Assessing the Influence of Contamination on Fixed-Effect Meta-Analysis for Continuous Outcomes: A Simulation Study

ASSESSING THE INFLUENCE OF CONTAMINATION ON FIXED-EFFECT META-ANALYSIS FOR CONTINUOUS OUTCOMES: A SIMULATION STUDY

BY

REGINA SHARON KAMPO

A THESIS

SUBMITTED TO THE DEPARTMENT OF MATHEMATICS & STATISTICS

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

© Copyright by Regina Sharon Kampo, December 23, 2014

All Rights Reserved

Master of Science (2014) McMaster University

(Mathematics & Statistics) Hamilton, Ontario, Canada

TITLE:	Assessing the Influence of Contamination on Fixed-Effect Meta-Analysis for Continuous Outcomes: A Simulation Study
AUTHOR:	Regina Sharon Kampo M.Sc., (Statistics) McMaster University, Canada
SUPERVISOR:	Dr. Joseph Beyene
NUMBER OF PAGES:	xiv, 57

I dedicate this thesis to my family for nursing me with affections and love and their dedicated partnership for success in my life.

Abstract

Important research questions are typically studied and analyzed more than once, often by different research teams in different locations. However, in many instances, the results of these multiple small studies are diverse and conflicting, which makes decision-making difficult. The need to arrive at decisions fostered the momentum towards synthesizing the results of these multiple studies. Therefore, meta-analysis, also referred to as the standard or traditional meta-analysis, is a statistical technique for combining the results or findings from multiple independent studies to address a specific research question. The applications of meta-analysis have been extended to many fields of research including medicine, psychology, ecology, education, business and many others.

Prior to carrying out a meta-analysis or statistically synthesizing data, a researcher must undertake a systematic review. Systematic review attempts to collate empirical evidence that fits pre-specified eligibility criteria to answer a specific research question. That is to determine which studies will be included or excluded from the analysis. Standard meta-analysis methods are used to obtain the relative efficacy (or safety) of a particular intervention versus a competing intervention in the

presence of a direct or head-to-head comparison. Thus only a pair-wise comparison can be made. The outcome of these interventions could be continuous, binary or count data.

A number of methodologies related to meta-analysis, assessments of underlying assumptions and strategies for the presentation of results have been proposed by several researchers. A commonly used model for estimating effect sizes in meta-analysis is the fixed-effect model. However, various factors can determine the performance of the model which needs to be considered before using the results for decision making.

This project aimed to investigate the performance of hypothesis properties and estimation properties on selecting data points from an underlying contaminated distribution under different scenarios for modeling a continuous outcome. Different levels of contamination, levels of significance, number of studies, number of individual study sample sizes, standard deviations and effect sizes were investigated in our simulation study for a continuous outcome.

The results of our simulation study shows that, the fixed-effect meta-analytic model does not perform well in the presence of contamination. As the level of contamination in the treatment group increases, the properties of estimators and hypothesis are greatly influenced. The method performs well as expected in the absence of contamination but performs poorly as we observe 50% contamination in the treatment group regardless of the individual sample size, the number of studies, the standard deviation and the effect size.

Acknowledgements

I am extremely grateful to Almighty God, for His grace and guidance from the beginning to the completion of this thesis. I would like to express my very great appreciation to my supervisor Dr. Joseph Beyene for his valuable and constructive suggestions during the planning and development of this research work. His willingness to give his time so generously has been very much appreciated. I would like to thank my supervisory committee members: Dr. Narayanaswamy Balakrishnan and Dr. Roman Viveros-Aguilera for their insightful inputs and useful suggestions.

I thank my fellow lab mates in the Statistics for Integrative Genomics and Meta-Analysis (SIGMA) Research Group for the stimulating discussions in our weekly meetings and for all the fun we have had during our studies. I would like to thank Ashley Bonner, Binod Neupane, Mateen Shaikh and Taddele Kibret, for an excellent guide and support through coding issues and our lovely talks. My immerse gratitude goes to the faculty and staff of the Department of Mathematics and Statistics for providing the resources I needed to complete this work.

I also want to thank my professors, Dr. Jemila Hamid, Dr. Aaron Childs, Dr. Fred Hoppe and Dr. Angelo Canty as they greatly supported me during my graduate

studies.

I am ever grateful to my dad (in loving memory), mum and siblings, for their prayers, encouragement, support and love in challenging moments. Last but not the least, my sincere thanks go to my Westside Baptist Church family, friends and colleagues who have helped to make my stay in Canada wonderful through their encouragement, support and love.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Prior to the 1990s, the task of combining data from multiple studies had been primarily subjected to the narrative review (Borenstein et al. (2009)). The approach which turns out to be mainly descriptive, does not involve a systematic search of the literature, and thereby often focuses on a subset of studies in an area chosen based on availability or author's selection. Some accompanying limitations to this approach include lack of transparency, coupled with inherent subjectivity on the part of the reviewers. Also, narrative reviews become less useful as more information becomes available. Due to these caveats associated with the narrative reviews, researchers across several fields have been moving away from this approach which lead to the adoption of systematic reviews (Borenstein et al. (2009)). As the name implies, systematic reviews typically involve a detailed and comprehensive plan and search

strategy, with the goal of reducing bias by identifying, appraising and synthesizing all relevant studies on a particular topic. A key element in most systematic reviews is the statistical synthesis of the data, or meta-analysis.

Since its introduction, meta-analysis, the statistical pooling of results from independent but combinable studies, has been established as an influential branch of clinical epidemiology and health services. There have been hundreds of meta-analysis publications each year in the medical literature demonstrating how important the method has become in the field of clinical trials (Sterne et al. (2001)). The statistical basis of meta-analysis dates back to the 17^{th} century when it was suggested that combinations of data might be better than attempts to select amongst them (Plackett (1958)). The distinguished statistician Karl Pearson was, in 1904, seen as the first researcher to address questions of combining clinical trials from different studies using formal techniques and in 1976, the psychologist, Gene Glass coined up the term 'meta-analysis' (Egger et al. (2002); O'Rourke (2007)). However, the technique was not recognized and implemented in medicine for several years to come. But it was rather embraced in the social sciences, particularly in educational and psychological research as a method of synthesizing research findings. In the 1980s, meta-analysis became increasingly popular in medicine, especially in the fields of clinical trials which resulted in the establishment of the Cochrane Collaboration in the early 1990s as an international network of health care professionals who prepare and regularly update systematic reviews. Thus, the 'Cochrane Reviews' expedited the conduct of meta-analysis in all areas of health care (Egger et al. (2002)).

Traditionally, systematic reviews compare only two interventions by using pairwise meta-analysis methods. The meta-analysis component involves using statistical techniques to synthesize the data from several studies into a single quantitative estimate or summary effect size. The effect size measures the strength of the relationship between two variables, hence providing information about the magnitude of the treatment effect. The types of effect size, including odds ratio, relative risk, correlation coefficient, standardized mean difference, generally depend on the type of outcome being examined.

When studies have consistent treatment effect, meta-analysis can be used to estimate their common effect size. However, sometimes the generalizability of treatment effects could be of interest, which focuses on the variability of measures. When the effect size differs across studies, meta-analysis can be used to estimate the average effect size as accurately as possible and can also be used to explore the sources of heterogeneity and its implications (Borenstein et al. (2010)). There are two commonly used statistical models for meta-analysis, namely the fixed-effect model and the random-effects model (Borenstein et al. (2009)). Each of these models makes different assumptions about the nature of the studies which lead to different definitions for the combined effect and different mechanisms for assigning weights (Hedges and Vevea (1998)). When combining studies, it is important to assign more weight to studies that carried more information (Borenstein et al. (2010)).

The fixed effect model assumes that there is one true effect size that underlies all the studies in the analysis implying that any variation in the observed effects is due to sampling error (Borenstein et al. (2009)).

By contrast, the random-effects meta-analysis model allows the true effect size underlying the studies to vary. Thus, the random-effect allows for unexplained heterogeneity, the variation across studies. The random-effects model assumes that the true effects are normally distributed. That is, the model is mainly used to estimate the mean of the distribution of the true effects using the true effects of individual studies.

1.2 Scope of the Study

Over the past decades, there has been a rapid growth in the publications of meta-analysis papers. Topics ranging from the effects estimates in meta-analysis, methods of result presentation in meta-analysis, tests of assumptions in meta-analysis, etc. have been explored in these publications. To fully understand the performance of the statistical methods in meta-analysis, comprehensive simulation studies need to be carried out. However, only a limited number of these publications have considered a continuous outcome. A few of the simulation studies includes Banerjee et al. (2008), where they explored some caveats in meta-analysis for a continuous outcome. From their results, it has been observed that the sample sizes and the number of studies greatly influence the significance of the weighted mean differences extracted from the meta-analysis in the scenarios they considered. Friedrich et al. (2008) also conducted a simulation study to investigate the performance of the ratio of means (RoM) method among other commonly used methods for analyzing a continuous outcome. Their study showed that, the ratio of means method exhibits comparable

performance characteristics to the other methods i.e., the mean difference (MD) and the standardized mean difference (SMD).

In the simulation report published by Friedrich et al. (2008), it was pointed out that, the number of participants per trial, the number of trials, the effect size and standard deviation, can affect the properties of estimation and the properties of hypothesis tests. A meta-analysis report published by Clifton et al. (2014) showed that, when treatment group is assigned intervention it is almost unlikely to observe 100% adherence. In most simulation studies that considered a continuous outcome, the data for both the experimental and control groups are generated from a normal distribution assuming 100% compliance in the treatment group. Whilst in real life scenarios, there could be some level of contamination in the data.

This project thus aimed to investigate the influence of contamination on fixed-effect meta-analysis on a continuous outcome. Estimation properties such as coverage probability, bias, mean square error and width of confidence interval are explored. Also, properties of hypothesis tests i.e., statistical power and probability of type I error are investigated. A Monte Carlo simulation analysis was performed using the R statistical software with the meta-analysis carried out by using the *"metafor"* package. The number of experimental and control patients per trial arm, the number of trials, the effect size, standard deviations and level of contamination were varied during the simulation.

The organization of the remainder of this thesis is as follows: The methods section, Chapter 2, briefly describes the methods in meta-analysis. Chapter 3, the simulation results section, presents the performance of the fixed-effect meta-analysis

in the presence of contamination. Finally, Chapter 4 includes discussion and future directions.

Chapter 2

Methods

2.1 Systematic Review and Meta-Analysis

Systematic Review

Systematic reviews form a crucial method for overcoming the challenges faced by researchers when trying to combine the results of multiple studies. A systematic review is a scientific tool that is used in bringing together a number of separately conducted studies, sometimes with conflicting findings, and synthesizing their result. However, in order to reduce bias in systematic reviews, a formal and rigorous methodology has been developed which requires a number of steps (Green et al. (2005); Uman (2011)). The key characteristics of systematic reviews include:

- Clearly formulated review question.
- Explicit and justified criteria for the inclusion and exclusion of any study.

- Comprehensive strategy to search for studies that address the objectives of the review (relevant studies) that include published and unpublished studies.
- Comprehensive list of all studies identified.
- Data extraction process.
- Clear presentation of the characteristics of each study included and an analysis of methodological quality.
- Clear analysis of the results of the eligible studies using statistical synthesis of data (meta-analysis) if appropriate and possible.
- Structured report of the review clearly stating the aims, describing the methods and materials, and reporting the results.

Meta-analysis

In many medical specialties it is common to find that several trials have attempted to answer similar questions about clinical effectiveness. For example, whether the new treatment confer significant benefits compared to conventional treatment. Often many of the individual trials will fail to show a statistically significant difference between the two treatments. However, when the results from the individual studies are combined using appropriate techniques (meta-analysis), significant benefits of treatment may be shown. Meta-analysis thus refers to the statistical synthesis of results from a series of studies. It aims to combine results of comparable studies in order to obtain an overall estimate of effect thereby reducing uncertainty.

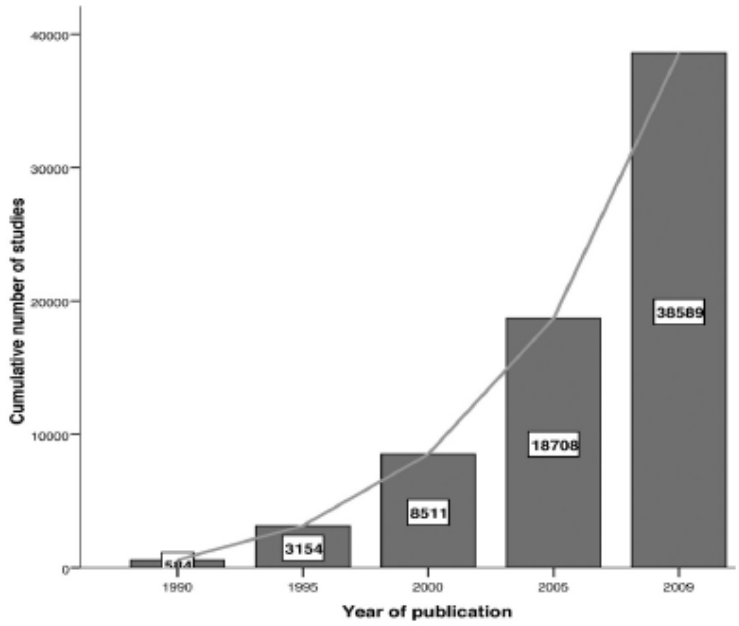

Figure 2.1: Cumulative number of publications about meta-analysis over time, until 17 December 2009 (results from Medline search using text meta-analysis), adapted from Haidich (2010).

In the last 30 years, meta-analysis has become an increasingly popular method of evidence synthesis in the field of medical research with an exponential growth over the years (Figure 2.1) (Haidich (2010)).

Some benefits of conducting meta-analysis include, more precise estimates (i.e., since more information is used to estimate the pooled effect) (Haidich (2010)), attaining higher power, the ability to address controversies that arise between conflicting studies, quantifying treatment effects and their uncertainties, the ability to explore differences between multiple studies, generate new hypotheses and to make future predictions (Harrison (2011)).

