The University Of Sheffield.

Non-inferiority Margin Setting from Indirect Comparison

Thesis submitted in Fulfilment for the Degree of Doctor of Philosophy

School of Health and Related Research (ScHARR)

Enass M. Duro

July 2019

Acknowledgements

This work would not have been completed without the support of my supervisors, my family and the Libyan Ministry of Higher Education. First, I would like to express my gratitude to my supervisors, Professor Steven Julious, Dr Shijie Ren and Dr Jean Sanderson, for their help and support, not only with my thesis but also outside of my studies.

I want to thank my parents for their prayers for me and my husband and my children for their patient support throughout my study period.

To my father (MOHAMED DURO) and my mother (FAWZIA WALI) for their unconditional love and support throughout my life.

To my husband (ABDUL KAREM), my son MOHAMED, my daughters FATAMA, and FARAH who are always asking when I will finish my studies.

Disclaimer and Author's Declaration

The views expressed in this research are those of the author only.

The author declares that this thesis is her original work and that none of the material contained in this thesis has previously been submitted for a degree to any awarding Institution. The work contained in this thesis has been undertaken by the author (DURO), with the support from those individuals or collaborators mentioned in the Acknowledgements section.

Abstract

Introduction: Non-inferiority trials (NI) test the efficacy of an experimental treatment in comparison to an active-controlled treatment and indirectly with the historical placebo to demonstrate that the new treatment is no worse than the active comparator. Setting the NI margin depends on the assumptions of constancy, assay sensitivity and the absence of placebo creep and bio-creep.

Research Question: This PhD research will investigate the changes in the efficacy of the placebo and active control over time. It will show how this could affect the setting of the NI margin and the conclusion of non-inferiority. The context is where there is a wish to make a retrospective indirect comparison of the experimental treatment with historical placebo.

Methods: An overview of Cochrane reviews of placebo-controlled trials was conducted to measure the correlations between the placebo, active treatment and the treatment difference with the year of publication. From the constructed dataset from the Cochrane reviews, a weighted regression model was built to investigate factors affecting the estimate of the future trial from a meta-analysis of historical trials, followed by proposing a method for the use of meta-regression to adjust for time while setting the NI margin.

Results: The correlations between the placebo, active treatment and treatment difference and the year of publication varied from strong negative to strong positive correlations. The median correlation for the treatment difference = - 0.1. The estimate of any future trial could be predicted from a meta-analysis of historical trials with coefficient of 0.92 and range from 0.75 to 1.047 of the historical trials. Moreover, increasing the year of prediction and increasing the year difference in the meta-analysis will reduce the predicted estimate by 0.015 and 0.005 respectively. Pairwise meta-regression and network meta-regression can be used to assess the constancy, to set the adjusted non-inferiority margin and to analyse the non-inferiority trial when the constancy assumption does not hold.

Conclusion: In NI trials, the constancy assumption needs to be assessed not assumed. Adjusting for the time will reduce the chance of the conclusion of non-inferiority of an inferior test treatment regardless of the constancy assumption.

Table of Contents

List of Tables

List of Figures

Table of Abbreviations

AD	Aggregated Data
BMJ	British Medical Journal
C	Active control
CABAG	Coronary Artery Bypass Grafting
CHMP	Committee for Medical Products for Human use
CI	Confidence Interval
CONSORT	The Consolidated Standard of Reporting Trials
CPMP	Committee for Proprietary Medical Products
Cr I	Credible Interval
DIC	Deviance Information Criterion
EMA	The European Medicines Agency
FDA	Food and Drug Administration
GAO	Governmental Accountability Office
HESDE	Historical Evidence of Sensitivity to Drug Effects
ICH- E9	International Conference on Harmonisation (statistical principle for medical trials)
ICH- E 10	International Conference on Harmonisation (choice of control group)
IPD	Individual Patient Data
ITT	Intent to Treat analysis
JAMA	Journal of American Medical Association
Kdl	Number of trials in the historical meta-analysis excluding last
LMWH	Low Molecular Weight Heparin
LOA	Limit of Agreement
LOOCV	Leave one out cross validation
M1	Statistical non-inferiority Margin
M2	Clinical non-inferiority Margin
MAIC	Matching-Adjusted Indirect Comparison
MCMC	Markov Chain Monte Carlo
MI	Myocardial Infarction
Ndl	Sample size of historical trials excluding last trial

NEJM	New England Journal of Medicine
NMA	Network Meta-analysis
NMR	Network Meta-regression
NI	Non-inferiority
NSTEMI	non- ST myocardial Infarction
P	Placebo
PCI	Percutaneous Intervention
PP	Per protocol analysis
RCT	Randomised Controlled Trial
RMSE	Root Mean Squared Error
SMD	Standardised Mean Difference
SMDdl	Standardised Mean difference from historical meta-analysis
SMDlt	Standardised Mean difference of the predictive trial
STC	Simulated Treatment Comparison
T	Test Treatment
UFH	Unfractionated Heparin
VIF	Variance inflation factor

Chapter 1 Introduction

1.1 Background

The gold standard in evidence-based medicine is a randomised clinical trial (RCT) (D'Agostino, Massaro, & Sullivan, 2003). For RCTs, two types of control could be used, a placebo, which for this thesis would include a placebo, no treatment, or usual care (if usual treatment is no treatment), or an active treatment which could be a comparator treatment or current treatment. RCTs are not only drug trials. They could compare different treatments, procedures or protocols.

Placebo-controlled trials are the main RCTs that are conducted to evaluate the efficacy and safety of the new treatment. Placebo-controlled trials are considered ethical if no standard treatment exists or if there will be no harm to the patients from delaying treatment. Placebo-controlled trials are considered unethical if they prevent or delay patients from getting access to an effective treatment, which may lead to harm (D'Agostino et al., 2003). In such a situation, active-controlled trials are undertaken. In active-controlled trials, the new treatment is compared with an established treatment rather than a placebo.

RCTs can be broadly divided into superiority trials that aim to conclude that the test treatment is better than the comparator and non-inferiority (NI) trials that aim to show that the test treatment is not worse than the comparator (FDA, 2016). Placebo-controlled trials are the most closely associated with the superiority trials, while non-inferiority trials are the most closely associated with the active-controlled trials. Usually, non-inferiority trials are efficacy trials that aim to prove that the efficacy of the new experimental treatment is not inferior to the current treatment and could promise maybe better safety or adherence or be less expensive. However, in recent years, NI trials have been used to evaluate the safety of the test treatment with placebo controlled trials (Mauri & D'Agostino, 2017).

The terminology of active-controlled trials and non-inferiority trials has become more popular since the 1990s (Rothmann et al., 2003). The concept of a better substitute to superiority placebo-controlled trials was the rationale for the introduction of non-inferiority trials (Mauri & D'Agostino, 2017). The number of non-inferiority trials that have been published has increased by a factor of six in a decade (Mauri & D'Agostino, 2017).

According to the GAO (Governmental Accountability Office, USA) report in 2010, between 2002 and 2009 a total of 175 new drugs were submitted for FDA approval, 43 of them based on at least one non-inferiority trials (GAO, 2010). A review of 583 non-inferiority trials published between 1989 and 2009 showed an increasing trend of publication of NI trials, with a third of these trials being infectious diseases or cardiology trials (Suda, Hurley, McKibbin, & Moroney, 2011). For this thesis, a search in PubMed for the term "active control or non-inferiority trials" revealed only one manuscript in 1990 compared with 510 papers in 2018; this reflects the growing interest in the active control and NI trials. Figure 1.1 demonstrates the growing interest in the active control and NI trials from 1990 to 2018. This search was conducted in 2016 and updated in January 2019.

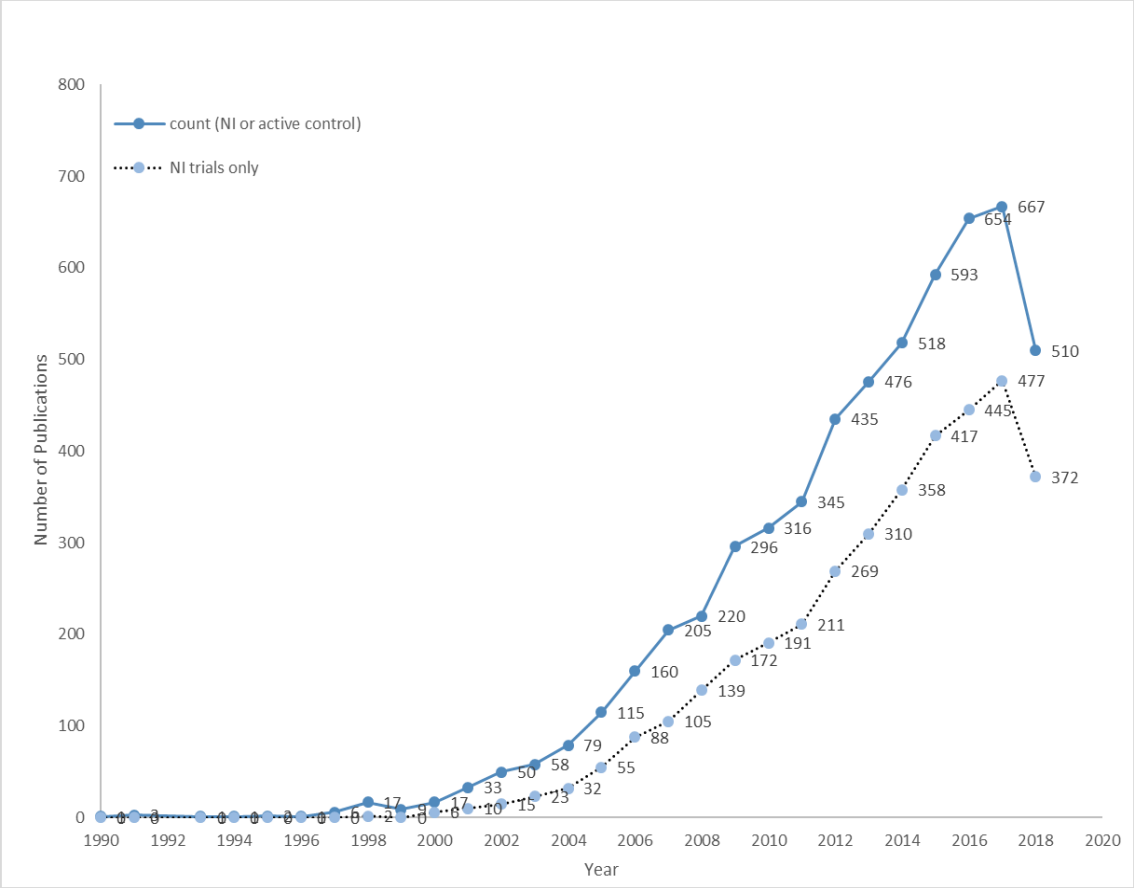

Figure 1-1 Number of Published papers of NI trials or Active control trials per year
(Note: the search done PubMed on April/ 2016 and updated in January/ 2019 with search terms: Search non-inferior* OR noninferior* OR ("active-controlled")* Filters: Clinical Trial; Humans)

The International Conference on Harmonization (ICH-E9) produced the first published regulatory guidelines on conducting clinical trials across regulatory jurisdictions (ICH, 1998). Due to the growing interest in active-controlled and non-inferiority trials, regulatory guidelines have been established to advise on the conducting and reporting of active control and NI trials. There is guidance on the choice of control in a study from ICH E-10 (ICH, 2001), the Committee for Medicinal Products for Human Use (CHMP, 2005). There are also the US Food and Drug Administration (FDA) guidelines regarding setting, conducting and analysis of NI trials (FDA, 2016). However, none of these guidelines establishes any enforceable responsibilities (FDA, 2016). Instead, they give only advice and guidance. To note also, for reporting, the Consolidated Standards of Reporting Trials (CONSORT) statement on the appropriate reporting of NI trials in medical journals has been released (Piaggio et al., 2012).

