sample size into consideration. Since both have no significant effect size in the studies, the effect size estimate can hardly be calculated. The next session discusses the methods to resolve the drawbacks mentioned above.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

3.2.3 Confidence Intervals Based on Equal Sample Sizes

Vote-counting method in this section presumes that there is an equal sample size n for a set of k independent studies. When the independent variable is continuous, we say the sample size n is the same for all k studies. If the independent variable is dichotomous, then we denote $n_E = n_C = n$ where n_E and n_C are the experimental and control group sample sizes for all k studies, respectively.

i. Confidence interval for Population Standardized Mean Difference δ

If the independent variable is dichotomous whilst dependent variable is continuous, the common effect size is the population standardized mean difference δ. Say for *i*th study,

$$\delta_i = \frac{\mu_{Ei} - \mu_{Ci}}{\sigma_i} \qquad i = 1,2,\ldots,k$$

where μ_{Ei} is the population mean for experimental group in *i*th study, μ_{Ci} is the population mean for control group in *i*th study, and σ_i is the population standard deviation in *i*th study. Sample estimation of δ,

$$d_i = \frac{\overline{X}_{Ei} - \overline{X}_{Ci}}{S_i} \qquad i = 1,2,\ldots,k$$

$\overline{X}_{Ei}$ is sample mean for experimental group in the *i*th study, $\overline{X}_{Ci}$ is sample mean for control group in the *i*th study, S_i is pooled within-group sample standard deviation in the *i*th study.

Suppose there is a set of k independent studies to estimate a common parameter θ. So throughout the study we have a collection of parameter $\theta_1,\ldots, \theta_k$. In each study there is an estimator T (test statistics) of the parameter θ which yield $T_1,\ldots,T_k$, respectively. If *i*th study has no effect, θ_i will produce value zero. To test for *i*th hypothesis,

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

$H_0 : \theta_i = 0$

Reject H_0 if $T_i > C_\alpha$ where C_α is a critical value at significance level α from T_i distribution. Each experimental study has two possible outcomes: either $T_i > C_\alpha$ as a "success" or $T_i \leq C_\alpha$ as a "failure". Denote the probabilities for each outcome:

$$\Pr (T_i > C_\alpha) = \pi \qquad \text{and}$$

$$\Pr (T_i \leq C_\alpha) = 1 - \pi$$

To find the confidence interval for π, two primary approaches can be used. The first is the binomial distribution for large sample size to the normal distribution. A good approximation for binomial distribution is when $k\pi \geq 5$ and also $k(1-\pi) \geq 5$. If these conditions are fulfilled, the sample proportion p is normally distributed with mean π and variance $\frac{\pi(1-\pi)}{k}$. Thus,

$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{k}}}$$

has a standard normal distribution. Here k is the number of studies.

By estimating the variance $\frac{\pi(1-\pi)}{k}$ by $\frac{p(1-p)}{k}$, the confidence interval for π can be estimated. Then, a 100 $(1-\alpha)$ percent confidence interval (π_L, π_U) for π is given by :

$$\pi = p \pm Z_{\alpha/2} \sqrt{\frac{p(1-p)}{k}}, \qquad (3.1)$$

where $Z_{\alpha/2}$ is a two-sided critical value of the standard normal distribution. For $\alpha = 0.05$, $Z_{0.025} = 1.96$.

The second approach is an alternative method for getting confidence interval for π, the square of the standard normal distribution is equivalent to a chi-square distribution with 1 degree of freedom, d.f.=1 (DeGroot, 1986).

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

For large k,

$$z^2 = \left\{ \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{k}}} \right\}^2$$

has a chi-square distribution, d.f.=1.

To find the lower and upper bound for the 100(1- α) percent confidence interval for π, the quadratic equation is required (Hedges & Olkin, 1985).

$$\pi = \frac{(2p+b) \pm \sqrt{b^2 + 4bp(1-p)}}{2(1+b)}, \qquad (3.2)$$

where b= $\chi^2_\alpha(1)$ / k. For α=0.05 with d.f.=1, $\chi^2_{0.05}(1)$=3.841.

Both methods are normal approximation from a large sample size to yield confidence intervals for significant positive outcomes. Generally these two methods give different confidence intervals even though they are based on large-sample approximation.

Example:

δ_i is said to be the effect size namely population standardized mean difference in *i*th study. To test if the effect sizes are homogeneous, the hypothesis can be written as

H_0: $\delta_1 = \delta_2 = \ldots = \delta_k = 0$

H_1: $\delta_1 = \delta_2 = \ldots = \delta_k > 0$

From Table B.1 (Appendix B), the experimental and control group sample size is $n_E = n_C = 30$ in each study. Suppose that we observed the number of studies of the open education group mean exceeded the control group mean. The calculation showed that there were only two of the seven studies with positive mean differences.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

The estimate of π is $p=\frac{U}{k}=2/7=0.29$. There are two methods for finding the confidence interval for π.

Using equation in (3.1), the 90% confidence interval of the standard normal distribution for π is

$$\pi_L = 0.29 - 1.645\sqrt{\frac{0.29(1-0.29)}{7}} \le \pi \le 0.29 + 1.645\sqrt{\frac{0.29(1-0.29)}{7}} = \pi_U$$

Since the sample size n=30 is not given in Table A.1 (Appendix A), we use linear interpolation between the values for n=25 and n=50. Thus, the values $\pi_L = 0.01$ and $\pi_U = 0.57$, a 90% confidence interval for δ:

$$\delta_L = -0.69, \qquad \delta_U = 0.05$$

Using the equation in (3.2), the 90% confidence interval of the chi-square distribution for π:

$$\frac{(2(0.29)+0.387) \pm \sqrt{(0.387)^2 + 4(0.387)(0.29)(0.71)}}{2(1.387)}$$

where b=2.706/7=0.387 and p=0.29. Using Table A.1, with n=30, $\pi_L = 0.10$ and $\pi_U = 0.59$, the 90% confidence interval for δ:

$$\delta_L = -0.37, \qquad \delta_U = 0.02$$

Apparently, the chi-square distribution is more suitable to be used as it has a narrower confidence interval compared to the standard normal distribution. Since we know the binomial distribution is symmetrical for p=0.500, values that are smaller than p=0.500 can use 1-p which yield a negative values for δ as shown in Table A.1. For standard normal distribution, the confidence interval for δ (-0.69, 0.05) while for chi-square distribution, the confidence interval for δ (-0.37, 0.02). We notice that both confidence intervals contain the value zero. Hence we can conclude that the entire effect sizes are zero (homogeneous) for this example.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

ii. Confidence interval for Population Correlation Coefficient ρ

The effect size usually would be the population coefficient of correlation if both independent variable(X) and dependent variable(Y) are continuous. Simplifying, We have the following table:-

Table 3.2 Common effect size based on distribution of independent and dependent variables

Independent variable	Dependent variable	Effect size
Dichotomous	Continuous	Population standardized mean difference δ
Continuous	Continuous	Population Correlation coefficient *ρ*

Say population correlation coefficient of X and Y:

$$\rho_i = \mathrm{Cov}(X_i, Y_i) / \sigma_{Xi}\, \sigma_{Yi} \qquad i = 1,2,\ldots,k$$

where $\mathrm{Cov}(X_i, Y_i)$ is the population covariance of X and Y for the *i*th study, σ_{Xi} is the population standard deviation of X for the *i*th study, and σ_{Yi} is the population standard deviation of Y for the *i*th study. Meanwhile, sample estimation of ρ_i is known as r Pearson product moment correlation coefficient

$$r_i = \frac{\sum_{j=1}^{n_i}(X_j - \overline{X})(Y_j - \overline{Y})}{[\sum_{j=1}^{n_i}(X_j - \overline{X})^2][\sum_{j=1}^{n_i}(Y_j - \overline{Y})^2]} \qquad i = 1,2,\ldots,k$$

To test if the effect sizes are homogeneous, the hypothesis can be written as

H_0: $\rho_1 = \rho_2 = \ldots = \rho_k = 0$

H_1: $\rho_1 = \rho_2 = \ldots = \rho_k > 0$

The common population correlation ρ is calculated by the amount of times r_i greater than the critical value r_α at significance level α. If α=0.5, we assume that the

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

number of studies with positive correlations. On the other hand, when α=0.05, the number of studies with significant positive correlations will be taken into account.

Example:

From Table B.2 (Appendix B), the sample sizes differ with range between 15 and 17. For the sake of our analysis, we consider the sample size on average to be 16 for each study. In this case, we observe only the number of studies with positive correlations U with respect to vote-counting method. There are six positive correlations, so U=6 and the estimate of π is p=U/k=6/7=0.86. Referring to Table A.2 (Appendix A) with n=16, the estimate $\hat{\rho}$ =0.28.

Confidence intervals for ρ can be obtained by either of the two methods mentioned in equation (3.1) or (3.2). The first method as in Equation (3.1), 90% confidence interval for π using standard normal distribution is

$$\pi_L = 0.86 - 1.645\sqrt{\frac{0.86(0.14)}{7}} \le \pi \le 0.86 + 1.645\sqrt{\frac{0.86(0.14)}{7}} = \pi_U$$

or (0.64, 1.07). For sample size n=16 from Table A.2, ρ_L = 0.09 is estimated from π_L= 0.64. Correlation is ranged between $0 \le \rho \le 1$, when ρ_U >1, we assume it to be equal to one as it provides no information for upper limit. The 90% confidence interval for ρ using the standard normal distribution:

$$0.09 \le \rho \le 1.$$

The second method as in Equation (3.2), 90% confidence interval for π using chi-square distribution:

$$\frac{(2(0.86)+0.387) \pm \sqrt{(0.387)^2 + 4(0.387)(0.86)(0.14)}}{2(1.387)}$$

or (0.55, 0.97). Using Table A.2 with n=16, (0.55, 0.97) is converted to ρ_L =0.03 and

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

$\rho_U = 0.48$. Then, 90% confidence interval for ρ using chi-square distribution:

$$0.03 \leq \rho \leq 0.48.$$

Normally, the one with a shorter interval is superior. From this example we notice that the second method is preferable. However, this will not always happen. Due to the absence of the value zero in the interval, we reject H_0 by saying that the existence of positive correlation between teacher indirectness (praise, criticism or acceptance) and student achievement.

3.2.4 Confidence Intervals Based on Unequal Sample Sizes

Suppose for each study, we observe whether $T_i > C_i$. Here, we have different critical values for entire k studies with different sample sizes $n_1,\ldots,n_k$. The possible outcomes for each study will be either $T_i > C_i$ or $T_i \leq C_i$.

To determine whether each study has $T_i > C_i$ from k studies, we can use maximum likelihood to estimate parameter θ (MLE). Somehow, the calculation based on MLE could be complicated. A better way to calculate the estimate θ is by using log-likelihood function $L(\theta)$ where the coarse grid of θ values were used to find the region with maximum value of $L(\theta)$. Later a fine grid of θ values is created to provide a better estimate parameter θ. Denote the probabilities for each outcome as

$$\Pr(T_i > C_i) = \pi_i \qquad \text{and}$$

$$\Pr(T_i \leq C_i) = 1 - \pi_i$$

Normal approximation $\hat{\theta}$ is required when the number of k studies combined is large.

The expression for large sample variance of $\hat{\theta}$:

$$Var(\hat{\theta}) = \left(\sum_{i=1}^{k} \frac{[D_i^{(1)}]^2}{\pi_i(1-\pi_i)} \right)^{-1}, \qquad (3.3)$$

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI

where $D_i^{(1)} = \partial\pi_i / \partial\theta$, is evaluated at $\theta = \hat{\theta}$. If k is large, 100(1- α)% confidence interval is

$$\hat{\theta} - C_{\alpha/2}\sqrt{Var(\hat{\theta})} \leq \theta \leq \hat{\theta} + C_{\alpha/2}\sqrt{Var(\hat{\theta})}, \qquad (3.4)$$

where $C_{\alpha/2}$ is a two-sided critical value of the standard normal distribution.

i. Confidence interval for Population standardized mean difference δ

Suppose we observe if there is any positive value of sample mean difference between experimental (E) and control (C) group for each study. In this case, the sample sizes for both groups may vary across studies. If we say the effect sizes are homogeneous, then $\delta_1 = \ldots = \delta_k = \delta$ implies that

$$\overline{X}_{Ei} - \overline{X}_{Ci} \sim N(\delta\sigma_i, \sigma_i^2 / \tilde{n}_i),$$

where $\tilde{n}_i = n_{Ei}n_{Ci} / (n_{Ei} + n_{Ci})$. The probability of positive result π_i is

$$\pi_i = \Pr(\overline{X}_{Ei} - \overline{X}_{Ci} > 0) = 1 - \Phi(-\delta\sqrt{\tilde{n}_i}) \qquad (3.5)$$

where $\Phi(-\delta\sqrt{\tilde{n}_i})$ is equivalent to $P(Z \leq -\delta\sqrt{\tilde{n}_i})$.

The log-likelihood function is

$$L(\delta) = \sum_{i=1}^{k}[x_i \ln(1 - \Phi(-\delta\sqrt{\tilde{n}_i})) + (1 - x_i)\ln\Phi(-\delta\sqrt{\tilde{n}_i})], \qquad (3.6)$$

and $D_i^{(1)} = \dfrac{\partial\pi_i}{\partial\theta} = \sqrt{\dfrac{\tilde{n}_i}{2\pi}}\exp\left(-\dfrac{1}{2}\hat{\delta}^2\tilde{n}_i\right)$ (3.7)

where derivative, $D_i^{(1)}$, is evaluated at $\theta = \hat{\theta}$, and x is the indicator variable such that $x_i = 1$ if $T_i > C_i$ or $x_i = 0$ otherwise.

Example:

From Table B.3 (Appendix B), there is a set of 18 studies with their outcome of open versus traditional education on student self-concept. Table 3.3 comprises of

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

the sample sizes n_E, n_C and $\tilde{n}$, and parametric effect size estimate d. However, we neglect the parametric effect size estimates and focus on common effect size δ.

Firstly, the coarse grid (left-hand side) with δ values range between –0.50 and 0.50 differ in 0.10 units from each other are given. We calculate the log-likelihood of δ values based on the equation (3.6). It shows the δ=0.00 is larger than δ= ± 0.10. Thus we can say that the maximum value of $L(\delta)$ occurs between δ= ± 0.10. We need a fine grid (right-hand side) to expand the region which lies between δ= ± 0.10 by steps of 0.01 in order to improve our estimate of $\hat{\delta}$.

Finally, we notice that the largest likelihood of δ value is at 0.03. We conclude that MLE of δ is therefore $\hat{\delta} = 0.03$.

Table 3.3 The likelihood Function for δ Based on 18 studies of the Effects of Open Education on Self-Concept.

Coarse grid		Fine grid	
δ	$L(\delta)$	δ	$L(\delta)$
-0.50	-59.236	-0.10	-15.153
-0.40	-43.518	-0.09	-14.726
-0.30	-30.842	-0.08	-14.334
-0.20	-21.327	-0.07	-13.978
- 0.10	-15.153	-0.06	-13.658
0.00	-12.477	-0.05	-13.372
0.10	-13.294	-0.04	-13.122
0.20	-17.367	-0.03	-12.908
0.30	-24.318	-0.02	-12.729
0.40	-33.813	-0.01	-12.585
0.50	-45.665	0.00	-12.477
		0.01	-12.403
		0.02	-12.365
		0.03	-12.361
		0.04	-12.393
		0.05	-12.458
		0.06	-12.558
		0.07	-12.692
		0.08	-12.859
		0.09	-13.060
		0.10	-13.294

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Table 3.4 Computation for obtaining the Large Sample Variance of the Estimator of Effect Size

Study	π_i	$D_i^{(1)}$	$[D_i^{(1)}]^2 / [\pi_i (1 - \pi_i)]$
1	0.5948	3.1009	9.9242
2	0.5970	3.1685	10.4847
3	0.5334	1.7681	2.8809
4	0.5533	1.7681	2.8809
5	0.5663	2.1904	4.4161
6	0.5520	1.7241	2.7493
7	0.5631	2.0846	3.9815
8	0.5771	2.5370	6.1093
9	0.5642	2.1205	4.1250
10	0.5653	2.1558	4.2699
11	0.5674	2.2245	4.5636
12	0.5642	2.1205	4.1250
13	0.5596	1.9724	3.5588
14	0.5293	0.9746	1.1263
15	0.5316	1.0522	1.2499
16	0.5378	1.2559	1.6163
17	0.5948	3.1009	9.9242
18	0.5642	2.1205	4.1250
Total			82.1085

Using the values in Table 3.4, substitute π_i from equation (3.5) and $D_i^{(1)}$ from equation (3.7) into the equation (3.3), the estimated large-sample variance of δ

$\text{Var}(\hat{\delta}) = 1/\ 82.1085 = 0.0122$. Hence, 95% confidence interval for δ from equation (3.4) is as follows:

$$0.03 - 1.96\sqrt{0.0122} \leq \delta \leq 0.03 + 1.96\sqrt{0.0122}$$

or $-0.19 \leq \delta \leq 0.25$.

ii. Confidence interval for the population correlation coefficient ρ

From a collection of k studies, assume that we observe r_i is positive in each study. The effect sizes are normally distributed when they are homogeneous.

$$r_i \sim N(\rho,\ [(1-\rho^2)^2 / n_i]).$$

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

The probability of positive outcome is

$$\pi_i = \Pr(r_i > 0) = 1 - \Phi[-\rho\sqrt{n_i}\ /(1-\rho^2)]. \qquad (3.8)$$

The log-likelihood function is

$$L(\rho) = \sum_{i=1}^{k}\left\{x_i \ln\left(1-\Phi\left[-\frac{\rho\sqrt{n_i}}{(1-\rho^2)}\right]\right)+(1-x_i)\ln\left(\Phi\left[-\frac{\rho\sqrt{n_i}}{1-\rho^2}\right]\right)\right\} \qquad (3.9)$$

and the derivative $D_i^{(1)}$ is

$$D_i^{(1)} = \sqrt{\frac{n_i}{2\pi}}\left(\frac{(1+\rho^2)}{(1-\rho^2)^2}\right)\exp\left(-\frac{1}{2}\frac{n_i\rho^2}{(1-\rho^2)^2}\right). \qquad (3.10)$$

Example:

The following Table 3.5 shows values of n, Pearson product moment correlation coefficient r, and X for a total of 20 studies as shown in Table B.4 (Appendix B). Substituting the values between $\rho = \pm 0.50$ differ in 0.10 units from each other into equation (3.9), the log-likelihood function can be obtained as given in Table 3.6.

Table 3.5 Sample sizes, Correlations, and Indicator Variable Values (X=1 if r is positive) from 20 Studies of the Relation Between Student Ratings of the Instructor and Student Achievement

Study	n	r	X
1	10	0.68	1
2	20	0.56	1
3	13	0.23	1
4	22	0.64	1
5	28	0.49	1
6	12	-0.04	0
7	12	0.49	1
8	36	0.33	1
9	19	0.58	1
10	12	0.18	1
11	36	-0.11	0
12	75	0.27	1
13	33	0.26	1
14	121	0.40	1
15	37	0.49	1
16	14	0.51	1
17	40	0.40	1
18	16	0.34	1
19	14	0.42	1
20	20	0.16	1

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

Table 3.6 Log-likelihood Function Values of ρ from 20 Studies of the relationship between Student Ratings of the Instructor and Student Achievement

Coarse grid		Fine grid	
ρ	$L(\rho)$	ρ	$L(\rho)$
- 0.50	-159.677	0.10	-8.961
-0.40	-95.780	0.11	-8.640
-0.30	-59.089	0.12	-8.346
-0.20	-36.659	0.13	-8.079
-0.10	-22.568	0.14	-7.837
0.00	-13.863	0.15	-7.621
0.10	-8.961	0.16	-7.430
0.20	-6.900	0.17	-7.263
0.30	-7.108	0.18	-7.119
0.40	-9.552	0.19	-6.998
0.50	-14.990	0.20	-6.900
		0.21	-6.824
		0.22	-6.770
		0.23	-6.738
		0.24	-6.727
		0.25	-6.737
		0.26	-6.769
		0.27	-6.821
		0.28	-6.895
		0.29	-6.991
		0.30	-7.108

On the left, the coarse grid provides the region with maximum value of $L(\rho)$. We can notice that the MLE lies between ρ= 0.10 and ρ= 0.30. Then, we expand the region ranging from 0.10 to 0.30 in steps of 0.01. Since the log-likelihood is the largest at ρ= 0.24, thus MLE of $\hat{\rho} = 0.24$.