Meta-analysis can be accomplished using either individual patient data (collected

directly from trial units) or summary data (collected from published literature). Depending on the pool of studies being used or the measures extracted from each study, a meta-analysis with summary data might sometimes yield different results than meta-analysis with individual patient data (Olkin and Sampson (1998)). However, since the interest of the meta-analyst lies in the overall result, both the individual patient data and summary data should yield similar findings (Olkin and Sampson (1998)).

In performing meta-analysis, firstly, the results of each study via numerical indicators such as odds ratio, mean difference, or correlation coefficient has to be specified depending on the outcome measure. These effect size estimates reflect the degree of the association of interest in each study. The second stage involves pooling the effect size estimates from the individual studies to produce a single effect size that summarizes the relationship of interest across the different studies.

The *Forest* plot, shown in Figure 2.2 (adapted from R package, metafor), is the most common graph in meta-analysis reports. It indicates the effect estimate together with its confidence interval for each study and the overall summary effect. The rows on the plot represents each study. The effect size of each study is represented by the box in the midpoint of the line and bounded by the confidence interval, the length of the line, reflecting the precision with which the effect size has been estimated in the study. Thus for greater precision, the confidence interval has to be narrower as reflected by the summary effect (Borenstein et al. (2009)). The size of the box indicates the weight given to that study. The summary effect is usually shown by the diamond on the bottom row. The location of the diamond represents

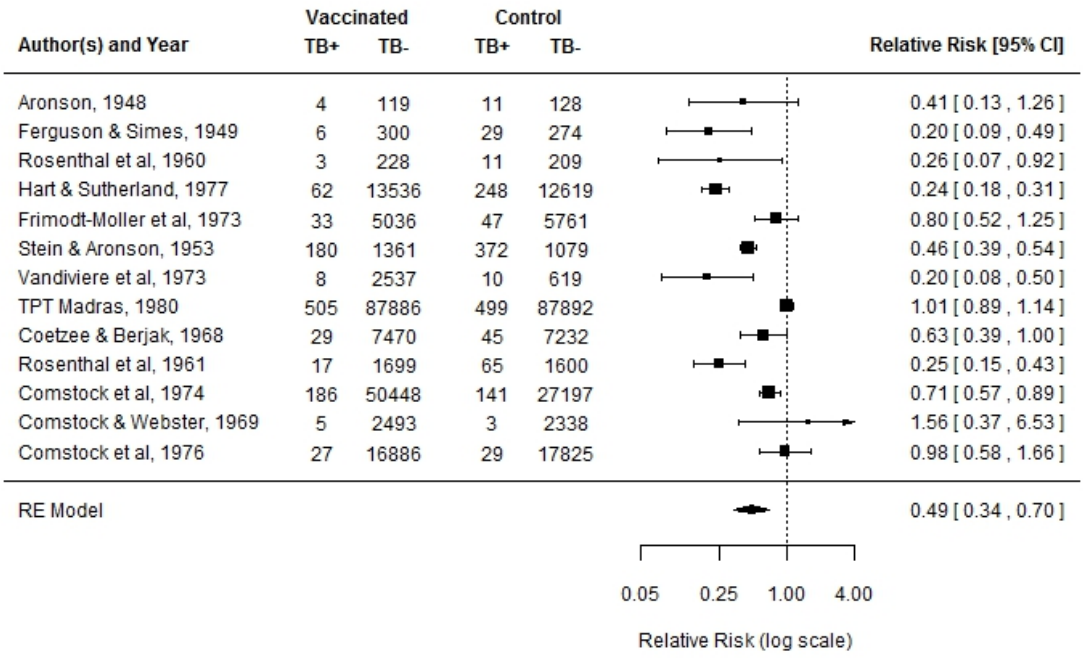

Figure 2.2: Forest plot showing the results of 13 studies examining the effectiveness of the BCG vaccine for preventing tuberculosis (adapted from "Metafor Package").

the effect size. The effect estimate from each individual study should ideally be on the same side as the summary effect and all their confidence intervals should overlap. The absence of overlap in confidence intervals may indicate the presence of heterogeneity in the included studies.

Exploring Heterogeneity

Systematic reviews and meta-analysis aim to capture the overall effects of an intervention or treatment when it has been tested in multiple trials. Ideally, if multiple

studies are testing the same intervention in similar patients, the effects of the intervention should be consistent across all studies. However, this is rarely the case because many factors can affect the results of a trial, such as researcher bias, problems with data collection, etc. (Higgins and Thompson (2002)). Exploring heterogeneity, the between studies variation, is thus inevitable for any meta-analysis before pooling the estimates of individual studies into a summary effect. Both statistical tests and graphical tools have been developed for detecting and quantifying inconsistency across studies and assessing the impact of heterogeneity on the meta-analysis.

Statistical tests for assessing heterogeneity are available but depend on the number of studies in the meta-analysis (Higgins et al. (2003)). The classical measure of heterogeneity is Cochran's chi-squared test of homogeneity, commonly referred to as Cochran's Q-test. Let Y_i denote the observed effect of the i^{th} study with its variance v_i, for $i = 1, 2, ..., k$, then the Q test statistic is given by:

$$Q \;=\; \sum_{i=1}^{k} w_i (Y_i - \hat{\theta})^2 \quad \sim \quad \chi^2_{k-1} \;\; \text{(under the null hypothesis of no heterogeneity)} \tag{2.1}$$

where $w_i = \frac{1}{v_i}$, is the weight of the i^{th} study, $\hat{\theta} = \frac{\sum_i^k w_i Y_i}{\sum_i^k w_i}$, is the estimate of the pooled effect size and k, the number of studies. P-value is obtained by comparing the statistic to a chi-squared distribution with $(k-1)$ degrees of freedom. The Q-test has a low power for a meta-analysis with small number of studies. Conversely, for larger number of studies, the Q-test has too much power. This means that, a non-significant result must not be taken as an evidence of no heterogeneity, whilst a statistically significant result may indicate the presence of variation between-studies

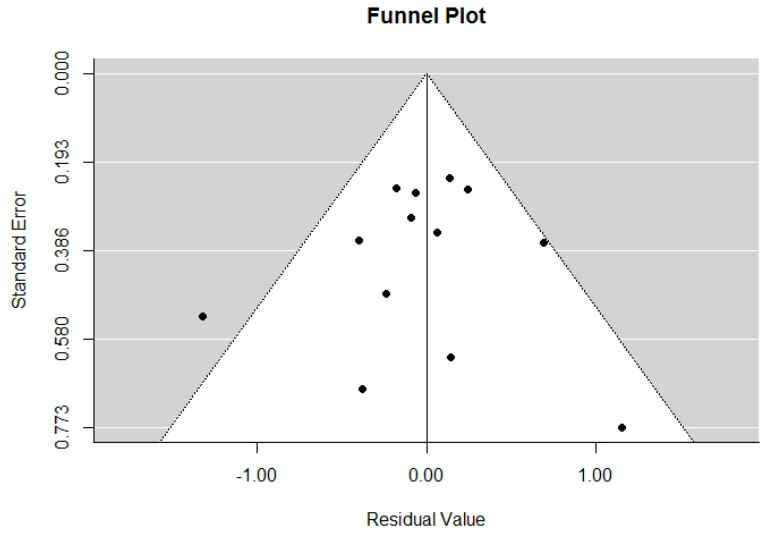

Figure 2.3: Funnel plot for a meta-analysis model (adapted from "Metafor Package").

(Higgins et al. (2003)). Thus, care must be taken in the interpretation of the Q-test.

However, meta-analysts are interested in the extent to which the heterogeneity affects the conclusions of the meta-analysis, but not just the indication of its presence or absence. A statistic proposed by Higgins et al. (2003) to quantify the degree of heterogeneity in meta-analysis is the value of the I^2 statistic given by:

$$I^2 = \left(\frac{Q - df}{Q}\right) \times 100\% \quad = \quad \begin{cases} 1, & \text{if } Q - (k-1) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{2.2}$$

where Q is the Cochran's Q test statistic, $df = k - 1$ is the degrees of freedom. I^2 describes the percentage of total variation across studies that is due to heterogeneity rather than chance (sampling error). A value of 0% indicates no observed heterogeneity, whilst larger values shows increasing variation across studies.

The funnel plot is a visual tool commonly employed in meta-analysis for investigating the presence of publication bias. A funnel plot shows the treatment effects estimated from individual studies (e.g. mean difference, log odds ratio, risk difference) on the horizontal axis and a measure related to the within-study variance (e.g. standard error, precision (inverse of standard error), variance) on the vertical axis. In practice, results from small studies are accompanied with larger variance compared to larger studies. That is, smaller studies have estimates that are more scattered and further away from the summary estimate. This creates a funnel-like plot, Figure 2.3, with the effect estimates from small studies scattered widely at the bottom of the plot. This spread is narrower for larger studies. The vertical line represents the null hypothesis value of no effect. The plot also indicates the effect summary based on the model with its accompanied confidence interval. The expected shape in the absence of publication bias corresponds to a symmetrical funnel. However, asymmetry could also result from heterogeneity of treatment effects. The funnel plot can be used to check for points lying outside of the confidence interval which indicates possible heterogeneity as seen in Figure 2.3.

Dealing with heterogeneity among study treatment effects is indispensable in meta-analysis. The most common approaches used in meta-analysis to address heterogeneity include:

- Fitting a meta-regression model that explains the heterogeneity in terms of study-level covariates.
- Stratify the studies into homogeneous subgroups and then fit a separate fixed-effect estimate.

- Construct a random-effects estimate across all studies.

Depending on the outcome of both the statistical and graphical tests of heterogeneity, a meta-analyst can then decide on fitting a fixed-effect model or a random-effects model in the meta-analysis.

2.2 Effect Size

In a meta-analysis, the measure reflecting the magnitude of the treatment effect or the relationship between two variables is termed the effect size. The effect size, unit of currency in meta-analysis (Borenstein et al. (2009)), is computed for each study and used to assess the consistency of the effect across studies and also to compute a summary effect. The effect size could be a representation of the impact of an intervention or could represent any relationship between two variables. The effect size is basically the estimate of a single value for each study which is usually accompanied with a confidence interval, demonstrated in the forest plot in Figure 2.2. The summary effect is the weighted mean of the individual effects. However, the method used to assign the weights depends on our assumption about the distribution of effect sizes from which the studies were sampled. In practice the type of outcome considered in the primary studies determines which effect measure to be employed in the meta-analysis.

Continuous Outcome

When summary data is reported on a numerical scale, the results are usually summarized as means and standard deviations. The outcome is thus referred to as continuous. The effect sizes computed for a continuous outcome such as weight change or blood pressure includes the mean difference, standardized mean difference and ratio of means.

	Number	Mean	Standard Deviation
Experiment	n_e	$\overline{X}_e$	s_e
Control	n_c	$\overline{X}_c$	s_c

Table 2.1: Data based on continuous outcome.

- Mean Difference (MD): Mean difference or raw mean difference is used when the outcome is reported on a meaningful scale and all the studies in the analysis use the same scale. An advantage of this effect measure is that it is intuitively meaningful either inherently or because of widespread use. Its estimate with standard deviation, S_{MD}, is given by:

$$MD = \overline{X}_e - \overline{X}_c, \quad S_{MD} = \left(\frac{s_e^2}{n_e} + \frac{s_c^2}{n_c}\right)^{1/2} \tag{2.3}$$

- Standardized Mean Difference (SMD): When different studies use different instruments to assess the same outcome, the scale of measurement will differ from study to study. Thus combining the raw mean differences would not be meaningful. In such instances, the standardized mean difference is employed. The estimate of the standardized mean difference with its standard deviation,

S_{SMD}, is given by:

$$SMD = f \times \frac{\bar{X}_e - \bar{X}_c}{s}, \quad S_{SMD} = \left(\frac{n_e + n_c}{n_e n_c} + \frac{SMD^2}{2(n_e + n_c)} \right)^{1/2} \tag{2.4}$$

where $s = \left(\frac{(n_e-1)s_e^2+(n_c-1)s_c^2}{n_e+n_c-2} \right)^{1/2}$ is the pooled standard deviation across the two groups and $f = \frac{4(n_e+n_c-2)-4}{4(n_e+n_c-2)-1}$.

- Ratio of Means (RoM): Ratio of means is preferred in research domains where the outcome is measured on a physical scale (e.g. length, area or mass) and is unlikely to be zero. Its estimate is given by:

$$RoM = \frac{\bar{X}_e}{\bar{X}_c} \tag{2.5}$$

The estimate of the natural logarithm of RoM with its standard deviation, $S_{ln(RoM)}$, is given by:

$$ln(RoM) = ln\left(\frac{\bar{X}_e}{\bar{X}_c} \right), \quad S_{ln(RoM)} = \left(\frac{1}{n_e} \left(\frac{s_e}{\bar{X}_e} \right)^2 + \frac{1}{n_c} \left(\frac{s_c}{\bar{X}_c} \right)^2 \right)^{1/2} \tag{2.6}$$

2.3 Effect Models

Meta-analysis usually involves describing the results of each study by means of a numerical index, an estimate of effect size such as mean difference, standardized mean difference, odds ratio or correlation coefficient. These estimates are then combined across studies to obtain a summary effect. Two different statistical models have

been developed for inference about average effect size, called fixed-effect and random-effects models. A third model known as the mixed-effects model, is an alternative model which arises in conjunction with analyses involving study-level covariates or moderator variables (Hedges et al. (1992)).

However, the choice between fixed-effect and random-effects procedure has been framed as entirely a question of homogeneity of effect size parameters (Hedges and Vevea (1998)). If all the studies estimate a common effect size parameter, then fixed-effect model is appropriate. The use of the random-effects model has been advocated if there is heterogeneity between study results, caused by different study populations across studies or methodological differences.