Compared to traditional superiority trials, NI trials present methodological and regulatory challenges that can influence the analysis and inference of their results (D'Agostino et al., 2003). These include choosing an appropriate active comparator (it could be the best available treatment or could be the standard of care), the subjectivity in the setting of the non-inferiority margin both statistically and clinically, and the use of an indirect comparison to compare the efficacy of the test treatment with the historical placebo.

The non-inferiority margin is a pre-specified amount (M), which is used to demonstrate that the test treatment is no worse than the active control (D'Agostino et al., 2003; FDA, 2016). It is the amount the active control can exceed the test treatment and for a conclusion of the test treatment being non-inferior to the active control to be made (D'Agostino et al., 2003). If in the past the active control had been compared to placebo then this could be used to determine the non-inferiority margin so that through the active control an indirect comparison could be made for the test treatment to show superiority over placebo (indirect comparison).

An indirect comparison is a comparison that is made between two treatments that have never been tested in the same trial but are used treat the same disease in the same patient population, sharing a common control treatment (Julious & Wang, 2008).

To demonstrate the meaning of indirect comparison, suppose two trials are conducted:

Trial 1: - Compared treatment (A) with treatment (B)

Trial 2: - Compared treatment (C) with the active control (B)

treatment A could be indirectly compared with treatment C since both of them had a common comparator B

$$(A - B) - (C - B) = A - C, \qquad (1.1)$$

Where A is the effect size of treatment A,

B is the effect size of treatment B**,**

C is the effect size of treatment C.

The situation in non-inferiority trials is:

Trial 1: (historical placebo-controlled trial): compares the active control (C) with the placebo (P),

Trial 2: (non-inferiority trial in present time): - compares Test treatment (T) with the active control (C),

$$(T - C) - (P - C) = T - P, \qquad (1.2)$$

where T is the effect size of test treatment,

C is the effect size of the active control**,** and

P is the effect size of the placebo,

The aim of trial 2 is to show that the test treatment is not inferior to the active control and indirectly superior to the historical placebo. This comparison is not straightforward, and several regulatory and methodological challenges accompany this comparison.

For an NI trial, the first step will be choosing the appropriate active control. Once this has been sorted, then a non-inferiority margin should be determined. According to ICH-E10 (ICH, 2001), designing and conducting non-inferiority trials can be summarised in four steps:

1- **Determining that historical evidence of sensitivity to drug effects exists (HESDE):** This means that the historical trials that were used in the past can distinguish the effective treatment from an ineffective one. It should be specified that the treatment that will be used as an active control was found reliably superior to the placebo in the historical placebo-controlled trials. HESDE should be determined before the beginning of the NI trial (ICH, 2001).
2- **Designing a trial with a detailed protocol**: The NI trial should be designed with a detailed protocol about inclusion and exclusion criteria, population, primary endpoints, and type of statistical analysis that will be used.
3- **Defining a non-inferiority margin (M):** as mentioned earlier, M is a pre-specified amount which is used to demonstrate that the test product is no worse than the comparator by more than this amount (FDA, 2016). The NI margin should be defined, taking into account the historical data that were used to estimate the effect of the active control, clinical judgement, and statistical considerations like regression to the mean bias and presence of the effect modifiers (Rothmann, Wiens, Chan, Crc, & Group, 2012). FDA defined two margins that should be specified: M1, the statistical NI margin, and M2, the clinically determined margin (FDA, 2016).
4- **Conduct of the trial:** The NI trial should be conducted according to regulatory and statistical guidelines. An NI trial should be similar to the historical trials that were used in determining the NI margin (ICH, 2001).

1.2 The research rationale, aims and objectives

As highlighted in Section 1.1, conducting and interpreting NI trials is accompanied by several methodological and regulatory challenges. The research rationale behind this thesis is to investigate how the changes in the placebo and active treatment effect over time could affect the estimation of the NI margin and NI trials conducting in general. In addition, what are the methods that can be used to adjust for a time while setting the non-inferiority margin?

Objectives:

The objectives of this thesis are to investigate:

- The methodological and regulatory challenges associated with the planning, conducting and reporting of non-inferiority trials,
- The changes in the placebo and active treatment effects over time and their impact on the design and analysis of NI trials,
- To quantify and model placebo and active treatment responses over time with recommendations for retrospective comparison back to placebo.
- Propose a method for adjusment for time from indirect comparison while setting the NI margin (in the design phase of NI trial).

1.3 Outlines of the thesis

Aiming to answer the objectives for this research, this thesis will be divided into three parts. Part one (Chapters 2, 3, and 4) will include the review chapters that will review the conducting, regulation and reporting of non-inferiority trials. The second part will investigate the changes in the placebo and active control over time (Chapter 5 and 6), and the final part will introduce new methods for setting the adjusted non-inferiority margin (Chapters 7 and 8), This research will be concluded in Chapter 9 with the summary, discussion, and main conclusion. Recommendations will be provided on how to adjust for a time in NI trials. Figure 1.2 illustrates the thesis road map.

Review Chapters:

- Chapter 2 will review the literature on the designing of NI trials, focusing on the choice of appropriate active control and the main assumptions, considerations and limitations of the NI trials. Moreover, it will present the methods used for setting the NI margin and the methods used for the analysis on NI trials; both Frequentist and Bayesian methods will be presented.
- Chapter 3 will review the regulatory guidelines that deal with NI trials and the differences and the similarities between these guidelines.

- Chapter 4 will conduct a systematic review of the published NI trials in 2015 in the top four medical journals to investigate the quality of the published NI trials in the clinical practice.

Chapters investigate the changes in the placebo and active treatment effect over time:

- Chapter 5 will provide an overview of Cochrane reviews published in 2015-2016 on placebo-controlled trials. Correlations and partial correlations between the year of publication and the sample size, placebo, active treatment and treatment difference will be reported to measure the changes in the treatment effect over time.
- In Chapter 6, data collected in Chapter 5 will be used to build a weighted regression model to investigate the predictors of a treatment effect on the trial based on the available historical trials. The relations between the year of publication, time difference, and type of model (fixed or random) will be studied.

Chapters 7 and 8 will review and propose a new method for adjusting for time in NI trials:

- Chapter 7 will review the available possible methods for adjusting for time while setting the NI margin.
- In Chapter 8, based on the review from chapter 7, a new method will be proposed to adjust for time while setting the NI margin from indirect comparison. Two possible scenarios of setting the NI margin will be presented: the first example willinvolve setting the margin when the constancy assumption cannot hold, and the second example will involve checking the validity of the proposed method for both cases when constancy is assumed.

Chapter 9 will present the final discussion, conclusion, and recommendations.

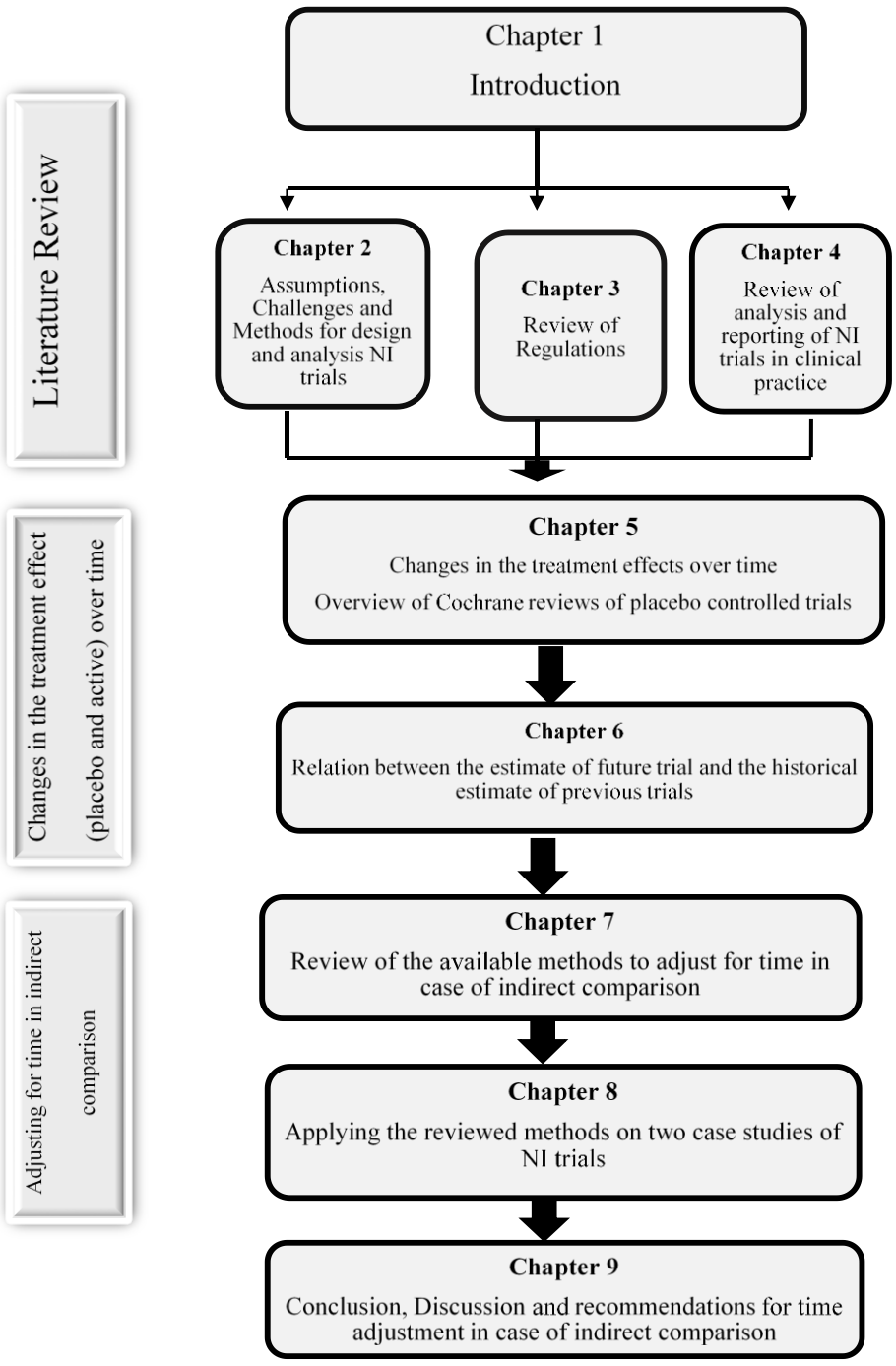

Figure 1-2 Thesis Road Map

Chapter 2 Literature Review: What is a Non-inferiority Trial?