Using the values in Table 3.7, substitute π_i from equation (3.8) and $D_i^{(1)}$ from equation (3.10) into the equation (3.3). The estimated large-sample variance of ρ $\text{Var}(\hat{\rho})$ = 1/ 2026.900= 0.000493. Hence, 95% confidence interval for ρ from Equation (3.4) is as follows:

$$0.24 - 1.96\sqrt{0.000493} \leq \rho \leq 0.24 + 1.96\sqrt{0.000493}$$

or $0.196 \leq \rho \leq 0.283$.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDI
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI

Table 3.7 Computation for Large-Sample Variance of $\hat{\rho}$ from 20 studies between Student Ratings of the Instructor and Student Achievement

Study	π_i	$D_i^{(1)}$	$[D_i^{(1)}]^2 / [\pi_i (1 - \pi_i)]$
1	0.790	3.413	70.118
2	0.873	3.489	109.552
3	0.821	3.530	84.708
4	0.884	3.430	114.607
5	0.911	3.185	125.273
6	0.811	3.503	80.133
7	0.811	3.503	80.133
8	0.937	2.787	131.041
9	0.867	3.513	106.710
10	0.811	3.503	80.133
11	0.937	2.787	131.041
12	0.986	1.136	95.356
13	0.928	2.940	129.840
14	0.997	0.325	41.491
15	0.939	2.735	131.219
16	0.830	3.547	89.008
17	0.946	2.580	131.152
18	0.846	3.553	96.821
19	0.830	3.547	89.008
20	0.873	3.489	109.552
Total			2026.900

3.3 Combining Estimates of Effect Sizes

Contrary to vote-counting method, the effect sizes from a great amount of studies can be calculated in order to combine estimates of effect sizes. From various studies, the combination of estimates effect size is said to be easy to conduct if studies are sharing the same methodological selection: within-study sample size, design, and construct. Fixed effects model is the simplest way to investigate a single population effect size θ. The observed studies of estimate effect sizes differ from the population effect size due to its sampling error. We assume this variance as conditional variance.

Under random effects models, we are no longer estimate a single population effect size since the existence of sampling error and even the difference among the

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIK
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PE

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

true distribution of population effect sizes. There is always discrepancy in methodological ways and construct to carry out a collection of studies.

3.3.1 Fixed Effects Models

Suppose we combine k independent studies where the *i*th study has effect size t_i, variance v_i, and population effect size θ_i. Under the fixed effects model, we assume all the parameters from k studies have a common parameter θ, data to be combined possess k effect sizes $t_1,\ldots,t_k$, and variances $v_1,\ldots,v_k$.

$\overline{T}\bullet$ represents the average effect size which is estimated by

$$\overline{T}\bullet = \frac{\sum_{i=1}^{k} \omega_i t_i}{\sum_{i=1}^{k} \omega_i}. \qquad (3.11)$$

The symbol ω_i is a weighting factor to represent variance of each study or specifically the inverse of individual conditional variance:

$$\omega_i = \frac{1}{v_i}$$

The purpose of this weighting factor is to decrease the error variances across the small-sample size studies, so that the individual study effect size can be adjusted to provide greater weight to the studies.

The conditional variance ($V\bullet$) is defined as the variance of distribution effect sizes. Square root of the conditional variance yields the standard error of the sampling distribution. In meta-analysis, the variance estimates are used to find the confidence interval of average effect size ($T\bullet$). When the sample size is increased, the variance will be decreased, and so more accurate estimate of effect size can be obtained.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

$$V\bullet = \frac{1}{\sum_{i=1}^{k} \frac{1}{v_i}}. \qquad (3.12)$$

(1) Combining Correlations

A wide range of researches have concluded that the correlation coefficient *r* has greater tendency to be biased (Hays, 1973). The first approach is called the *r* to Z_i transformation. Since the value of *r* initially is not normally distributed, the Z_i is generated to obtain a normal distribution for *r*.

$$Z_i = 0.5 \{\ln [(1+ r_i) / (1- r_i)]\}$$

The data set in Table 3.8 is acquired from Table B.4 (Appendix B). The calculation has been performed using the spreadsheet program Excel. The values of *r* are provided throughout the studies between the mean student rating of the instructor and the mean achievement score for the students in the class. The effect size is measured by using product moment correlation.

The mean value for Z=0.417 is fairly close to the average of *r* =0.378. Each study is weighted by the sample size to form a standardized effect size. Overall, due to weak correlations (e.g. $|r| < 0.5$), z_i is close to r_i. Therefore the variance of r_i is also close to variance of z_i. The variance of z_i is:

$$v_i = 1 / (n_i - 3)$$

For instance, first study the value $v_i = 1/ (10 - 3) = 0.143$ and the same for the rest of the studies. Total of the values of v_i is 1.050. We proceed to calculate the weighting factor by finding the inverse of v_i for each study.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI

Table 3.8 Meta-Analysis on Correlation Coefficients

Study	n_i	r_i	Z_i	v_i	$\omega_i=1/v_i$	$t_i^* \omega_i$	$t_i^* t_i^* \omega_i$
1	10	0.680	0.829	0.143	7.000	5.804	4.811
2	20	0.560	0.633	0.059	17.000	10.758	6.812
3	22	0.640	0.758	0.053	19.000	14.405	10.917
4	13	0.230	0.234	0.100	10.000	2.342	0.548
5	28	0.490	0.536	0.040	25.000	13.402	7.182
6	36	0.330	0.343	0.030	33.000	11.313	3.882
7	19	0.580	0.662	0.063	16.000	10.599	7.012
8	12	0.180	0.182	0.111	9.000	1.638	0.298
9	36	-0.110	-0.110	0.030	33.000	-3.645	0.399
10	75	0.270	0.277	0.014	72.000	19.934	5.524
11	33	0.260	0.266	0.033	30.000	7.983	2.123
12	37	0.490	0.536	0.029	34.000	18.226	9.768
13	14	0.510	0.563	0.091	11.000	6.190	3.487
14	40	0.400	0.424	0.027	37.000	15.675	6.652
15	16	0.340	0.354	0.077	13.000	4.603	1.629
16	14	0.420	0.448	0.091	11.000	4.925	2.208
17	20	0.160	0.161	0.059	17.000	2.744	0.441
Sum	445	6.430	7.096	1.050	394.000	146.896	73.692
Mean	26.18	0.378	0.417	0.062	23.176	8.641	4.335

To construct a standardized weighted effect size, we multiply each effect size by the weighting factor to form $t_i^* \omega_i$. Note that the Z transformations are the tabulated effect size (t) in this example. The total of the standardized weighted effect size is 146.896 as shown in Table 3.8. Next, the sum of the adjusted effect size will be divided by the sum of the weights ω with a total of 394 to calculate the overall average as in equation (3.11). Thus, the weighted average effect size $T\bullet$

$$\overline{T}\bullet = \frac{\sum_{i=1}^{k} \omega_i t_i}{\sum_{i=1}^{k} \omega_i} = \frac{146.896}{394} = 0.373$$

In fact, the average effect size 0.373 is very close to the mean of correlation 0.378. As the individual correlations are small, we realise that the effect size of r is almost similar as the effect size of Z.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIK
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Conditional variance ($V\bullet$) exists in every sampling distribution. The value of this variance is the inverse of the sum of the ω_i as shown in equation (3.12). For this example, a total of ω_i is 394, so the conditional variance yields

$$V\bullet = \frac{1}{\sum_{i=1}^{k}\frac{1}{v_i}} = \frac{1}{394} = 0.003$$

Calculating the square root of 0.003 we get the estimated standard error (standard deviation) value of 0.0504.

The confidence interval of the population effect size δ with respect to the average effect size $T\bullet$ is given as

$$\overline{T}\bullet - C_{\alpha/2}\sqrt{V\bullet} \leq \delta \leq \overline{T}\bullet + C_{\alpha/2}\sqrt{V\bullet} \qquad (3.13)$$

$$0.373 - 1.96\ (0.0504) \leq \delta \leq 0.373 + 1.96\ (0.0504)$$

$$(0.274,\ 0.472)$$

where $C_{\alpha/2}$ is the two-tailed critical value in the sampling distribution, the value α=0.05 in a general form. $C_{0.025} = 1.96$ is the value of Z in a normal distribution. 95% confidence interval for the population effect size lies between 0.274 and 0.472. We conclude that the greater the ratings of the instructor, the better in student achievement (positive correlation).

The following formula Q statistics which is chi-squared distributed with degree of freedom d.f.= k-1 is used to test if a series of effect sizes is heterogeneous:

$$Q = \sum_{i=1}^{k}\omega_i t_i^{\,2} - \left[\frac{\left(\sum_{i=1}^{k}\omega_i t_i\right)^2}{\sum_{i=1}^{k}\omega_i}\right] \qquad (3.14)$$

$$= 73.690 - [(146.896)^2 / 394] = 18.924$$

where k is the number of studies.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

From the χ^2 table (Appendix A), χ^2 = 26.30 with d.f. =16 at alpha 0.05. The computed value Q= 18.924 < $\chi^2_{0.05,16}$= 26.30 is not a significant result, homogeneity is not rejected indicates that all the effect sizes are homogeneous and they can be combined as a whole.

Z test or t-test is used for testing the statistical significance on the average effect size. The value of Z here is not the value of the Z transformation in the earlier part. The absolute mean effect size value that is divided by the standard deviation to yield the formula of Z-test is as follows:

$$Z = \frac{|\overline{T}\bullet|}{\sqrt{V\bullet}} = \frac{|0.373|}{\sqrt{0.003}} \qquad (3.15)$$

where the tabulated Z=7.4 > $Z_{0.025}$=1.96 or p-value < 0.001. Hence the p-value is small enough to reject the hypothesis with a statistical significant result. This average effect size is significantly different from zero (correlation=0.00) showing that there is a correlation between student ratings of the performance of their instructors and the students' grades in the class.

(2) Combining Standardized Mean Differences

Mean differences is the most fundamental statistic used in many research studies. The example in Table 3.9 is a data set involving hypothetical scores (denote as X) on a functional status instrument which are normally distributed such that

$$X \sim N(50, 5.0)$$

In Table 3.9, the first three columns are labelled n, M and SD (sample size, mean, and the standard deviation) for control groups. The same variables of the next three columns belong to the treatment groups. The treatment groups were the specialized exercise programs respectively.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Each of the effect size is calculated by using its control and treatment group throughout the studies. Suppose the effect size statistic is the standardized mean difference *d*:

$$d_i = \frac{\overline{X}_{Ti} - \overline{X}_{Ci}}{S_i}$$

where S_i is the pooled standard deviation for both groups

$$S_i = \sqrt{\frac{(n_{Ti}-1)S_{Ti}^2 + (n_{Ci}-1)S_{Ci}^2}{(n_{Ti}+n_{Ci}-2)}},$$

n_{Ti} is the within-study sample size in the treatment group, n_{Ci} is the within-study sample size in the control group, $\overline{X}_{Ti}$ is the mean of treatment group in *i*th study, and $\overline{X}_{Ci}$ is the mean of control group in *i*th study. The column labelled "DIFF" indicates the difference between the mean of the treatment group and the mean of the control group.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Table 3.9 Computational Meta-Analysis of Standardized Mean Differences

Study	Control			Treatment			Diff	S_i	d_i	v_i	ω_i	$t_i * \omega_i$	$t_i * ti * \omega_i$
	n	M	SD	n	M	SD							
1	122	42	6.23	120	71	5.23	29	5.75592	5.0383	0.06898	14.4976	73.0430	368.0115
2	16	56	8.48	16	64	6.55	8	7.5767	1.0559	0.14242	7.0215	7.4138	7.8280
3	80	48	5.54	82	44	5.25	-4	5.39514	-0.7414	0.02639	37.8907	-28.0925	20.8280
4	42	40	3.22	42	32	4.48	-8	3.9012	-2.0506	0.07265	13.7647	-28.2265	57.8826
5	219	52	6.94	220	68	7.92	16	7.44726	2.1484	0.01437	69.5950	149.5209	321.2368
6	54	49	4.31	48	53	4.22	4	4.26794	0.9372	0.04366	22.9055	21.4675	20.1198
7	26	38	6.46	23	44	5.55	6	6.0511	0.9916	0.09197	10.8728	10.7810	10.6900
8	11	37	7.66	12	59	8.67	22	8.20457	2.6814	0.33055	3.0253	8.1121	21.7520
9	14	66	6.63	14	58	5.4	-8	6.04636	-1.3231	0.17412	5.7432	-7.5989	10.0542
10	66	44	2.38	70	66	5.78	22	4.4666	4.9254	0.11863	8.4297	41.5199	204.5043
11	44	40	3.49	44	48	4.33	8	3.93249	2.0343	0.06897	14.4993	29.4964	60.0055
12	54	54	7.72	50	56	5.69	2	6.82063	0.2932	0.03893	25.6859	7.5318	2.2085
Sum	748	566	69.06	741	663	69.07	97	69.065	419	366.033	233.931	284.9684	1105.12124
Mean	62.33	47.167	5.755	61.75	55.25	5.7558	8.0833	5.7554	34.9167	30.5027	19.4943	23.7474	92.0934

UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Each study creates a standardized effect size that is weighted by the variance estimate within-study. The conditional variance of each effect size is estimated as:

$$v_i = \frac{(n_{Ti} + n_{Ci})}{(n_{Ti} n_{Ci})} + \frac{d_i^2}{[2(n_{Ti} + n_{Ci})]}$$

Let take first study in Table 3.9 to calculate the formula above:

$$v_1 = \frac{(120+122)}{(120)(122)} + \frac{(5.0383)^2}{[2(120+122)]}$$

$$= 0.06898$$

The inverse of the conditional variance produces the weighting factor such that $\omega_i = \frac{1}{v_i}$. For study 1, the weighting factor $\omega_1 = \frac{1}{0.06898} = 14.4976$ and the same procedure for the rest of the studies.

When the effect size t_i is multiplied by the weighting factor ω_i, the "adjusted effect size" with the value of 73.0430 for study 1 is obtained. Using the same method, the remaining studies produce the values of "adjusted effect size". Then sum up those individual $t_i * \omega_i$ (284.968) and divide by the summation of weighting factors across all of the studies(233.9312). Finally the average effect size as in equation (3.11) is given as:

$$\left|\overline{T}\bullet\right| = 284.968 / 233.9312 = 1.21817$$

There are two equivalent ways to test for the significance of the average effect size in a meta-analysis: calculate a confidence interval for average adjusted effect size, and Z-test statistic. Before this, we need to find the conditional variance, and the standard error. As in equation (3.12),

$$V\bullet = 1/(233.9312) = 0.00427,$$

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Finding the square root of 0.00427, which, the standard error in Table 3.9, would be 0.0654.

Set α=0.05 for the normal distribution which yields $C_{\alpha/2}$= 1.96. Thus, the confidence interval for $T\bullet$ as in equation (3.13):

$$[\overline{T}\bullet - C_{\alpha/2}\sqrt{V\bullet} \leq \delta \leq \overline{T}\bullet + C_{\alpha/2}\sqrt{V\bullet}]$$

$$[1.21817 - (1.96)(0.00427)^{1/2} \leq \delta \leq 1.21817 + (1.96)(0.00427)^{1/2}]$$

$$(1.09002, 1.34632).$$

We can conclude that the 95% confidence interval for the population effect size parameter ranges from 1.09002 to 1.34632, which excludes zero in the interval.

By using Z-test for testing the significance on the average effect size as in equation (3.15),

$$Z = \frac{|\overline{T}\bullet|}{\sqrt{V\bullet}} = \frac{|1.218|}{\sqrt{0.00427}} = 18.6317,$$

which exceeds the 1.96 critical value at α= 0.05. Furthermore there is no zero value in 95% confidence interval. We can reject the null hypothesis by saying there is a significance effect on specialized exercise program.

To investigate whether the effect sizes are heterogeneous, the Q statistics as mentioned in (3.14) explain if the data set comes from the same population.

$$Q = \sum_{i=1}^{k} \omega_i t_i^2 - \frac{\left(\sum_{i=1}^{k} \omega_i t_i\right)^2}{\sum_{i=1}^{k} \omega_i}$$

$$= 1105.1212 - [(284.968)^2/233.931]$$

$$= 757.981$$

with d.f= k – 1 distributed as chi-squared χ^2. As the example from Table 3.9, the calculated Q= 757.981 > $\chi^2_{0.05,11}$= 19.68. So we reject homogeneity of effect sizes at

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

significance level 0.05 implying that the data set is not from the same population. As a result, other variables will be required to explain further on the variance in these effect sizes. The random effects method comes into play if the source of variability could hardly be identified in the fixed effect model.

(3) Combining differences between proportions

A number of studies report the proportion of the treatment (exposed) group with a condition (a treatment success or a disease) and the proportion of the comparison (unexposed) group with that condition. There are advantages and disadvantages of using proportions discussed by Hasselblad & Hedges (1995) and Fleiss (1994).

The following data set on a meta-analysis (Table 3.10) compare two different treatments (experimental designs, and quasi-experimental designs) on patients outcomes. The proportions of successful aminoglycoside therapy for gram-negative infections are compared for a total of 17 studies.

For treatment group, the drug therapy is managed by a coordinated team of nurses, physicians, clinical staff, and pharmacists to observe at an appropriate time with a proper dose. In the control group, a standard clinical practice is carried out. Both groups are compared among the studies to find out if the treatment conditions are more effective than the control group.

To calculate the effect size for each study, we need the first four columns "treatment success", "treatment failure", "control success", and "control failure". The effect size refers to the difference between proportions $t_i = p_{i1} - p_{i2}$ where p_{i1} is the proportion of treatment success whilst p_{i2} is the proportion of control success.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI

The p_{i1} is the treatment success which is divided by the total treatment (p_{i1} = treatment success / total treatment). The same method is applied for p_{i2} where the control success is divided by the total controls (p_{i2} = control success / total controls). Example for study 1, p_{11} = 22 / (22+58) = 0.275 while p_{12} = 2 / (2+71) = 0.027. The same procedure is used for the remaining 16 studies.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

Table 3.10 Computation of Meta-analysis of differences between proportions

Study	n	Treatment success	Treatment failure	Control success	Control failure	P_{i1}	P_{i2}	t_i	v_i	ω_i	$t_i * \omega_i$	$t_i * t_i * \omega_i$
1	153	22	58	2	71	0.27500	0.02740	0.24760	0.00286	349.99176	86.65892	21.45699
2	41	8	10	5	18	0.44444	0.21739	0.22705	0.02111	47.36087	10.75343	2.44160
3	148	21	40	12	75	0.34426	0.13793	0.20633	0.00507	197.33648	40.71669	8.40113
4	41	17	5	10	9	0.77273	0.52632	0.24641	0.02110	47.38401	11.67596	2.87709
5	143	58	12	44	29	0.82857	0.60274	0.22583	0.00531	188.35170	42.53578	9.60593
6	110	49	26	8	27	0.65333	0.22857	0.42476	0.00806	124.10408	52.71469	22.39119
7	50	9	15	2	24	0.37500	0.07692	0.29808	0.01250	80.02163	23.85260	7.10991
8	145	49	26	11	59	0.65333	0.15714	0.49619	0.00491	203.58389	101.01639	50.12337
9	262	97	4	52	109	0.96040	0.32298	0.63741	0.00173	576.45095	367.43829	234.21056
10	60	14	16	3	27	0.46667	0.10000	0.36667	0.01130	88.52459	32.45902	11.90164
11	46	10	13	6	17	0.43478	0.26087	0.17391	0.01907	52.44397	9.12069	1.58621
12	54	20	7	12	15	0.74074	0.44444	0.29630	0.01626	61.50938	18.22500	5.40000
13	2011	1098	429	189	295	0.71906	0.39050	0.32856	0.00062	1602.43946	526.49928	172.98719
14	217	108	60	10	39	0.64286	0.20408	0.43878	0.00468	213.60386	93.72414	41.12386
15	89	33	33	11	12	0.50000	0.47826	0.02174	0.01464	68.32049	1.48523	0.03229
16	435	125	78	137	95	0.61576	0.59052	0.02525	0.00221	452.94342	11.43515	0.28870
17	77	45	7	17	8	0.86538	0.68000	0.18538	0.01094	91.37201	16.93896	3.14022
Sum	4082	1783	839	531	929	10.29232	5.44606	4.84626	0.16237	4445.74253	1447.25022	595.07786
Mean	240.118	104.882	49.353	31.235	54.647	0.60543	0.32036	0.28507	0.00955	261.51427	85.13237	35.00458

UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

The variance estimate within each study is used to create a standardized effect size. Formula for variance within-study is

$$v_i = \frac{p_{i1}(1-p_{i1})}{n_{i1}} + \frac{p_{i2}(1-p_{i2})}{n_{i2}} \qquad (3.16)$$

As for study 1, $v_1 = \frac{0.275(1-0.275)}{80} + \frac{0.027(1-0.027)}{80}$

$= 0.00286.$

The next step is to find the weighting factor, $\omega_i = \frac{1}{v_i}$. If we consider the weighting factor for the study 1, $\omega_1 = \frac{1}{0.00286} = 349.992.$

We will also need to find the "adjusted effect size" in the form of multiplication between effect size and weighting factor ($t_i * \omega_i$). For study 1, $t_1 * \omega_1 = 86.659.$

We find the average effect size from the given equation (3.11). We total up all the $t_i * \omega_i$ and divide the summation by the sum of the mean variance estimate across 17 studies.

$$\bar{T}\bullet = \frac{\sum_{i=1}^{k} \omega_i t_i}{\sum_{i=1}^{k} \omega_i} = \frac{1447.250}{4445.742} = 0.325$$

When testing for significance, the first step is to compute the conditional variance and then square root the value to get the standard error. The conditional variance as in (3.12)

$$V\bullet = \frac{1}{\sum_{i=1}^{k} \frac{1}{v_i}} = 0.002,$$

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

so the standard error is $(0.002)^{1/2} = 0.015$.