2.3.1 Fixed Effect Model

Under the fixed-effect model, there are basically two assumptions. The first assumption is that all the studies in the meta-analysis share a common (true) effect size. Secondly, it is assumed that, the difference between studies is due to the random error inherent in each study. Borenstein et al. (2009) adopted the wording *fixed-effect* without an 's' since there is only one true effect. Fixed-effect procedures are appropriate for making conditional inferences and thus the model conditions on the true effects and therefore provides a conditional inference about the set of k studies included in the meta-analysis (Hedges and Vevea (1998)).

For a set of $i = 1, ..., k$ independent studies, let Y_i denote the observed value of the effect in the i^{th} study and let θ represent the single true (unknown) effect. Then the fixed-effect model assumes:

$$Y_i|\theta \sim N(\theta, \sigma_i^2). \tag{2.7}$$

Equivalently, it can be written as:

$$Y_i = \theta + \epsilon_i \tag{2.8}$$

where ϵ_i is the error term for study i and its distribution is assumed to be normal, that is, $\epsilon_i \sim N(0, \sigma_i^2)$ and σ_i^2 denotes the within-study variance for study i. The pooled effect size can be obtained by weighting the observed effect in each study by their corresponding inverse variance in such a way that more weight is assigned to the studies that carry more information (i.e. studies with less variance). The weight for the i^{th} study is given as $w_i = \frac{1}{\sigma_i^2}$. Thus an estimate of the weighted average of the true effect size is given as:

$$\hat{\theta} = \frac{\sum_{i=1}^{k} w_i Y_i}{\sum_{i=1}^{k} w_i} \tag{2.9}$$

The variance of $\hat{\theta}$ is simply the reciprocal (inverse) of the sum of the weights, thus; $var(\hat{\theta}) = \frac{1}{\sum_{i=1}^{k} w_i}$, and the standard error $se(\hat{\theta})$ of $\hat{\theta}$ is just the square root of $var(\hat{\theta})$. Since the Y_i's are assumed normally distributed, it follows that $\hat{\theta}$ is also normally distributed. Thus a $100(1-\alpha)\%$ confidence interval for θ can be calculated as $\hat{\theta} \pm Z_{\alpha/2} se(\hat{\theta})$, where $Z_{\alpha/2}$ is the two tailed critical value of the standard normal distribution. The Z-value could be computed using $Z = \hat{\theta}/se(\hat{\theta}) \sim N(0,1)$ (Borenstein

et al. (2009)) to test the assumption of no treatment effect.

2.3.2 Random Effects Model

In general, when combining a group of studies in meta-analysis, it is assumed that the studies have enough in common which makes it logical for the information to be synthesized. In practice, studies will differ in the characteristics of the included samples and in the implementations of intervention or methods, thus there may be different effect sizes underlying different studies. Hence it is not realistic to assume that the studies are identical in the sense that the true effect size is exactly the same across all studies.

Random-effects model allows the effect sizes to vary between studies by assuming that they have a normal distribution around the mean effect. The measure of the between studies variance (variation among effect parameters across studies) is referred to as the amount of heterogeneity and is denoted by τ^2 .

There are basically two sources of variability in a set of studies in a meta-analysis. The within-study variability due to sampling error is inevitable in meta-analysis, since every study uses a different finite sample (as with fixed-effect approach). The other source of variability is heterogeneity. The between-studies variability can be accounted for by many factors. Some of these factors include:

- Clinical variation: difference in characteristics of participants, conditions under investigation, eligibility criteria for trials, dose variations, difference in outcome, etc.

- Methodological variation: diversity in design, (randomized versus non-randomized, cross over versus parallel, group versus cluster randomized, etc.), follow up durations, cut-off points on scales, etc.

It is appropriate to use the random-effects model when interest lies in generalizing the results of the analysis to a range of scenarios and the included studies are unlikely to be functionally equivalent (Borenstein et al. (2009)).

Random-effects analysis procedures are designed to facilitate unconditional inference about a larger set of studies from which the k studies included in the meta-analysis are assumed to be a random sample (Hedges and Vevea, 1998). Under the random-effects model the goal is not to estimate one true effect, but to estimate the mean of a distribution of effects. Treatment in each study is assumed to be randomly selected from a normal distribution with mean θ and variance τ^2. That is,
$Y_i|\theta_i \sim N(\theta_i, \sigma_i^2)$, where $\theta_i \sim N(\theta, \tau^2)$. Hence the random effects model is given by:

$$Y_i \sim N(\theta, \sigma_i^2 + \tau^2) \tag{2.10}$$

or equivalently,

$$Y_i = \theta + u_i + \epsilon_i \tag{2.11}$$

where $u_i \sim N(0, \tau^2)$ and $\epsilon_i \sim N(0, \sigma_i^2)$. Y_i is the observed effect for the i^{th} study, u_i is the deviation of the true effects in individual study's from the grand mean, θ, and ϵ_i the sampling error, is the deviation of the study's observed effect, Y_i from the study's true effect, θ_i. The overall weighted mean effect size, θ can be obtained similar to the fixed-effect model but the weight differs due to the additional between

study variance. Thus,

$$\hat{\theta} = \frac{\sum_{i=1}^{k} w_i^* Y_i}{\sum_{i=1}^{k} w_i^*} \tag{2.12}$$

where, $w_i^* = \frac{1}{\sigma_i^2 + \tau^2}$.

In order to compute a study's variance under the random-effects model, both the within-study variance (σ_i^2) and the between-study variance (τ^2) need to be computed. A practical method for computing the estimate of τ^2 (i.e. $\hat{\tau}^2$) is the method of moments (or the DerSimonian and Laird method), which proceeds as follows:

$$\hat{\tau}^2 = \begin{cases} \dfrac{Q - (k-1)}{c} & , if \quad Q > k-1 \\ 0 & , if \quad Q \leq k-1 \end{cases} \tag{2.13}$$

where, $c = \sum_{i=1}^{k} w_i - \frac{\sum_{i=1}^{k} w_i^2}{\sum_{i=1}^{k} w_i}$ and $w_i = \frac{1}{\sigma_i^2}$ for $i = 1, 2, ..., k$.

2.3.3 Mixed Effects Model

When study-level covariates or variables are incorporated in the random-effects model, the resulting model is known as mixed-effects model. These moderators may account for at least part of the heterogeneity in the true effects. The mixed-effects model is given by:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + u_i + \epsilon_i \tag{2.14}$$

where $u_i \sim N(0, \tau^2)$, $\epsilon_i \sim N(0, \sigma_i^2)$, X_{ij} denotes the value of the j^{th} moderator variable for the i^{th} study. τ^2 is the amount of heterogeneity among the true effects (i.e. the variability among the true effects that is not accounted for by the moderators included in the model).

2.4 Concept of Contamination

When investigating the effectiveness of a new intervention or treatment in a trial, not all subjects in the treatment group might adhere completely with the treatment regimen. Hence fitting a normal distribution to the data is not the best option in such cases.

Thus, despite the popularity and the simplicity of employing the use of the Normal distribution, it is too restrictive for modeling real datasets. In practice, when dealing with real applications, there is often a possibility of having a small to moderate percentage of contamination. Contaminated normal distributions have been adopted in many areas of application such as medicine, engineering, population genetics, biology, etc. (Figueiredo and Gomes (2013)).

Contaminated Normal Distribution (CND)

Contaminated normal distribution is the composite of two normal distributions. Thus, a contaminated normal distribution is given by the equation:

$$f = \lambda f_1 + (1 - \lambda) f_2 \tag{2.15}$$

with $f_1 \equiv N(\mu_1, \sigma_1^2)$, $f_2 \equiv N(\mu_2, \sigma_2^2)$ and $0 \leq \lambda \leq 1$.
Where λ represents the proportion of subjects that complied with the treatment or intervention and $(1 - \lambda)$ represents the proportion that did not comply with the treatment.

2.5 Frequentist Approach versus Bayesian Approach

Up until now, the methods discussed are based on conventional Frequentist approach. That is, the Frequentists use the sampling distribution as the basis of statistical inference that is proportional to the likelihood function. However, Bayesian methods involve a formal combination of a prior probability distribution that reflects a prior belief of the possible values of the pooled effect with a (likelihood) distribution of the pooled effect based on the observed data to obtain a posterior probability distribution of the pooled effect.

The posterior distribution, as obtained with the Bayesian approach, can be interpreted in terms of probabilities, which allows for a more intuitive interpretation (e.g. "There is an $x\%$ probability that treatment A results in a larger cholesterol reduction than treatment B") which is in contrast to findings with Frequentist approach. Another major advantage of the Bayesian approach is that the method naturally leads into a decision framework to support decision-making. Also, Bayesian meta-analysis includes a straightforward way of making predictions, and the possibility of incorporating different sources of uncertainty.

Bayesian Approach

Performing a Bayesian meta-analysis under a fixed effect model, the observed treatment effect Y_i is assumed to come from a normal distribution, $Y_i \sim N(\mu, s_i^2)$, where $E(Y_i) = \mu$ with known variance, $var(Y_i) = s_i^2$, for $i = 1, ..., k$ independent studies.

However, in random-effects model, where we assume a study specific mean effect θ_i, two levels are to be considered in modeling. The first level models the variability

of Y_i given θ_i and s_i^2. That is,

$$Y_i/\theta_i, s_i^2 \sim N(\theta_i, s_i^2) \tag{2.16}$$

The second level describes the variability of the study level parameters θ_i given the mean, μ and the between variance, τ^2. That is,

$$\theta_i/\mu, \tau^2 \sim N(\mu, \tau^2) \implies Y_i \sim N(\theta_i, \sigma_i^2 + \tau^2) \tag{2.17}$$

Placing priors on hyper-parameters (μ and τ^2) constitutes the hierarchical Bayesian model. In the absence of more specific information, we use non-informative (vague or flat) normal priors for μ and non-informative inverse-gamma or uniform prior for τ^2 is frequently used.

2.6 Data Format

In traditional meta-analysis, the input data are usually the summary statistics which is extracted from published literature (i.e. aggregate data (AD) or study-level data), rather than the primary or original data collected directly from trial units (i.e., patient-level data or individual patient data (IPD)). However, the study-level data is available in two formats namely:

1. Arm-level summary data: where the effect measures are reported for each arm (i.e., mean, hazard, absolute risk or odds).

2. Contrast-level summary data: where results are presented as the difference in

effect between arms (i.e., relative effect measures such as mean difference, hazard ratios, odds ratio or risk ratio).

Both the Bayesian and Frequentist approaches can be used to specify models based on either of the two data formats (Franchini et al. (2012)). One advantage of the arm-level summary approach is that, it is possible to adopt the exact likelihood for the data (i.e., binomial for binary data, Poisson for count data) rather than its normal approximation, as for the contrast-level summary (Greco et al. (2013)).

2.7 R-package: 'metafor'

Metafor is a statistical package implemented in the R programming language. The package evolved out of the *mima*() function written by Viechtbauer (2006). The package provides a comprehensive collection of functions for conducting meta-analysis. Functions for calculating various effect sizes or outcome measures, fitting meta-analytic fixed-, random- and mixed-effects are available. The package also includes various meta-analytical plot functions (e.g., forest plot, funnel plot) as well as functions to carry out moderator and meta-regression analyses. Functions for assessing model fit, tests of publication bias and obtaining case diagnostics are incorporated in the package.

The *metafor* package was used to carry out the fixed-effect meta-analysis in the simulation study presented in the next Chapter.

Chapter 3

Simulation Studies

3.1 Constructing the Simulations

In this section, we will describe how our data was generated and how the different simulation scenarios were obtained. The parameters to vary, the simulations settings considered and the number of simulation iterations considered in this thesis are also outlined. In order to enable a flexible and user-friendly coding environment to test many of the settings considered, the statistical package *metafor*, in R, was used to carry out the meta-analysis part of the simulations.

Certain outcomes (e.g., weight, blood pressure, etc.) are expressed as continuous data. These outcomes are reported by study investigators either as final (end of treatment) values or as change (change from baseline) values. Meta-analysis can then be performed to derive summary estimates for such outcomes. In this simulation study, randomized control trials with one experimental group and one control

group were simulated by generating change values for a continuous outcome. Contamination was introduced in the experimental group in such a way that we have a proportion (λ) adhering to treatment and a proportion ($1 - \lambda$) that did not adhere to treatment, constituting the experimental group. These change values are generated taking into consideration a number of different parameters. The change values were simulated by generating normal random variables. Randomized control trial data were simulated by generating change values for outcomes for each participant in both the experimental and control groups. Thus the change value for the experimental group comprises of both the adherence and the non-adherence groups. No effect was assumed for the control group and a small, medium, or large effect was assumed for the experimental group.

For the analysis in this simulation study, it was assumed that all the outcome measurements were on the same scale. Thus, a fixed-effect meta-analysis was performed on the mean differences to estimate the weighted mean difference (WMD) with its accompanied $100(1 - \alpha)\%$ confidence interval. A Monte Carlo simulation analysis was performed using R version 3.1.1. The *rnorm* function in R was used to generate random numbers from a normal distribution. Different seeds were used for each simulation setting.

In this simulation study, we aim to investigate the effect of selecting data points from a contaminated normal distribution on properties of estimators and properties of hypothesis tests in a traditional meta-analysis. The contaminated normal distribution was constructed empirically by mixing a combination of two normal distributions. Thus, a contamination of level ($1 - \lambda$) (proportion of subjects that

do not conform to treatment) was introduced in the mixing distribution. We also investigate how the number of studies and individual study sample size affect the properties of hypothesis tests (i.e., the power of the test and the size of the test) and the properties of estimators (i.e., the coverage probability, the width of confidence intervals, the bias and the mean square error (MSE)) for varying levels of contamination. Also, the effect of observing low compliance, moderate compliance and high compliance in the non-adherence group was investigated.