2.1 Introduction

As mentioned in Chapter 1, designing and conducting non-inferiority trials can be summarised in four steps (ICH, 2001): determining that historical evidence of sensitivity to treatment effects exists (HESDE); designing a trial with the detailed protocol; defining a non-inferiority margin; and finally, conducting the trial**.**

This chapter will review the general considerations and assumptions for designing NI trials, setting the NI margin, and methods for analysing the non-inferiority trials. Section 2.3 will present the main considerations and assumptions regarding choice of the appropriate active control, determining its sensitivity and constancy assumption, the placebo creep and bio-creep as main challenges in the non-inferiority trial and other challenges in conducting the non-inferiority trial. This will be followed in Section 2.4 by presentation of the setting of the non-inferiority margin and the role of meta-analysis in the setting of the non-inferiority margin. The available methods for the analysis of non-inferiority trials will then be reviewed in Section 2.5. An example of the analysis of the non-inferiority trials using the different presented methods will be illustrated in Section 2.6. The chapter will close with a summary of the findings in Section 2.7.

2.2 Aims and Objectives:

- Review the literature regarding designing of NI trials
- Review the assumptions, considerations associated with NI trials
- Address the definitions and the differences between the two types of non-inferiority margins.
- Review methods for the analysis of NI trials

2.3 Challenges, considerations and assumptions of non-inferiority trials

For an NI trial, the first step will be to choose the appropriate active control. Once that has been sorted, then a non-inferiority margin should be determined. In general, when possible, the most effective available standard treatment should be used as the active control in the NI trial (Rothmann et al., 2012). That means appropriately designed and conducted trials in the past that used a specific active treatment and regularly showed this active control to be superior to placebo. These findings allow for a reliable estimate of the effect size of the active control compared to the placebo in the historical trials, and this will form a base to estimate the effectiveness of active control in the current NI trial (FDA, 2016).

Fleming defined the appropriate "suitable" active control as a widely used treatment whose efficacy was proven by well-designed randomised controlled trials that documented its superiority and which is expected to have the same efficacy in the current active-controlled trial (Fleming, 2008).

The effectiveness of active control could be concluded from two determinations:

1. **HESDE**: the historical trials that were used in the indirect comparison should be similar to the non-inferiority trial in efficacy endpoint and population and should be evaluated before the beginning of the NI trial (FDA, 2016). The conclusion from these trials should be that the active control is reliably superior to the placebo in these historical trials (CHMP, 2005).
2. **Proper NI trial conducting:** the NI trial should be conducted under the regulatory guidelines to ensure its ability to distinguish effective treatment from less effective ones (FDA, 2016).

There are some considerations regarding the estimate of the effect size of active control from previous studies and applying it in the current NI trial. These include assay sensitivity, constancy assumption, bias minimising (regression to mean bias, publication bias, and the bio-creep and placebo creep) (D'Agostino et al., 2003; FDA, 2016; Rothmann et al., 2003).

In this section, these considerations and assumptions will be discussed in more detail, as well as how the violation of these assumptions could affect the setting of the NI margin.

2.3.1 Assay sensitivity of the active control

ICH- E10 defined assay sensitivity as:
"A property of a clinical trial to distinguish an effective treatment from a less effective or ineffective treatment"; the trial should provide assurance that if a placebo is included in that trial, the active-control will show superiority to the placebo (ICH, 2001).

Assay sensitivity is essential in any trial (superior or non-inferior). In a superiority trial, assay sensitivity is established once the superiority of the test treatment is concluded (conclusion of efficacy achieved). However, assay sensitivity cannot be established directly from NI trials (FDA, 2016; Snapinn, 2000). The efficacy in NI trials is demonstrated by showing that a test treatment is no worse (non-inferior) than the active control. As a result of this, even if the trial's assumption of assay sensitivity does not hold, the trial may find an ineffective treatment to be non-inferior to active control (which is ineffective against placebo), and thus a biased conclusion of efficacy could be made (FDA, 2016; ICH, 2001).

2.3.2 Constancy Assumption

The difference between the active control and the placebo in the historical trial is assumed to hold in the designing of the NI trial; this is referred to as the "constancy assumption" (D'Agostino et al., 2003; FDA, 2016). Fleming considered the assumption of constancy as the most critical challenge in designing and conducting NI trials (Fleming, 2008).

Proving that the effect size of the difference between the active control and the placebo is constant over time (same in historical and NI trial) is difficult, especially with the rapid changes in medical practice and standard of care in many therapeutic areas (Fleming, 2008). Changes in medical practice over the years could reduce the efficacy of the active control and improve standard care (LeLorier, Grégoire, Benhaddad, Lapierre, & Derderian, 1997).

An example of how medical practice can change can be taken from the therapeutic area of antibiotics resistance. Vancomycin was considered an effective treatment for urinary tract infection compared to no treatment (placebo). However, the development of vancomycin resistant enterococci in recent years reduced the efficacy of vancomycin in treating urinary tract infections. In this situation, using vancomycin as an active control in an NI trial to establish the non-inferiority of any new treatment compared to vancomycin will be sub-optimal, since the assumption of the constant effect of vancomycin cannot be held. Even though the superiority of vancomycin to placebo was established in previous historical trials, the constancy assumption cannot be held due to change in the infective agent itself, not the active control (Fleming, 2008).

Supporting the constancy assumption is difficult to achieve, not only in anti-infective NI trials but in NI trials in general (FDA, 2016). The presence of effect modifiers like differences between the historical trials and the NI trial in the population, in the definition of the endpoint, changes in procedures, and changes in causes of the disease could affect the constancy assumption and lead to false favourable rates of the effect of active control compared to placebo. This will lead to approval of non-effective new treatments (K. Odem-Davis & Fleming, 2015).

Including a placebo arm in an NI trial design will establish both assay sensitivity and constancy of active control without the need for indirect comparison between two different trials that were conducted at a different time point. However, this is not feasible most of the time for clinical and ethical reasons (FDA, 2016). Another possible way to secure the constancy assumption in the NI trial is to ensure the similarity between both the past trials and the new NI trial. Both trials should be as close as possible in all essential respects, including the primary outcome, study population, and structure of the study. However, the similarity of these trials may not be possible to assess fully until the NI study is completed (ICH, 2001).

2.3.3 Variability of historical trials

Another problem facing the determination of the effectiveness of the active control is the presence of different historical trials with different sample sizes, different methods of analysis and different conclusions. This variability between historical trials could affect the measurement of the efficacy of the active control based on these trials (FDA, 2016). Even though the use of meta-analysis to estimate the effect size of active control from historical trials could resolve part of the problem regarding the sample size and conclusion, meta-analysis cannot address the effect of the time difference between the trials in its estimate since it does not take into consideration the time changes (Julious & Wang, 2008).

Determining the effectiveness of the active control based on a single randomised placebo-controlled trial is also an issue that could affect the precision of effectiveness of the active control (FDA, 2016). The heterogeneity of the effect of the active control cannot be assessed if there was only one historical study (Rothmann et al., 2003). Lelorier et al. stated that using a single large randomised controlled trial is more accurate and less biased than the use of traditional meta-analysis methods (LeLorier et al., 1997). However, according to FDA regulations, the use of only one randomised controlled trial as historical evidence is possible in only one situation, namely where both the active control and the experimental treatment belong to the same pharmacological family (FDA, 2016).

The sample size of historical trials that are used to estimate the effect of the active control will affect the width of the confidence interval that is used to estimate the effect of the active control. Studies with a small sample size will produce a wide CI; hence, a large sample size will be required for an NI trial to achieve the non-inferiority (Rothmann et al., 2003).

The variability in the effect of the active control across the studies could lead to an inconsistent estimate of the actual active control effect. In this case, using the random effect model in the pairwise meta-analysis could account for the between trials variability (Rothmann et al., 2003). However, a random effect model will give more weight for smaller trials, which are usually older, and with extreme results. Another problem with assessing the efficacy of active control using historical trials is the publication bias; historical trials with positive results are published more frequently than trials with negative results, which

could lead to overestimating of the effect size of active control compared to placebo (Rothstein, Sutton, & Borenstein, 2005).

2.3.4 Regression to the mean

Everitt (2002) defined regression to the mean as:

"*The phenomenon that a variable that is extreme on its first measurement will tend to be closer to the centre of the distribution for a later measurement".*

As mentioned earlier, the most effective available treatment is chosen to be the active control in an NI trial. Estimation of the effect of active control could have the potential for regression toward the mean bias since the effect of active control is based on the maximum performance of the active control in the historical trials, not on its random effect, which would lead to overestimation of the effect of active control in the NI trial (Rothmann et al., 2012).

As an example of regression to the mean, suppose in the therapeutic area of cardiovascular there were three placebo-controlled trials with three different drugs for reducing the total blood cholesterol level (drug A, drug B, and drug C). In these trials, drug B showed a higher reduction in the total cholesterol level compared to drug A and C. Drug B is now used as the active control in any new NI trial. However, due to the regression toward the mean phenomenon, the efficacy of the drug B in any future trials (including NI trial) will be less than its efficacy in the first trial. Moreover, setting the NI margin depending on its efficacy in the first trial will lead to overestimation of its effectiveness and possibly to concluding non-inferiority of an ineffective drug. Making the appropriate adjustments for the population age or structural changes to the effect of the active control will mitigate the effect of the regression to the mean (FDA, 2016; Rothmann et al., 2003).

2.3.5 Changes in the treatment effect over time (placebo and active treatment)

The use of the word "placebo" in medicine goes back to the end of the 18th century, when it was used to describe a kind of treatment to make a patient comfortable (Kerr, Milne, & Kaptchuk, 2008). The word placebo has been used since 1811 to mean a medicine given

more to please than to benefit the patient (Thomas, 2001). Shapiro & Morris (Shapiro, 1978) defined placebo as

"*a placebo is any therapy or component of therapy used for its nonspecific, psychological, or psychophysiological effect, or that is used for its presumed specific effect but is without specific activity for the condition being treated*."

Although the use of the word placebo to refer to a control treatment in clinical trials started in the 20th century, its use to describe a control group with no treatment can be traced as far back as the first trial conducted by James Lind in 1740 (Bown, 2003). The placebo-controlled trial has usually been considered as the gold standard for testing the efficacy of new treatments. The placebo in these trials is usually used as a control to test the effect of the active treatment due to its inert contents.

The placebo effect has accompanied the practice of medicine from its very beginning, but interest in placebo effects only began with the widespread adoption of the randomised controlled trial (RCT) after world war II (Koshi, E., & Short, 2007). In 1955, Beecher published his paper "The powerful placebo" (Beecher, 1955). Beecher used the words "placebo effect" to describe the positive effect of placebo in a clinical trial (Beecher, 1955). He claimed that in the 15 clinical trials he studied, placebo groups showed clinical improvement and the placebo had a therapeutic effect on the patients (Beecher, 1955). Beecher's article was reanalysed by Kienle in 1997 with the surprising result that no evidence was found of any change in the placebo effect in any of the studies cited by Beecher (Kienle & Kiene, 1997). Kienle claimed that the reported improvements in patients in these trials were due to other factors like a spontaneous improvement, fluctuation of symptoms, regression to the mean, additional treatments, irrelevant response variables, but not due to the therapeutic effect of the placebo itself (Kienle & Kiene, 1997).