Finding 95% confidence interval for mean effect size, the critical value for the normal distribution is $C_{\alpha/2} = 1.96$. As in formula (3.13),

$$\bar{T}\bullet - C_{\alpha/2}\sqrt{V\bullet} \leq \delta \leq \bar{T}\bullet + C_{\alpha/2}\sqrt{V\bullet}$$

$$[0.325 - (1.96)(0.0002)^{1/2}] \leq \delta \leq [0.325 + (1.96)(0.0002)^{1/2}]$$

$$(0.296, 0.355)$$

From the findings, we are 95% confident that the population of the effect size with a range from 0.296 to 0.355.

Another approach to test for the significance on the average effect size is the Z or t-test. As shown in Formula (3.15), the absolute value of the mean effect size is divided by the standard deviation,

$$Z = \frac{|\bar{T}\bullet|}{\sqrt{V\bullet}} = \frac{|0.325|}{\sqrt{0.0002}} = 22.981$$

The Z value and the confidence interval excludes zero have proven that the mean effect size is significantly different from zero (proportion = 0.00), thus the null hypothesis is false.

To make sure if the effect sizes in the data set of Table 3.10 is homogeneous and come from the same population, the formula as in (3.14) is tested across the studies,

$$Q = \sum_{i=1}^{k} \omega_i t_i^2 - \left[\frac{\left(\sum_{i=1}^{k} \omega_i t_i \right)^2}{\sum_{i=1}^{k} \omega_i} \right]$$

$$= (595.078 - \{(1447.250)^2 / 4445.742\})$$

$$= 123.945.$$

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Since the Q statistics is distributed as chi-squared with d.f. = k – 1, so we use χ^2 table to search for the value of $\chi^2_{0.05,16}$ = 124.430 > tabulated Q=123.945, thus it is statistically significant. We reject the null hypothesis of which the effect sizes are homogeneous. This suggests that the studies do not come from the same population. Perhaps other methods would be more appropriate to explain the source of variability among the studies either attempt using other variables or choose a random rather than a fixed effect model.

(4) Combining Quality-Weighted Approach

Each study is rated based on its research quality scaling from 0 to 1.0. Each effect size is adjusted by the quality index, by combining quality-weighted effect sizes throughout the studies we manage to calculate for mean effect size eventually. The example of this method is re-created from Table 3.10 with additional variable q_i (quality-weighted). The procedures are almost identical to the previous part (combining differences between proportions) except a few computations differ.

The first three columns of Table 3.10 with "Treatment success", "treatment failure", and "control success" are used to calculate for weighted-quality. The effect size is the proportion of treatment success p_{i1} deducts the proportion of control success p_{i2}, which similar as the one calculated in Table 3.10.

Formula for calculating the variance estimate within-study is exactly the same as the equation (3.16)

$$v_i = \frac{p_{i1}(1-p_{i1})}{n_{i1}} + \frac{p_{i2}(1-p_{i2})}{n_{i2}}$$

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI

The weighting factor $\omega_i = \frac{1}{v_i}$ is included in the next column of Table 3.11.

The quality index ranges between 0 and 1.0. The quality is in line with the scores obtained from the studies. The higher the scores it is, the better the quality of the study. The "adjusted effect size" can be obtained by the term $\omega_i * q_i * t_i$. For study 1 in Table 3.11, $\omega_1 * q_1 * t_1 = (0.2476)(281.0434) = 69.5863$.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Table 3.11 Computational Meta-analysis of Quality-Weighted Approach

Study	n	Quality, q_i	t_i	v_i	ω_i	$\omega_i * q_i$	$\omega_i * q_i * t_i$	$\omega_i * q_i * q_i$
1	153	0.80300	0.24760	0.00286	349.99176	281.04338	69.58634	225.67784
2	41	0.66800	0.22705	0.02111	47.36087	31.63706	7.18319	21.13356
3	148	0.66800	0.20633	0.00507	197.33648	131.82077	27.19858	88.05627
4	41	0.70700	0.24641	0.02110	47.38401	33.50050	8.25486	23.68485
5	143	0.71200	0.22583	0.00531	188.35170	134.10641	30.28525	95.48376
6	110	0.57200	0.42476	0.00806	124.10408	70.98753	30.15266	40.60487
7	50	0.76000	0.29808	0.01250	80.02163	60.81644	18.12816	46.22049
8	145	0.38900	0.49619	0.00491	203.58389	79.19413	39.29534	30.80652
9	262	0.33600	0.63741	0.00173	576.45095	193.68752	123.45836	65.07901
10	60	0.26900	0.36667	0.01130	88.52459	23.81311	8.73155	6.40573
11	46	0.47600	0.17391	0.01907	52.44397	24.96333	4.34137	11.88254
12	54	0.44200	0.29630	0.01626	61.50938	27.18715	8.05555	12.01672
13	2011	0.72100	0.32856	0.00062	1602.43946	1155.35885	379.60470	833.01373
14	217	0.41300	0.43878	0.00468	213.60386	88.21839	38.70847	36.43420
15	89	0.41300	0.02174	0.01464	68.32049	28.21636	0.61342	11.65336
16	435	0.57700	0.02525	0.00221	452.94342	261.34835	6.59905	150.79800
17	77	0.55800	0.18538	0.01094	91.37201	50.98558	9.45171	28.44995
Sum			4.84626	0.16237	4445.74253	2676.88488	809.64858	1727.40140
Mean			0.28507	0.00955	261.51427	157.46382	47.62639	101.61185

UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

The quality-and-variance-weighted average effect size is estimated in Equation (3.11) (here weighted for quality) as

$$\bar{T}\bullet = \frac{\sum_{i=1}^{k} \omega_i q_i t_i}{\sum_{i=1}^{k} \omega_i q_i} = \frac{809.6508}{2676.8849} = 0.3025$$

Finding for conditional variance, firstly the column of ω_i *q_i multiplies column labelled q_i as reported in the column of ω_i*q_i*q_i (i.e. ω_1 *q_1*q_1 = 225.6778), next we sum up all of the $q_i^2 \omega_i$, this is the numerator of conditional variance with a value of 1727.4014. Next, we find the total of all the ω_i*q_i in 17 studies with a value of 2676.8849, then square this figure to become the denominator of conditional variance. The conditional variance is where the numerator divided by the denominator as given

$$V\bullet = \frac{\sum_{i=1}^{k} q_i^2 \omega_i}{\left(\sum_{i=1}^{k} q_i \omega_i\right)^2} = \frac{1727.4014}{(2676.8849)^2}$$

$$= 0.0002$$

For the significance testing, both the confidence interval and the Z-test use the same Formula as in (3.13) and (3.15)

$$\bar{T}\bullet - C_{\alpha/2}\sqrt{V\bullet} \leq \delta \leq \bar{T}\bullet + C_{\alpha/2}\sqrt{V\bullet}$$

$$[0.3025 - 1.96(0.0002)^{1/2} \leq \delta \leq 0.3025 + 1.96(0.0002)^{1/2}]$$

(0.2720, 0.3329), and

$$Z = \frac{|\bar{T}\bullet|}{\sqrt{V\bullet}} = \frac{|0.3025|}{\sqrt{0.0002}} = 19.4805$$

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

We have 95% confidence that the quality-unweighted confidence interval in part (3) lies between 0.296 and 0.355. The interval is narrower than the quality-weighted confidence interval scaling from 0.2720 to 0.3329. Based on quality-weighted statistics, the confidence interval excludes the value of zero and the Z-test is statistically significant at alpha 0.05.

The Q statistics (χ^2 distributed) is used for homogeneity testing on effect sizes.

$$Q = \sum_{i=1}^{k} \omega_i q_i t_i^2 - \frac{\left(\sum_{i=1}^{k} q_i \omega_i t_i\right)^2}{\sum_{i=1}^{k} q_i \omega_i}$$

$$= 299.734 - [(809.6508)^2 / 2676.8849]$$

$$= 54.8468 > \chi^2_{0.05,16} = 26.30$$

We conclude that the data set is not from the same population (heterogeneous) as it is statistically significant.

3.3.2 Random Effects Model

In fixed effects model, we assume that all the studies are taken from the same population having a common population despite they were randomly sampled from different populations. In fact, in meta-analysis, different variables in each populations yield different effect size. To deal with the variation (treatment or populations) effectively, we consider the random effects model. Here, we need to estimate the between- and within-study variances while the fixed effects approach only deals with estimating the within-study variances.

Reuse the example from part (3) as presented in Table 3.10 and demonstrate the meta-analysis in Table 3.12 by using random effects model.

In finding the effect size estimate t_i, the formula is identical to the previous part (3) in Table 3.10. For instance, $t_1 = 0.24760$.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Table 3.12 Computational Meta-analysis of Random Effects Approach

Study	t_i	v_i	ω_i	$t_i * \omega_i$	$\omega_i * \omega_i$	v_i^a	ω_i^a	$t_i * \omega_i^a$
1	0.24760	0.00286	349.99176	86.65892	122494.23207	0.03226	30.99814	7.67514
2	0.22705	0.02111	47.36087	10.75343	2243.05201	0.05051	19.79806	4.49515
3	0.20633	0.00507	197.33648	40.71669	38941.68634	0.03447	29.01073	5.98578
4	0.24641	0.02110	47.38401	11.67596	2245.24440	0.05050	19.80198	4.87941
5	0.22583	0.00531	188.35170	42.53578	35476.36289	0.03471	28.81014	6.50619
6	0.42476	0.00806	124.10408	52.71469	15401.82267	0.03746	26.69514	11.33903
7	0.29808	0.01250	80.02163	23.85260	6403.46127	0.04190	23.86635	7.11408
8	0.49619	0.00491	203.58389	101.01639	41446.40027	0.03431	29.14602	14.46196
9	0.63741	0.00173	576.45095	367.43829	332295.69776	0.03113	32.12335	20.47575
10	0.36667	0.01130	88.52459	32.45902	7836.60303	0.04070	24.57002	9.00909
11	0.17391	0.01907	52.44397	9.12069	2750.36999	0.04847	20.63132	3.58799
12	0.29630	0.01626	61.50938	18.22500	3783.40383	0.04566	21.90101	6.48927
13	0.32856	0.00062	1602.43946	526.49928	2567812.22297	0.03002	33.31113	10.94470
14	0.43878	0.00468	213.60386	93.72414	45626.60901	0.03408	29.34272	12.87500
15	0.02174	0.01464	68.32049	1.48523	4667.68935	0.04404	22.70663	0.49364
16	0.02525	0.00221	452.94342	11.43515	205157.74172	0.03161	31.63556	0.79880
17	0.18538	0.01094	91.37201	16.93896	8348.84421	0.04034	24.78929	4.59544
Sum	4.84626	0.16237	4445.74253	1447.25022	3442931.44378	0.66217	449.13760	131.72643
Mean	0.28507	0.00955	261.51427	85.13237	202525.37905	0.03895	26.41986	7.74861

UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

Variance estimate for random effects approach has an additional between-study variance σ_θ^2 besides the within-study variance v_i as in (3.16).

$$v_i^a = \hat{\sigma}_\theta^2 + v_i \qquad (3.17)$$

where the between-study variance is estimated by

$$\hat{\sigma}_\theta^2 = [Q - (k-1)] / c$$

The value of Q = 123.9455 is obtained from Table 3.10 showing that the effect size is heterogeneous. The value of c is estimated as

$$c = \sum_{i=1}^{k} \omega_i - \left(\frac{\sum_{i=1}^{k} \omega_i^2}{\sum_{i=1}^{k} \omega_i} \right)$$

$$= 4445.7425 - [(3442931.4518) / (4445.7425)]$$

$$= 3671.3091$$

The total of the column labelled ω_i is 4445.7425 while column labelled $\omega_i * \omega_i$ is 3442931.4518, thus the value of c is 3671.3091. The values of v_i, t_i, ω_i and $\omega_i * t_i$ as presented in Table 3.10 are repeated in the following Table 3.12. Now, with the values of Q and c, the between-study variance can be estimated

$$\hat{\sigma}_\theta^2 = (123.9455 - 16) / 3671.3091 = 0.0294$$

Finally, substitute the between-and within-study variance into Equation (3.17) yields the variance estimate. We take study 1 as an example $v_1^a = 0.0294 + 0.00286 = 0.03226$.

The way to compute weighting factor in random effects model is similar as the fixed effects model. It is defined as the inverse of variance estimate. However,

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

the variance now is a combination of between-and within-study variances v_i^a. For study 1, $\omega_1^a = \frac{1}{v_1^a} = 1/0.03226 = 31.0008$.

Then, the effect size will be adjusted by the weighting factor. For random effects approach we use $t_1 * \omega_i^a$ instead of $t_1 * \omega_i$ that has been applied in fixed effects model. For study 1, $t_1 * \omega_1^a = 0.2476 * 31.00082 = 7.6759$.

Referring to the Formula (3.11), the average effect size in random effects model is the same as the fixed effects model except the ω_i is now replaced by ω_i^a.

$$\overline{T}\bullet = \frac{\sum_{i=1}^{k} t_i \omega_i^a}{\sum_{i=1}^{k} \omega_i^a}$$

$$= 131.6759/449.362 = 0.29328$$

Conditional variance is the inverse of the total weighting factors across the studies as follows:

$$V\bullet = \frac{1}{\sum_{i=1}^{k} \omega_i^a} = 1/449.13617 = 0.00223$$

We can find the square root of this figure to obtain the standard error as $(0.00223)^{1/2}$ = 0.04719. Thus, 95% confidence interval for $T\bullet$ as in Formula (3.13):

$$[\overline{T}\bullet - C_{\alpha/2}\sqrt{V\bullet} \leq \delta \leq \overline{T}\bullet + C_{\alpha/2}\sqrt{V\bullet}]$$

$$[0.2933 - (1.96)(0.04719) \leq \delta \leq 0.2933 + (1.96)(0.04719)]$$

$$(0.2008, 0.3858)$$

The range for interval in random effects method (0.2008, 0.3858) is greater than in the fixed effects method (0.296, 0.355).

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

Similarly to equation (3.15), Z-test is used to estimate the significance test on average effect size,

$$Z = \frac{|\bar{T}\bullet|}{\sqrt{V\bullet}} = \frac{|0.2933|}{\sqrt{0.0022}} = 6.2155$$

where z is statistically significant which is the same as the outcome in fixed effects model. This similar outcome between fixed and random effects model would be varied if the sample size is smaller, or the effect sizes are smaller in the meta-analysis.

(1) Combining Difference in Odds Ratios

There are two methods in this section for testing random effects model in both cases investigated by Fleiss (1981), Shadish & Haddock (1994) – log odds ratio, and odds ratio using Mantel-Haenszel methods. Log odds ratio is applied for large sample sizes in a small number of studies. On the other hand, Mantel-Haenszel method is used when a big amount of studies each has a small sample size.

Table 3.13 shows artificial data based on the randomized controlled clinical trials in meta-analysis. These clinical trials aimed to test the effectiveness of the treatment of a new bacteriocidal antibiotic (D-71) for adult and elderly patients (55 to 95 years old) with chronic *E.coli pyelonephritis*. For control condition, it was operated by a traditional bacteriostatic *coli pyelonephritis*.

The column labelled "Infection cleared" shows that no growth in the catheterized urine 2 weeks after 16-weeks of treatment and also at the 3- and 6 months posttreatment period. "Infection same" means the urine grew more than 100,000 colonies per millilitre (ml) of the same coliform strain anytime 2 weeks after the 16-week treatment and before the 6-month posttreatment follow-up.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

Table 3.13 Computational Meta-analysis for Odds Ratios for Hypothetical Pyelonephritis Population (RCT) Testing New Antibiotic

		Treatment (New Antibiotic D-71)		Control (Traditional Antibiotic Y)							
		Infection cleared	Infection same	Infection cleared	Infection same						
Study	n	A	B	C	D	Odds ratio (OR)	Log Odds $t_i = \ln(OR)$	v_i	ω_i	$\omega_i * t_i$	$\omega_i * t_i * t_i$
1	153	22	58	2	71	13.4655	2.6001	0.5768	1.7338	4.5080	11.7214
2	41	8	10	5	18	2.8800	1.0578	0.4806	2.0809	2.2012	2.3284
3	148	21	40	12	75	3.2813	1.1882	0.1693	5.9072	7.0190	8.3402
4	41	17	5	10	9	3.0600	1.1184	0.4699	2.1280	2.3799	2.6618
5	143	58	12	44	29	3.1856	1.1586	0.1578	6.3377	7.3432	8.5081
6	110	49	26	8	27	6.3606	1.8501	0.2209	4.5268	8.3751	15.4950
7	50	9	15	2	24	7.2000	1.9741	0.7194	1.3900	2.7439	5.4167
8	145	49	26	11	59	10.1084	2.3134	0.1667	5.9978	13.8751	32.0982
9	262	97	4	52	109	50.8317	3.9285	0.2887	3.4636	13.6069	53.4552
10	60	14	16	3	27	7.8750	2.0637	0.5043	1.9830	4.0922	8.4450
11	46	10	13	6	17	2.1795	0.7791	0.4024	2.4850	1.9360	1.5084
12	54	20	7	12	15	3.5714	1.2730	0.3429	2.9167	3.7128	4.7263
13	2011	1098	429	189	295	3.9949	1.3850	0.0119	83.8744	116.1675	160.8940
14	217	108	60	10	39	7.0200	1.9488	0.1516	6.5977	12.8574	25.0561
15	89	33	33	11	12	1.0909	0.0870	0.2348	4.2581	0.3705	0.0322
16	435	125	78	137	95	1.1113	0.1055	0.0386	25.8758	2.7299	0.2880
17	77	45	7	17	8	3.0252	1.1070	0.3489	2.8661	3.1727	3.5122
Sum	4028	1783	839	531	929	130.2413	25.9383	5.2856	164.4226	207.0916	344.4871
Mean	240.118	104.88235	49.352941	31.235294	54.647059	7.6613	1.5258	0.3109	9.6719	12.1819	20.2639

UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

Each study has an effect size in the form of natural log (ln) of the odds ratios but is not the odds ratios itself (Shadish & Haddock, 1994). The formula for calculating odds ratios using 2-way contingency table as in Table 3.13:

Odds Ratio= (A*D) / (B*C)

For study 1, OR= (22*71) / (58*2) = 13.466. We obtain the effect size t_i for each study by ln transformation such as t_1 = ln (13.466) = 2.6001 for study 1 in Table 3.13.

To calculate the variance estimate of the odds ratio for each study, we use

$$v_i = \frac{1}{A_i} + \frac{1}{B_i} + \frac{1}{C_i} + \frac{1}{D_i}$$

For study 1, v_1 = (1/22) + (1/58) + (1/2) + (1/71) = 0.5768. To find the weighting factor, it is just the inverse of the variance estimate of odds ratio $\omega_i = \frac{1}{v_i}$. As for study 1, ω_1 = 1/0.5768 = 1.7338.

We adjust the effect size with the variance estimate by log odds ratio multiplies with the weighting factor such as study 1, $t_1 * \omega_1$ = 2.6001*1.7338 = 4.5080.

The procedures to find for the mean effect size is identical to the Equation (3.11),

$$\overline{T}\bullet = \frac{\sum_{i=1}^{k} \omega_i t_i}{\sum_{i=1}^{k} \omega_i}$$

As for the conditional variance shown in Formula (3.12), we can calculate the standard error by square root of its figure

$$V\bullet = \frac{1}{\sum_{i=1}^{k} \frac{1}{v_i}} = \frac{1}{164.42} = 0.006$$

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

such that the standard error is $(0.006)^{1/2} = 0.0780$.

It is the similar as the previous equation (3.13) and (3.15) to test for the significance on average effect size. Set the $C_{\alpha/2} = 1.96$ where $\alpha = 0.05$.

$$[\bar{T}\bullet - C_{\alpha/2}\sqrt{V\bullet} \leq \delta \leq \bar{T}\bullet + C_{\alpha/2}\sqrt{V\bullet}]$$

$$[1.260 - 1.96(0.006)^{1/2} \leq \delta \leq 1.260 + 1.96(0.006)^{1/2}]$$

$$(1.108, 1.411)$$

Thus, 95% confidence interval for the odds ratio ranges from 1.108 to 1.411. For Z-test,

$$Z = \frac{|\bar{T}\bullet|}{\sqrt{V\bullet}} = \frac{|1.259|}{\sqrt{0.0780}} = 16.15$$

From the findings above, we can conclude that the null hypothesis is false which indicates the treatment is more effective than the control condition in this study.

To test whether the effect size is homogeneous, Q statistics as given in Equation (3.14) will be considered,

$$Q = \sum_{i=1}^{k} \omega_i t_i^2 - \left[\frac{\left(\sum_{i=1}^{k} \omega_i t_i \right)^2}{\sum_{i=1}^{k} \omega_i} \right]$$

$$= (344.4871 - \{(207.0916)^2 / 164.4226\})$$

$$= 83.6356.$$

since Q is χ^2 distributed with d.f. = k – 1, the value of $Q = 83.6356 > \chi^2_{0.05,16} = 26.30$ shows it is statistically significant. The 17 studies are not sampled from the same population and so they are heterogeneous.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

(2) Combining Differences in Odds Ratios with small sample size (Mantel-Haenszel)

This approach is more appropriate for small sample size for within-study in a large number of studies. Likewise, Table 3.13 shows the values of A, B, C, and D, the following Table 3.14 reuses these figures by using the MH approach in this example.