The parameters of interest and their assigned values considered for the various scenarios are indicated in Table 3.1. The assigned values considered were chosen such that they are representative of a typical meta-analysis (Clifton et al. (2014)). The level of contamination $(1 - \lambda)$ ranged from 0% to 50% to indicate an increasing proportion of patients who do not conform to treatment in the treatment group. The level of significance (α) was set at 0.01, 0.05 and 0.10. The number of studies per meta-analysis (k), ranged from 5 to 30. The sample size per trial arm (n), was varied between 10 to 100 in each of the experimental and control groups, and all trials within each simulation were assumed to have the same number of patients. The range of standard deviation for the control and the experimental groups was varied between 2 and 10 to reflect a narrow to broad distribution around the mean value. Also, the standard deviations of the experimental and the control groups were assumed to be equal for all simulation scenarios. The effect size was set at 0 under the null hypothesis and $-1, -2$, and -5 under the alternate hypotheses to reflect the increasing effect of the effect sizes from small to large. The effect measure considered is the mean difference (MD). Several different scenarios were simulated by varying

Table 3.1: Parameter values used in the simulated data sets.

Parameters	Assigned Values
Level of Contamination $(1 - \lambda)$	0%, 10%, 20%, 30%, 40%, 50%
Level of Significance (α)	0.01, 0.05, 0.10
Number of Studies (k)	5, 10, 30
Sample Size Per Arm (n)	10, 50, 100
Standard Deviation (SIG)	2, 5, 10
Effect Size (MD)	0, -1, -2, -5

the values of the parameters in Table 3.1.

3.2 Performance Evaluation

For each scenario, data points were generated and analyzed 1000 times and performance characteristics of the effect measure (mean difference) for the various contamination levels were assessed. These consisted of bias, computed as the difference between the average of the estimated mean difference from the 1000 datasets, and the true mean difference used to generate the data, mean square error calculated as the sum of the square of the average standard deviation of the 1000 samples and the square of the bias, the coverage probability of the $100(1 - \alpha)\%$ confidence interval of the simulated result i.e., the proportion of the time that the true parameter value falls within the $100(1 - \alpha)\%$ confidence interval of the simulated result were also calculated. The width of the confidence interval is computed as the average of the difference between the upper confidence limit (UCL) and the lower confidence limit (LCL) of the 1000 simulated datasets.

The statistical power is computed as the proportion of the time that the simulated

result yields a significant treatment effect under the alternative hypothesis. The size of the test was assessed by computing the statistical power when data is generated under the null hypothesis (i.e. when the mean difference is assumed to be zero).

3.3 Simulation Results

Concise yet complete simulation results are presented in this section. Each table and figure presents simulation results for specific scenarios determined by the set of parameters in Table 3.1.

Results Based on Estimation Properties

The values in each table include estimates of the bias, mean square error estimated from the meta-analysis model for the data points analyzed 1000 times. The width of the confidence intervals for the true mean difference are also included in the table. The graph of coverage probabilities are shown in the figures.

Figure 3.1 shows the plot of coverage probability for varying levels of contamination. The plot also includes varying mean differences while the other parameters are kept constant, i.e. level of significance at 5%, $n = 50$, $k = 10$, standard deviation at 5 and the change value for the non-adherence group kept at halfway between the adherence and the control group (moderate compliance). Figure 3.1 shows that the coverage probability for the true value is greatly influenced by contamination. Also, the coverage is affected by varying effect sizes. When we observe a small effect, moderate effect and high effect, the coverage is affected accordingly. The coverage

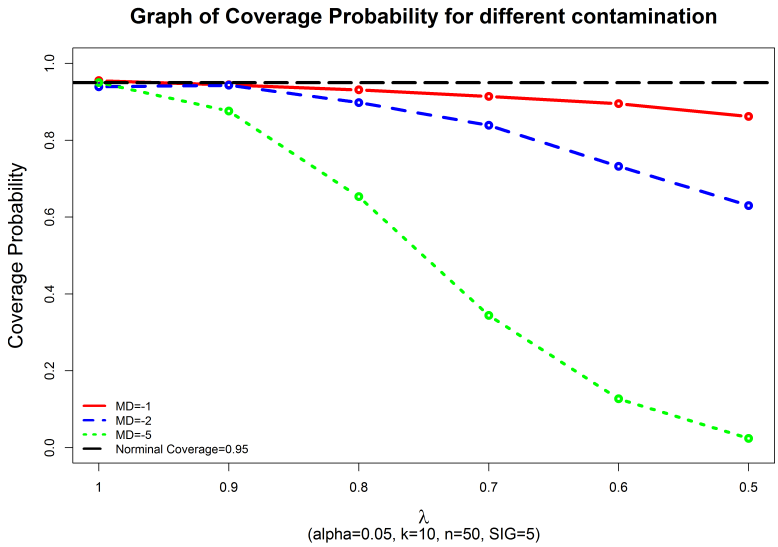

Figure 3.1: Effect of mean difference on coverage probability for varying levels of contamination for $\alpha = 0.05$, $k = 10$, $n = 50$, standard deviation $= 5$, moderate compliance.

probability for the scenario when there is no contamination is seen to have attained the nominal coverage of 95% and as the contamination level increases, the coverage is seen to be decreasing. The coverage probability is observed to be the least when we have 50% contamination (i.e. 50% of the subjects in the experimental group did not adhere to treatment). However, as the mean difference increases, the coverage probability tend to decrease.

Figure 3.2 depicts the effect of the standard deviation from 2 to 10 on the coverage probability when other parameters are kept constant (i.e. $\alpha = 0.05$, $n = 50$, $k = 10$, moderate compliance). When the standard deviation is at 2 (Figure 3.2 (a)), the coverage probabilities are very low as the level of contamination increases, compared to the higher coverage probabilities (Figure 3.2 (b)) for standard deviation at 10.

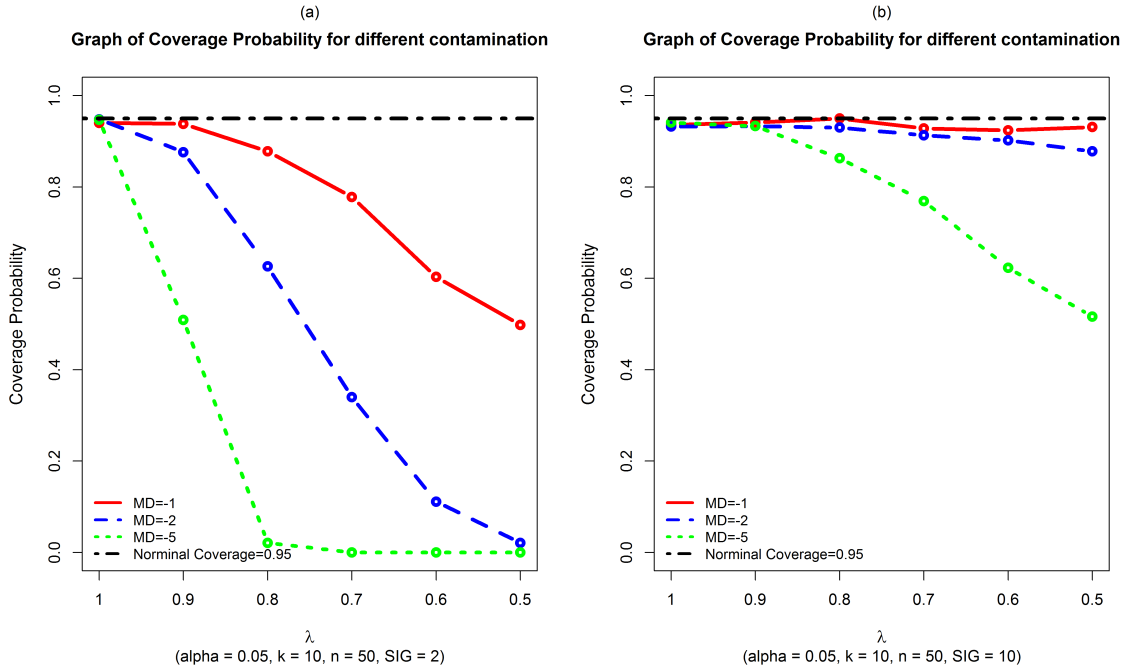

Figure 3.2: Effect of standard deviation (SIG=2 (a) and SIG=10 (b)) on coverage probability for varying levels of contamination for $\alpha = 0.05$, $k = 10$, $n = 50$, moderate compliance.

Figures 3.3 and 3.4 show the effect of number of studies, k, and sample size, n, respectively on the coverage probability as the level of contamination increases. A similar trend is seen in both figures. As both k and n increases (from Figure 3.3 (a) to Figure 3.3 (b) for k and from Figure 3.4 (a) to Figure 3.4 (b) for n), the coverage is seen to be decreasing. From Figure 3.4 (a), the coverage probability does not attain the nominal coverage despite 100% adherence. This is due to the small sample size considered, n=10.

We also investigated if the position of the change value in the non-adherence group has any effect on the coverage probability. Clearly, from Figure 3.5, when the

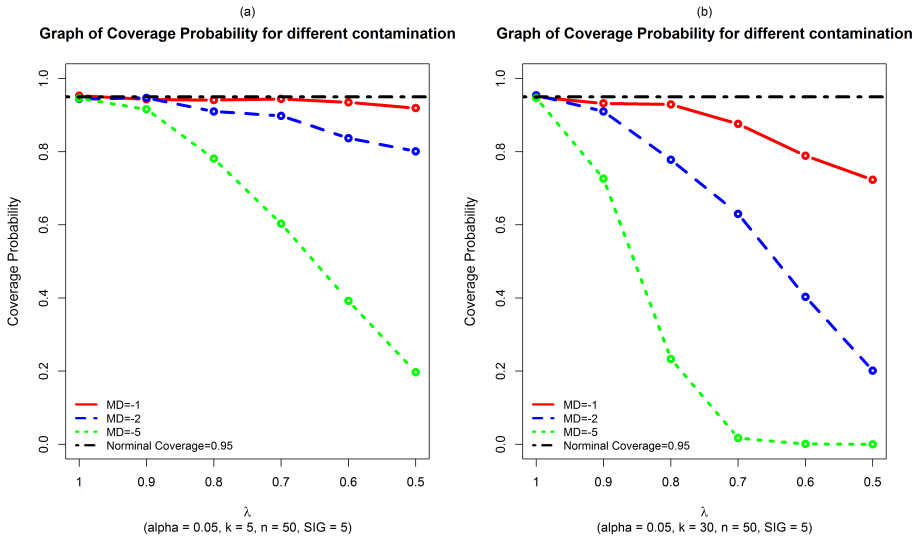

Figure 3.3: Effect of number of studies ($k = 5$ (a) and $k = 30$ (b)) on coverage probability for varying levels of contamination for $\alpha = 0.05$, $n = 50$, standard deviation = 5, moderate compliance.

change value in the non-adherence group is closer to the control group (low compliance) (Figure 3.5 (a)), the coverage probability is observed to decrease drastically as compared to Figure 3.5 (b), when the change value in the non-adherence group is closer to the adherence group (high compliance). The lowest coverage occurs when we observe 50% contamination across all scenarios.

Figures 3.6 (a) and (b) exhibits similar trend. The coverage probabilities decrease as the level of contamination increases for $\alpha = 0.01$ (3.6 (a)) and $\alpha = 0.1$ (3.6 (b)).

Table 3.2 presents simulation results for bias, MSE and interval width for the scenarios with underlying contaminated normal distribution of the data points. A significance level (α) at 0.01 with standard deviation at 2 for each combination of effect size $(-1, -2, -5)$, number of patients, n, (10,50,100) for both control and experimental groups and number of trials, k, (5,10,30). The position of the change

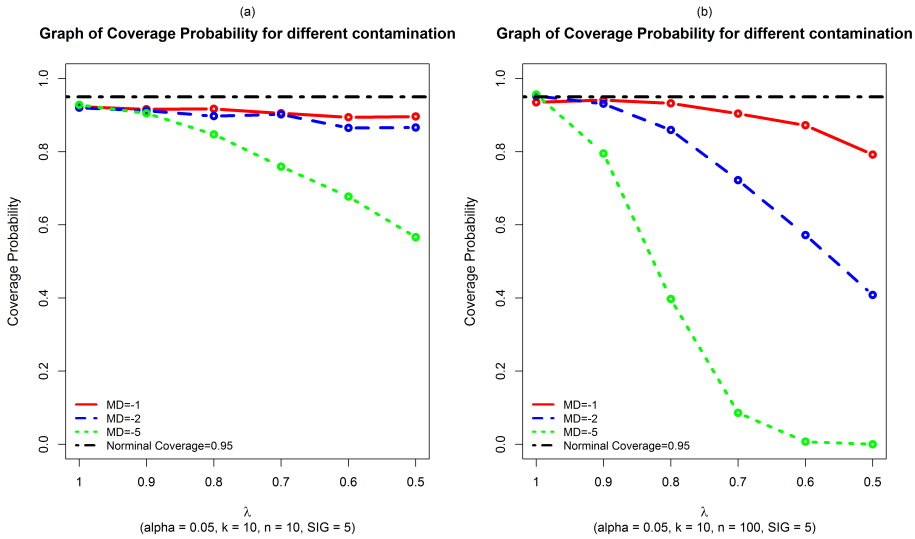

Figure 3.4: Effect of sample size ($n = 10$ (a) and $n = 100$ (b)) on coverage probability for varying levels of contamination for $\alpha = 0.05$, $k = 10$, standard deviation = 5, moderate compliance.

value for the non-adherence group fixed at halfway between the adherence group and the control group (moderate compliance). The "Bias" columns show the bias of each effect size for the various levels of contamination. The "MSE" columns show the mean square error and the "Width" columns show the width of the 90% confidence intervals for the true mean difference. Results for similar scenarios except with standard deviation at 5 are shown in Table 3.3. Tables 3.4 and 3.5 results were similar for $\alpha = 0.05$.