In 2000, Talbot wrote a cover article for The New York Times Magazine, concluding that placebos are very powerful, and medicine should regularly make use of "the powerful placebo" (Talbot, 2000). This article revived the dilemma about the placebo effect and triggered a wave of similar articles on the same theme. A year later, an article by Hróbjartsson and Gøtzsche, published in the New England Journal of Medicine, concluded that placebos have no effect on the objective outcomes of treatment and there is no

justification for the use of placebos outside the setting of clinical trials (Hróbjartsson & Gøtzsche, 2001). This article prompted a wave of articles that now question the very existence of the placebo effect (Koshi, E., & Short, 2007). In conclusion, there may be an improvement in the placebo group, which is less than the active treatment (if the active treatment has a therapeutic effect). However, these improvements are usually due to the nature of the disease and the characteristics of the participants and not due to the therapeutic effect of the placebo.

Improvement of placebo response over time (placebo creep) was documented by several systematic and narrative reviews which revealed a continuous improvement in the placebo response over the past decades and decrease in the difference between the placebo and active treatment, mainly in antidepressant, antipsychotic and pain trials (Dold & Kasper, 2015).

Increase in the placebo response in antidepressant trials is well documented and usually considered as the main reason for the rising number of failed antidepressant trials in recent years (Furukawa et al., 2018). Walsh et al. found a positive correlation of 0.43 between the publication year and placebo response in 53 antidepressant trials published from 1980-2000 (Walsh et al., 2002). Additionally, Julious et al. found a weighted correlation of - 0.39 between the placebo response in antidepressants and the year of publication from 1966 to 2001 (Julious & Wang, 2008).

Khan et al. investigated the placebo response in antidepressant clinical trials by reviewing FDA data from 1987 to 2013 and concluded that the placebo response had increased since 2000 by 6.4%. However, the difference in treatment response between the placebo and the active treatment has remained steady over time (Khan, Fahl Mar, Faucett, Khan Schilling, & Brown, 2017). Additionally, Furukawa et al. used meta-regression to study the changes in placebo response in antidepressants in both published and unpublished trials (Furukawa et al., 2016). The review concluded that the placebo response remained constant between 1987 and 2015, ranging between 35% and 40%, and the improvement in the placebo response was not due to the placebo effect itself but instead to other trial characteristics like length of the trial and number of study centres (Furukawa et al., 2016). The differences between the Furukawa review (Furukawa et al., 2016) and the Khan review (Khan et al., 2017) are that Furukawa et al. (Furukawa et al., 2016) used meta-regression weighted for

the sample size from both published and unpublished trials. Khan et al. (Khan et al., 2017), on the other hand, used a linear regression without weighing for sample size and used data reported from FDA reviews which usually involve trials with positive results (Furukawa et al., 2018). Moreover, Khan et al. (Khan et al., 2017) examined only three covariates, while Furukawa et al. (Furukawa et al., 2016)examined 14 different covariates (Furukawa et al., 2018).

The improvement of placebo response was also investigated in relation to antipsychotic medication. Leucht et al. conducted a meta-regression for 38 antipsychotics placebo-controlled trials, with year of publication as moderator, and found that the drug-placebo difference became smaller over time. However, this difference was not statistically significant (Leucht, Arbter, Engel, Kissling, & Davis, 2009).

Aiming to investigate the potential causes of increasing placebo response over time in antipsychotics, Agid and colleagues (Agid et al., 2013) analysed all placebo-controlled antipsychotic drug trials since 1970 with meta-regression. They found that placebo response had increased over time, and this increase was associated with multi-centre trials, in trials conducted by pharmaceutical companies, shorter trial duration, younger patients, short duration of illness, higher illness severity at baseline, and a lower percentage of patients assigned to the placebo group. The number of treatment arms, country, and duration of drug washout periods were not associated with increased placebo response over time (Agid et al., 2013).

To investigate the predictors of placebo response in negative symptoms in schizophrenia, Fraguas et al. (Fraguas, Díaz-Caneja, Pina-Camacho, Umbricht, & Arango, 2018) conducted a meta-regression of all double-blinded randomised placebo-controlled trials that reported the treatment and placebo effect on negative symptoms of schizophrenia. They concluded that even though the active treatment was more effective than placebo in reducing the negative symptoms, the placebo response was statistically significant and clinically relevant. The moderators of the placebo response were a more significant number of trial arms, larger number of study sites and being funded by apharmaceutical company (Fraguas et al., 2018).

Both antidepressants and antipsychotics trials are considered as trials with subjective measures and this could be the reason for the changes in the placebo response. However, the improvement of placebo response over time has been documented in therapeutic areas where an objective measure was used. For example, a meta-analysis of a large set of antiepileptic clinical trials (1987-2009) conducted by Rheims et al. (Rheims, Perucca, Cucherat, & Ryvlin, 2011) found an improvement in both the treatment and placebo effect by increasing the year of publication. However, the treatment effect (differences between the placebo and the active treatment) was not improved and remain stable over time (Rheims et al., 2011).

Khan et al. (Khan, Fahl Mar, Schilling, & Brown, 2018b) assessed the magnitude and the pattern of the placebo effect in antiepileptic medication by reviewing data from the FDA between 1996 to 2016. The review concluded that the placebo response was increased over the 20 years; the reduction in seizure frequency increased from 5% to 20% (Khan et al., 2018b).

Khan et al. documented the improvement of placebo response (placebo creep) in anti-hyperglycaemic agents (Khan, Fahl Mar, Schilling, & Brown, 2018c). They found improvement of placebo response by 0.5% HBA1c reduction in the placebo group with no change in the effect size in general (Khan et al., 2018c). The improvement of placebo response was statistically significant in an antihypertensive trial (Khan, Fahl Mar, Schilling, & Brown, 2018a). Improvement of placebo response has also been documented in other medical fields, such as in Crohn's disease (Gallahan, Case, & Bloomfeld, 2010) and in acupuncture trials (We, Koog, Park, & Min, 2012).

Despite the considerable amount of data concluding the improvement of the placebo effect in different therapeutic areas, some argue that these changes are not due to the placebo itself but instead to the changing quality of the trials conducted and improvement of the standard care (Furukawa et al., 2018; Kirsch, 2013). Kirsch argued that the observed placebo response is not actually a placebo effect; rather the changes are due to the regression to mean phenomenon (Kirsch, 2013) (Section 2.7). However, whether due to the effect of the placebo itself or due to changes of the medical setting and the improvement of the quality of the clinical trials it is nonetheless difficult to ignore the noticed changes in the placebo response. With all this evidence, the use of historical data for indirect comparison in NI trials will be accompanied by higher chances of concluding the effectiveness of inferior treatments.

2.3.6 Placebo creep and bio-creep

Placebo creep is a cyclic phenomenon that occurs when the effect of a placebo improves over time because of improvement in the standard of care and, at the same time, the effectiveness of the active control is slightly reduced over time due to drug resistance or shifting in the human population (Julious & Wang, 2008) or other unknown reasons. Few studies have investigated the presence of placebo creep. Julious and Wang presented evidence of improvement of placebo response over time in anti-depressant drug trials between 1966 and 2011 with a weighted correlation of - 0.39 (Julious & Wang, 2008).

Explanations as to the causes of placebo creep include improvement of standard (concomitant) care, population drifts and geographical differences (Julious & Wang, 2008). In the literature, the information about placebo creep is minimal, and most of the time, there is confusion between bio-creep and placebo creep.

D'Agostino et al. (2003) defined bio-creep as; "*The phenomenon that can occur when a slightly inferior treatment becomes the active control for the next generation of non-inferiority trials and so on until the active controls become no better than a placebo*." The main concern regarding the presence of bio-creep in NI trials was highlighted in the GAO report in 2010 (GAO, 2010) as "*A concern that successive generations of drugs approved*

based on non-inferiority trial, with the active control changing in each new generation, could lead to the adoption of decreasingly effective drugs."

Figure 2-1 illustrates the presence of placebo creep and the effect of bio-creep on the efficacy of active control:

- The placebo was used as the comparator with treatment one (T1) in the period (A). The efficacy of (T1) compared to the placebo (P) was established, in the period (A).
- T1 became the active comparator (as it was unethical to use the placebo any more) and was compared to treatment two (T2) in the period (B). The non-inferiority of T2, compared to T1, was established.
- With time T2 became the active comparator and was compared with treatment three (T3) in the present time (period C).
- By the end of the different trials, the conclusion was that T3 is non-inferior to T2, but it is inferior to T1, and its efficacy is almost the same as the efficacy of placebo, which is known as bio-creep. Using T2 as active comparator instead of T1 or placebo will lead to overestimation of the efficacy of the new treatment T3, which could lead to approval of an ineffective drug.

It is also clear from the figure that the effect of placebo improved over time, which is due to placebo creep.

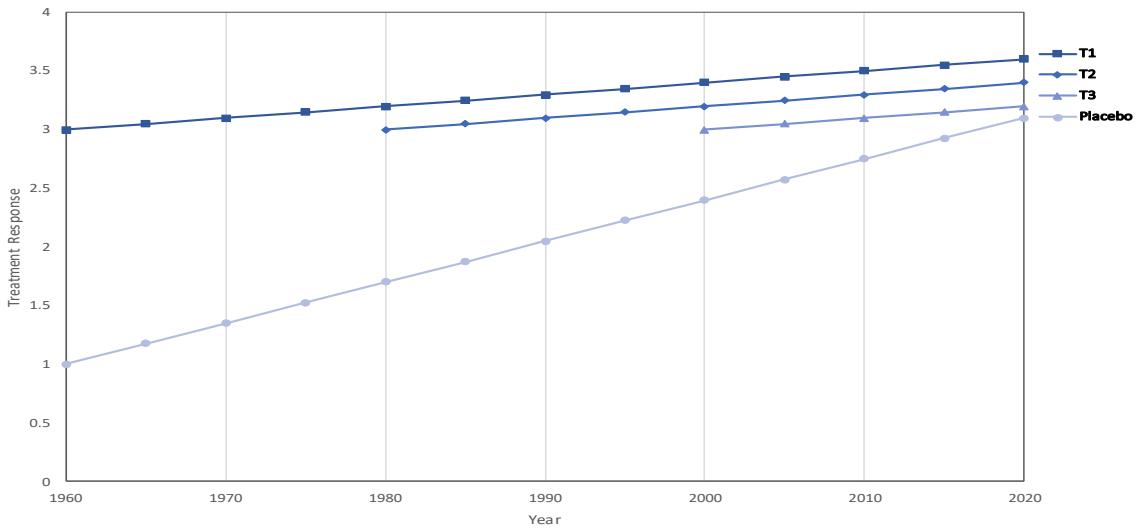

Figure 2-1 Graphical presentation of Placebo creep and Bio-creep in NI trials

Choosing the active control and estimating its effect size from both the historical trials and NI trials are the most important factors in affecting the occurrence of bio-creep (Everson-Stewart & Emerson, 2010; Fleming, 2008).

Fleming explained the hazard of bio-creep in anti-infective trials where generations of non-inferiority trials were conducted, leading to approval of antibiotics that may not be providing as large a clinical effect compared to the placebo as perceived and could induce safety risks and development of resistance (Fleming, 2008).

Addressing the possibility of bio-creep by choosing the best active control available is very important in any NI trial (D'Agostino et al., 2003). The efficacy of the active control should be protected and maintained in any NI trial.

Odem-Davis & Fleming (2015) reported several factors that could influence the risk of bio-creep in NI trials. Besides choosing the appropriate active control, these factors include the method for choosing the non-inferiority margin to account for publication bias and random high bias and regression to the mean.