The average effect size of the MH ($\overline{O}_{MH}$) is calculated according to the cell frequencies

$$\overline{O}_{MH} = \overline{T}\bullet = \frac{\sum_{i=1}^{k}(A_i D_i / n_i)}{\sum_{i=1}^{k}(B_i C_i / n_i)}$$

where n_i indicates the sample size within-study. The total for the numerator is 348.9514 whereas the summation of denominator is 92.7145. Thus the mean effect size of MH is 3.76 which means the probability of treatment more effective was 3.76 times as much as the control group in the data set.

The process of calculating the conditional variance could be complicated (Robins, Breslow, & Greenland, 1986). Conditional variance of MH is given by

$$V\bullet = \frac{\sum_{i=1}^{k} P_i R_i}{2(\sum_{i=1}^{k} R_i)^2} + \frac{\sum_{i=1}^{k}(P_i S_i + Q_i R_i)}{2(\sum_{i=1}^{k} R_i)(\sum_{i=1}^{k} S_i)} + \frac{\sum_{i=1}^{k} Q_i S_i}{2(\sum_{i=1}^{k} S_i)^2}.$$

However, the Q here is unlike the variable stated in Equation (3.14) for testing the heterogeneity. We define P, Q, R and S as follows:

$P_i = (A_i + D_i)/n$, $Q_i = (B_i + C_i)/n$,

$R_i = (A_i * D_i)/n$, $S_i = (B_i * C_i)/n$.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Let us take study 1 as an example,

$$P_1 = (22+71)/\ 153 = 0.608,$$

$$Q_1 = (58+2)\ /\ 153 = 0.392,$$

$$R_1 = (22*71)/153 = 10.209,$$

$$S_1 = (58*2)\ /\ 153 = 0.758$$

such that the conditional variance is

$$V\bullet = \frac{236.668}{2(348.951)^2} + \frac{170.339}{2(348.951)(92.715)} + \frac{34.660}{2(92.715)^2}$$

$$= 0.006$$

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Table 3.14 Computational Differences in Odds Ratios with Small Sample Sizes (Mantel-Haenszel)

Study	n	Risk Present (A)	Risk Absent (B)	Disease Present (C)	Disease Absent (D)	(A*D) / n	(B*C) / n	P (A+D) / n	Q (B+C) / n	R (A*D) / n	S (B*C) / n
1	153	22	58	2	71	10.2092	0.7582	0.6078	0.3922	10.2092	0.7582
2	41	8	10	5	18	3.5122	1.2195	0.6341	0.3659	3.5122	1.2195
3	148	21	40	12	75	10.6419	3.2432	0.6486	0.3514	10.6419	3.2432
4	41	17	5	10	9	3.7317	1.2195	0.6341	0.3659	3.7317	1.2195
5	143	58	12	44	29	11.7622	3.6923	0.6084	0.3916	11.7622	3.6923
6	110	49	26	8	27	12.0273	1.8909	0.6909	0.3091	12.0273	1.8909
7	50	9	15	2	24	4.3200	0.6000	0.6600	0.3400	4.3200	0.6000
8	145	49	26	11	59	19.9379	1.9724	0.7448	0.2552	19.9379	1.9724
9	262	97	4	52	109	40.3550	0.7939	0.7863	0.2137	40.3550	0.7939
10	60	14	16	3	27	6.3000	0.8000	0.6833	0.3167	6.3000	0.8000
11	46	10	13	6	17	3.6957	1.6957	0.5870	0.4130	3.6957	1.6957
12	54	20	7	12	15	5.5556	1.5556	0.6481	0.3519	5.5556	1.5556
13	2011	1098	429	189	295	161.0691	40.3187	0.6927	0.3073	161.0691	40.3187
14	217	108	60	10	39	19.4101	2.7650	0.6774	0.3226	19.4101	2.7650
15	89	33	33	11	12	4.4494	4.0787	0.5056	0.4944	4.4494	4.0787
16	435	125	78	137	95	27.2989	24.5655	0.5057	0.4943	27.2989	24.5655
17	77	45	7	17	8	4.6753	1.5455	0.6883	0.3117	4.6753	1.5455
Sum	4028	1783	839	531	929	348.9514	92.7145	11.0034	5.9966	348.9514	92.7145
Mean	240.118	104.882	49.3529	31.2353	54.6471	20.5266	5.4538	0.6473	0.3527	20.5266	5.4538

UNIVERSITI PENDIDIKAN SULTAN IDRIS

Table 3.14 Continued

(PS+QR)	(Q*R)	(P*R)	(Q*S)	E	(A+C)	(B+D)	(A+B)	(C+D)	$N^2(N-1)$	V	(O-E)
4.4644	4.0036	6.2056	0.2973	12.5490	24.00	129.00	80.00	73.00	3558168	5	9.4510
2.0583	1.2849	2.2272	0.4462	5.7073	13.00	28.00	18.00	23.00	67240	2	2.2927
5.8428	3.7390	6.9028	1.1395	13.6014	33.00	115.00	61.00	87.00	3219888	6	7.3986
2.1386	1.3653	2.3664	0.4462	14.4878	27.00	14.00	22.00	19.00	67240	2	2.5122
6.8526	4.6062	7.1560	1.4459	49.9301	102.00	41.00	70.00	73.00	2903758	7	8.0699
5.0240	3.7175	8.3098	0.5845	38.8636	57.00	53.00	75.00	35.00	1318900	6	10.1364
1.8648	1.4688	2.8512	0.2040	5.2800	11.00	39.00	24.00	26.00	122500	2	3.7200
6.5567	5.0876	14.8503	0.5033	31.0345	60.00	85.00	75.00	70.00	3027600	9	17.9655
9.2497	8.6255	31.7295	0.1697	57.4389	149.00	113.00	101.00	161.00	17916084	15	39.5611
2.5417	1.9950	4.3050	0.2533	8.5000	17.00	43.00	30.00	30.00	212400	3	5.5000
2.5217	1.5265	2.1692	0.7004	8.0000	16.00	30.00	23.00	23.00	95220	3	2.0000
2.9630	1.9547	3.6008	0.5473	16.0000	32.00	22.00	27.00	27.00	154548	3	4.0000
77.4265	49.4981	111.5710	12.3903	977.2496	1287.00	724.00	1527.00	484.00	8128683210	85	120.7504
8.1344	6.2613	13.1488	0.8919	91.3548	118.00	99.00	168.00	49.00	10171224	9	16.6452
4.2620	2.1997	2.2497	2.0164	32.6292	44.00	45.00	66.00	23.00	697048	4	0.3708
25.9165	13.4925	13.8063	12.1416	122.2667	262.00	173.00	203.00	232.00	82123650	26	2.7333
2.5210	1.4572	3.2181	0.4817	41.8701	62.00	15.00	52.00	25.00	450604	3	3.1299
170.3386	112.2836	236.6678	34.6596	1526.7631	2314.00	1768.00	2622.00	1460.00		191.8571	256.2369
10.0199	6.6049	13.9216	2.0388	89.8096	136.12	104.00	154.24	85.88		11.2857	15.0728

UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

The confidence interval of MH below is according to the procedures by Shadish and Haddock (1994). First find the lower limit of 95% confidence interval which is given by

$$\varpi = \exp[\ln(T\bullet) \pm 1.96\sqrt{V\bullet}\ ]$$

where the "exp" is the inverse of the natural log (ln) or rather is an antilog to obtain the lower limit confidence interval at 5% significance level such that

$$\varpi_L = \exp\{\ln(3.763) - 1.96(0.006)^{1/2}\} = 3.23.$$

Next, the upper limit of MH confidence interval,

$$\varpi_U = \exp\{\ln(3.763) + 1.96(0.006)^{1/2}\} = 4.38.$$

95% confidence interval of MH for average effect size ranging from 3.23 to 4.38 excluded the value of 1.00 which proved to us the null hypothesis can be rejected. By saying so, we conclude that the probability of success in treatment is different from the control group.

The Q statistics for MH method makes no difference from the equation in (3.14). By substituting the figures throughout the studies, we manage to get the value to test the heterogeneity of the effect sizes.

Instead of using Z-test for significance test in the previous section, we use the following formula to test the significance of the overall MH odds ratio,

$$\chi^2 = \frac{\left[\left|\sum_{i=1}^{k}(O_i - E_i)\right| - 0.5\right]^2}{\sum_{i=1}^{k} V_i}.$$

According to Shadish & Haddock (1994), finding the values for O_i, E_i, and V_i are as follows:

$$E_i = \frac{(A_i + C_i)(A_i + B_i)}{n_i}$$

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

In order to have a better understanding of the expression presented above, we take an example, (A+C)/n is the proportion of the patients at risk who later had the disease, (A+B) is the total of people who had the disease. Say there is no significant result in this case. Thus the proportion of patients at risk to have the disease is A / (A+B) will have the same proportion as those who were not at risk to have the disease is C / (C+D).

$$\frac{A_i}{(A_i+B_i)}=\frac{C_i}{(C_i+D_i)}=\frac{(A_i+C_i)}{n_i}$$

E estimates the number of success in treatment patients, while $(O-E)$ is the difference between the observed and expected values. V_i is the variance within-study of $(O_i - E_i)$.

$$V_i=\frac{(A_i+C_i)(B_i+D_i)(A_i+B_i)(C_i+D_i)}{n_i^2(n_i-1)}$$

The value of $\sum_{i=1}^{k}(O_i-E_i)$ is deducted by 0.5 to obtain a more accurate χ^2 approximation. If we reject null hypothesis, then we can say that the value $\sum_{i=1}^{k}(O_i-E_i)$ is relatively large.

Through this example, estimated $\chi^2 = 340.8858 > \chi^2_{(0.05,16)} = 26.30$, the result is statistically significant and thus we reject null hypothesis.

As a conclusion, use vote counting method if the effect size would not be able to be calculated. Vote-counting method has its drawback as it yields no significant effect if a low power in the statistical tests (Hedges and Olkin, 1980). However, if the effect sizes exist in the studies, we may combine estimates of effect sizes either in the form of fixed effects model or random effects model which depends on the heterogeneity testing of effect sizes throughout the studies.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

CHAPTER 4
CASE STUDY

4.1 Introduction

Breast cancer is the most common cancer especially in women throughout the world. It is the fifth most common cause of cancer death worldwide. In Malaysia, breast cancer is the leading cause of death among women with cancer.

Early breast cancer mainly occurs in the breast for node-positive patients as well as the local lymph nodes. Perhaps it will spread outside the breast or armpit area. It can hardly be detected and has no symptoms either. For early breast cancer cases, the tumour basically is not larger than 5 cm (2 inches) in size.

The risk factors are normally age, gender (Giordano *et al.*, 2004) and no breastfeeding (Collaborative Group on Hormonal Factors in Breast Cancer, 2002). On top of these factors, race, heredity, alcohol intake, and environmental factors would be the causes for breast cancer as well. In view of the risk of breast cancer, it is an imperative issue in getting early diagnosis (e.g. screening mammogram or ultrasound) and thus more likely for a cancer patient to obtain better treatment in advance.

Treatment for early breast cancer such as radiotherapy, surgery, chemotherapy, and hormone therapy remove the breast cancer or the cancer cells located at undetectable parts of the body. For breast surgery, it can be categorised into breast conserving surgery and mastectomy. Upon surgical treatment, radiotherapy is always the following treatment for the early breast cancer patients. In

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

case the cancer cells spread outside the breast or armpit for early breast cancer, chemotherapy may be taken into consideration. Hormone therapy is applied for women with oestrogen receptor positive (ER+) where the drug (e.g. Tamoxifen) is given to reduce the rate of breast cancer recurrence.

Adjuvant tamoxifen is an effective anti-oestrogen therapy for early and ER+ breast cancer. It slows down the growth of contralateral (in the opposite breast) cancer. However, tamoxifen may cause several side effects like substantial hair drop, fatigue, depression, dry skin and the like.

There are randomised trials confined to adjuvant tamoxifen (treatment) and no adjuvant tamoxifen (control) among women with early breast cancer on recurrence and death cases, respectively. Unlike previously, the effectiveness of adjuvant tamoxifen is observed in two ways: recurrence cases, and combined recurrence and mortality cases.

Several randomised trials for 1 year, 2 year and a median of 5 years of patients who consumed tamoxifen were investigated. Throughout the test, Early Breast Cancer Trialists' Collaborative Group (EBCTCG, 1998) has identified that adjuvant tamoxifen help in reducing the risk of recurrence and mortality.

In this case study, our objective is to determine how long the provision of adjuvant tamoxifen for women with early breast cancer is.

4.2 Methodology

EBCTCG has implemented meta-analysis across the randomised trials of various treatment of early breast cancer 5 years alternately since 1985. The data collection was completed in 1995 with a total of 37000 women in 55 randomised trials which were carried out before 1990. The trials were categorized into three

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

duration levels: approximate 1 year, 2 years and a median of 5 years of tamoxifen. Based on previous studies, a significant result shows the rise of 10-year survival of patients who consumed of tamoxifen.

For every woman, data was randomly collected at any range of age. It is necessary to determine whether oestrogen receptor (ER) or progesterone receptor (PR) is present in the primary tumour. Then we observe the breast cancer cells to see whether they spread to other local lymph nodes (either node-positive or node-negative). A follow-up from 1985 to 1995 is recorded which comprises the date of randomization, the given treatment, test of any contralateral breast cancer, recurrence, as well as mortality. Information relative to the cause of death is available for women of local recurrence but not the distance recurrence cases. Data was gathered, compiled and a meeting was held to discuss among the researchers and members of EBCTCG before analysing the data.

Based on the information given by EBCTCG (1998), women in node-positive category, approximately 85% with surgically assured nodal involvement, 10% with unknown nodal status, 5% of nodal involvement without a standard measurement, and less than 1% merely clinical evidence of involvement. On the other hand, 80% node-negative women had no nodal involvement upon axillary clearance (the entire lymph nodes below the arm are removed if the malignant cells exist in a lymph gland), 12% of negative axillary sampling (patients who were node negative upon axillary sample), 7% with no nodal involvement based on undetermined measurement, and 2% merely clinical evidence without involvement.

ER status can be separated into 3 categories: (1) ER-positive (at least l0fmol ER per mg cytosol protein), (2) ER-poor means the ER status was given, and (3) ER unknown is when the ER status is not mentioned. If ER status is not reported, then

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA

DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID

PR status also is not included as well (and not vice versa). Sometimes, ER-positive can also have two subgroups: ER++ ($\geq$ 100 fmol per mg) and ER+ ($10 \leq ER \leq 99$ fmol per mg) for particular analyses.

There were more than 42000 women in total with 63 randomised controlled trials before 1990 (see Table 4.1). However, eight of the trials were not available for the meta-analysis. Hence, the entire study of tamoxifen treatment for early breast cancer patients had around 37000 women in 55 randomised trials, or about 87% of the population.

Table 4.1 Data collected for adjuvant tamoxifen and no adjuvant tamoxifen treatment via randomization before 1990

	Available		Not available	
Duration of treatment	Trials	Total women	Trials	Total women
$\approx$ 1 year	14	9128	1	100
$\approx$ 2 years	32	19212	4	1400
Median of 5 years	9	8349	3	4200
Total	55	36689	8	5700

(Source: EBCTCG, 1998)

Women from the same trial only will be used to compare with each other directly, and not from other different trials. The outcome which is combined from different trials produce an overall estimate of treatment effect across 55 trials. The true treatment effect in each trial is not similar to the weighted average effect size in each different trial.

In data analysis, there are both recurrence and mortality using odds ratio to measure the effect sizes. If we obtain the odds ratio for death is 0.80, then it would be equivalent to 20% reduction in the death rate.

Recurrence also means reoccur or relapse of cancer detected either in local, distant, or contralateral breast. Mortality or death from breast cancer would take into

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

account other unspecified causes as long as breast cancer is the main factor. Those departed who did not report any recurrence and causes of death will be considered there was a relapse of cancer prior to their death.

The following secondary data set was obtained from the primary study of EBCTCG (1998). Table 4.2 illustrates 55 trials with a follow-up of 10-year studies for early breast cancer patients in recurrence and mortality events. These trials were divided into different duration of treatment studies: (a) around a year, (b) around 2 years, and (c) a median of 5 years, respectively. Every event was subdivided into two groups: treatment (Tamoxifen) and control (non-Tamoxifen) groups. Each study provides a total number of sample size shown in parenthesis for both events. For instance, the study "Case Western A" from (a) shows 55 recurrences and 52 deaths from a total of 97 patients in the treatment group whereas 67 recurrences and 69 deaths from a total of 101 patients in the control group.

Table 4.2 Data set of 55 randomised trials subdivided by duration of allocated treatment for both recurrence and mortality cases

(a) Treated for around a year

	Recurrence				Mortality (any cause)			
	Events (Total)				Events (Total)			
Study	Tamoxifen		Control		Tamoxifen		Control	
Case Western A	55	(97)	67	(101)	52	(97)	69	(101)
Christie B post	137	(282)	187	(306)	160	(282)	188	(306)
Danish BCG 77c post	505	(927)	590	(915)	693	(927)	720	(915)
UK MCCG 009	62	(123)	74	(140)	41	(123)	58	(140)
S Swedish BCG post	99	(239)	118	(236)	121	(239)	131	(236)
Innbruck *ER*+	50	(130)	49	(107)	30	(130)	48	(107)
NCCTG/Mayo Clinic	185	(311)	200	(319)	169	(311)	176	(319)
ECOG 5177/6177	186	(303)	187	(307)	161	(303)	168	(307)
SWOG7827ApoER+	148	(325)	178	(325)	141	(325)	156	(325)
Vienna Gyn. Post	25	(79)	38	(86)	26	(79)	33	(86)
DanishBCG 82b pre	223	(344)	224	(350)	193	(344)	200	(350)
ACETBC-1	183	(937)	185	(936)	166	(937)	166	(936)
Kawasaki 2	2	(12)	0	(8)	1	(12)	0	(8)
PetrovStPetersb'g	129	(434)	159	(449)	112	(434)	135	(449)
Subtotal	1989	(4543)	2256	(4585)	2066	(4543)	2248	(4585)

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Table 4.2 (continued)

(b) Treated for around 2 years

	Recurrence				Mortality (any cause)			
	Events	(Total)			Events	(Total)		
Study	Tamoxifen		Control		Tamoxifen		Control	
Copenhagen	60	(168)	74	(154)	91	(168)	102	(154)
Stockholm B post	116	(273)	126	(269)	136	(273)	138	(269)
NATO	219	(563)	269	(568)	249	(563)	292	(568)
NSABP B-09	596	(950)	625	(941)	561	(950)	586	(941)
Toronto-Edmo.post	128	(199)	154	(201)	129	(199)	141	(201)
GUN Naples	89	(206)	123	(227)	79	(206)	93	(227)
UK/Asia Collab.	103	(240)	132	(248)	89	(240)	102	(248)
ECOG EST1178 post	43	(91)	65	(90)	66	(91)	78	(90)
NKCC Japan	95	(648)	105	(587)	85	(648)	95	(587)
Ghent Univ. N-post	17	(73)	20	(65)	16	(73)	13	(65)
GABG/HD Germany	134	(308)	149	(306)	118	(308)	129	(306)
Helsinki	42	(81)	44	(78)	43	(81)	43	(78)
N Sweden BCG	69	(183)	89	(185)	90	(183)	101	(185)
CRC2,UK	375	(947)	469	(965)	376	(947)	427	(965)
Toulouse	57	(125)	55	(126)	60	(125)	54	(126)
GABG2Germanypost	29	(285)	36	(293)	12	(285)	15	(293)
Monpellier post	35	(101)	43	(102)	42	(101)	51	(102)
IBBordeaux N+ER+	75	(166)	76	(160)	54	(166)	60	(160)
NBCG 1a ER+	80	(190)	106	(179)	66	(190)	72	(179)
ACETBC-1	407	(1717)	512	(1719)	345	(1717)	413	(1719)
Kumamoto	9	(89)	14	(95)	8	(89)	7	(95)
Amsterdam C8209	205	(850)	**116	(410)	199	(850)	**108	(410)
Oita	11	(62)	12	(63)	13	(62)	7	(63)
GBSG 02 Germany	83	(187)	97	(194)	61	(187)	72	(194)
Austrian BCSG 4	25	(74)	35	(79)	24	(74)	30	(79)
SE Sweden BCG	20	(51)	28	(63)	14	(51)	18	(63)
S Swedish BCG	75	(213)	86	(215)	63	(213)	64	(215)
ACETBC-2	145	(461)	144	(444)	118	(461)	98	(444)
Osaka Japan	15	(122)	17	(129)	8	(122)	6	(129)
Wilconsin U.	6	(69)	7	(70)	2	(69)	4	(70)
Oita2	9	(136)	17	(138)	7	(136)	9	(138)
CRCRAMS Moscow	0	(10)	2	(11)	0	(10)	2	(11)
Subtotal	3372	(9838)	3963	(9784)	3224	(9838)	3538	(9784)

** Since the treatment for Amsterdam study is randomised evenly between 1 year and 3 years in control group, as a result it had been carried out twice for control patients

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI

Table 4.2 (continued)

(c) Treated for a median of 5 years

	Recurrence		Mortality (any cause)	
	Events (Total)		Events (Total)	
Study	Tamoxifen	Control	Tamoxifen	Control
Stockholm B post	275 (1104)	385 (1096)	263 (1104)	300 (1096)
Scottish	264 (666)	344 (657)	305 (666)	354 (657)
CRFB Caen C5 post	40 (89)	60 (90)	57 (89)	64 (90)
Marseille post	20 (37)	13 (34)	15 (37)	8 (34)
NSABPB-14 N-ER+	264 (1439)	456 (1453)	248 (1439)	278 (1453)
Osaka	15 (53)	18 (55)	10 (53)	12 (55)
GROCTA I Italy ER+	75 (171)	103 (168)	57 (171)	86 (168)
FASG GFEA 02	97 (375)	162 (368)	68 (375)	104 (368)
CRFB Caen 002	33 (250)	52 (244)	33 (250)	38 (244)
Subtotal	1083 (4184)	1593 (4165)	1056 (4184)	1244 (4165)

Total of cancer patients (a), (b), and (c):

$$4543 + 4585 + 9838 + 9784 + 4184 + 4165 - 410 \text{ (repeated)} = 36689$$

Total of recurrences (a), (b), and (c):

$$1989 + 2256 + 3372 + 3963 + 1083 + 1593 - 116 \text{ (repeated)} = 14140$$

Total of mortality (a), (b), and (c):

$$2066 + 2248 + 3224 + 3538 + 1056 + 1244 - 108 \text{ (repeated)} = 13268$$

According to EBCTCG (1998), a follow-up of 10-year study for 36689 patients yields 14140 recurrences and 13268 deaths on average. From 55 trials, 88% (32422) of the women were in contralateral breast cancer whilst 6% (839) of the recurrences discovered a new cancer in the opposite breast. On the other hand, 90% (32947) of the randomised trials identified deaths from breast cancer together with other causes whereas 14% of the deaths were caused by other causes (not from breast cancer) without recording any recurrence, we so-called non-breast-cancer mortality.