When we observe no contamination in the treatment group ($\lambda = 1$), we experienced minimal bias in almost all of the scenarios, as expected. In contrast, we observe maximum bias when there is 50% contamination in the experimental group. Thus the bias gradually increases as the level of contamination increases upto 50%. The bias is more prominent as the mean difference increases (from -1 to -5) in the

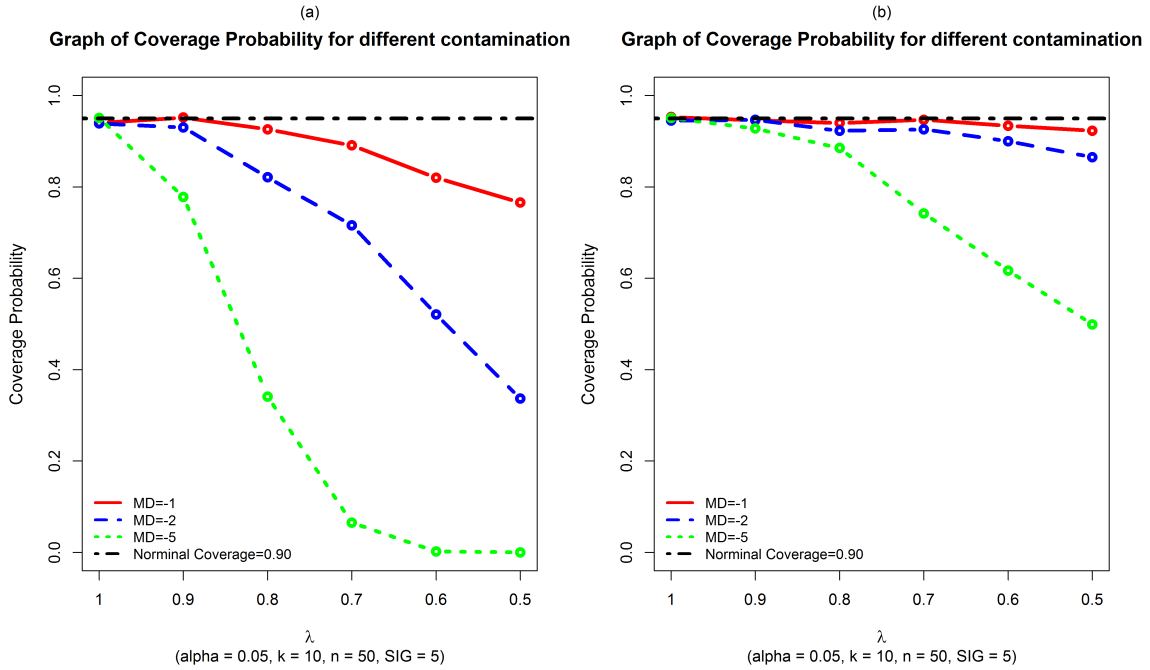

Figure 3.5: Effect of the position of the non-adherence change value (low compliance (a) and high compliance (b)) on coverage probability for varying levels of contamination for $\alpha = 0.05$, $k = 10$, $n = 50$, standard deviation $= 5$.

presence of contamination. The bias is accentuated by the change value in the experimental group being pooled towards the change value in the control group by the non-adherence group.

The mean square error, as expected, increases with increasing level of contamination across all tables. The mean square error is at highest when the number of trials, k, is 5 and the sample size, n, at 10 (i.e. least number of studies and sample size). The increase in the mean square error is more prominent by increase in the standard deviation demonstrated in Tables 3.2 and 3.3 when $\alpha = 0.01$ and Tables 3.4 and 3.5 when $\alpha = 0.05$. Increasing the mean difference, the mean square error increases accordingly in the presence of contamination due to the increase in the bias.

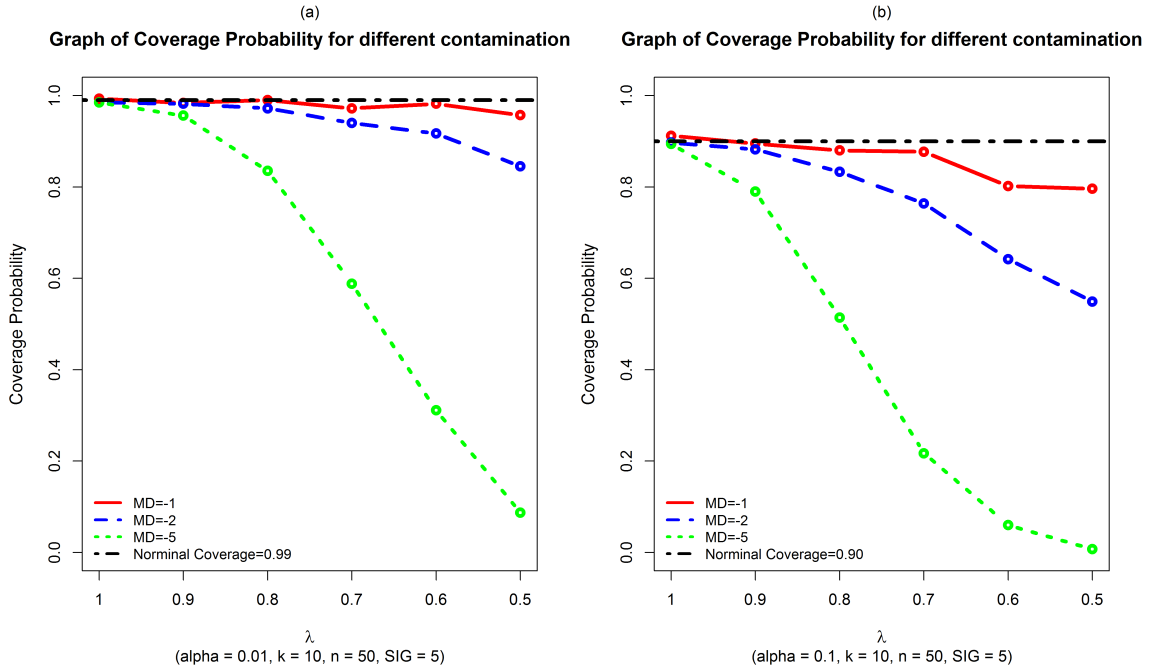

Figure 3.6: Effect of significance level ($\alpha = 0.01$ (a) and $\alpha = 0.1$ (b)) on coverage probability for varying levels of contamination for, $k = 10$, $n = 50$, standard deviation = 5, moderate compliance.

The width of the $100(1-\alpha)\%$ confidence interval increases for increasing levels of contamination. The width of the confidence interval is tighter when we have large number of trials (k=30) and large number of patients per trial arm (n=100) across all scenarios. As observed in Table 3.2 and Table 3.4 for standard deviation at 2, the width of the confidence interval for the true mean difference increases steadily as the level of contamination increases since we have less variance. In Tables 3.3 and 3.5 when the standard deviation is at 5, and with small mean difference, the effect of the width is not evident as the contamination level increases and this is due to the fact that confidence intervals are greatly influenced by standard deviation and thus the mean difference has to be large enough to detect the influence of contamination

Table 3.2: Simulation Results for bias, MSE and interval width (Significance level of 0.01, Standard deviation = 2, change value for non-adherence group fixed at halfway between the adherence group and the control group (moderate compliance)).

			Bias				MSE				Width			
MD	n	k	$\lambda = 1$	$\lambda = 0.9$	$\lambda = 0.7$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 0.9$	$\lambda = 0.7$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 0.9$	$\lambda = 0.7$	$\lambda = 0.5$
-1	10	5	-0.008	0.047	0.159	0.255	0.166	0.172	0.212	0.241	1.964	1.968	1.970	1.969
-1	10	10	-0.013	0.054	0.163	0.267	0.083	0.092	0.121	0.163	1.382	1.383	1.386	1.385
-1	10	30	0.010	0.048	0.154	0.259	0.030	0.033	0.054	0.095	0.794	0.796	0.798	0.798
-1	50	5	0.002	0.048	0.145	0.252	0.034	0.034	0.053	0.095	0.915	0.915	0.918	0.918
-1	50	10	-0.004	0.051	0.155	0.248	0.016	0.018	0.040	0.078	0.646	0.647	0.647	0.648
-1	50	30	-0.002	0.050	0.150	0.253	0.005	0.008	0.028	0.069	0.372	0.373	0.374	0.374
-1	100	5	0.000	0.056	0.152	0.251	0.016	0.019	0.040	0.078	0.649	0.651	0.651	0.651
-1	100	10	-0.001	0.048	0.147	0.252	0.008	0.010	0.030	0.071	0.459	0.459	0.460	0.461
-1	100	30	0.000	0.051	0.149	0.250	0.003	0.005	0.025	0.065	0.265	0.265	0.266	0.266
-2	10	5	-0.031	0.086	0.279	0.491	0.180	0.184	0.259	0.421	1.966	1.965	1.992	2.006
-2	10	10	-0.022	0.095	0.299	0.508	0.082	0.108	0.180	0.347	1.382	1.390	1.397	1.405
-2	10	30	-0.008	0.092	0.301	0.501	0.031	0.038	0.118	0.280	0.796	0.800	0.806	0.808
-2	50	5	0.003	0.109	0.310	0.505	0.034	0.045	0.126	0.289	0.912	0.919	0.926	0.926
-2	50	10	0.002	0.097	0.305	0.500	0.015	0.027	0.110	0.266	0.646	0.649	0.653	0.656
-2	50	30	0.003	0.106	0.300	0.500	0.005	0.017	0.095	0.255	0.372	0.375	0.378	0.379
-2	100	5	0.001	0.098	0.299	0.500	0.018	0.025	0.106	0.266	0.650	0.653	0.657	0.658
-2	100	10	0.002	0.098	0.303	0.503	0.008	0.017	0.099	0.261	0.458	0.461	0.465	0.466
-2	100	30	0.005	0.103	0.304	0.496	0.002	0.013	0.095	0.249	0.265	0.266	0.268	0.269
-5	10	5	-0.013	0.264	0.750	1.258	0.182	0.238	0.735	1.752	1.956	2.042	2.132	2.171
-5	10	10	-0.010	0.257	0.767	1.243	0.081	0.155	0.677	1.636	1.377	1.431	1.506	1.532
-5	10	30	0.000	0.239	0.743	1.253	0.030	0.086	0.581	1.601	0.795	0.825	0.864	0.879
-5	50	5	0.009	0.253	0.750	1.253	0.032	0.099	0.596	1.603	0.914	0.945	0.987	1.001
-5	50	10	-0.003	0.249	0.746	1.248	0.017	0.078	0.574	1.575	0.645	0.669	0.698	0.707
-5	50	30	0.003	0.247	0.750	1.249	0.006	0.066	0.567	1.566	0.373	0.386	0.402	0.408
-5	100	5	-0.006	0.252	0.748	1.253	0.017	0.079	0.576	1.586	0.648	0.671	0.701	0.710
-5	100	10	-0.003	0.256	0.747	1.249	0.009	0.074	0.567	1.569	0.459	0.474	0.496	0.501
-5	100	30	0.000	0.252	0.751	1.250	0.003	0.066	0.567	1.566	0.265	0.274	0.286	0.290

on the width of the confidence interval.

In Table 3.6, the change value in the non-adherence group was considered to be closer to the control group (i.e. low compliance) and in Table 3.7, the change value in the non-adherence group is closer to the adherence group (i.e. high compliance). For low compliance, the bias is higher, the MSE is higher and an increment in the width, as compared to when we observe high compliance.

Table 3.3: Simulation Results for bias, MSE and interval width (Significance level of 0.01, Standard deviation = 5, change value for non-adherence group fixed at halfway between the adherence group and the control group (moderate compliance)).

			Bias				MSE				Width			
MD	n	k	$\lambda = 1$	$\lambda = 0.9$	$\lambda = 0.7$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 0.9$	$\lambda = 0.7$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 0.9$	$\lambda = 0.7$	$\lambda = 0.5$
-1	10	5	0.020	0.024	0.154	0.266	1.123	1.125	1.137	1.207	4.912	4.909	4.894	4.894
-1	10	10	-0.001	0.054	0.143	0.259	0.527	0.596	0.567	0.597	3.463	3.454	3.459	3.461
-1	10	30	0.010	0.064	0.150	0.260	0.197	0.195	0.202	0.237	1.989	1.988	1.988	1.990
-1	50	5	-0.002	0.052	0.161	0.223	0.211	0.197	0.230	0.253	2.283	2.286	2.288	2.282
-1	50	10	-0.003	0.051	0.163	0.279	0.103	0.105	0.137	0.177	1.613	1.613	1.615	1.615
-1	50	30	0.014	0.045	0.156	0.252	0.033	0.040	0.057	0.098	0.931	0.932	0.933	0.931
-1	100	5	-0.009	0.038	0.138	0.265	0.100	0.110	0.117	0.177	1.619	1.625	1.624	1.623
-1	100	10	-0.005	0.065	0.154	0.244	0.049	0.054	0.073	0.107	1.147	1.147	1.147	1.146
-1	100	30	0.004	0.052	0.155	0.243	0.018	0.021	0.041	0.076	0.662	0.662	0.662	0.662
-2	10	5	-0.004	0.054	0.320	0.502	1.102	1.079	1.268	1.350	4.903	4.891	4.910	4.960
-2	10	10	-0.007	0.068	0.231	0.488	0.600	0.611	0.617	0.776	3.457	3.441	3.453	3.452
-2	10	30	-0.030	0.105	0.286	0.511	0.189	0.195	0.275	0.440	1.988	1.990	1.991	1.992
-2	50	5	0.011	0.125	0.297	0.523	0.206	0.228	0.273	0.483	2.280	2.286	2.283	2.290
-2	50	10	0.007	0.102	0.297	0.494	0.102	0.119	0.194	0.346	1.612	1.615	1.618	1.619
-2	50	30	0.004	0.108	0.299	0.499	0.035	0.046	0.125	0.281	0.932	0.932	0.933	0.933
-2	100	5	-0.009	0.092	0.321	0.488	0.097	0.101	0.198	0.332	1.620	1.622	1.626	1.624
-2	100	10	0.010	0.100	0.299	0.498	0.052	0.066	0.143	0.296	1.146	1.147	1.150	1.150
-2	100	30	0.003	0.100	0.300	0.501	0.017	0.029	0.107	0.268	0.662	0.662	0.663	0.664
-5	10	5	-0.030	0.267	0.747	1.250	1.104	1.105	1.656	2.592	4.922	4.952	5.000	4.988
-5	10	10	-0.002	0.250	0.778	1.231	0.523	0.601	1.137	2.080	3.452	3.456	3.501	3.507
-5	10	30	0.012	0.234	0.764	1.258	0.190	0.238	0.776	1.774	1.988	1.999	2.011	2.024
-5	50	5	-0.010	0.244	0.721	1.285	0.181	0.261	0.718	1.853	2.285	2.296	2.312	2.319
-5	50	10	-0.014	0.242	0.749	1.258	0.103	0.165	0.656	1.683	1.614	1.624	1.634	1.637
-5	50	30	0.014	0.259	0.751	1.247	0.034	0.102	0.602	1.589	0.932	0.936	0.943	0.946
-5	100	5	-0.005	0.226	0.755	1.260	0.104	0.153	0.660	1.687	1.621	1.630	1.642	1.649
-5	100	10	-0.006	0.256	0.738	1.241	0.047	0.114	0.594	1.590	1.147	1.154	1.163	1.164
-5	100	30	0.001	0.248	0.753	1.251	0.018	0.079	0.582	1.583	0.662	0.666	0.671	0.672

Results Based on Hypothesis Tests Properties

We now present results based on hypothesis tests properties. Statistical power (the proportion of scenarios yielding a significant treatment effect) are presented in Figures 3.7 to 3.12. The size of the test are also indicated on the figures.