Several methods were proposed to minimise the occurrence of bio-creep in NI trials (Fleming, 2008; Odem-Davis & Fleming, 2015; Odem-Davis & Fleming, 2013; Rothmann et al., 2003). The FDA guideline published in 2016 recommends the use of the 95% - 95% fixed margin method for this purpose (FDA, 2016).

2.4 Setting of the non-inferiority margin

As highlighted earlier in this chapter, once the effectiveness of treatment becomes such that placebo-controlled trials are no longer possible, this active treatment could then be the active control for further new treatments through NI trials. In this context, there will be a need to determine an acceptable non-inferiority margin that takes into account the historical evidence, the relevant statistical considerations, and medical judgement. The non-inferiority margin is usually established from the main estimate from the meta-analysis of the placebo-controlled trials that compare the active treatment to the placebo. In this section, the role of pairwise meta-analysis in the setting of the NI margin will be discussed, followed by the methods for setting the NI margin.

2.4.1 The role of pairwise meta-analysis in setting the Non-inferiority Margin

GLASS (1976) defines meta-analysis as

"*The statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings.*"

Meta-analyses are considered to be the top evidence-based medical studies and an important tool for treatment approval (Paul & Leibovici, 2014). In NI trials, a meta-analysis can be used to estimate the historical effect of the active control compared to the placebo to set the NI margin.

Pairwise meta-analysis depends on a direct comparison between two treatments. In NI trials, a pairwise meta-analysis is conducted to measure the effect size of the active comparator from historical placebo-controlled trials. Two comparator treatments are included (the placebo and the active control). The point estimate and the 95% CI extracted from this meta-analysis is used for setting the NI margin to indirectly compare the efficacy of the experimental treatment compared to the placebo either by the fixed margin approach or the synthesis approach.

There are several issues that accompany the planning, conducting and analysis of a meta-analysis. These include the source of data used, study selection (publication bias), differences between the studies (heterogeneity), and choosing the appropriate model for analysis (fixed versus random models). In the case of NI trials, violation of these assumptions will lead to either over or underestimation of the effect size of the active control compared to the placebo, which eventually leads to a biased NI margin.

2.4.1.1 Heterogeneity

Heterogeneity is defined as any variability among studies that are included in the meta-analysis. Heterogeneity could be in the form of clinical diversity, methodological diversity or statistical heterogeneity (Higgins & Green, 2008). Heterogeneity should be investigated initially by inspection of the 95% CI in the studies. Non-overlapping CI is an initial indicator of heterogeneity (Pinto, 2013). Cochran's *Q* is a chi-square distributed method that is used to measure the heterogeneity. It is calculated as the weighted sum of squared differences between individual study effects and the pooled effect across studies, with the

weights being those used in the pooling method. The interpretation of the results from Cochran's Q test should be treated with caution since it has low power to detect heterogeneity in a meta-analysis that contains studies with small sample size or when small numbers of studies are included in the meta-analysis (Higgins & Green, 2008). In contrast, if a large number of studies are included in the meta-analysis, Cochran's Q test will have a high power to detect a small amount of heterogeneity, which could have no clinical importance (Higgins & Green, 2008).

The I^2 statistic is an alternative method to measure heterogeneity. It describes the percentage of variation across studies that is due to heterogeneity rather than chance, and is a derivative from the Q statistics. I^2 is an intuitive and simple expression of the inconsistency of studies' results. Unlike Q it does not inherently depend upon the number of studies considered (Higgins & Green, 2008).

$I^2 > 50\%$ is indicative of considerable heterogeneity, $I^2 \geq 30$ % and ≤ 50 % is indicative of moderate heterogeneity, and $I^2 < 30$ per cent is indicative of mild heterogeneity (Whitehead, 2002).

In the case of NI trials, the variability (heterogeneity) between historical placebo-controlled trials could affect the measurement of the efficacy of the active control based on these trials (FDA, 2016). A high percentage of heterogeneity requires additional investigation to attempt to explain the heterogeneity. Sensitivity analysis, subgroup analysis, and including the possible effect modifiers should be undertaken.

2.4.1.2 Publication bias

Another issue that could be faced in a meta-analysis is the chance of reporting bias (publication bias). Studies with positive results are more likely to be published than studies with negative results. They are more likely to be rapidly published in high impact journals and more likely to be cited by others (Rothstein, Sutton, & Borenstein, 2006). A meta-analysis that contains only studies with positive results tends to have positively biased results. In the case of NI trials, the presence of publication bias could lead to concluding the efficacy of ineffective active control, which eventually will lead to the conclusion of

non-inferiority of the inferior experimental treatment. The funnel plot is the most commonly used visual method to assess the publication bias (Rothstein et al., 2006).To minimise the effect of publication bias, meta-analyses should include both published and unpublished studies. However, searching for unpublished studies is usually hard and challenging to perform, especially for older trials.

2.4.1.3 Fixed effect model (FE) versus random effects model (RE)

The fixed-effect model assumes all trials are to estimate a common treatment effect with any differences across trials in observed effects assumed to be due to sampling variation (within trial variations only). The random-effects approach allows for between trial and within trial variations. Selection of the model affects the overall effect size, mostly where the studies in the analysis include both small and large studies. A fixed effect model gives more weight to studies with larger sample size, while a random effects model gives more weight for smaller studies; this means under a random effect model studies with extreme results will have less influence if they are large and more influence if they are small (Borenstein, 2009). However, smaller studies tend to have more extreme results than large ones. The variance, standard error and the confidence interval are wider in the random effect model compared to the fixed model, since the random model accounts for both the sampling variance and the between-study variance (Borenstein, 2009).

The selection between random or fixed models depends on the nature of studies included, number of studies, and the assumption of heterogeneity. Usually, a random effect model is more appropriate for meta-analyses that include a large number of studies with different sample sizes. On the other hand, a fixed effect model is more appropriate if the meta-analysis includes studies with similar sample sizes and a small number of studies (Borenstein, 2009).

Even though the random effects model accounts more for the heterogeneity, there are multiple concerns regarding using it to estimate the effect of the active control in NI trials. First, the random effect model will give more weight to trials with smaller sample size compared to the fixed effect model; this will violate the assumption of the similarity between NI trial and the historical trials since NI trials tend to be conducted with larger sample size (Rothmann et al., 2012). In Chapter 5 of this thesis the changes in sample size

over time will be investigated, as well as their effect on the main estimate in the meta-analysis. In Chapter 6, the difference between fixed and random models will be investigated.

2.4.2 Non-inferiority Margin (M)

An NI trial is undertaken based on the quantification of a margin that in turn depends on evidence on the effectiveness of the active control in historical well-conducted placebo-controlled trials. This assessment of the effectiveness needs to account for any possible biases, effect modifiers and the clinical judgement.

There are different methods used to set the NI margin, most of which in general follow one of two approaches to set the appropriate NI margin (Rothmann et al., 2012). The first approach depends on making adjustments for any possible biases or uncertainty and then using a test procedure that targets a pre-specified type I like error rate (Rothmann et al., 2012). The second approach involves the use of methods for analysis with the hope that they will account for any possible biases that could arise from the use of unadjusted active control effect (Rothmann et al., 2012). This thesis will present the most common methods used to set the NI margin, namely the fixed margin method and the synthesis method, since these are the only methods identified by the regulations (FDA, 2016).

The NI margin is a pre-specified amount (M), which can be used to demonstrate that the test product is no worse than the active control (D'Agostino et al., 2003; FDA, 2016). It is the amount by which the active control can exceed the test treatment for it to be concluded that the test treatment is non-inferior to the active control and indirectly superior to the placebo (D'Agostino et al., 2003).

The null hypothesis means that the active control is superior to the test treatment, and the alternative hypothesis means that the test treatment is not inferior to the active control (D'Agostino et al., 2003). The determination of the non-inferiority margin is based on both statistical consideration and clinical judgement (ICH, 2001; CHMP, 2005; FDA, 2016).

According to FDA regulations, two NI margins should be identified in any NI trial: M1 and M2. **M1** "*is the whole effect of the active control relative to placebo*" (FDA, 2016). M1 is estimated indirectly from the historical placebo-controlled trials where active control

worked as a test treatment against placebo. The validity of the NI trial depends on the choice of M1 (ICH, 2001; CHMP, 2005; FDA, 2016).

M2 "*is the largest clinically acceptable difference (degree of inferiority) of the test* treatment *compared to an active control*" (FDA, 2016). M2 represents clinical judgement. M2 is a fraction of M1 that is judged the clinically acceptable difference between the active control and test treatment and should always be smaller than M1. Both the point estimate and the boundaries of the confidence interval (CI) are essential in the statistical analysis of non-inferiority. Different methods are used to set the NI margin, and these include the fixed margin method and the synthesis method.

The Hypothesis to be tested is:

$H_0: C - T \geq M2$, (active control is superior to test treatment) (2.1)

$H_a: C - T < M2$, (test treatment is not inferior to the active control) (2.2),

Where C is the effect size of the active control, T is the effect size of the test treatment
Figure 2.2 gives four different scenarios for the results of the NI trial method:

a- The point estimate is less than zero, which favours the test treatment. The upper bound of 95% CI is less than zero; the superiority of the test treatment over the active control is concluded.
b- The point estimate is equal to zero and the upper bound of 95% CI is less than M2; non-inferiority of test treatment is concluded.
c- The point estimate is equal to zero, but the upper bound of 95% CI is less than M1 and larger than M2; clinical judgement could lead to the conclusion of effectiveness (FDA, 2016).
d- The point estimate favours active control and the upper bound of 95% CI is greater than M1; non-inferiority cannot be established.

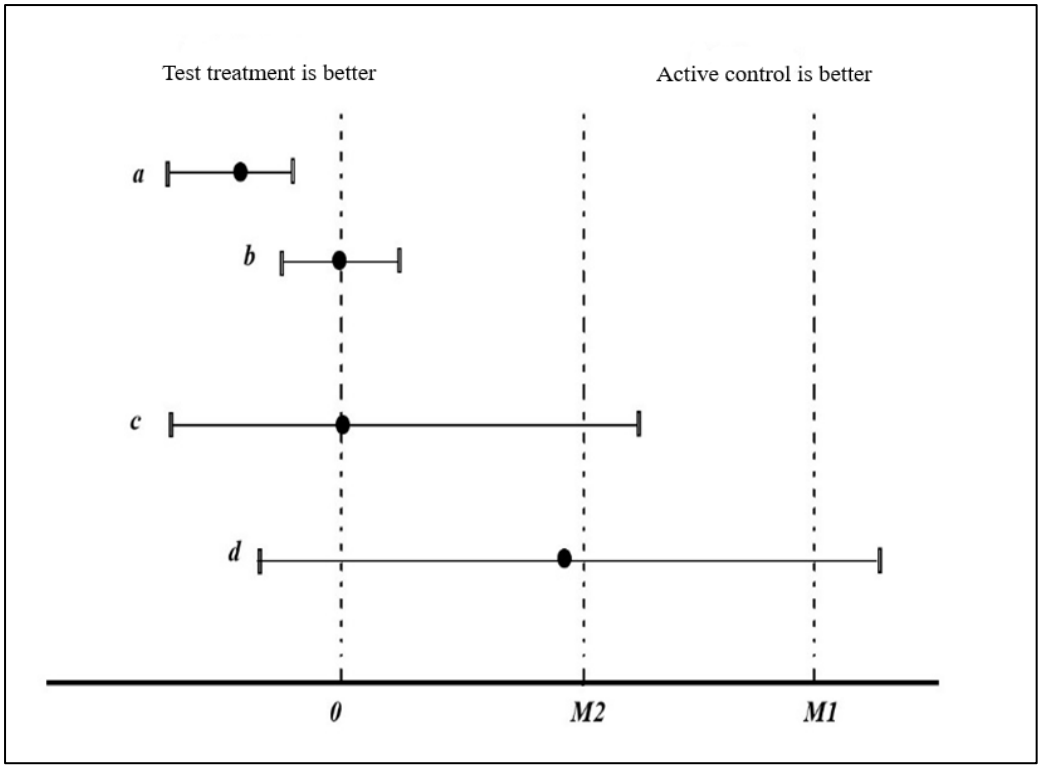

Figure 2-2 The possible outcomes in the NI trial
(adapted from FDA guidelines, 2016)

2.4.2.1 Fixed margin method for setting the NI margin

This is considered as the most common method for setting the NI margin. The fixed margin approach is also known as the two confidence intervals approach and or the 95% - 95% method. This method is the method recommended by the regulatory guidelines (FDA, 2016). It depends on choosing a fixed margin in the designing stage of the NI trial based on historical data.