Statistical software which is specifically designed by recognized experts in the fields of meta-analysis, namely, Comprehensive Meta-Analysis (CMA) version 2.0 (Borenstein *et al.*, 2005) is a powerful program which assists to ease unnecessary

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID

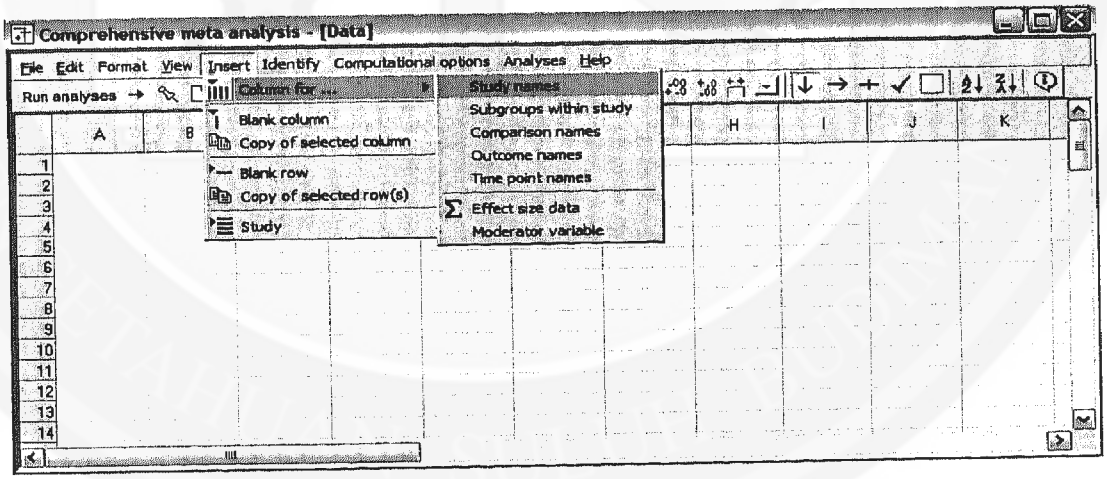
problems and difficulty in executing the data. It provides sophisticated graphics such as forest plot which can be customised based on researchers' requirement to get a better understanding on the results and this could be informative for the data interpretation.

Most importantly, CMA is a menu-driven program which can be user friendly to the beginner in meta-analysis research. The process is generally fast and accurate without any written commands needed. Despite Excel possesses menu-driven program as well, yet we ought to know all the effect size estimates formulas in order to test the analysis. Furthermore, Excel does not provide any forest plot in the output. Although Metawin is also a meta-analysis program, however, Metawin only will accept several data formats and provide a very basic forest plot whereas CMA accepts more than 100 formats of data.

The procedures of the meta-analysis method are as follows:

(1) Create a column for study names (55 studies altogether)

Select Insert → Column for → Study names

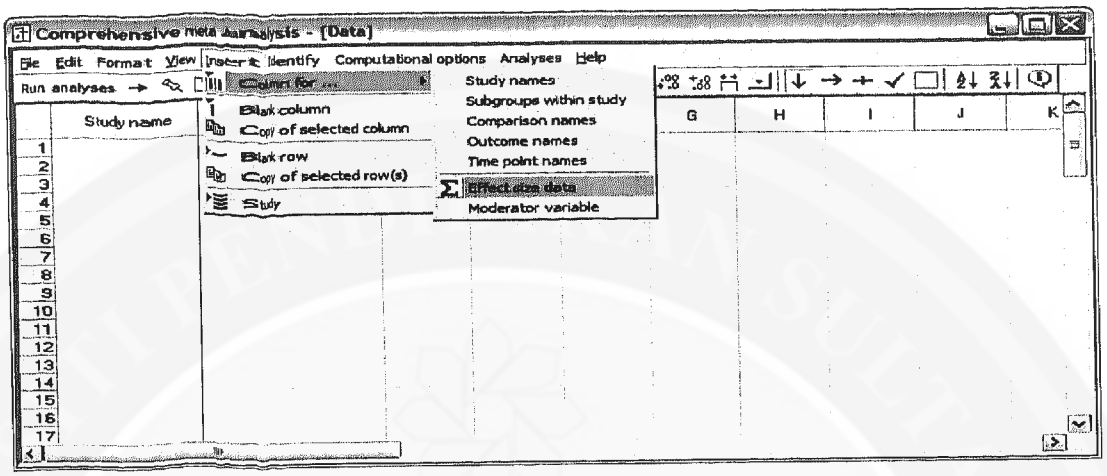


UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

(2) Create columns for effect size estimates

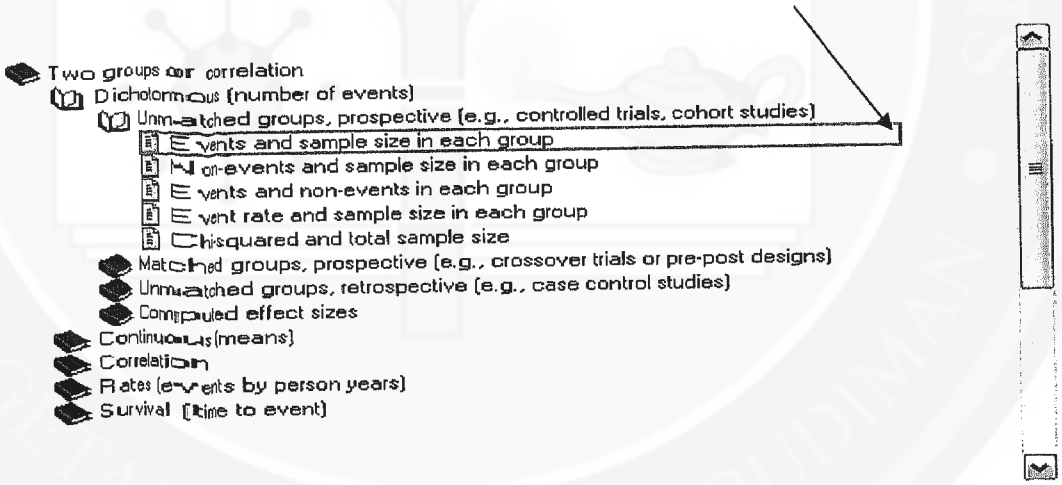
Select Insert → Column for → Effect size data.



A screen shows the types of studies included in meta-analysis, select "comparison of two groups" since we have treatment and control groups for the data set.

(3) Choose an appropriate data format

Our data comprise the number of events and total sample size in each group.

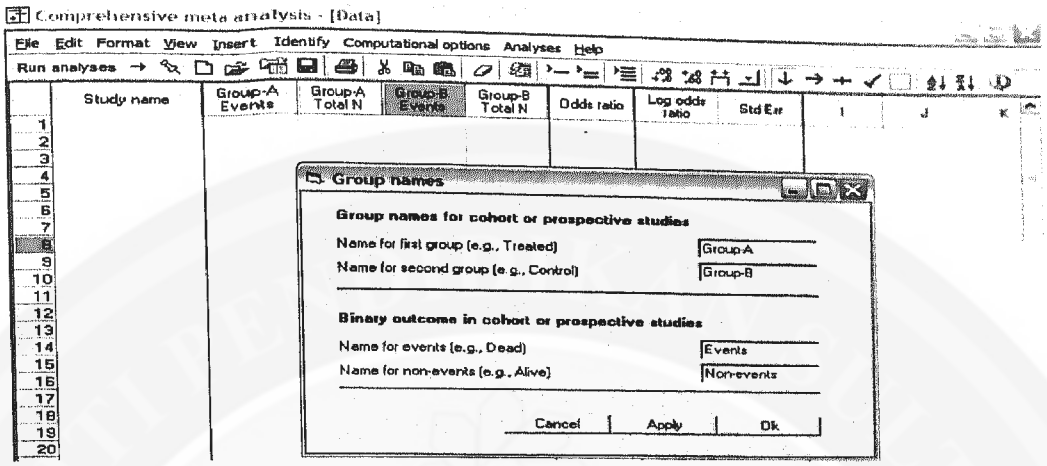


UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

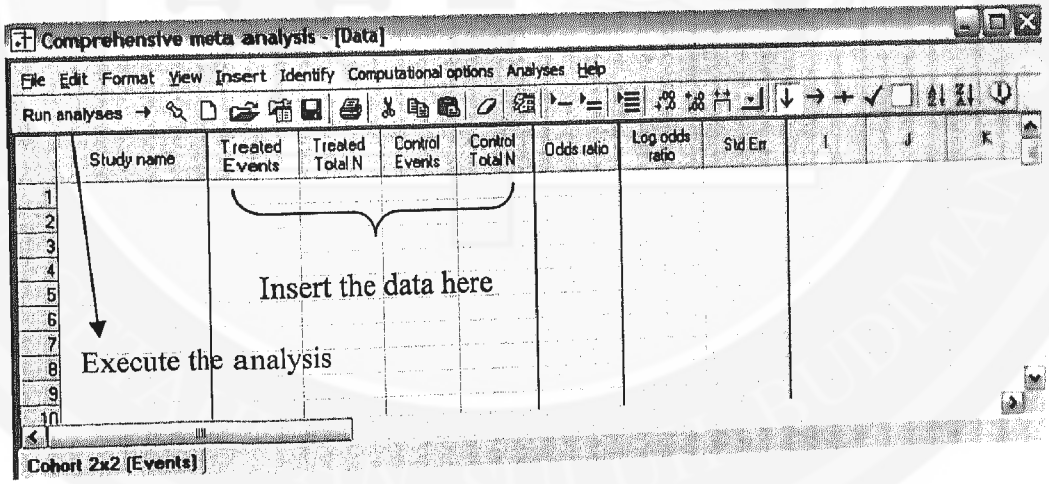
(4) Modify the names for groups columns (binary data)

Type Tamoxifen (first group) →Control (second group) then click "ok".



(5) Key-in effect size data

Finally the data entry is ready to be used, the coloured columns will show the figures automatically if all four values have been inserted in each study. We can check on the computational formula for the effect size estimates (odds ratio in our case study) via double-click on "Odds ratio". Once the data entry is completed, click on "Run analyses" to execute the analysis.



4.3 Results and Discussion

All the results are calculated by using fixed effects method since it is more powerful than random effects method which takes the overall null hypothesis into

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

consideration. Fixed effects method is useful for finding the treatment effects that diverge within-study since it does not consider for between-study variability. We set the data at 5% significance level throughout the studies to obtain the 95% confidence interval (CI) for the effect size estimate (odds ratio). The forest plots are shown in the following output ranging from 0.10 to 10.00 with odds ratio and 95% of CI.

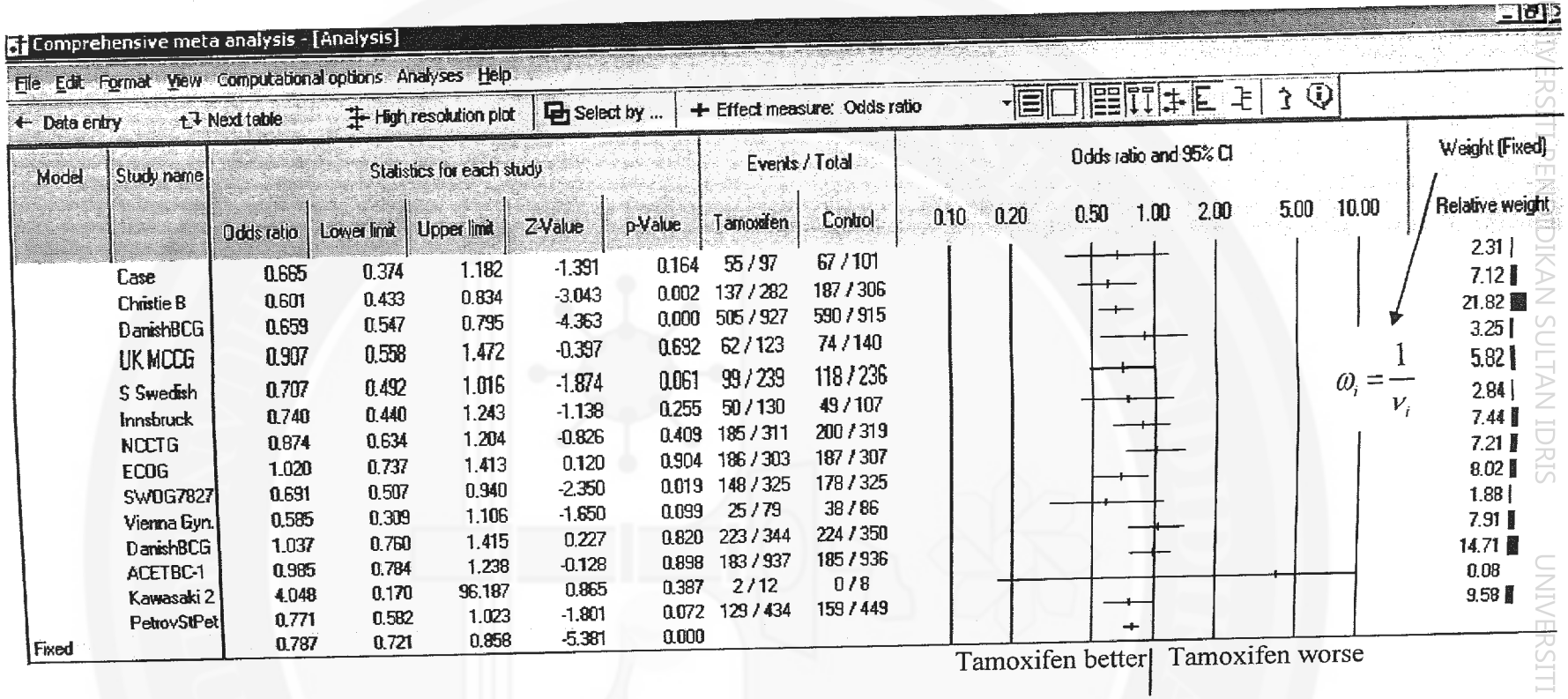
Regarding recurrence events (Figures 4.1-4.3), odds ratio for about 1 year, 2 years, and median 5 years were 0.787, 0.748, and 0.548, all were statistically significant ($2p < 0.00001$) based on fixed effects model. As the duration of treatment increases, the recurrence reduction is increased which indicates Tamoxifen is effective in treating recurrences.

CI for recurrence around 1 year is (0.721, 0.858), around 2 years is (0.704, 0.795), and around 5 years is (0.498, 0.603). All the CIs do not include the value zero which means all the CIs for odds ratio in recurrence are statistically significant.

For heterogeneity testing, Q statistics will be used generally. Here, we have a new measure, I^2 (0%-100%) with higher values indicating more significant in heterogeneity. When studies are compared within the same durations respectively, all were homogeneous ($p > 0.05$). If recurrences are compared between 55 trials (Figure 4.4), all were heterogeneous ($I^2 = 43.952$). If recurrences are compared between three subtotals (Figure 4.5), the heterogeneity is highly significant ($p < 0.00001$, $I^2 = 94.65$). The heterogeneity is due to the test for different durations of treatment.

The last column of the output (Figures 4.1-4.3 and Figures 4.6-4.8) indicates the relative weight based on fixed effects method. We can conclude that the bigger the sample size, the greater the weight of evidence. The purpose of estimating the eight, $\omega_i = \frac{1}{v_i}$, is to minimize the variance so that the most accurate effect size estimate can be obtained.

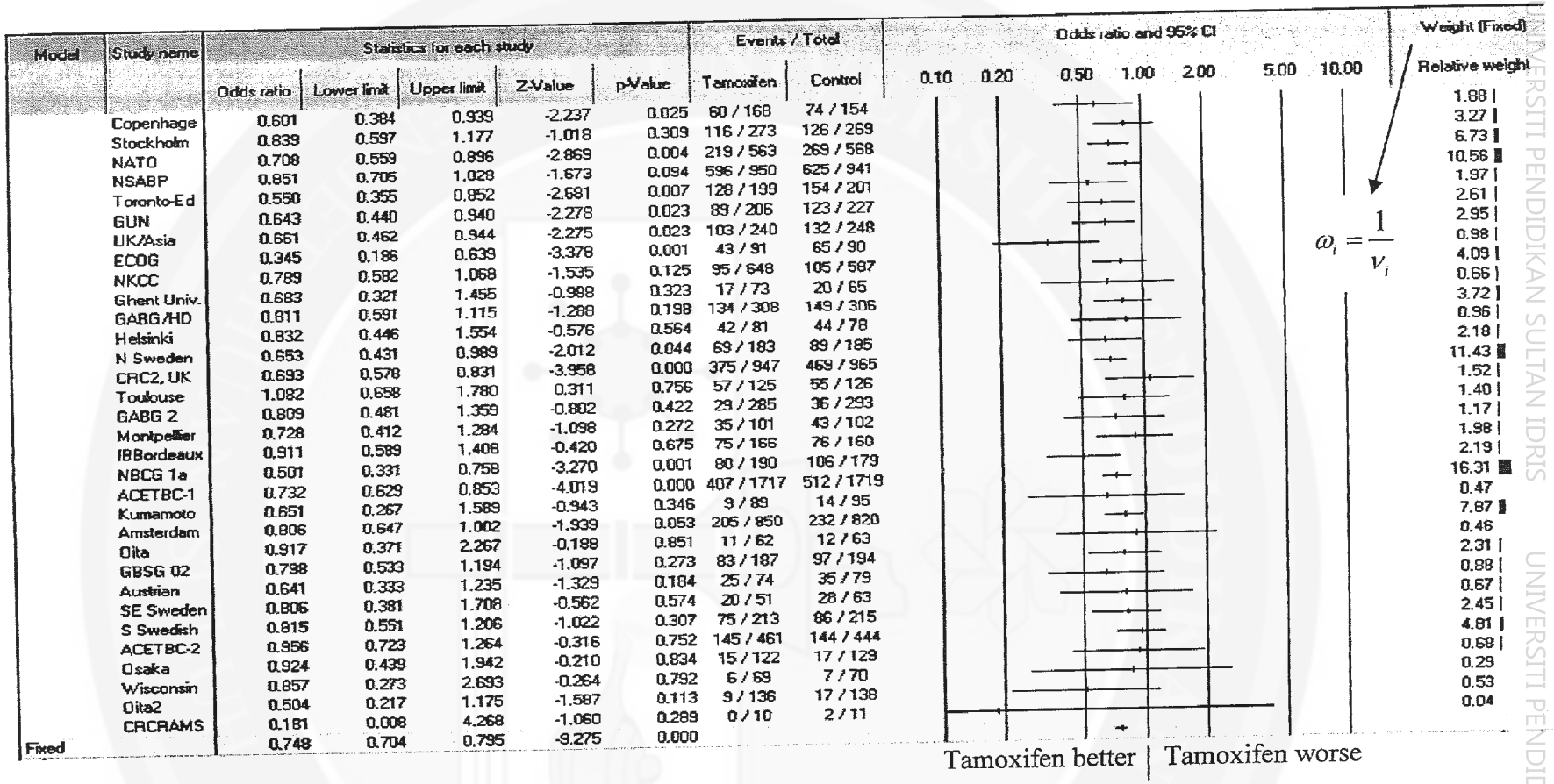
UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN



Model	Number Studies	Effect size and 95% interval: Point estimate	Lower limit	Upper limit	Test of null (2-Tail): Z-value	P-value	Heterogeneity: Q-value	df (Q)	P-value	I-squared
Fixed	14	0.787	0.721	0.858	-5.381	0.000	19.250	13	0.116	32.466
Random	14	0.791	0.706	0.886	-4.046	0.000				