Figure 3.7 shows the plot of power for varying levels of contamination. The plot also includes varying mean differences while keeping other parameters constant, i.e. level of significance at 5%, n=50, k=10, standard deviation at 5 and the change value for the non-adherence group kept at halfway between the adherence and the control

Table 3.4: Simulation Results for bias, MSE and interval width (Significance level of 0.05, Standard deviation = 2, change value for non-adherence group fixed at halfway between the adherence group and the control group (moderate compliance)).

			Bias				MSE				Width			
MD	n	k	$\lambda = 1$	$\lambda = 0.9$	$\lambda = 0.7$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 0.9$	$\lambda = 0.7$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 0.9$	$\lambda = 0.7$	$\lambda = 0.5$
-1	10	5	0.027	0.051	0.153	0.224	0.170	0.189	0.190	0.238	1.492	1.496	1.500	1.498
-1	10	10	0.004	0.051	0.153	0.268	0.090	0.091	0.118	0.157	1.051	1.053	1.054	1.053
-1	10	30	-0.004	0.046	0.151	0.248	0.032	0.034	0.052	0.092	0.605	0.606	0.607	0.607
-1	50	5	0.011	0.044	0.151	0.263	0.031	0.038	0.054	0.100	0.696	0.696	0.697	0.698
-1	50	10	-0.000	0.046	0.148	0.248	0.017	0.018	0.039	0.078	0.491	0.492	0.493	0.493
-1	50	30	0.000	0.048	0.151	0.251	0.005	0.007	0.028	0.068	0.284	0.284	0.284	0.285
-1	100	5	0.007	0.050	0.148	0.250	0.016	0.019	0.038	0.078	0.494	0.495	0.496	0.495
-1	100	10	-0.002	0.052	0.152	0.245	0.008	0.011	0.032	0.068	0.349	0.350	0.350	0.350
-1	100	30	0.001	0.050	0.150	0.249	0.003	0.005	0.025	0.065	0.201	0.202	0.202	0.202
-2	10	5	-0.012	0.111	0.302	0.510	0.165	0.187	0.277	0.449	1.493	1.512	1.515	1.519
-2	10	10	-0.016	0.107	0.310	0.501	0.090	0.101	0.190	0.341	1.053	1.056	1.068	1.067
-2	10	30	-0.003	0.103	0.298	0.502	0.028	0.041	0.116	0.284	0.606	0.608	0.613	0.614
-2	50	5	-0.002	0.095	0.293	0.499	0.032	0.040	0.119	0.284	0.696	0.700	0.704	0.707
-2	50	10	0.001	0.102	0.303	0.499	0.016	0.026	0.108	0.264	0.491	0.494	0.498	0.499
-2	50	30	-0.001	0.101	0.299	0.498	0.006	0.016	0.095	0.254	0.283	0.285	0.287	0.288
-2	100	5	-0.001	0.105	0.301	0.502	0.016	0.029	0.105	0.268	0.494	0.497	0.501	0.501
-2	100	10	-0.007	0.096	0.302	0.499	0.008	0.018	0.099	0.257	0.349	0.351	0.353	0.354
-2	100	30	0.002	0.100	0.298	0.502	0.003	0.013	0.091	0.254	0.201	0.203	0.204	0.205
-5	10	5	0.004	0.243	0.737	1.239	0.176	0.230	0.726	1.703	1.492	1.554	1.620	1.662
-5	10	10	0.006	0.261	0.743	1.242	0.091	0.161	0.643	1.636	1.046	1.093	1.144	1.162
-5	10	30	0.011	0.244	0.746	1.256	0.029	0.094	0.587	1.610	0.605	0.627	0.658	0.667
-5	50	5	0.007	0.252	0.748	1.258	0.035	0.094	0.594	1.614	0.695	0.720	0.751	0.760
-5	50	10	-0.003	0.252	0.751	1.258	0.016	0.081	0.580	1.600	0.491	0.508	0.531	0.538
-5	50	30	0.004	0.251	0.755	1.250	0.006	0.068	0.575	1.568	0.283	0.293	0.306	0.311
-5	100	5	0.007	0.253	0.755	1.249	0.016	0.079	0.585	1.575	0.494	0.511	0.533	0.541
-5	100	10	-0.002	0.255	0.747	1.250	0.008	0.073	0.565	1.571	0.349	0.361	0.377	0.382
-5	100	30	-0.001	0.247	0.750	1.248	0.003	0.064	0.565	1.559	0.201	0.209	0.218	0.220

group (moderate compliance).

Figure 3.7 shows that the statistical power is influenced by contamination. Also, the power is affected by varying effect sizes. When we observe small effect, moderate effect and high effect, the power is being affected accordingly as seen in Figure 3.7. The power for the scenario when there is no contamination is seen to be high as compared to when contamination is introduced. Thus the power is observed to be the least when we have 50% contamination (i.e. 50% of the subjects in the experimental group did not adhere completely to treatment). However, as the mean difference increases, the power tends to increase.

Table 3.5: Simulation Results for bias, MSE and interval width (Significance level of 0.05, Standard deviation = 5, change value for non-adherence group fixed at halfway between the adherence group and the control group (moderate compliance)).

			Bias				MSE				Width			
MD	n	k	$\lambda = 1$	$\lambda = 0.9$	$\lambda = 0.7$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 0.9$	$\lambda = 0.7$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 0.9$	$\lambda = 0.7$	$\lambda = 0.5$
-1	10	5	0.030	0.095	0.206	0.228	1.138	1.034	1.150	1.151	3.715	3.744	3.731	3.755
-1	10	10	-0.002	0.060	0.143	0.219	0.526	0.560	0.608	0.618	2.628	2.630	2.628	2.620
-1	10	30	-0.023	0.036	0.140	0.247	0.189	0.179	0.208	0.247	1.513	1.510	1.512	1.513
-1	50	5	-0.006	0.051	0.137	0.231	0.196	0.209	0.220	0.256	1.737	1.736	1.739	1.741
-1	50	10	-0.001	0.034	0.151	0.255	0.099	0.103	0.126	0.167	1.228	1.228	1.228	1.227
-1	50	30	0.001	0.059	0.141	0.247	0.033	0.040	0.051	0.095	0.708	0.708	0.709	0.709
-1	100	5	0.012	0.059	0.153	0.247	0.112	0.110	0.117	0.155	1.235	1.235	1.234	1.236
-1	100	10	-0.002	0.039	0.154	0.248	0.053	0.055	0.074	0.114	0.873	0.872	0.872	0.872
-1	100	30	-0.001	0.053	0.148	0.248	0.018	0.020	0.039	0.079	0.503	0.504	0.504	0.504
-2	10	5	-0.037	0.098	0.315	0.437	1.127	1.204	1.195	1.367	3.752	3.748	3.740	3.743
-2	10	10	-0.011	0.104	0.235	0.475	0.539	0.571	0.603	0.806	2.634	2.624	2.637	2.642
-2	10	30	0.010	0.114	0.319	0.524	0.187	0.204	0.299	0.476	1.511	1.511	1.515	1.515
-2	50	5	0.007	0.073	0.314	0.493	0.212	0.214	0.291	0.448	1.734	1.738	1.739	1.742
-2	50	10	-0.012	0.109	0.290	0.508	0.109	0.102	0.187	0.362	1.229	1.229	1.230	1.232
-2	50	30	-0.001	0.095	0.297	0.510	0.032	0.044	0.124	0.294	0.708	0.709	0.710	0.710
-2	100	5	-0.009	0.108	0.308	0.491	0.093	0.111	0.201	0.342	1.232	1.236	1.238	1.237
-2	100	10	0.001	0.104	0.300	0.494	0.051	0.060	0.144	0.296	0.872	0.874	0.874	0.874
-2	100	30	0.004	0.103	0.296	0.505	0.016	0.027	0.103	0.274	0.504	0.504	0.504	0.505
-5	10	5	0.039	0.211	0.801	1.227	1.109	1.133	1.665	2.577	3.729	3.757	3.792	3.802
-5	10	10	-0.006	0.273	0.770	1.200	0.528	0.613	1.149	1.986	2.624	2.647	2.665	2.679
-5	10	30	-0.014	0.243	0.767	1.247	0.181	0.248	0.772	1.745	1.512	1.520	1.534	1.537
-5	50	5	-0.016	0.275	0.754	1.259	0.211	0.270	0.777	1.789	1.739	1.748	1.761	1.767
-5	50	10	0.004	0.259	0.748	1.251	0.096	0.168	0.657	1.659	1.227	1.234	1.245	1.247
-5	50	30	-0.007	0.239	0.739	1.250	0.034	0.092	0.578	1.595	0.709	0.712	0.717	0.720
-5	100	5	-0.007	0.252	0.729	1.244	0.100	0.167	0.634	1.650	1.236	1.242	1.251	1.255
-5	100	10	-0.003	0.250	0.753	1.251	0.049	0.109	0.618	1.615	0.872	0.877	0.884	0.886
-5	100	30	-0.004	0.240	0.746	1.256	0.016	0.075	0.573	1.594	0.504	0.507	0.510	0.511

Figure 3.8 depicts the effect of the standard deviation from 2 to 10 on the statistical power when other parameters are kept constant (i.e. $\alpha = 0.05$, n=50, k=10, moderate compliance). When the standard deviation is at 2, Figure 3.8 (a), we observe very high power compared to Figure 3.8 (b) for standard deviation at 10.

Figures 3.9 and 3.10 show the effect of number of studies, k, and sample size, n, respectively on the power as the level of contamination increases. A similar trend is seen in both figures. As both k and n increases (from Figure 3.9 (a) to Figure 3.9 (b) for k and from Figure 3.10 (a) to Figure 3.10 (b) for n), the power is seen to be increasing. From Figure 3.10 (a), it has been observed that, we do not attain the size

Table 3.6: Simulation Results for bias, MSE and interval width (Significance level of 0.05, Standard deviation = 2, change value for non-adherence group closer to the control group (low compliance)).

			Bias				MSE				Width			
MD	n	k	$\lambda=1$	$\lambda=0.9$	$\lambda=0.7$	$\lambda=0.5$	$\lambda=1$	$\lambda=0.9$	$\lambda=0.7$	$\lambda=0.5$	$\lambda=1$	$\lambda=0.9$	$\lambda=0.7$	$\lambda=0.5$
-1	10	5	0.008	0.087	0.237	0.376	0.189	0.189	0.224	0.298	1.490	1.499	1.509	1.504
-1	10	10	0.003	0.084	0.223	0.372	0.088	0.105	0.145	0.229	1.051	1.055	1.059	1.061
-1	10	30	-0.004	0.068	0.219	0.376	0.030	0.036	0.076	0.171	0.604	0.606	0.609	0.609
-1	50	5	-0.005	0.076	0.227	0.372	0.034	0.039	0.084	0.171	0.696	0.697	0.700	0.701
-1	50	10	0.005	0.079	0.228	0.375	0.017	0.023	0.069	0.156	0.491	0.493	0.495	0.496
-1	50	30	0.000	0.072	0.223	0.378	0.006	0.010	0.055	0.148	0.283	0.284	0.286	0.286
-1	100	5	-0.001	0.073	0.221	0.372	0.016	0.020	0.065	0.154	0.494	0.496	0.497	0.498
-1	100	10	-0.005	0.073	0.222	0.376	0.008	0.013	0.058	0.149	0.349	0.350	0.352	0.352
-1	100	30	-0.000	0.078	0.225	0.374	0.003	0.009	0.054	0.142	0.201	0.202	0.203	0.203
-2	10	5	0.003	0.144	0.476	0.755	0.178	0.204	0.406	0.751	1.497	1.513	1.536	1.550
-2	10	10	0.004	0.157	0.458	0.749	0.097	0.113	0.296	0.655	1.050	1.065	1.087	1.091
-2	10	30	-0.001	0.150	0.450	0.757	0.030	0.052	0.233	0.602	0.605	0.613	0.624	0.628
-2	50	5	-0.001	0.143	0.450	0.741	0.032	0.051	0.234	0.582	0.695	0.704	0.716	0.720
-2	50	10	0.001	0.142	0.448	0.748	0.017	0.037	0.217	0.575	0.491	0.498	0.506	0.508
-2	50	30	0.004	0.151	0.448	0.749	0.005	0.028	0.206	0.566	0.283	0.287	0.292	0.293
-2	100	5	0.000	0.150	0.452	0.753	0.016	0.038	0.220	0.584	0.494	0.500	0.509	0.511
-2	100	10	0.000	0.149	0.448	0.750	0.008	0.030	0.209	0.570	0.349	0.353	0.359	0.361
-2	100	30	-0.001	0.148	0.449	0.752	0.003	0.025	0.204	0.568	0.202	0.204	0.207	0.208
-5	10	5	-0.016	0.403	1.142	1.878	0.175	0.342	1.468	3.684	1.494	1.627	1.780	1.831
-5	10	10	0.013	0.374	1.128	1.874	0.089	0.225	1.365	3.599	1.047	1.140	1.250	1.290
-5	10	30	0.005	0.377	1.119	1.872	0.030	0.174	1.283	3.533	0.604	0.657	0.722	0.742
-5	50	5	0.002	0.364	1.132	1.883	0.032	0.166	1.314	3.578	0.695	0.749	0.816	0.838
-5	50	10	-0.004	0.375	1.130	1.878	0.016	0.158	1.294	3.544	0.492	0.529	0.577	0.592
-5	50	30	-0.005	0.372	1.126	1.877	0.005	0.144	1.273	3.526	0.284	0.305	0.333	0.341
-5	100	5	-0.003	0.377	1.127	1.877	0.016	0.158	1.287	3.541	0.494	0.533	0.579	0.593
-5	100	10	0.002	0.371	1.120	1.872	0.008	0.146	1.261	3.511	0.349	0.376	0.409	0.419
-5	100	30	-0.001	0.376	1.126	1.873	0.003	0.144	1.271	3.511	0.201	0.217	0.236	0.242

of the test (approximately 0.05) despite 100% adherence. This is due to the small sample size considered (i.e. n=10).