To obtain this margin, the estimate of the active control effect from the historical placebo-controlled trials needs to be obtained. The lower bound of the CI of the historical placebo-controlled trials will be defined as M1. Both the variability of the active control effect and constancy assumption should be addressed in this stage.

M2 will be taken as a fraction of M1 depending on the clinical judgement. Using a fraction of the lower bound of the confidence interval as the NI margin (M2) is a common practice and is recommended by FDA regulations. It is especially important if the primary endpoint

is mortality or irreversible morbidity, and this is referred to as “preservation of effect”, and it guarantees that some fraction of the effect of the active control is preserved (FDA, 2016; Rothmann et al., 2012).

From historical trials: M1 = lower bound of 95% CI of (C-P) (2.3)

M2 = percentage of M1 (2.4)

From the NI trial: The upper bound of 95% CI (T-C) > M2 (2.5)

Where C is the active control, P is the placebo, T is the test treatment, M1 is the statistical NI margin, and M2 is the clinical NI margin

M2 is the fixed margin in this method, not the M1. Using M2 instead of M1 will account for any effect modifiers, regression to the mean bias or deviation from the constancy assumption (Rothmann et al., 2012). M1 and M2 are used to demonstrate that the test treatment is superior to placebo and is not unacceptably worse than the active control (Rothmann et al., 2012). The determination of M2 should always be implied after the choosing of M1. In cardiovascular diseases, M2 is usually 50% of M1 (FDA, 2016). In anti-infective trials, M2 is usually set at 10-15% of an absolute risk difference scale between treatments (FDA, 2016). Figure 2.2 illustrates the different outcomes using the fixed margin method.

There is an argument that using a fraction of the lower bound of the confidence interval is uniformly conservative (chance of concluding the non-inferiority of an inferior treatment is low) (Sankoh, 2008). Sankoh recommends the use of a fraction of the point estimate instead of the lower bound of the CI (Sankoh, 2008). However, using a fraction of the lower bound of CI is not conservative if the constancy assumption is violated or in cases where the regression to the mean and other biases are major problems (Rothmann et al., 2012). Moreover, choosing a fraction of M1 can provide an allowance for the deviation from the constancy assumption (Rothmann et al., 2012).

The 95% CI is the most commonly used CI with this approach. It is known as the 95% - 95% approach because two different 95% CIs, one from the historical placebo-controlled

trial and the other from the NI trial, are used to estimate the non-inferiority margin (Rothmann et al., 2012).

The advantages of using this method are: separation of the calculation, justification, and determination of the NI margins from the NI analysis stage (since the NI margins will be determined in the design phase of NI trial). The separation will keep the variability of estimated treatment effect from the past trials and variability of observed treatment effect from the NI trial separate (Wangge et al., 2010). The pre-specified margin will be used in determining the sample size of the NI trial needed to provide sufficient power for testing the NI hypothesis and controlling for type I error. Choosing a fraction of M1 can provide an allowance for the deviation from the constancy assumption.

2.4.2.2The synthesis method

The synthesis method is usually used in the analysis phase of the trial, where both the main estimates from the meta-analysis of placebo-controlled trials are used to set the M1 (instead of the 95% CI boundaries). Then a fraction (percentage) from the active control effect will be determined to be the M2.

2.5 Methods for analysis of NI trials

The most common approaches used are the fixed margin method and the synthesis method (the regulatory approaches), and the network meta-analysis (the predictive approach).

2.5.1 Fixed margin approach

As mentioned in section (2. 4.2.1), a NI margin M2 will be specified in the designing phase of the NI trial. This margin represents a fraction of the effectiveness of active control. In the analysis phase of the NI trial, the effectiveness of the test treatment is judged by the upper boundary of the confidence interval (CI) from the conducted NI trial.

2.5.2 Synthesis method

The synthesis method combines the estimate of treatment effect relative to the control from the NI trial with the estimate of the control effect from the historical trials (FDA, 2016). It treats both sources of data as if they came from the same source (which is opposite to the

separate approach in the fixed margin method) to establish the placebo effect in the NI trial. A single confidence interval is then used, combining the results from both the NI trials and the historical trials to test the null hypothesis that the treatment is non-inferior to the active control, without actually specifying any fixed NI margin based on the control effect (FDA, 2016).

This approach assumes that the constancy assumption holds for any NI trial, i.e. that there is no between trials variability. As only one 95% CI is used, which in turn is derived from the historical placebo-controlled trials, this approach is both less conservative (concluding the non-inferiority of an inferior treatment) and less accessible to measure the difference between test treatment and placebo (Rothmann et al., 2012). This approach can be used with both Frequentist and Bayesian approaches (Rothmann et al., 2012).

This method compares test statistics based on the estimates from the NI trial and from the historical trials with their cross ponding error (FDA, 2016).

$$Z = \frac{\hat{\Delta}TP}{\sqrt{SE^2(TC) + SE^2(CP)}} \qquad (2.7)$$

Where C is the effect size of the active control, T is the effect size of the test treatment,

P is the effect size of the placebo, $\hat{\Delta}TP =$ (T - C) - (P - C) is the difference between test treatment and placebo, TC is the difference between the effect size of test treatment and active control, CP is the difference between the effect size of the placebo and active control. $SE\ (CP)$ is the standard error of the difference between C and P (from the meta-analysis of historical trials), and $SE(TC)$ is the standard error on the NI trial. If the Z is smaller than the predetermined Z value (for Type I error), the NI will be concluded.

The main difference between the synthesis and fixed margin methods is in the standard error measure, the fixed margin method assuming the standard error of the indirect comparison is the sum of the standard error of the meta-analysis of historical trials and the standard error of the NI trial

$$Z = \frac{((\hat{\Delta}TP)}{\sqrt{(SE^2\ (TC)} + \sqrt{SE^2(CP)}} \qquad (2.8)$$

In the synthesis method, M1 is the main estimate of the meta-analysis of the placebo-controlled trial (instead of the 95% CI boundaries in the fixed method), M2 is the percentage of the preserved active treatment effect that will be pre-specified (based on the clinical judgement) of the main estimate (FDA, 2016).

According to the FDA regulations, using synthesis method will lead to smaller sample size and greater power for a given sample size compared to the fixed margin approach if the constancy assumption holds (FDA, 2016). That is because the synthesis method uses a smaller standard error compared to the fixed margin approach method.

2.5.3 Network meta-analysis (NMA) (Predictive Approach)

Network meta-analysis (NMA) is a meta-analysis where multiple treatments are compared both directly and indirectly based on the common comparator. The idea behind the use of NMA is that for many diseases, there are many interventions possible for treatment, and so there is a need to compare these treatments (in the case of NI trials comparing placebo, active treatment and the experimental treatment). However, in clinical trials, it is difficult and costly to compare more than two treatments in the same trial. A network meta-analysis was introduced just over 20 years ago as a solution to this problem (Tonin, Rotta, Mendes, & Pontarolo, 2017). A network meta-analysis allows synthesis, estimation and comparison of the effectiveness of several treatments in one setting (Donegan, Williamson, D'Alessandro, & Smith, 2013). It uses all direct and indirect evidence to produce relative effects of all compared treatments.

Lumley first introduced the network meta-analysis (NMA) in 2002 (Lumley, 2002). The model introduced by Lumley was extended by Lu and Ades in 2004 who included multiple treatment comparisons in the model through Bayesian and were able to rank the included treatments from best to worst (Lu & Ades, 2004).

Figure 2.3 illustrates the simplest form of a network meta-analysis in NI trials that include only three treatments. The common comparator in the model is the active control (C), the nodes represent the interventions (the included treatments).

The use of NMAs allows for both head-to-head comparison and indirect comparison in the same model, which is considered as an advantage compared to traditional indirect

comparisons. Moreover, NMA reduces the cost of conducting additional clinical trials and offers an overview of the entire set of the clinical condition (available treatments, possible outcomes, side effects of each treatment).

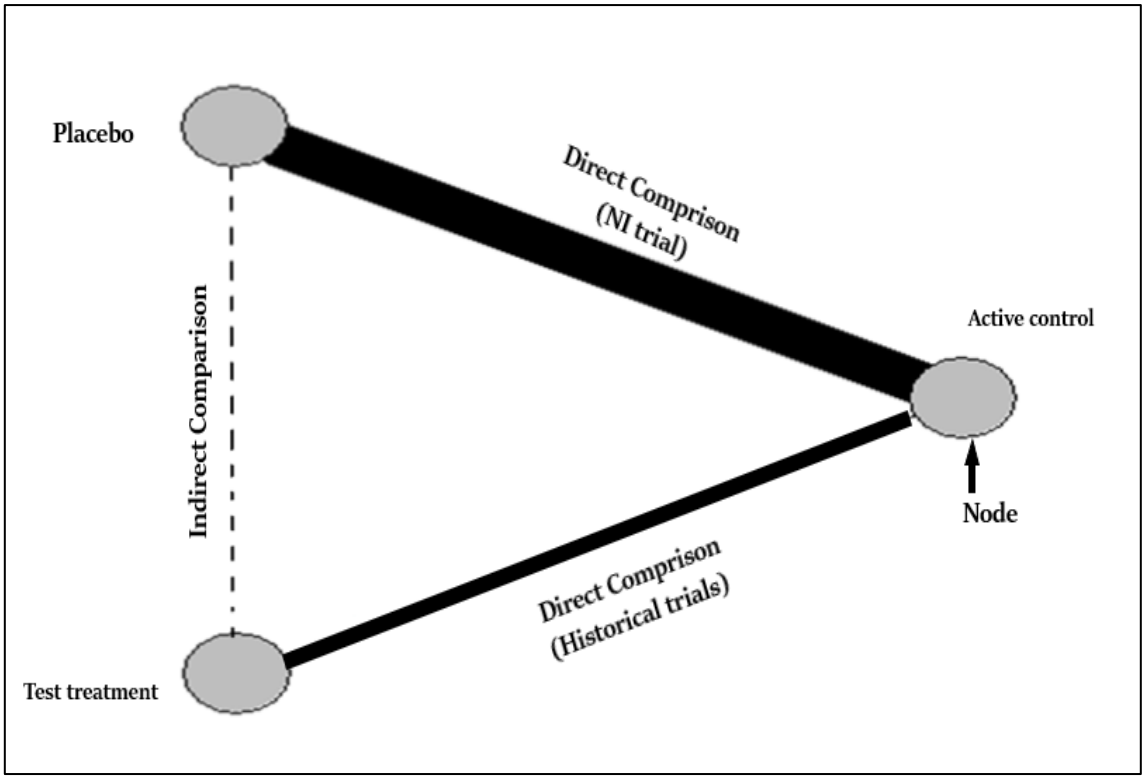

Figure 2-3 Network meta-analysis
The thickness of the lines represents the sample size, nodes represent treatments, and dashed line represents the indirect comparison

NMA allows for the determination of the amount of agreement between results from different comparisons for the same treatment (Tonin et al., 2017). Since 2008, the number of published studies that include NMA has increased, mostly in pharmacological interventions and mostly in the therapeutic areas of cardiovascular, oncology, mental disorders and infectious diseases (Tonin et al., 2017). NMA models are available for all types of data in both the Frequentist and Bayesian framework with different software available for analysis including R, STATA, SAS and Win bugs (Tonin et al., 2017).