Figure 4.1 Recurrences for adjuvant tamoxifen versus no adjuvant tamoxifen treatment about 1 year

UNIVERSITI PENDIDIKAN SULTAN IDRIS

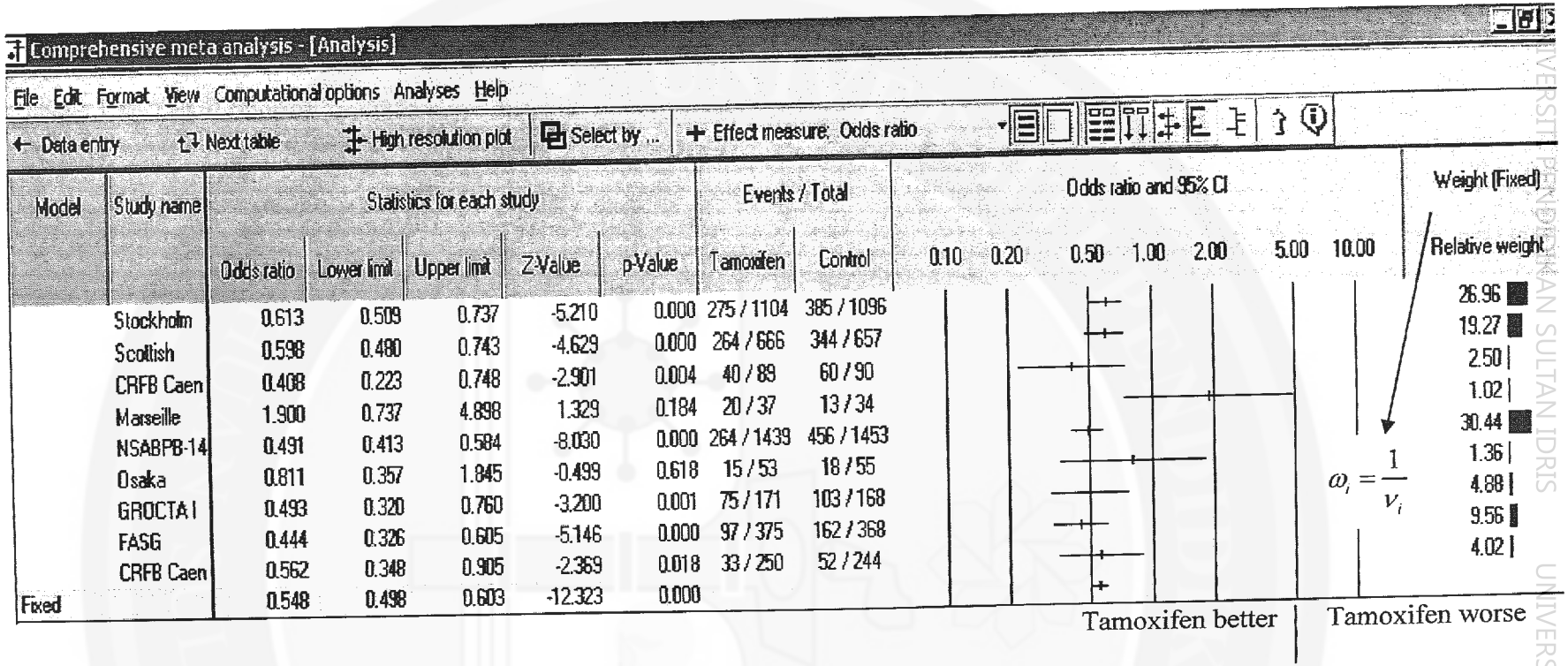


Model	Study name	Statistics for each study					Events / Total		Weight (Fixed)
		Odds ratio	Lower limit	Upper limit	Z-Value	p-Value	Tamoxifen	Control	Relative weight
	Copenhage	0.601	0.384	0.939	-2.237	0.025	60 / 168	74 / 154	1.88
	Stockholm	0.839	0.597	1.177	-1.018	0.309	116 / 273	126 / 269	3.27
	NATO	0.708	0.559	0.896	-2.869	0.004	219 / 563	269 / 568	6.73
	NSABP	0.851	0.705	1.028	-1.673	0.094	596 / 950	625 / 941	10.56
	Toronto-Ed	0.550	0.355	0.852	-2.681	0.007	128 / 199	154 / 201	1.97
	GUN	0.643	0.440	0.940	-2.278	0.023	89 / 206	123 / 227	2.61
	UK/Asia	0.661	0.462	0.944	-2.275	0.023	103 / 240	132 / 248	2.95
	ECOG	0.345	0.186	0.639	-3.378	0.001	43 / 91	65 / 90	0.98
	NKCC	0.789	0.582	1.068	-1.535	0.125	95 / 648	105 / 587	4.09
	Ghent Univ.	0.683	0.321	1.455	-0.988	0.323	17 / 73	20 / 65	0.66
	GABG/HD	0.811	0.591	1.115	-1.288	0.198	134 / 308	149 / 306	3.72
	Helsinki	0.832	0.446	1.554	-0.576	0.564	42 / 81	44 / 78	0.96
	N Sweden	0.653	0.431	0.989	-2.012	0.044	69 / 183	89 / 185	2.18
	CRC2, UK	0.693	0.578	0.831	-3.958	0.000	375 / 947	469 / 965	11.43
	Toulouse	1.082	0.658	1.780	0.311	0.756	57 / 125	55 / 126	1.52
	GABG 2	0.809	0.481	1.359	-0.802	0.422	29 / 285	36 / 293	1.40
	Montpellier	0.728	0.412	1.284	-1.098	0.272	35 / 101	43 / 102	1.17
	IBBordeaux	0.911	0.589	1.408	-0.420	0.675	75 / 166	76 / 160	1.98
	NBCG 1a	0.501	0.331	0.758	-3.270	0.001	80 / 190	106 / 179	2.19
	ACETBC-1	0.732	0.629	0.853	-4.019	0.000	407 / 1717	512 / 1719	16.31
	Kumamoto	0.651	0.267	1.589	-0.943	0.346	9 / 89	14 / 95	0.47
	Amsterdam	0.806	0.647	1.002	-1.939	0.053	205 / 850	232 / 820	7.87
	Oita	0.917	0.371	2.267	-0.188	0.851	11 / 62	12 / 63	0.46
	GBSG 02	0.798	0.533	1.194	-1.097	0.273	83 / 187	97 / 194	2.31
	Austrian	0.641	0.333	1.235	-1.329	0.184	25 / 74	35 / 79	0.88
	SE Sweden	0.806	0.381	1.708	-0.562	0.574	20 / 51	28 / 63	0.67
	S Swedish	0.815	0.551	1.206	-1.022	0.307	75 / 213	86 / 215	2.45
	ACETBC-2	0.956	0.723	1.264	-0.316	0.752	145 / 461	144 / 444	4.81
	Osaka	0.924	0.439	1.942	-0.210	0.834	15 / 122	17 / 129	0.68
	Wisconsin	0.857	0.273	2.693	-0.264	0.792	6 / 69	7 / 70	0.29
	Oita2	0.504	0.217	1.175	-1.587	0.113	9 / 136	17 / 138	0.53
	CRCRAMS	0.181	0.008	4.268	-1.060	0.289	0 / 10	2 / 11	0.04
Fixed		0.748	0.704	0.795	-9.275	0.000			

Model		Effect size and 95% interval			Test of null (2-Tail)		Heterogeneity			
Model	Number Studies	Point estimate	Lower limit	Upper limit	Z-value	P-value	Q-value	df (Q)	P-value	I-squared
Fixed	32	0.748	0.704	0.795	-9.275	0.000	26.872	31	0.679	0.000
Random	32	0.748	0.704	0.795	-9.275	0.000				

Figure 4.2 Recurrences for adjuvant tamoxifen versus no adjuvant tamoxifen treatment about 2 years

UNIVERSITI PENDIDIKAN SULTAN IDRIS



Model		Effect size and 95% interval			Test of null (2-Tail)		Heterogeneity			
Model	Number Studies	Point estimate	Lower limit	Upper limit	Z-value	P-value	Q-value	df (Q)	P-value	I-squared
Fixed	9	0.548	0.498	0.603	-12.323	0.000	13.976	8	0.082	42.759
Random	9	0.551	0.476	0.638	-7.931	0.000				

Figure 4.3 Recurrences for adjuvant tamoxifen versus no adjuvant tamoxifen treatment about a median of 5 years

Model		Effect size and 95% interval			Test of null (2-Tail)		Heterogeneity			
Model	Number Studies	Point estimate	Lower limit	Upper limit	Z-value	P-value	Q-value	df (Q)	P-value	I-squared
Fixed	55	0.709	0.678	0.741	-15.185	0.000	96.345	54	0.000	43.952
Random	55	0.714	0.668	0.764	-9.881	0.000				

Figure 4.4 Recurrences for adjuvant tamoxifen versus no adjuvant tamoxifen treatment within 55 trials

Model		Effect size and 95% interval			Test of null (2-Tail)		Heterogeneity			
Model	Number Studies	Point estimate	Lower limit	Upper limit	Z-value	P-value	Q-value	df (Q)	P-value	I-squared
Fixed	3	0.728	0.698	0.760	-14.691	0.000	37.381	2	0.000	94.650
Random	3	0.704	0.581	0.853	-3.579	0.000				

Figure 4.5 Recurrences for adjuvant tamoxifen versus no adjuvant tamoxifen treatment among 3 subtotals (1,2, and median 5 years)

UNIVERSITI PENDIDIKAN SULTAN IDRIS

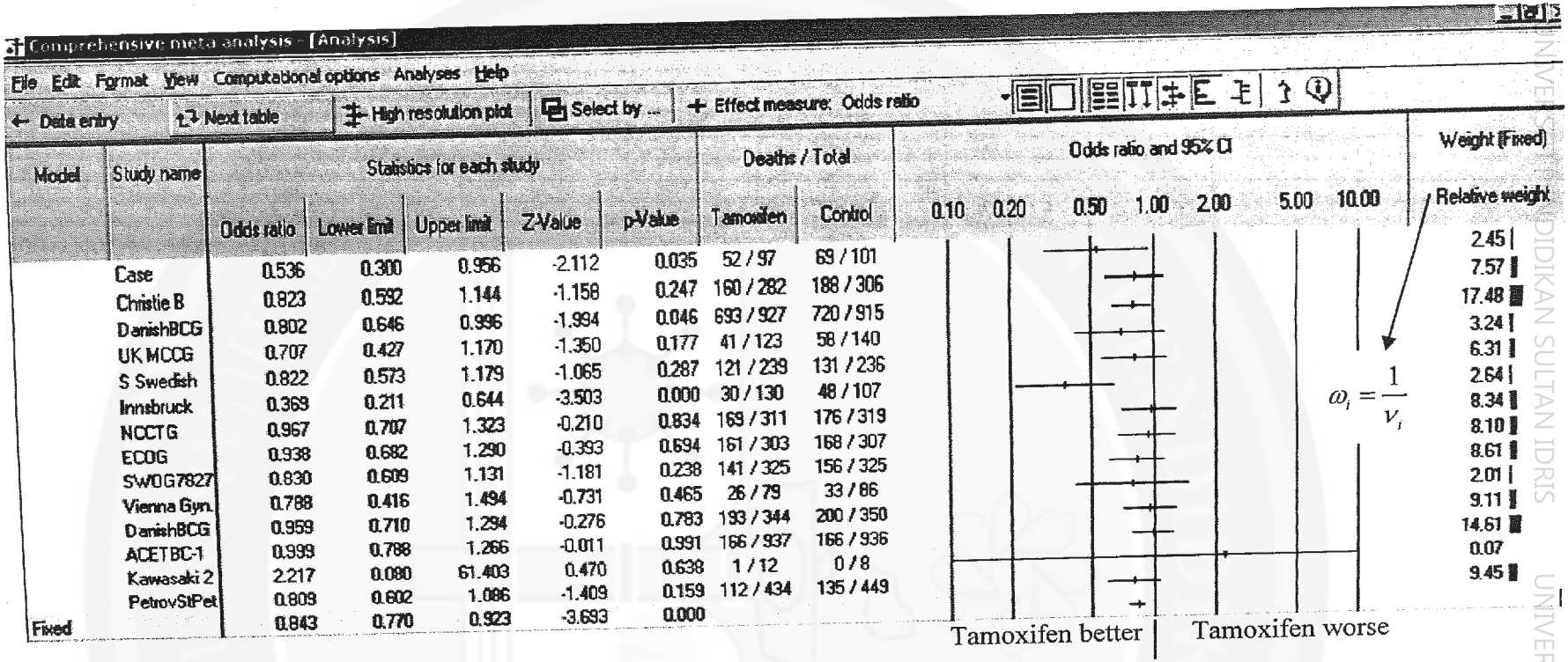
UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID

For mortality events (Figures 4.6-4.8), odds ratio for about 1 year, 2 years, and median 5 years were 0.843, 0.845, and 0.774, all were highly statistically significant ($2p < 0.00001$) based on fixed effects model. Thus, we can conclude that the longer the adjuvant tamoxifen treatment, the greater mortality reduction it would be.

CI for mortality around 1 year is (0.770, 0.923), around 2 years is (0.793, 0.900), and around 5 years is (0.700, 0.856). All the CIs do not include the value zero which means all the CIs for odds ratio in mortality are statistically significant.

For heterogeneity testing, when studies are compared within the same durations respectively, there was no significant heterogeneity ($p > 0.05$). If recurrences are compared between 55 trials (Figure 4.9), there was no significant heterogeneity (p-value = 0.384, $I^2 = 4.360$). If deaths are compared between three subtotals (Figure 4.10), the heterogeneity is not significant ($p = 0.300$, $I^2 = 16.828$). Hence, for mortality cases, all the randomized trials were homogeneous to death, this heterogeneity test is considered to have great power.

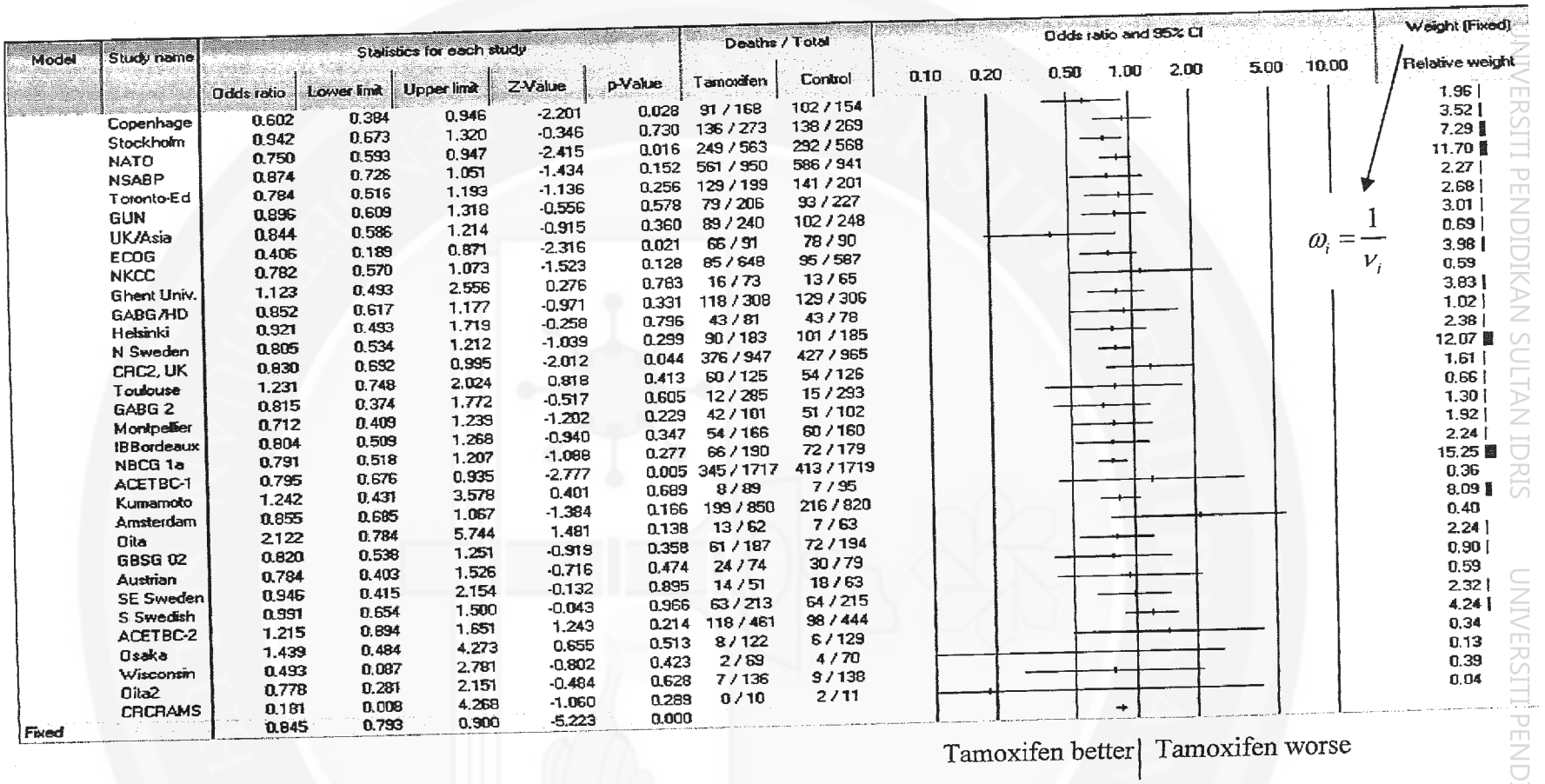
UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS



Model	Effect size and 95% interval			Test of null (2-Tail)		Heterogeneity			
Model	Number Studies	Point estimate	Lower limit	Upper limit	Z-value	P-value	Q-value	df (Q)	P-value	I-squared
Fixed	14	0.843	0.770	0.923	-3.693	0.000	15.778	13	0.261	17.609
Random	14	0.836	0.754	0.927	-3.396	0.001				

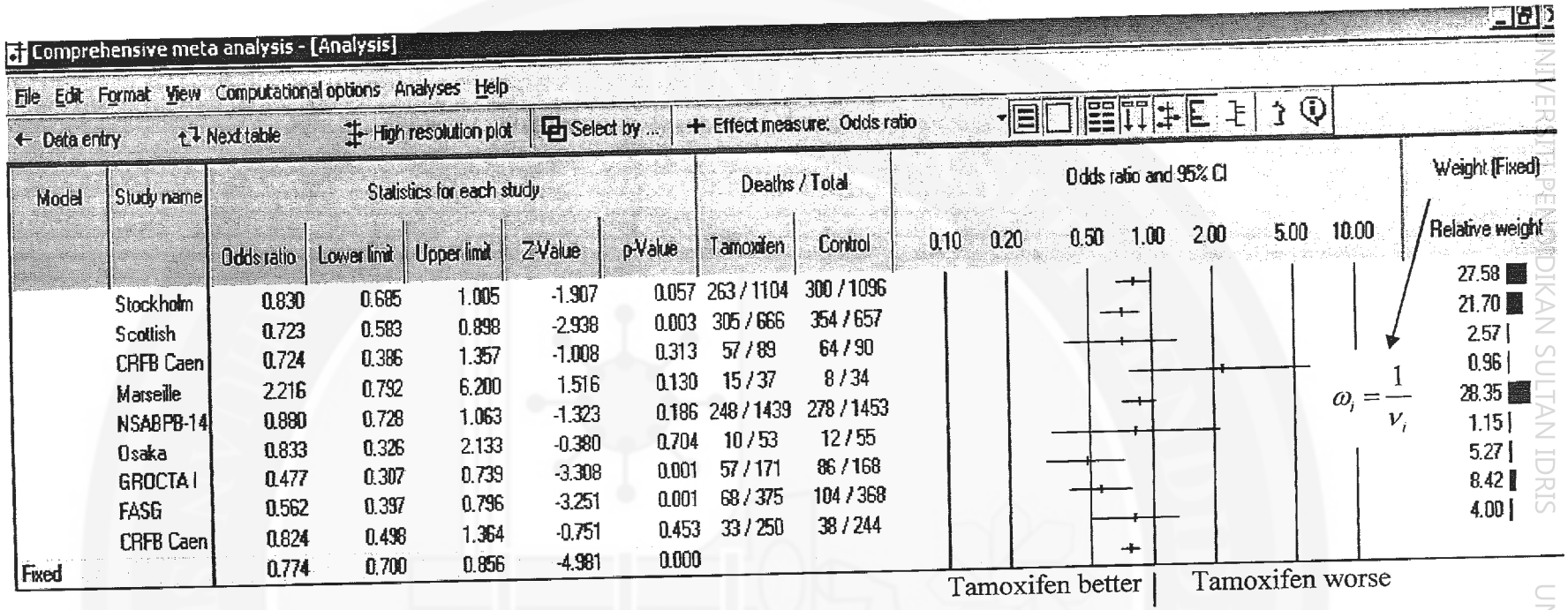
Figure 4.6 Deaths for adjuvant tamoxifen versus no adjuvant tamoxifen treatment about 1 year

UNIVERSITI PENDIDIKAN SULTAN IDRIS



Model		Effect size and 95% interval			Test of null (2-Tail)		Heterogeneity			
Model	Number Studies	Point estimate	Lower limit	Upper limit	Z-value	P-value	Q-value	df (Q)	P-value	I-squared
Fixed	32	0.845	0.793	0.900	-5.223	0.000	23.683	31	0.823	0.000
Random	32	0.845	0.793	0.900	-5.223	0.000				

Figure 4.7 Deaths for adjuvant tamoxifen versus no adjuvant tamoxifen treatment about 2 years



Model		Effect size and 95% interval			Test of null (2-Tail)		Heterogeneity			
Model	Number Studies	Point estimate	Lower limit	Upper limit	Z-value	P-value	Q-value	df (Q)	P-value	I-squared
Fixed	9	0.774	0.700	0.856	-4.981	0.000	14.738	8	0.064	45.718
Random	9	0.752	0.641	0.882	-3.490	0.000				

Figure 4.8 Deaths for adjuvant tamoxifen versus no adjuvant tamoxifen treatment about median of 5 years

UNIVERSITI PENDIDIKAN SULTAN IDRIS

Model		Effect size and 95% interval			Test of null (2-Tail)		Heterogeneity			
Model	Number Studies	Point estimate	Lower limit	Upper limit	Z-value	P-value	Q-value	df (Q)	P-value	I-squared
Fixed	55	0.829	0.792	0.868	-7.967	0.000	56.462	54	0.383	4.360
Random	55	0.829	0.790	0.870	-7.635	0.000				

Figure 4.9 Deaths for adjuvant tamoxifen versus no adjuvant tamoxifen treatment within 55 trials

Model		Effect size and 95% interval			Test of null (2-Tail)		Heterogeneity			
Model	Number Studies	Point estimate	Lower limit	Upper limit	Z-value	P-value	Q-value	df (Q)	P-value	I-squared
Fixed	3	0.848	0.813	0.886	-7.514	0.000	2.405	2	0.300	16.828
Random	3	0.847	0.807	0.889	-6.774	0.000				

Figure 4.10 Deaths for adjuvant tamoxifen versus no adjuvant tamoxifen treatment among 3 subtotals (1,2, and median 5 years)

UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

4.4 Conclusion

The randomised controlled trial is the most powerful tool in research as it reduces the selection bias (time interval or sampling error) in meta-analysis. A great collection of large samples is required since each treatment from different studies generally has relatively small effect size. As a result, meta-analysis would be a useful and proper approach to assess the effectiveness of the allocated treatment for certain cancer. The duration of treatment affects the risk reduction among the trials. Meta-analysis with a longer follow-up possesses stronger evidence for the response variable (duration of treatment) to show how effective the allocated treatment is. Despite both recurrence and the death yield highly statistically significant odds ratio for a 10-year follow up, proportional reduction of recurrence is greater than the mortality cases. This result proves that Tamoxifen treatment is much more effective for recurrence cases compared to death cases.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

CHAPTER 5
CONCLUSION

5.1 Summary

This chapter summarizes the dissertation by determining how meta-analysis approach has fulfilled the objective as stated in Chapter 1. Next, it discusses the limitation of the review and the future enhancement of the meta-analyses.