Figures 3.11 (a) and (b) exhibits similar trend. Thus, the power decreases as the level of contamination increases and increases as the mean difference increases for alpha=0.01 (Figure 3.11 (a)) and alpha=0.1 (Figure 3.11 (b)).

We also investigated if the position of the change value in the non-adherence group has any effect on the power. Clearly, from Figure 3.12, when the change value in the non-adherence group is closer to the control group (low compliance) (Figure 3.12 (a)), the power is observed to decrease as compared with Figure 3.12 (b) , when

Table 3.7: Simulation Results for bias, MSE and interval width (Significance level of 0.05, Standard deviation = 2, change value for non-adherence group closer to the adherence group (high compliance)).

			Bias				MSE				Width			
MD	n	k	$\lambda = 1$	$\lambda = 0.9$	$\lambda = 0.7$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 0.9$	$\lambda = 0.7$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 0.9$	$\lambda = 0.7$	$\lambda = 0.5$
-1	10	5	-0.008	0.052	0.082	0.128	0.179	0.162	0.183	0.194	1.494	1.493	1.497	1.495
-1	10	10	0.015	0.038	0.086	0.112	0.095	0.101	0.101	0.094	1.050	1.050	1.051	1.055
-1	10	30	-0.011	0.020	0.068	0.129	0.030	0.029	0.033	0.047	0.604	0.605	0.605	0.605
-1	50	5	-0.001	0.034	0.074	0.130	0.032	0.034	0.037	0.048	0.695	0.694	0.695	0.694
-1	50	10	-0.006	0.026	0.072	0.128	0.015	0.017	0.021	0.032	0.491	0.491	0.491	0.492
-1	50	30	0.000	0.027	0.076	0.124	0.005	0.006	0.011	0.021	0.284	0.283	0.284	0.284
-1	100	5	-0.004	0.030	0.074	0.126	0.016	0.017	0.021	0.032	0.493	0.493	0.495	0.494
-1	100	10	0.000	0.028	0.075	0.118	0.008	0.009	0.014	0.022	0.349	0.349	0.349	0.349
-1	100	30	-0.001	0.026	0.077	0.124	0.002	0.003	0.009	0.018	0.202	0.201	0.202	0.202
-2	10	5	-0.001	0.049	0.137	0.259	0.175	0.183	0.195	0.237	1.497	1.492	1.505	1.497
-2	10	10	0.005	0.053	0.141	0.241	0.092	0.094	0.104	0.151	1.048	1.050	1.052	1.057
-2	10	30	-0.004	0.044	0.145	0.259	0.032	0.032	0.050	0.097	0.604	0.606	0.607	0.608
-2	50	5	-0.002	0.053	0.150	0.242	0.032	0.037	0.054	0.091	0.695	0.696	0.698	0.698
-2	50	10	-0.004	0.054	0.148	0.251	0.017	0.019	0.038	0.079	0.490	0.492	0.493	0.493
-2	50	30	0.003	0.047	0.150	0.253	0.005	0.008	0.028	0.069	0.283	0.284	0.284	0.285
-2	100	5	0.000	0.045	0.147	0.247	0.017	0.018	0.039	0.076	0.494	0.495	0.496	0.495
-2	100	10	-0.002	0.050	0.152	0.253	0.008	0.010	0.031	0.072	0.349	0.349	0.350	0.351
-2	100	30	0.002	0.047	0.150	0.250	0.003	0.005	0.025	0.065	0.202	0.202	0.202	0.202
-5	10	5	-0.025	0.111	0.351	0.609	0.184	0.196	0.300	0.534	1.491	1.517	1.532	1.534
-5	10	10	0.011	0.126	0.373	0.619	0.089	0.098	0.232	0.471	1.050	1.058	1.076	1.083
-5	10	30	0.005	0.120	0.373	0.618	0.031	0.045	0.170	0.414	0.604	0.611	0.617	0.622
-5	50	5	0.003	0.128	0.371	0.632	0.037	0.048	0.171	0.431	0.695	0.701	0.710	0.713
-5	50	10	-0.000	0.120	0.375	0.628	0.016	0.032	0.158	0.411	0.491	0.496	0.501	0.503
-5	50	30	0.004	0.122	0.375	0.622	0.006	0.020	0.146	0.393	0.284	0.286	0.289	0.290
-5	100	5	0.003	0.123	0.372	0.626	0.015	0.032	0.155	0.408	0.493	0.497	0.503	0.505
-5	100	10	-0.001	0.125	0.377	0.628	0.008	0.024	0.150	0.402	0.349	0.352	0.356	0.357
-5	100	30	-0.001	0.123	0.374	0.625	0.003	0.018	0.143	0.393	0.201	0.203	0.206	0.206

the change value in the non-adherence group is closer to the adherence group (high compliance). The lowest power occurs when we observe 50% contamination across all scenarios.

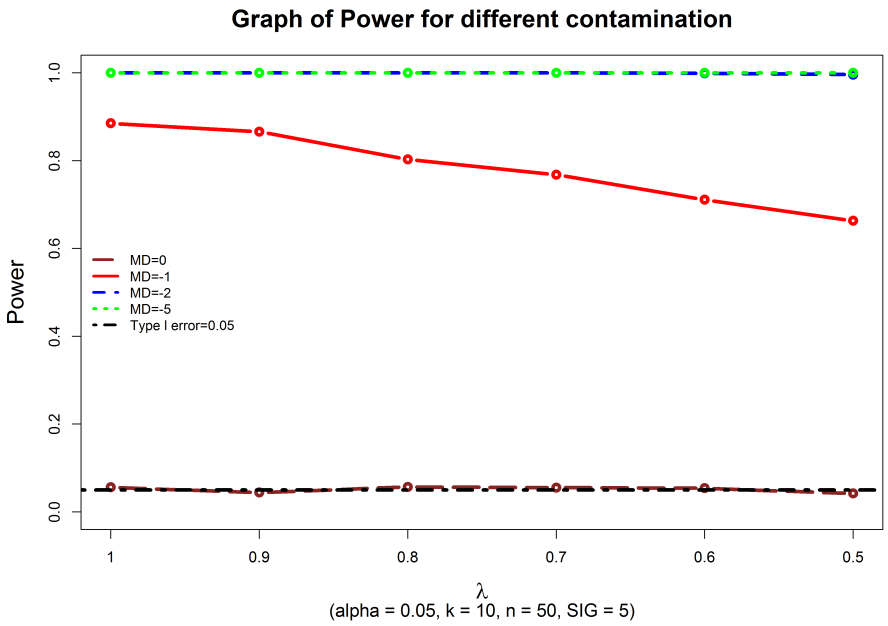

Figure 3.7: Effect of the mean difference on power for varying levels of contamination when other parameters are kept constant, $\alpha = 0.05$, $k = 10$, $n = 50$, standard deviation $= 5$, moderate compliance.

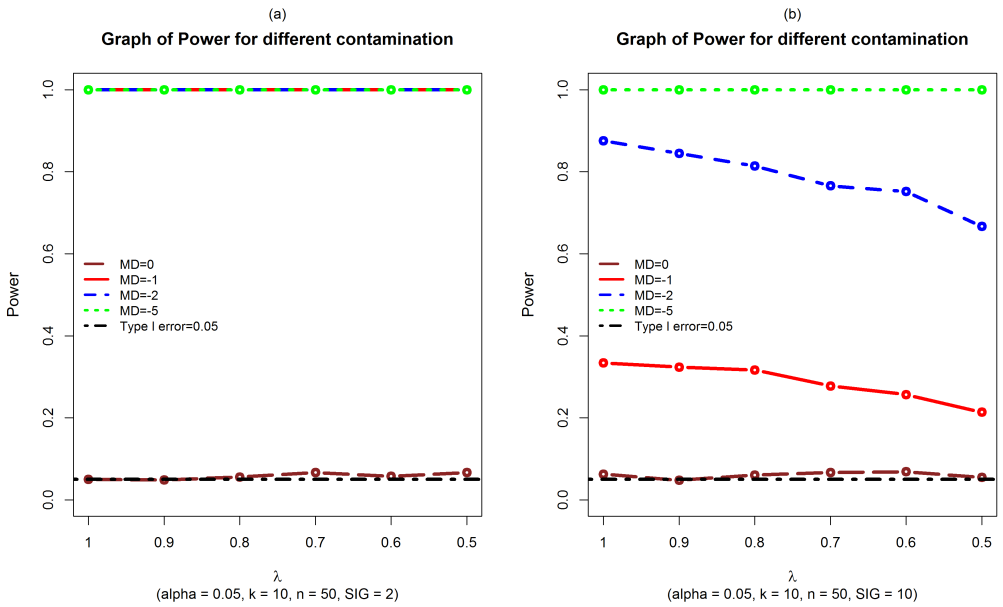

Figure 3.8: Effect of standard deviation (standard deviation = 2 (a) and standard deviation = 10 (b)) on power for varying levels of contamination when other parameters are kept constant, $\alpha = 0.05$, $k = 10$, $n = 50$, moderate compliance.

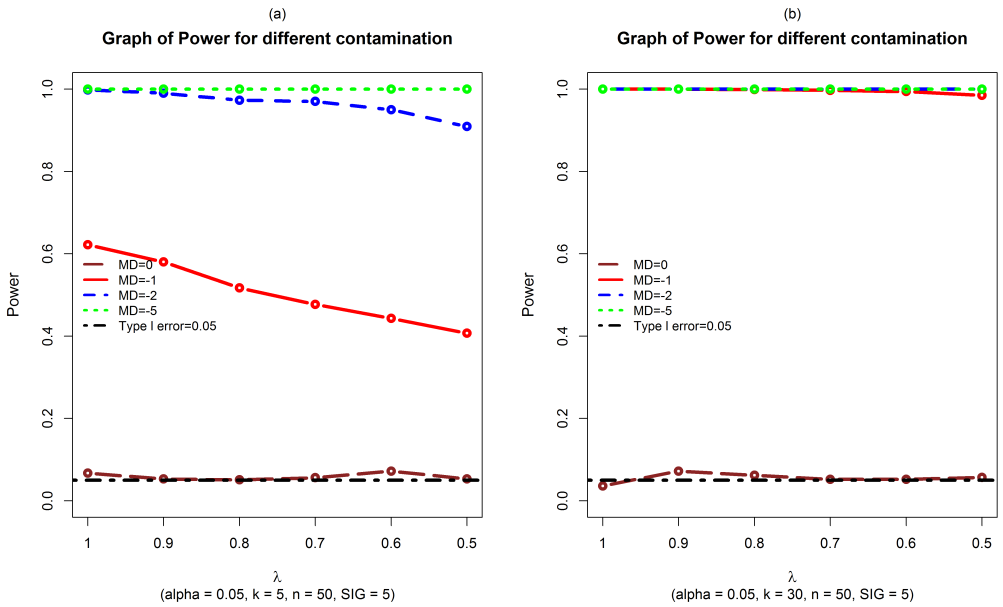

Figure 3.9: Effect of number of studies ($k = 5$ (a) and $k = 30$ (b)) on power for varying levels of contamination when other parameters are kept constant, $\alpha = 0.05$, $n = 50$, standard deviation $= 5$, moderate compliance.

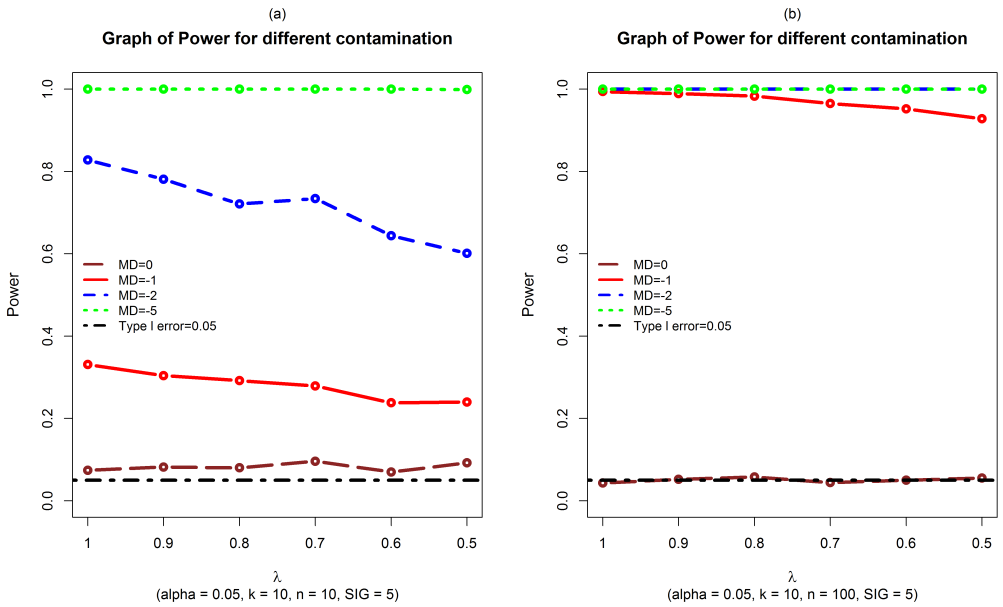

Figure 3.10: Effect of sample size ($n = 10$ (a) and $n = 100$ (b)) on power for varying levels of contamination when other parameters are kept constant, $\alpha = 0.05$, $k = 10$, standard deviation = 5, moderate compliance.