As mentioned earlier, the conduct and analysis of NI trials depend on the indirect comparison between the test treatment and placebo. For this reason, NMA can be used in the designing phase to compare all possible active controls with placebo and to set an NI

margin. Moreover, it can be used in the analysis phase of NI trials to provide both direct and indirect comparison between all possible active controls, placebo and the test treatment in the same model. NMA allows for a comparison of multiple treatments in the same model, which increases the reliability of the comparisons and ensures the selection of the best available active control to compare with the test treatment (Schmidli & Wandel, 2011). The direct comparisons were between the placebo and the active treatment in the historical trials and between the active treatment and test treatment in the NI trial. The indirect comparison is between the test treatment and the placebo. The strength of the network depends on the treatments in the network, how they are presented in the model, and the evidence they carry (Tonin et al., 2017). The analysis framework can be implemented using either Frequentist or Bayesian approaches. As in pairwise meta-analysis, fixed and random effects models can be used in NMA. In addition to the assumptions of homogeneity, the consistency assumption is also essential in NMA.

These assumptions must be based as far as possible on both statistical and clinical judgement (Tonin et al., 2017). These assumptions include the homogeneity assumption, where trials in the network meta-analysis that are directly compared must be sufficiently similar, and the similarity assumption, where the trials included in NMA should be selected based on well-defined criteria that ensure the similarity between the trials. The study population, study design, efficiency measures and the effect modifiers should be comparable to reduce the chance of bias in the pooled estimate (Tonin et al., 2017).

Finally, for the consistency (transitivity) assumptions, which are specific for network meta-analysis, there should be an agreement between direct and indirect evidence (White, Barrett, Jackson, & Higgins, 2012). When direct and indirect evidence are combined for a particular comparison, it is vital that the indirect estimate is not biased, and there is no discrepancy between the direct and indirect comparisons (Tonin et al., 2017). The statistical manifestation of the consistency is called transitivity (Tonin et al., 2017).

Hoaglin describes both fixed and random effect models, implemented using both Frequentist and Bayesian equation frameworks (Hoaglin et al., 2011).

2.5.3.1Fixed effect network meta-analysis

In the following results, A is the primary reference treatment (active control), B is the placebo, and C is the test treatment. AB trials are the historical trials, and AC trial is the NI trial. The indirect comparison will be between B (placebo) and C (test treatment) (Hoaglin et al., 2011). The fixed effect model is given by:

$$\eta_{jk} = \begin{cases} \mu jb & b = A,B,C, if\ k = b \\ \mu jb + d_{bk} = \mu_{jb} + d_{AK} - d_{Ab} & k = A,B,C,D\ if\ the\ k\ is\ after\ b \end{cases} \quad (2.9)$$
$$d_{AA=0}$$

Where η_{jk} is the outcome of treatment *k* in study *j*, μ_{jb} is the outcome for treatment *b* in study *j*, d_{bk} is the fixed effect of treatment *k* relative to treatment *b*. The $d_{bk}$are identified by expressing them in terms of effects relative to treatment A: $d_{bk} = d_{Ak-d_{Ab}}$with d_{AA}=0 (the order of the subscripts on d_{bk} is conventional, but counterintuitive). For the underlying effects, this relation is a statement of consistency: the "direct" effect $d_{bk}$and the "indirect" effect $d_{Ak}-d_{Ab}$are equal.

2.5.3.2Random effect network meta-analysis

The random effect model takes into consideration both within trial variation (sample variation) and within between-trial variation. The study-specific treatment effects δ_{jbk} are assumed to follow a Normal distribution $\delta_{jbk} \sim N(d_{bk}, \sigma^2)$, where σ^2 is the random effect variance and when $\sigma^2 = 0$ (i.e. there is no between-study heterogeneity) a fixed effect model is specified (Hoaglin et al., 2011).

The random effects model can be written as:

$$\eta_{jk} = \begin{cases} \mu jb & b = A,B,C, if\ k = b \\ \mu jb + \delta_{jbk} & k = A,B,C,D\ if\ the\ k\ is\ after\ b \end{cases}$$

$$\delta_{jbk} \sim N\left(d_{bk} + \beta_{bk}X_j, \sigma^2\right) = N\left(d_{Ak} - d_{Ab} + (\beta_{Ak} - \beta_{Ab})X_j, \sigma^2\right) \quad (2.10)$$
$$d_{AA} = 0, \beta_{AA} = 0$$

The d_{bk} are identified by expressing them in terms of effects relative to treatment A: $d_{bk} = d_{Ak-d_{Ab}}$ with d_{AA}=0 (the order of the subscripts on d_{bk} is conventional, but counterintuitive). For the underlying effects, this relation is a statement of consistency: the "direct" effect d_{bk} and the "indirect" effect $d_{Ak}-d_{Ab}$ are equal.

The random effect model allows for heterogeneity between and within the trials, but it cannot explain it (Hoaglin et al., 2011). Using network meta-regression models that take into account the covariates in the model may account for and explain heterogeneity, and therefore reduce both inconsistency and biases (Jansen et al., 2011). However, Rothmann et al. argue that the use of the random effect model could lead to biased results since it gives more weight for smaller trials which usually tend to have more extreme results (Rothmann et al., 2012).

Both Frequentist and Bayesian approaches could be used for NMA. The Frequentist approach measures the probability that the observed results occurred under specific sampling distribution of the hypothesised values of the parameters (Tonin et al., 2017). This approach applies traditional statistical methods to make the comparison. A network meta-analysis belongs to a category of generalised linear mixed models that use a likelihood-based function to estimate model parameters (point estimate) and estimate the confidence interval (CI) (Schmidli & Wandel, 2011). In the case of NI trials, the primary interest is the indirect prediction of the efficacy of the test treatment compared to placebo.

Bayesian Network Meta-analysis combines the likelihood function with prior information about these parameters to obtain a posterior distribution for these parameters (Hoaglin et al., 2011). The Bayesian approach can lead to a straightforward prediction of the treatment effect. However, a common criticism is that the results could be biased if an inappropriate prior was chosen (Hoaglin et al., 2011).

Using Bayesian methods in the analysis of NI trials became more popular due to the advances and the availability of Bayesian software, also because Bayesian methods provide an intuitive framework for accounting more for the heterogeneity between the trials (Lin et al., 2016). In the Bayesian model, the likelihood function represents the extent to which different values for the parameter of interest are supported by the data (Hoaglin et al., 2011). The posterior distribution (the outcome) can be interpreted regarding probabilities of which treatment from the compared treatments is the best and also other probabilities can be definined (Jansen et al., 2011).

An example of a Bayesian random effect model with log odds as the outcome measure (Hoaglin et al., 2011) is given below:

Likelihood: (2.11)

$$r_{jk} \sim binomial\left(P_{jk}, n_{jk}\right)$$

Model:

$$logit(p_{jk}) = \begin{cases} \mu_{jb} & b = A, B, C, if\ k = 0 \\ \mu_{jb} + \delta_{jbk}, & k = B, C, D, if\ k\ after\ b \end{cases}$$

$$\delta_{jbk} \sim N(d_{bk}, \sigma^2) \sim N(d_{AK} - d_{Ab}, \sigma^2)$$

$$d_{AA} = 0$$

Priors:

$$d_{Ak} \sim normal\ (0, 10^6)\ k = B, C, D$$

$$\sigma \sim uniform\ [0, 2]$$

The main challenge in using a Bayesian approach is choosing the appropriate prior distribution (Hoaglin et al., 2011; Lin et al., 2016; Schmidli & Wandel, 2011). The choice of prior should be based on the nature of the studies that are included in the network, as well as the purpose of the analysis.

The choice of the prior depends on the distribution of the data (e.g. continuous or binomial) and the structure of the network (the number of treatments included in the network). A stronger prior is needed if not enough data is available (Hoaglin et al., 2011).

In the analysis of the Essence trial, Schmidli and colleagues used a standard reference normal prior for the random effect means and a half normal prior for the between-trial standard deviation (Schmidli, Wandel, & Neuenschwander, 2012). This approach was the same approach as used by Hoaglin et al. and Lin et al. (Hoaglin et al., 2011; Lin et al., 2016). There is a need to check the assumptions in the Bayesian approach in the same way as in the Frequentist approach (Hoaglin et al., 2011). Several methods were proposed for evaluating the consistency assumption (Lu & Ades, 2006; Lumley, 2002).

Both Frequentist and Bayesian methods have their pros and cons. A Frequentist approach will result in point estimated confidence intervals, while a Bayesian approach provides a posterior distribution of the parameters from which summaries such as median and 95% credible intervals (CrI) can be taken (Hoaglin et al., 2011). Checking the assumptions is very important in both approaches.

Addressing inconsistency is the main challenge in the Frequentist approach. In the Bayesian approach, the subjectivity in choosing the prior distribution is the main issue. The Bayesian approach covers the uncertainty in the study parameters and makes direct probability statements regarding interested parameters, i.e. it has a straightforward way to make predictions with more flexible prediction models (Tonin et al., 2017).

As mentioned in Chapter 1 (1.2.1) the main aim of this thesis is to investigate how the adjustment for the time scale could improve the estimate used for setting the NI margin; using network meta-analysis in the analysis of NI trials while adjusting for co-variables (network meta-regression) could be one possible solution.

2.6 Illustrated Example

Until now, this thesis has described three possible methods for analysing the NI trial: two that are described as regulatory approaches (fixed effect method and synthesis method) and the network meta-analysis (the predictive approach). This example, will illustrate how these different methods could be used to analyse an NI trial.

The OASIS-5 was a multicentre double-blinded randomised controlled trial that investigated the non-inferiority of fondaparinux (test treatment) compared to enoxaparin (Active control) (low molecular heparin LMWH). It included 20,078 patients from 576 centres from forty-one countries. The primary endpoint was the triple endpoint of death, myocardial infarction (MI) or refractory ischemia (OASIS investigators, 2006)

There were no placebo-controlled trials to compare the enoxaparin and placebo. The investigators used a meta-analysis of historical placebo-controlled trials that compared either heparin (unfractionated heparin UFH) or other LMWH to placebo (Eikelboom et al., 2000) to establish the efficacy of enoxaparin compared to placebo. The estimate from the historical meta-analysis shows that the odds of death or myocardial infarction in the UFH and LMWH groups compared to placebo group were 0.52 [0.37; 0.72]. This means the odds of death in the placebo group compared to active control (UFH or LMWH) were 1.92 [1.38; 2.70]. Figure 4.2 illustrates the forest plot for the difference between the LMWH and UFH compared to placebo.