In this dissertation, the whole process of the research work was implemented using the menu-driven programs of Excel and Comprehensive Meta-Analysis V2. Of late, meta-analysis used Randomised Controlled Trials (RCT) in medical research. Investigators have explored the application of meta-analyses to test the effectiveness of the allocated treatments based on different kind of diseases in order to increase the statistical power and the accuracy of effect size estimates.

Meta-analysis is employed substantially in medical, psychological, social, education, and economics research. In meta-analysis, RCT and collaborative methods will be applied for medical, pre-test and post-test method for organizational intervention, cooperative learning method for educational studies, and meta-regression analysis for economics. Each of them has a model to evaluate the effect size estimates respectively.

There are two possibilities in finding the effect size estimate. Firstly, the possibility that no effect size is applicable (all effect size were not significant) throughout the studies, hence the vote-counting method will be applied in this aspect. Secondly, the possibility that all the effect sizes are applicable for meta-analysis,

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI F

then we combine effect sizes from different studies either by using fixed effects or random effects model.

5.2 Contribution

We have discovered the latest computer program specifically designed for meta-analysis with its applicability to a wide range of research of studies. Despite a series of statistical programs such as Stata, SPSS, SAS, Metawin or even Excel can be utilised for meta-analysis, however, they are not as powerful as Comprehensive Meta-Analysis (CMA) version 2.0 which is a user-friendly and an interactive tool for meta-analysis purposes.

Up to now, there has not been any research using CMA 2.0 in early breast cancer treatment such as adjuvant tamoxifen for oestrogen receptor positive (ER+) cancer patients or radiotherapy after breast-conserving surgery to investigate the odds ratio of recurrence and mortality. It computes the effect size automatically in a fast and accurate process without any written command (macros) as required in Stata, SPSS or SAS while Excel needs the researcher to calculate all the effect sizes manually which can be time-consuming. Although Metawin and CMA provide a forest plot in their output, Metawin has merely a very basic plot compared to CMA which shows more information instead. The essence of the CMA 2.0 is it can mix and match different formats of data (some studies provide events and sample size while some studies report in odds ratio and confidence interval) in the same meta-analysis, but it is not available in other software.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

5.3 Limitation

As we know meta-analysis requires great effort in accumulating a large collection of samples. Thus we face one of the major problems in getting the data set within a limited duration. In view of the validity of meta-analysis, we have to gather data with the same characteristics. Combined studies must be comparable and independent of each other. The outcome of meta-analysis would be futile if any of the selection bias is not stated or identified in advance. Hence it is vital to have a good knowledge in handling data selection. The procedures are complicated while time constraint is another issue. A good meta-analysis holds a longer follow-up on the treatment effect studies to produce significant trend in the results.

Most of the research papers did not publish the data especially if there were outputs showing no statistical significance. Furthermore, there is no information about which statistical software was utilised in the research. Thus it is unlikely we can review the papers comprehensively and efficiently.

5.4 Recommendation for future research

Meta-analysis is known as a powerful tool in statistical research to enhance the accuracy of data-analysis in independent trials. Future research ought to focus on providing a better quality study by getting larger sample sizes so that more accurate results of the treatment effect can be obtained.

The majority of meta-analysis research papers only publish the significant results rather than the not significant ones. Those "uninteresting" results (not significant) would not be published, and so lead to publication bias. In order to improve meta-analysis study, we should consider both published and unpublished research materials for the purpose of literature review.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Due to the efficiency of Comprehensive Meta-Analysis 2.0 in analyzing the meta-analysis, we should apply this program instead of written-command (macros) which can be very time-consuming and prone to erroneous results.

It is truly subjective to draw a conclusion as there is no single methodology which can guarantee the validity of meta-analysis. Since a wide range of primary studies are used nowadays, it is crucial to proceed to re-examine the results, get the systematic reviews up-to-date, use homogeneous population (age, diagnosis or cancer duration), and determine the choice between random effects and fixed effects model appropriately for the sake of validity of meta-analysis in future studies.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA

DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

BIBLIOGRAPHY

Armstrong, N., Chang, S. M, & Brickman, M. (2007). Cooperative Learning in Industrial-sized Biology Classes. *Life Sciences Education, 6,*163-171.

Arvanitis, T. N. (2004). An Introduction to Meta-Analysis. In Evangelia Chryssafidou, Khalid Khan & Arri Coomarasamy (Eds.), *E-health & learning*. United Kingdom: The University of Birmingham.

Baum, M. L., Anish, D. S., Chalmers, T. 1., Sacks, H. S., Smith, H., & Fagerstrom, R. M. (1981). A survey of clinical trials of antibiotic prophylaxis in colon surgery: evidence against further use of no-treatment controls. *New Engl J Med,* 505(14), 795-799.

Bayraktar, S. (2001-2002). A meta-analysis of the effectiveness of computer-assisted instruction in science education. *Journal of Research on Technology in Education, 34,* 173-188.

Becker, H. S. (1992). Foci and Bases of Commitment : Are they distinctions worth making? In C. C. R. H.S.Becker (Ed.), *What is a case? Exploring the foundations of social inquiry*. Chicago: University of Chicago Press.

Bernard, R. M., & Abrami, P. C. (2005). The Basics Of Effect Size Extraction And Statistical Applications For Meta-Analysis: Concordia University.

Bjelakovic, G., Nikolova, D., Gluud, L. L., Simonetti, R. G., & Gluud, C. (2008). Antioxidant supplements for prevention of mortality in healthy participants and patients with various diseases. *Cochrane Database of Systematic Reviews* (2).

Bolton, B., Bonge, D., & Marr, J. (1979). Ratings of instruction, examination performance, and subsequent enrolment in psychology courses. *Teaching of Psychology, 6,* 82-85.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive Meta-Analysis Version 2*. Englewood, NJ: Biostat.

Boruch, R. (2005). Better Evaluation for Evidence-Based Policy: Place Randomized Trials in Education, Criminology, Welfare, and Health. *American Academy of Political and Social Science, **599.***

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Bryson, R. (1974). Teacher evaluations and student learning: A reexamination. *Journal of Educational Research, 68,* 12-14.

Cavanaugh, C. (2001). The effectiveness of interactive distance education technologies in K-12 learning: A meta-analysis. *International Journal of Educational Telecommunications, 7*(1), 73-88.

Centra, J. A. (1977). Student ratings of instruction and their relationship to student learning. *American Educational Research Journal, 14,* 17-24.

Chalmers, I., & Altman, D. G. (1995). *Systematic Reviews.* London: BMJ Publications.

Chalmers, T. C, & Lau. J. (1993). Meta-analytic stimulus for changes in clinical trials. *Statistical Methods in Medical Research, 2,* 161 -172.

Chalmers, T. C, Matta, R. J., Smith, H., & Kunzler, A. M (1977). Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. *New EnglJMed,297,\09\-\096.*

Choi, J. K., Yu, U. S., Kim, S. S., & Yoo, O. J. (2003). Combining multiple microarray studies and modeling interstudy variation. *Bioinformatics, 19,* 84-90.

Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Journal of Royal Statistical Society Supplement, 4(1),* 102-118.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45,*1304-1312.

Cohen, J. (1998). *Statistical Power Analysis for the Behavioral Sciences.* Mahwah, N.J.: Lawrence Erlbaum.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research, 51,* 281-309.

Cohen, P. A. (1983). Comment on "A selective review of the validity of student ratings of teaching." *Journal of Higher Education, 54,* 449-458.

Collaborative Group on Hormonal Factors in Breast Cancer. (2002). Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50 302 women with breast cancer and 96 973 women without the disease. *The Lancet, 360,* 187-195.

Conover, W. J. (1980). *Practical nonparametric statistics* (2nd ed.). New York: Wiley.

Cooper, H., & Hedges, L. V. (1993). *The Handbook of Research Synthesis.* New York: RUSSELL SAGE FOUNDATION.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Crook, T. J., & Smock, H. R. (1974). *Student ratings of instructors related to student achievement.* Urbana: University of Illinois, Office of Instructional Resources.

DeGroot, M. H. (1986). A conversation with Erich L. Lehmann. *Statistical Science, 1(2),* 243-258.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Control Clin Trials,* 7, 177-188.

Doyle, K. 0., & Crichton, L. I. (1978). Student, peer, and self-evaluations of college instructors. *Journal of Educational Psychology, 70,* 815-826.

Doyle, K. O., & Whitely, S. E. (1974). Student ratings as criteria for effective teaching. *American Educational Research Journal, **11,*** 259-274.

Early Breast Cancer Trialists' Collaborative Group. (1988), Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. *New Engl J Med, **319,*** 1681-1692.

EBCTCG. (1998). Tamoxifen for early breast cancer: an overview of the randomised trials. *The Lancet, **351,*** 1451-1467.

Elliott, D. N. (1950). Characteristics and relationships of various criteria of college and university teaching. *Purdue University Studies in Higher Education, 70,* 5-61.

Ellis, N. R., & Rickard, H. C. (1977). Evaluating the teaching of introductory psychology. *Teaching of Psychology, 4,*128-132.

Elwood, P. (2006). The first randomized trial of aspirin for heart attack and the advent of systematic overviews of trials. *Journal Of The Royal Society Of Medicine, **99,*** 586-588.

Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions.* New York: John Wiley & Sons, Inc.

Fleiss, J. L. (1994). Measures of effect size for categorical data. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (p.p. 245-260). New York: Russell Sage Foundation.

Frey, P. W., Leonard, D. W., & Beatty, W. W. (1975). Student ratings of instruction: Validation research. *American Educational Research Journal, 12,*435-447.

Gage, N. L. (1978). *The scientific basis of the art of teaching.* New York: Teachers College Press.

Giordano, S. H., Cohen, D. S., Buzdar, A. U, Perkins, G., & Hortobagyi, G. N. (2004). Breast carcinoma in men: a population-based study. *PubMed 101(1),* 51-57.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Laboratory of Educational Research,* 3-8.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Newbury Park: Sage Publications.

Greenwood, G. E., Hazelton, A., Smith, A. B., & Ware, W. B. (1976). A study of the validity of four types of student ratings of college teaching assessed on a criterion of student achievement gains. *Research in Higher Education,* 5,171-178.

Guzzo, R. A., Jackson, S. E., & Katzell, R. A. (1986). Meta-analysis analysis. *Research in organizational behavior, 9,* 407-442.

Haddock C. K., Rindskopf D., & Shadish, W. R. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods,* 5(3), 339-353.

Hasselblad, V., & Hedges, L. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin 1995, 117,* 167-178.

Hays, W. L. (1973). *Statistics.* New York: Holt Rinehart & Winston.

Hedges, L. (1981). Distributional theory for Glass' estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 107-128.

Hedges, L. V., Giaconia, R. M., & Gage, N. L. (1981). *The empirical evidence on the effectiveness of open education.* Stanford, CA: Stanford University School of Education.

Hedges, L. V., & Olkin, I. (1980). Vote-counting methods in research synthesis. *Psychological Bulletin, 88,* 359-369.

Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis.* USA: Academic Press.

Hoffman, R. G. (1978). Variables affecting university student ratings of instructor behavior. *American Educational Research Journal, IS,* 287-299.

Holmbeck, G. N. (1997). Toward Terminological, Conceptual, and Statistical Clarity in the Study of Mediators and Moderators: Examples From the Child-Clinical and Paediatric Psychology Literatures. *Journal of Consulting and Clinical Psychology, 65(4),* 599-610.

Hunt, M. (1997). *How science takes stock: The story of meta-analysis.* New York: Russell Sage Foundation.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Hunter, J., & Schmidt, F. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Beverly Hills California: Sage.

Hunter, J. E., & Gerbing, D. W. (1982). tridimensional measurement, second order factor analysis, and causal models. *Research in Organizational Behaviour, 4*,267-320.

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8 (4),* 275-292.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating findings across research.* Beverly Hills, California: Sage.

Johnson, D. W., Johnson, R., & Holubec, E. (1992). *Advanced cooperative learning.* Edina, MN: Interaction Book Company.

Johnson, D. W., Johnson, R., & Holubec, E. (1993). *Cooperation in the classroom.* Edina, MN: Interaction Book Company.

Johnson, D. W., Johnson, R., & Holubec, E. (1998). *Cooperation in the classroom* (7th ed.). Edina, MN: Interaction Book Company.

Kraemer, H. C, Stice, E., Kazdin, A., Offord, D., & Kupfer, D. (2001). How Do Risk Factors Work Together? Mediators, Moderators, and Independent, Overlapping, and Proxy Risk Factors. *The American Journal of Psychiatry, 158(6),* 854.

Lee, J. (1999). Effectiveness of computer-based instructional simulation: A meta-analysis. *International Journal of Instructional Media, 26,* 71-85.

Light, R. J., & Smith, P. V. (1971). Accumulating evidence: Procedures for resolving contradictions among different research studies. *Harvard Educational Review, 41,* 429-471.

Lindley, D. V., & Scott, W. F. (1984). *New Cambridge Statistical Tables* (2nd. ed.). Cambridge, U.K.: Cambridge University Press

Lipsey, M. W., & Wilson, D. B. (1993). The Efficacy of Psychological, Educational, and Behavioral Treatment : Confirmation From Meta-Analysis. *American Psychologist, 48(12),* 1181-1209.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-analysis* (Vol. 49). USA: SAGE.

Livingston, M., Messura, J., Dellinger, T., Holder, R., & Hyde, J. (2008). Meta-analysis: an introduction into a research process. *Spec Care Dentist, 28(A),* 125-130.

Loughin, T. M. (2004). A systematic comparison of methods for combining p-values from independent tests. *Computational Statistics & Data Analysis, 47*,467-485.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

McKeachie, W. J., Lin, Y. G., & Mann, W. (1971). Student ratings of teacher effectiveness: Validity studies. *American Educational Research Journal, 8,*435-445.

Moher D, & Olkin I. (1995). Meta-analysis of randomized controlled trials. A concern for standards. *JAMA, 274(24),* 1962-1964.

Morsh, J. E., Burgess, G. G., & Smith, P. N. (1956). Student achievement as a measure of instructor effectiveness. *Journal of Educational Psychology, 47,* 79-88.

National Research Council. (1992). *Combining information: Statistical issues and opportunities for research.* Washington, DC: National Academy of Sciences Press.

Olkin, I. (1990). History and goals. In K. W. Wachter & M. L. Straf (Eds.), *The future of meta-analysis* (pp. 3-10). New York: Russell Sage Foundation.

Osborne, J. (2008). Sweating the small stuff in educational psychology: How effect size and power reporting failed to change from 1969 to 1999,and what that means for the future of changing practices. *Educational Psychology, 28,*151-160.

Paik, H., & Comstock, G. (1994). The effects of television violence on antisocial behavior: A meta-analysis. *Communication Research, 21(4),* 516-546.

Pinello, D. R. (1999). Linking Party to Judicial Ideology in American Courts: A Meta-analysis. *The Justice System Journal, 20(3).*

Remmers, H. H., Martin, F. D., & Elliott, D. N. (1949). Are students ratings of instructors related to their grades? *Purdue University Studies in Higher Education, 66,* 17-26.

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

Rosenthal R. (1991). *Meta-analytic procedures for social research* (2nd ed. Vol. 6). London: Sage Publications.

Rosenthal R. (1961). *On the social psychology of the psychological experiment: With particular reference to experimenter bias.* Paper presented at the meeting of the American Psychological Association, New York.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44,*1276-1284.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental & Quasi-Experimental Designs for Generalized Causal Inference.* Boston: Houghton Mifflin.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of research synthesis* (pp. 261-281). New York: Russell Sage Foundation.

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *Journal of Experimental Education, 61,* 334-349.

Springer, L., Stanne, M. E., & Donovan, S. (1999). Measuring the success of small-group learning in college level SMET teaching: a meta-analysis. *Review of Educational Research, 69,* 21 -51.

Stanley, T. D. (2001). Wheat From Chaff: Meta-analysis As Quantitative Literature Review. *Journal of Economic Perspectives, 15(3),* 131-150.

Stanley, T. D. (2005). Beyond publication bias. *Journal of Economic Surveys, 19,* 309-345.

Stanley, T. D. (2006). Meta-regression methods for detecting and estimating empirical effect in the presence of publication selection. *School Working Paper, Economics Series 2006/20.* Deakin University.

Stanley, T. D., & Jarrell, S. B. (1989). Meta-regression analysis: a quantitative method of literature surveys. *Journal of Economic Surveys, 3,* 54-67.

Sullivan, A. M., & Skanes, G. R. (1974). Validity of student evaluation of teaching and the characteristics of successful instructors *Journal of Educational Psychology, 66,* 584-590.

Thomas, J. R., Salazar, W., & Landers, D. M. (1991). What is missing in p less than .05? Effect size. *Research quarterly for exercise and sport, 62(3),* 344-348.

Tippett, L. H. C. (1931). *The method of statistics.* London: Williams & Norgate.

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counselling and Psychology, 51,* 473-481.

Vasquez, Linda, M., Gangstead, Kay, S., Henson, & Robin, K. (2000). *Understanding and Interpreting Effect Size Measures in General Linear Model Analyses.* Paper presented at the Southwest Educational Research Association.

Wherry, R. J. (1952). *Control of bias in ratings* (PRS Reports Nos. 914, 915, 919, 920, and 921). Washington, DC: Department of the Army, Adjutant General's Office.

Williamson, J. W., Weir, C. R., Turner, C. W,, Lincoln, M. J., & Cofrin, K. M. W. (2002). *Healthcare Informatics and Information Synthesis.* U.S.A: Sage Publications.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis.* Beverly Hills California: Sage.