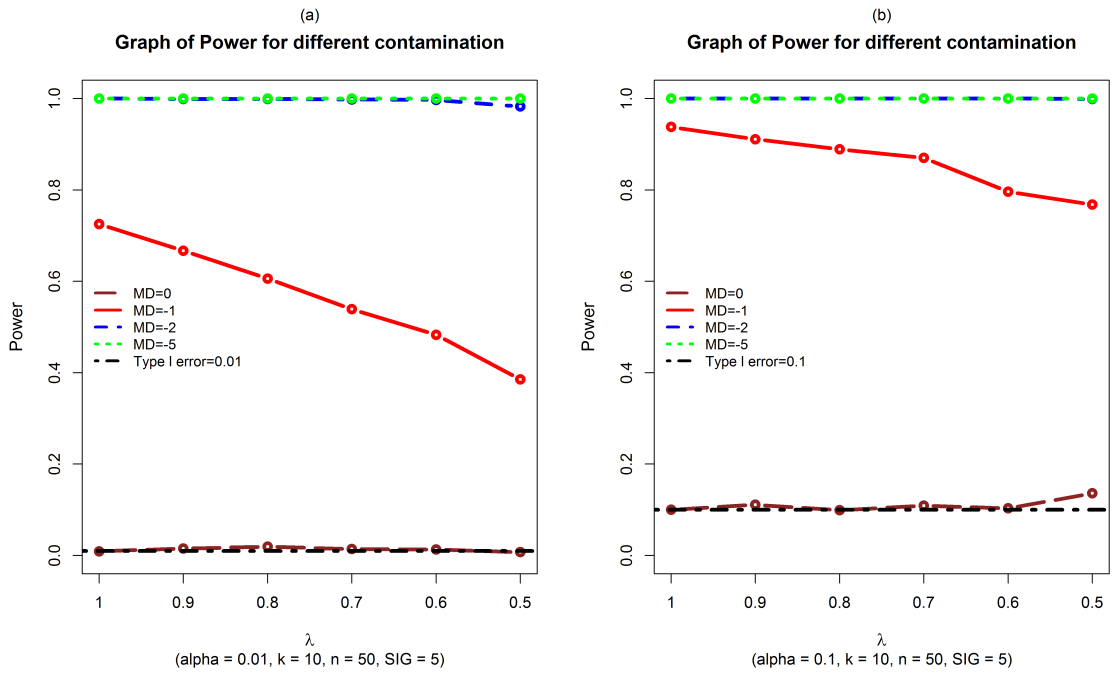

Figure 3.11: Effect of significance level ($\alpha = 0.01$ (a) and $\alpha = 0.1$ (b)) on power for varying levels of contamination when other parameters are kept constant, $k = 10$, $n = 50$, standard deviation $= 5$, moderate compliance.

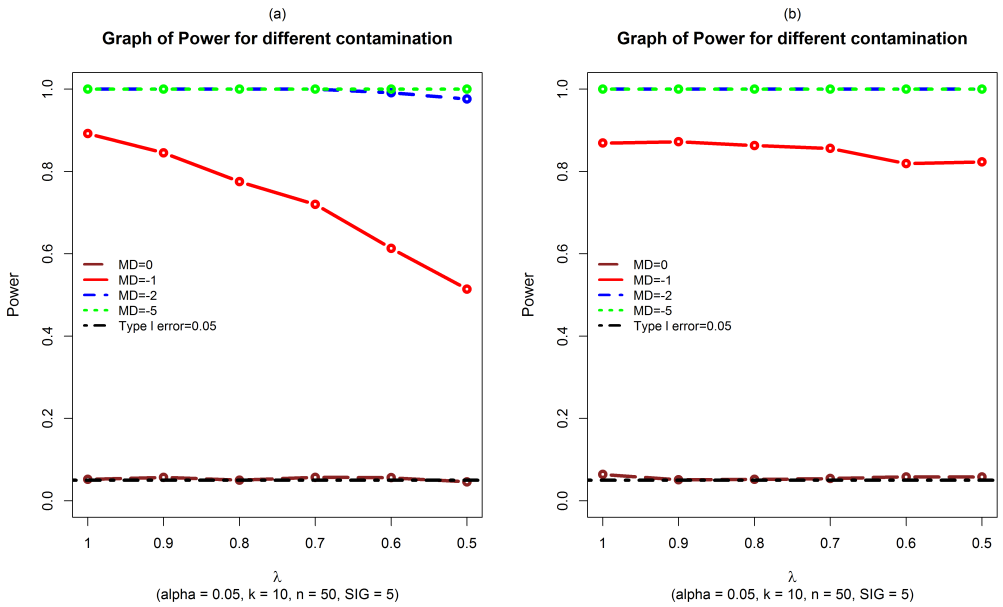

Figure 3.12: Effect of the position of non-adherence change value (low compliance (a) and high compliance (b)) on power for varying levels of contamination when other parameters are kept constant, $\alpha = 0.05$, $k = 10$, $n = 50$, standard deviation $= 5$.

Chapter 4

Summary, Discussion and Future Directions

4.1 Summary

Meta-analysis refers to the statistical synthesis of results from a series of studies concerned with similar interventions. The standard meta-analysis involves only a pairwise comparisons of two treatments. The applications of meta-analysis cut across several research fields and are observed to be an influential branch of clinical epidemiology and health services.

We have noticed in our methodology section, Chapter 2, that the type of effect size generally depends on the type of outcome being examined. Also, there are basically two types of statistical models employed in meta-analysis termed as fixed-effect and random-effects model. In practice, when investigating the effectiveness of a new

intervention, it is common not to observe 100% adherence in the treatment or intervention group. That is, it is plausible to observe contamination in the intervention group. This thesis thus focused on investigating the impact of contamination on fixed-effect meta-analysis for a continuous outcome.

In Chapter 3, we have outlined the design of our simulation. The change values were simulated by generating contaminated normal random variables. The effect measure considered is the mean difference. A fixed-effect meta-analysis was carried out on the mean differences to estimate the weighted mean difference from the *metafor* package in R.

4.2 Discussion

This simulation study was undertaken to assess the influence of contamination on fixed-effect meta-analysis of continuous outcome. The results of this study provide valuable insights into the potential problems and limitations of the method of analysis chosen in different scenarios.

In the absence of contamination, the fixed-effect meta-analysis performed well with the estimation properties within the simulated parameters with low bias and appropriate coverage even in scenarios with low and high standard deviations. Conversely, the method does not perform well in the presence of contamination.

The proportion of the scenarios for which the $100(1-\alpha)\%$ confidence interval contains the true effect size is greatly influenced by the presence of contamination for all the scenarios. The coverage probability attains the nominal coverage in the

absence of contamination except for when the sample size is relatively small which is as expected. Undercoverage was observed in almost all the scenarios as contamination increases with increasing bias. The method performs worst when 50% of the subjects in the treatment group did not adhere completely to treatment (i.e. 50% contamination) in all of the scenarios.

With increasing effect measure, coverage probability decreases as the level of contamination increases. Similar trend is observed when other parameter values are varied. This reflects the fact that in the presence of contamination, as the proportion of the non-adherence group increases in the treatment group, the mean difference between the treatment group and the control group moves further away from the true mean difference and hence the confidence interval computed for the true mean difference based on the estimated mean difference results in missed coverage.

In our simulation study, we found that the coverage is greatly influenced by the standard deviation which is consistent with theoretical considerations. Also, both the sample size and the number of studies affects the coverage. Again, this is intuitive from a theoretical perspective.

The mean square error increases as the contamination level increases in all the scenarios. The mean square error is at the lowest when the sample size and the number of studies is at the highest ($k = 30$, $n = 100$) for the various levels of contamination. The width of the $100(1-\alpha)\%$ confidence interval becomes narrower as we observe less contamination.

As expected, the statistical power decreases with increasing level of contamination. The power increases with increasing effect size, increasing number of patients,

and increasing number of trials but decreases as the standard deviation increases.

We also investigated the effect of the position of the change value in the non-adherence group. As the change value gets closer to the change value in the adherence group (high compliance), the method performs better in all of the scenarios compared to when the change is closer to the change in the control group (low compliance).

4.2.1 Conclusion

The results of our fixed-effect meta-analytic simulation study suggests that, in the presence of contamination, the properties of estimators and hypothesis are not optimal. That is, contamination in the treatment or intervention group must not be ignored when performing meta-analysis. It must be taken into consideration since it greatly affects the results of the method. For better results, there should be almost 100% adherence. Thus benefits are greater with better compliance (Clifton et al. (2014)).

4.3 Future Directions

In this study, our focus was based mainly on the traditional meta-analysis with a continuous outcome. In reality, the primary outcome could be binary. Binary outcome is a summary data that was originally reported as the number of events and non-events in two groups. The appropriate effect sizes computed by researches based on this outcome is the odds ratio, relative risk or the risk difference. Thus we

would like to extend our simulation study to investigate the performance of fixed-effect meta-analysis in the presence of contamination considering a binary outcome.

Meta-analysis is the statistical tool for combining the results or findings from multiple independent studies for similar interventions. For many years, only the direct and pairwise comparisons have been made using the standard meta-analysis methods. However, in recent years, network meta-analysis (NMA) also known as mixed treatment comparisons (MTC) or multiple treatment meta-analysis (MTM), has been presented as an extension of the traditional meta-analysis by including multiple different pairwise comparisons across the range of different interventions. Network meta-analysis allows for indirect comparisons and thus can provide very useful information in the absence of head-to-head evidence via a common comparator. We plan to extend our simulation study to some critical issues related to network meta-analysis such as rank probability (the probability of ranking a treatment to be the best among other competing treatments), relative treatment effect, the heterogeneity parameter (the I^2) considering both continuous and binary outcomes.

Bibliography

Banerjee, S., Wells, G., and Chen, L. (2008). *Caveats in the Meta-analyses of Continuous Data: A Simulation Study*. Canadian Agency for Drugs and Technologies in Health.

Borenstein, M., Hedges, L. V., Higgins, J., and Rothstein, H. R. (2009). Fixed-effect versus random-effects models. *Introduction to Meta-analysis*, pages 77–86.

Borenstein, M., Hedges, L. V., Higgins, J., and Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2):97–111.

Clifton, P. M., Condo, D., and Keogh, J. B. (2014). Long term weight maintenance after advice to consume low carbohydrate, higher protein diets–a systematic review and meta analysis. *Nutrition, Metabolism and Cardiovascular Diseases*, 24(3):224–235.

Egger, M., Ebrahim, S., and Smith, G. D. (2002). Where now for meta-analysis? *International Journal of Epidemiology*, 31(1):1–5.

Figueiredo, F. and Gomes, M. I. (2013). The skew-normal distribution in spc. *REVSTAT–Statistical Journal*, 11(1):83–104.

Franchini, A., Dias, S., Ades, A., Jansen, J., and Welton, N. (2012). Accounting for correlation in network meta-analysis with multi-arm trials. *Research Synthesis Methods*, 3(2):142–160.

Friedrich, J. O., Adhikari, N. K., and Beyene, J. (2008). The ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis: a simulation study. *BMC Medical Research Methodology*, 8(1):32.

Greco, T., Landoni, G., Biondi-Zoccai, G., D'Ascenzo, F., and Zangrillo, A. (2013). A bayesian network meta-analysis for binary outcome: how to do it. *Statistical Methods in Medical Research*.

Green, S. et al. (2005). Systematic reviews and meta-analysis. *Singapore Medical Journal*, 46(6):270.

Haidich, A. (2010). Meta-analysis in medical research. *Hippokratia*, 14(Suppl 1):29.

Harrison, F. (2011). Getting started with meta-analysis. *Methods in Ecology and Evolution*, 2(1):1–10.

Hedges, L. V., Cooper, H., and Bushman, B. J. (1992). Testing the null hypothesis in meta-analysis: A comparison of combined probability and confidence interval procedures. *Psychological Bulletin*, 111(1):188.

Hedges, L. V. and Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods*, 3(4):486.

Higgins, J. and Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11):1539–1558.

Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal*, 327(7414):557.

Olkin, I. and Sampson, A. (1998). Comparison of meta-analysis versus analysis of variance of individual patient data. *Biometrics*, pages 317–322.

O'Rourke, K. (2007). An historical perspective on meta-analysis: dealing quantitatively with varying study results. *Journal of the Royal Society of Medicine*, 100(12):579–582.

Plackett, R. L. (1958). Studies in the history of probability and statistics: Vii. the principle of the arithmetic mean. *Biometrika*, 45(1-2):130–135.

Sterne, J. A., Egger, M., and Smith, G. D. (2001). Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ: British Medical Journal*, 323(7304):101.

Uman, L. S. (2011). Systematic reviews and meta-analyses. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 20(1):57.

Viechtbauer, W. (2006). Mima: an s-plus/r function to fit meta-analytic mixed-, random-, and fixed-effects models. *URL http://www.wvbauer.com*.