Yusuf, S., Peto, R., Lewis, J., Collins, R,, & Sleight, P. (1985). Beta blockade during and after myocardial infarction: an overview of the randomized trials. *PubMed, 27(5),* 335-371.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDID
N IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI P

APPENDICES

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKA
DRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PEN

Appendix A : A list of tables from Table A.1 to Table A.4

Table A.1 Probability That the Sample Mean (Experimental Group) Exceeds the Sample Mean (Control Group)

	Population Standardized Mean Difference δ										
n	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
1	0.500	0.503	0.506	0.508	0.511	0.514	0.517	0.520	0.523	0.525	0.528
2	0.500	0.504	0.508	0.512	0.516	0.520	0.524	0.528	0.532	0.536	0.540
3	0.500	0.505	0.510	0.515	0.520	0.524	0.529	0.534	0.539	0.544	0.549
4	0.500	0.506	0.511	0.517	0.523	0.528	0.534	0.539	0.545	0.551	0.556
5	0.500	0.506	0.513	0.519	0.525	0.532	0.538	0.544	0.550	0.557	0.563
6	0.500	0.507	0.514	0.521	0.528	0.534	0.541	0.548	0.555	0.562	0.569
7	0.500	0.508	0.515	0.522	0.530	0.537	0.545	0.552	0.560	0.567	0.574
8	0.500	0.508	0.516	0.524	0.532	0.540	0.548	0.556	0.564	0.571	0.579
9	0.500	0.508	0.517	0.525	0.534	0.542	0.551	0.559	0.567	0.576	0.584
10	0.500	0.509	0.518	0.527	0.536	0.544	0.553	0.562	0.571	0.580	0.588
11	0.500	0.509	0.519	0.528	0.537	0.547	0.556	0.565	0.574	0.584	0.593
12	0.500	0.510	0.520	0.529	0.539	0.549	0.558	0.568	0.578	0.587	0.597
13	0.500	0.510	0.520	0.530	0.541	0.551	0.561	0.571	0.581	0.591	0.601
14	0.500	0.511	0.521	0.532	0.542	0.553	0.563	0.574	0.584	0.594	0.604
15	0.500	0.511	0.522	0.533	0.544	0.554	0.565	0.576	0.587	0.597	0.608
16	0.500	0.511	0.523	0.534	0.545	0.556	0.567	0.578	0.590	0.600	0.611
17	0.500	0.512	0.523	0.535	0.546	0.558	0.569	0.581	0.592	0.604	0.615
18	0.500	0.512	0.524	0.536	0.548	0.560	0.571	0.583	0.595	0.606	0.618
19	0.500	0.512	0.525	0.537	0.549	0.561	0.573	0.585	0.597	0.609	0.621
20	0.500	0.513	0.525	0.538	0.550	0.563	0.575	0.588	0.600	0.612	0.624
21	0.500	0.513	0.526	0.539	0.552	0.564	0.577	0.590	0.602	0.615	0.627
22	0.500	0.513	0.526	0.540	0.553	0.566	0.579	0.592	0.605	0.617	0.630
23	0.500	0.514	0.527	0.540	0.554	0.567	0.581	0.594	0.607	0.620	0.633
24	0.500	0.514	0.528	0.541	0.555	0.569	0.582	0.596	0.609	0.622	0.636
25	0.500	0.514	0.528	0.542	0.556	0.570	0.584	0.598	0.611	0.625	0.638
50	0.500	0.520	0.540	0.560	0.579	0.599	0.618	0.637	0.655	0.674	0.692
100	0.500	0.528	0.556	0.584	0.611	0.638	0.664	0.690	0.714	0.738	0.760
200	0.500	0.540	0.579	0.618	0.655	0.692	0.726	0.758	0.788	0.816	0.841
400	0.500	0.556	0.611	0.664	0.714	0.760	0.802	0.839	0.871	0.898	0.921

UNIVERSITI PENDIDIKAN SULTAN IDRIS

Table A.1 (continued)

Population Standardized Mean Difference δ												
0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4
0.556	0.584	0.611	0.638	0.664	0.690	0.714	0.738	0.760	0.782	0.802	0.821	0.839
0.579	0.618	0.655	0.692	0.726	0.758	0.788	0.816	0.841	0.864	0.885	0.903	0.920
0.597	0.643	0.688	0.730	0.769	0.804	0.836	0.865	0.890	0.911	0.929	0.944	0.957
0.611	0.664	0.714	0.760	0.802	0.839	0.871	0.898	0.921	0.940	0.955	0.967	0.976
0.624	0.682	0.736	0.785	0.829	0.866	0.897	0.923	0.943	0.959	0.971	0.980	0.986
0.636	0.698	0.756	0.807	0.851	0.887	0.917	0.940	0.958	0.972	0.981	0.988	0.992
0.646	0.713	0.773	0.825	0.869	0.905	0.933	0.954	0.969	0.980	0.988	0.992	0.996
0.655	0.726	0.788	0.841	0.885	0.919	0.945	0.964	0.977	0.986	0.992	0.995	0.997
0.664	0.738	0.802	0.856	0.898	0.931	0.955	0.972	0.983	0.990	0.994	0.997	0.998
0.673	0.749	0.814	0.868	0.910	0.941	0.963	0.978	0.987	0.993	0.996	0.998	0.999
0.680	0.759	0.826	0.880	0.920	0.950	0.970	0.983	0.990	0.995	0.998	0.999	1.000
0.688	0.769	0.836	0.890	0.929	0.957	0.975	0.986	0.993	0.996	0.998	0.999	1.000
0.695	0.778	0.846	0.899	0.937	0.963	0.979	0.989	0.995	0.998	0.999	1.000	
0.702	0.786	0.855	0.907	0.944	0.968	0.983	0.991	0.996	0.998	0.999	1.000	
0.708	0.794	0.863	0.914	0.950	0.972	0.986	0.993	0.997	0.999	1.000		
0.714	0.802	0.871	0.921	0.955	0.976	0.988	0.994	0.998	0.999	1.000		
0.720	0.809	0.878	0.928	0.960	0.979	0.990	0.996	0.998	0.999	1.000		
0.726	0.816	0.885	0.933	0.964	0.982	0.992	0.996	0.999	1.000			
0.731	0.822	0.891	0.938	0.968	0.984	0.993	0.997	0.999	1.000			
0.736	0.829	0.897	0.943	0.971	0.987	0.994	0.998	0.999	1.000			
0.742	0.834	0.902	0.947	0.974	0.988	0.995	0.998	0.999				
0.746	0.840	0.908	0.951	0.977	0.990	0.996	0.999	1.000				
0.751	0.846	0.912	0.955	0.979	0.991	0.997	0.999	1.000				
0.756	0.851	0.917	0.958	0.981	0.992	0.997	0.999	1.000				
0.760	0.856	0.921	0.962	0.983	0.993	0.998	0.999	1.000				
0.841	0.933	0.977	0.994	0.999	1.000	1.000	1.000					
0.921	0.983	0.998	1.000	1.000								
0.977	0.999	1.000										
0.998	1.000											

(source: Hedges and Olkin, 1985)

UNIVERSITI PENDIDIKAN SULTAN IDRIS

Table A.2 Probability that a Correlation Coefficient from a sample of size n will be positive when the population Correlation Coefficient is ρ

	Population Correlation Coefficient ρ									
n	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
3	0.500	0.505	0.510	0.515	0.520	0.525	0.530	0.535	0.540	0.545
4	0.500	0.506	0.513	0.519	0.525	0.532	0.538	0.545	0.551	0.557
5	0.500	0.508	0.515	0.522	0.530	0.537	0.545	0.552	0.560	0.567
6	0.500	0.508	0.517	0.525	0.534	0.542	0.551	0.559	0.568	0.576
7	0.500	0.509	0.519	0.528	0.537	0.547	0.556	0.565	0.575	0.584
8	0.500	0.510	0.520	0.531	0.541	0.551	0.561	0.571	0.581	0.591
9	0.500	0.511	0.522	0.533	0.544	0.555	0.565	0.576	0.587	0.598
10	0.500	0.512	0.523	0.535	0.546	0.558	0.570	0.581	0.592	0.604
11	0.500	0.512	0.525	0.537	0.549	0.561	0.573	0.586	0.598	0.610
12	0.500	0.513	0.526	0.539	0.552	0.564	0.577	0.590	0.602	0.615
13	0.500	0.514	0.527	0.541	0.554	0.567	0.581	0.594	0.607	0.620
14	0.500	0.514	0.528	0.542	0.556	0.570	0.584	0.598	0.612	0.625
15	0.500	0.515	0.529	0.544	0.558	0.573	0.587	0.602	0.616	0.630
16	0.500	0.515	0.530	0.546	0.561	0.576	0.590	0.605	0.620	0.634
17	0.500	0.516	0.531	0.547	0.563	0.578	0.593	0.609	0.624	0.639
18	0.500	0.516	0.532	0.549	0.565	0.581	0.596	0.612	0.628	0.643
19	0.500	0.517	0.533	0.550	0.566	0.583	0.599	0.615	0.631	0.647
20	0.500	0.517	0.534	0.551	0.568	0.585	0.602	0.618	0.635	0.651
21	0.500	0.518	0.535	0.553	0.570	0.587	0.605	0.622	0.638	0.655
22	0.500	0.518	0.536	0.554	0.572	0.590	0.607	0.625	0.642	0.659
23	0.500	0.518	0.537	0.555	0.574	0.592	0.610	0.627	0.645	0.662
24	0.500	0.519	0.538	0.557	0.575	0.594	0.612	0.630	0.648	0.666
25	0.500	0.519	0.539	0.558	0.577	0.596	0.615	0.633	0.651	0.669
50	0.500	0.528	0.555	0.583	0.610	0.636	0.662	0.687	0.712	0.735
100	0.500	0.540	0.579	0.617	0.654	0.690	0.724	0.757	0.787	0.815
200	0.500	0.556	0.611	0.664	0.714	0.760	0.801	0.838	0.871	0.898
400	0.500	0.579	0.655	0.725	0.788	0.841	0.885	0.919	0.945	0.964

UNIVERSITI PENDIDIKAN SULTAN IDRIS

Table A12 (continued)

Population Correlation Coefficient ρ								
0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950
0.564	0.626	0.688	0.748	0.805	0.858	0.906	0.948	0.981
0.575	0.648	0.718	0.784	0.844	0.896	0.939	0.972	0.993
0.584	0.666	0.743	0.813	0.873	0.923	0.960	0.985	0.997
0.593	0.683	0.765	0.837	0.896	0.942	0.973	0.991	0.998
0.601	0.697	0.784	0.857	0.915	0.956	0.982	0.995	1.000
0.608	0.710	0.800	0.874	0.929	0.967	0.988	0.997	1.000
0.615	0.722	0.815	0.889	0.941	0.975	0.992	0.998	1.000
0.621	0.733	0.828	0.901	0.951	0.980	0.994	0.999	1.000
0.627	0.744	0.840	0.912	0.959	0.985	0.996	0.999	1.000
0.633	0.754	0.851	0.922	0.966	0.988	0.997	1.000	
0.639	0.763	0.861	0.930	0.971	0.991	0.998	1.000	
0.644	0.771	0.871	0.938	0.976	0.993	0.999	1.000	
0.649	0.779	0.879	0.944	0.980	0.995	0.999	1.000	
0.654	0.787	0.887	0.950	0.983	0.996	0.999	1.000	
0.658	0.794	0.894	0.955	0.985	0.997	1.000		
0.663	0.801	0.901	0.960	0.988	0.997	1.000		
0.667	0.808	0.907	0.964	0.990	0.998	1.000		
0.671	0.814	0.913	0.967	0.991	0.998	1.000		
0.675	0.820	0.918	0.971	0.992	0.999	1.000		
0.679	0.826	0.923	0.974	0.994	0.999	1.000		
0.683	0.831	0.927	0.976	0.995	0.999	1.000		
0.687	0.836	0.932	0.979	0.995	0.999	1.000		
0.757	0.920	0.984	0.998	1.000	1.000			
0.840	0.978	0.999	1.000					
0.921	0.998	1.000						
0.977	1.000							

(Source: Hedges and Olkin, 1985)

UNIVERSITI PENDIDIKAN SULTAN IDRIS

Table A.3 Table of the Standard Normal Distribution Function $\Phi(z)$

	z									
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936

UNIVERSITI PENDIDIKAN SULTAN IDRIS

Table A.3 (continued)

2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997

(Source: Lindley & Scott, 1984)

UNIVERSITI PENDIDIKAN SULTAN IDRIS

Table A.4 Table of the χ^2 distribution

p	0.995	0.975	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.002	0.001
d.f = 1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124
9	1.735	2.700	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588
11	2.603	3.816	14.631	17.275	19.675	21.920	22.618	24.725	26.757	29.354	31.264
12	3.074	4.404	15.812	18.549	21.026	23.337	24.054	26.217	28.300	30.957	32.909
13	3.565	5.009	16.985	19.812	22.362	24.736	25.472	27.688	29.819	32.535	34.528
14	4.075	5.629	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123
15	4.601	6.262	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697
16	5.142	6.908	20.465	23.542	26.296	28.845	29.633	32.000	34.267	37.146	39.252
17	5.697	7.564	21.615	24.769	27.587	30.191	30.995	33.409	35.718	38.648	40.790
18	6.265	8.231	22.760	25.989	28.869	31.526	32.346	34.805	37.156	40.136	42.312
19	6.844	8.907	23.900	27.204	30.144	32.852	33.687	36.191	38.582	41.610	43.820
20	7.434	9.591	25.038	28.412	31.410	34.170	35.020	37.566	39.997	43.072	45.315
21	8.034	10.283	26.171	29.615	32.671	35.479	36.343	38.932	41.401	44.522	46.797
22	8.643	10.982	27.301	30.813	33.924	36.781	37.659	40.289	42.796	45.962	48.268
23	9.260	11.689	28.429	32.007	35.172	38.076	38.968	41.638	44.181	47.391	49.728
24	9.886	12.401	29.553	33.196	36.415	39.364	40.270	42.980	45.559	48.812	51.179

UNIVERSITI PENDIDIKAN SULTAN IDRIS

Table A.4 (continue)

25	10.520	13.120	30.675	34.382	37.652	40.646	41.566	44.314	46.928	50.223	52.620
26	11.808	13.844	31.795	35.563	38.885	41.923	42.856	45.642	48.290	51.627	54.052
27	11.808	14.573	32.912	36.741	40.113	43.195	44.140	46.963	49.645	53.023	55.476
28	12.461	15.308	34.027	37.916	41.337	44.461	45.419	48.278	50.993	54.411	56.892
29	13.121	16.047	35.139	39.087	42.557	45.722	46.693	49.588	52.336	55.792	58.301
30	13.787	16.791	36.250	40.256	43.773	46.979	47.962	50.892	53.672	57.167	59.703
32	15.134	18.291	38.466	42.585	46.194	49.480	50.487	53.486	56.328	59.899	62.487
34	16.501	19.806	40.676	44.903	48.602	51.966	52.995	56.061	58.964	62.608	65.247
36	17.887	21.336	42.879	47.212	50.998	54.437	55.489	58.619	61.581	65.296	67.985
38	19.289	22.878	45.076	49.513	53.384	56.896	57.969	61.162	64.181	67.966	70.703
40	20.707	24.433	47.269	51.805	55.758	59.342	60.436	63.691	66.766	70.618	73.402
50	27.991	32.357	58.164	63.167	67.505	71.420	72.613	76.154	79.490	83.657	86.661
60	35.534	40.482	68.972	74.397	79.082	83.298	84.580	88.379	91.952	96.404	99.607
70	43.275	48.758	79.715	85.527	90.531	95.023	96.388	100.425	104.215	108.929	112.317
80	51.172	57.153	90.405	96.578	101.879	106.629	108.069	112.329	116.321	121.280	124.839
100	67.328	74.222	111.667	118.498	124.342	129.561	131.142	135.807	140.169	145.577	149.449

(Source: Lindley & Scott, 1984)

UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

Appendix B : A list of data sets from Table B.1 to Table B.4

1. DATA SET I: Studies of Effects of Open Education on Student Independence and Self-Reliance

A very large data set derived from studies of the effectiveness of open education programs was assembled by Hedges, Giaconia, and Gage (1981). Their analysis included nearly 200 studies that examined the effects of open education on student outcomes by comparing students in experimental open classroom schools with students from traditional schools. A variety of outcome variables were classified into 16 different dependent variable clusters for the statistical analyses:

Achievement motivation, adjustment, anxiety, attitude toward school, attitude toward teacher, cooperativeness, creativity, curiosity, general mental ability, independence and self-reliance, language achievement, locus of control, mathematics achievement, reading achievement, science and social studies achievement, and self-concept.

A data set of seven studies were selected, in part, since they had the same sample size $n_E = n_C = 30$. The results of these studies are summarized in Table B.1

Table B.1 Studies of the Effects of Open Education on Student Independence and Self-Reliance

Study	Sample size	Standardized mean difference, δ	Unbiased standardized mean difference, d
1	30	0.708	0.699
2	30	0.092	0.091
3	30	-0.059	-0.058
4	30	-0.080	-0.079
5	30	-0.238	-0.235
6	30	-0.500	-0.494
7	30	-0.595	-0.587

(Source: Hedges, Giaconia, and Gage, 1981).

2. DATA SET II: The Relationship Between Teacher Indirectness and Student Achievement

Studies of the relationship of teaching variables and student achievement were reviewed by Gage (1978). Several of the studies that reported correlations between an observational measure of teacher indirectness and student achievement are

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

reported in Table B.2. In these data the sample size is the number of teachers. The correlation coefficient reflects the relationship between a score on teacher indirectness (praise, criticism, or acceptance) derived from an observational measure and mean class achievement.

Table B.2 Studies of the Relationship between an Observational Measure of Teacher Indirectness and Student Achievement

Study	Grade level	No. of teachers	Correlation coefficient, r
1	2	15	-0.073
2	4	16	0.308
3	7	15	0.481
4	8	16	0.428
5	10-12	15	0.180
6	12	17	0.290
7	4	15	0.400

(Source: Gage, 1978).

3. DATA SET III: Studies of the Effects of Open versus Traditional Education on Student Self-Concept

Table B.3 Studies of the Effects of Open Education on Student Self-Concept

		Sample size			
Study	Grade level	Open education group, n_E	Traditional school group, n_C	Standardized mean difference, δ	Unbiased standardized mean difference, d
1	4-6	100	180	0.100	0.100
2	4-6	131	138	-0.162	-0.162
3	4-6	40	40	-0.091	-0.090
4	4-6	40	40	-0.049	-0.049
5	4-6	97	47	-0.046	-0.046
6	K-3	28	61	-0.010	-0.010
7	K-3	60	55	-0.434	-0.431
8	4-6	72	102	-0.262	-0.261
9	4-6	87	45	0.135	0.134
10	K-3	80	49	0.019	0.019
11	K-3	79	55	0.176	0.175
12	4-6	40	109	0.056	0.056
13	4-6	36	93	0.045	0.045
14	K-3	9	18	0.106	0.103
15	K-3	14	16	0.124	0.121
16	4-6	21	22	-0.491	-0.482
17	4-6	133	124	0.291	0.290
18	K-3	83	45	0.344	0.342

A set of 18 studies was derived from the large collection of data set from studies of the effectiveness of open education programs was assembled by Hedges, Giaconia, and Gage (1981) as mentioned in DATA SET I. The sample sizes for both open education, n_E and traditional group, n_C are probably different or unequal in this studies.

UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS UNIVERSITI PENDIDIKAN SULTAN IDRIS

4. *DATA SET IV: Validity Studies Correlating* Student Ratings of the Instructor with Student Achievement

The studies were all conducted in multiple courses which had various instructors but used a common final examination. Cohen (1981, 1983) reviewed studies of the validity of student ratings of instructors. The validity had index (correlation coefficient) between the mean instructor ratings and the mean examination score. Table B.4 provides the effect sizes (product moment correlation coefficient) with each study at least ten in a sample size.

Table B.4 Validity Studies Correlating Student Ratings of the Instructor with Student Achievement

Study	Sample	Ability Control	n	r
Bolton et al. 1979	General psychology	Random assignment	10	0.68
Bryson 1974	College algebra	CIAT pretest	20	0.56
Centra 1977 [1]	General biology	Final exam for prior semester of same course	13	0.23
Centra 1977 [2]	General psychology		22	0.64
Crooks & Smock 1974	General physics	Math pretest	28	0.49
Doyle & Crichton 1978	Introductory communications	PSAT verbal score	12	-0.04
Doyle & Whitely 1974	Beginning French aptitude test	Minnesota Scholastic	12	0.49
Elliot 1950	General chemistry	College entrance exam scores	36	0.33
Ellis & Richard 1977	General psychology	Random Assignment	19	0.58
Frey, Leonard & Beatty 1975	Introductory calculus	SAT math scores	12	0.18
Greenwood et al. 1976	Analytic geometry and calculus	Achievement pretest	36	-0.11
Hoffman 1978	Introductory math	SAT math scores	75	0.27
McKeachie et al. 1971	General psychology	Intelligence test scores	33	0.26
Morsh et al. 1956	Aircraft mechanics	Hydraulics pretest	121	0.40
Remmer et al. 1949	General chemistry	Orientation test scores	37	0.49
Sullivan & Skanes 1974 [1]	First-year science	Random assignment	14	0.51
Sullivan & Skanes 1974 [2]	Introductory psychology	Random assignment	40	0.40
Sullivan & Skanes 1974 [3]	First-year math	Random assignment	16	0.34
Sullivan & Skanes 1974 [4]	First-year biology	Random assignment	14	0.42
Wherry 1952	Introductory psychology	Ability acores	20	0.16

(Source: Cohen ,1981, 1983)

HAK MILIK PERPUSTAKAAN UNIVERSITI PENDIDIKAN SULTAN IDRIS