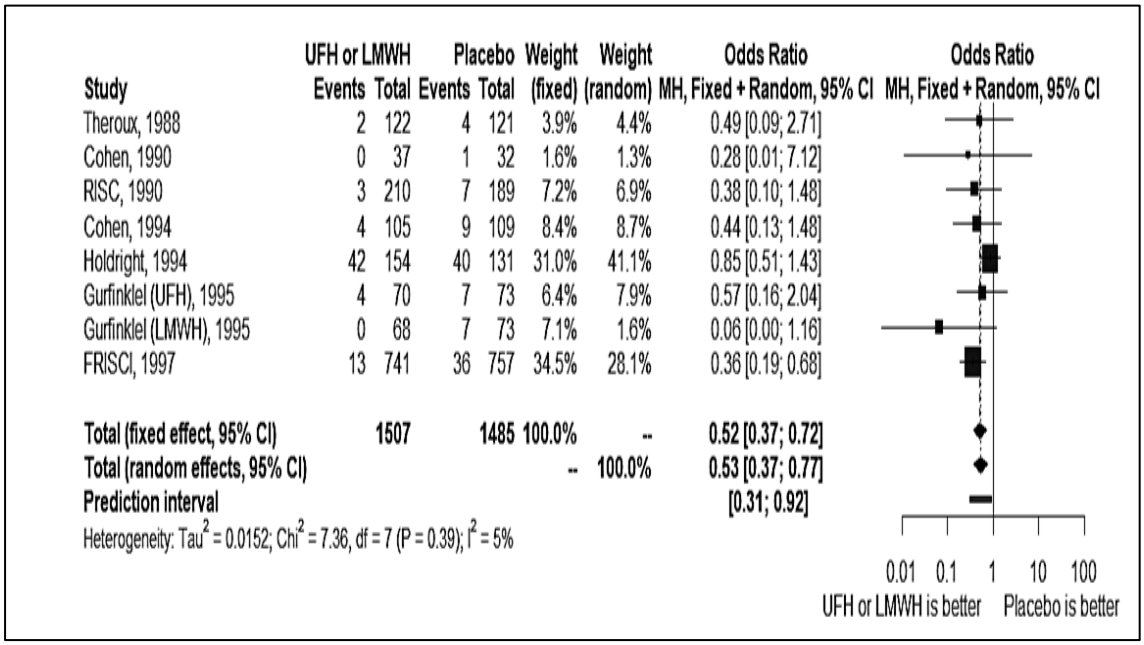

Figure 2-4 Forest plot of comparison between Placebo vs LMWH or UFH

2.6.1 .Using the fixed margin method for analysis of OASIS trial

From Figure (2.4), the M1 will be the lower limit of the 95% CI of placebo versus (LMWH and UFH) =1.38. M2 will be the ½ log odds at the lower limit of the 95% CI = 1.18. The reason for chosing the 50% of the M1 is based on clinical judgement and the recommendation from the FDA (OASIS investigators, 2006; FDA, 2016). In the OASIS trial, the odds of death or MI (95 % CI) = 0.9 (0.81; 1.01), the upper limit of the 95% CI was less than the M2 (1.18). Based on these results, the non-inferiority of the fondaparinux compared to enoxaparin can be concluded.

2.6.2 Using the Synthesis method for the analysis of OASIS trial

Under the synthesis method, a 50% fraction from the main estimate from the meta-analysis of placebo-controlled trial (LMWH and UFH versus placebo) will be used as M2 instead of the whole estimate M1. Test statistics (Z) will be used for the analysis (FDA, 2016). The predetermined Z< -1.96.

Log odds (T-C) (fondaparinux versus enoxaparin from the NI trial) = log 0.9

Log odds (C-P) (LMWH & UFH versus Placebo from the meta-analysis) = log (1/1.8)

$$(T\ (\text{Fondaparinux}) - P(Placebo)) = \frac{(\log odds\ (T-C))+{}^{1}/_{2}(\log odss\ (C-P))}{\sqrt{(SE(T-C))^{2}+({}^{1}/_{2}SE\ (C-P))^{2}}} \qquad (2.12),$$

The observed test statistics is

$$= \frac{(-0.1)+{}^{1}/_{2}(-0.63)}{\sqrt{0.003+[{}^{1}/_{4}(0.0042)]}} = -\ 6.5$$

The (- 6.5) is less than (more negative) -1.96. The non-inferiority of the fondaparinux compared to enoxaparin is concluded.

2.6.3 Using the network meta-analysis for the analysis of OASIS trial

The NMA will be conducted using the Frequentist package "netmeta" R Package (Guido Schwarzer, 2015). The historical placebo-controlled trials used by the OASIS investigator to set the non-inferiority margin will be incorporated in a network with the OASIS trial to investigate the efficacy of the test treatment fondaparinux compared to enoxaparin (LMWH). The network is composed of the placebo, the active control (LMWH and UFH) and the test treatment (fondaparinux) (Figure 2.5).

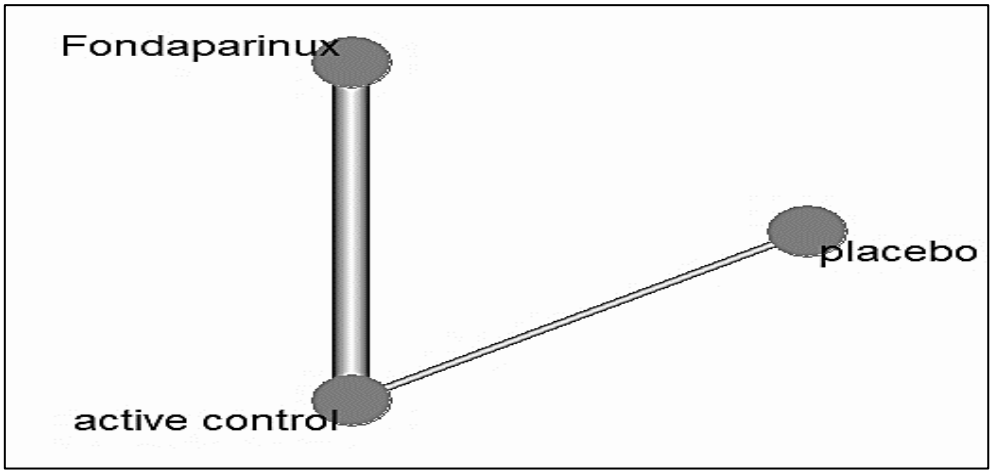

Figure 2-5 Evidence of network of the three included treatments

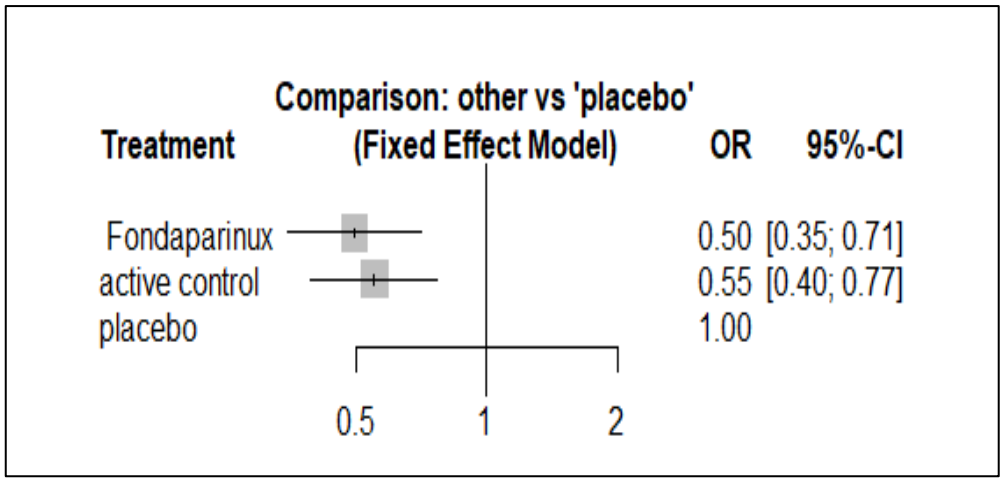

Figure 2-6 Forest plot of the network meta-analysis with comparison to placebo

Table 2.1 Comparison of the odds ratio of the three treatments in the network

	Fondaparinux, odds (95% CI)	Active control, odds (95% CI)	Placebo, odds (95% CI)
Fondaparinux	1.00	0.90(0.80; 1.01)	0.49(0.34; 0.71)
Active control	1.10(0.99; 1.25)	1.00	0.55(0.39; 0.77)
Placebo	2.01(1.41; 2.86)	1.81(1.29; 2.52)	1.00

Active control; LMWH or UFH

From Table (2.1) and Figure (2.6), the conclusion is that both the active control and the fondaparinux were superior to placebo. Moreover, NMA can provide the rank of the best possible treatments based on the probability. The probability fondaparinux will be ranked as the best treatment was 98.47%, the remaining 1.53% is the probability of being ranked second or third. The probability that the active control will be ranked first is 51%. For placebo, the probability of it being ranked the best is 0.0%.

The results from the three methods are the same. The conclusion is that the fondaparinux was superior to placebo and non-inferior to the active control (enoxaparin). It should be noticed that the investigators used placebo-controlled trials that did not include the active control of enoxaparin specifically. Instead, they used placebo-controlled trials of other treatments similar to enoxaparin.

2.7 Summary

This chapter has described the concepts, assumptions and challenges associated with the design of NI trials, the setting of NI margin, and the methods used for the analysis of NI trials. Compared to traditional superiority trials, NI trials have many methodological and regulatory challenges that can influence proper analysis and inference of the results.

A major challenge in designing and conducting any NI trial is the choice of the appropriate active control and obtaining its efficacy indirectly from historical studies. Assay sensitivity and the constancy of active control effect size over time and controlling for bio-creep and placebo creep are the main issues in choosing appropriate active control. Including a placebo arm in the NI trial will control for most of these challenges. However, this is not feasible most of the time for clinical reasons. Other methods like population homogeneity

and similarity between trials could reduce the risk of bio-creep and placebo creep but cannot control it.

This thesis will investigate the changes in the treatment difference between the active control and placebo and how adjusting for these changes could adjust the constancy assumptions and reduce the chance of placebo creep. The bio-creep and the assay sensitivity assumptions will be included in the context of the investigation of the constancy assumption and placebo creep.

This chapter has reviewed the different available methods for analysing NI trials. These include the fixed margin and synthesis approach (regulatory approaches), network meta-analysis (predictive approach). None of these methods adjusts for changes in the time, and it is evident that setting the NI margin by any of these methods depends on the availability of the historical trials (Table 2.2).

Table 2.2 Comparison between the different methods for the analysis of NI trials

Comparison	Fixed margin method	Synthesis method	NMA
Active control	Only one active control can be included	Only one active control can be included	Can include more than one active control
M1	Specified in the designing phase using the boundaries of the 95% CI of historical placebo-controlled trials	Specified in the designing phase using the main estimate of the 95% CI of historical placebo-controlled trials	Cannot be specified in the designing phase
M2	Specified in the designing phase as a fraction of M1	Specified in the designing phase as a fraction of M1	Cannot be specified in the designing phase
Ranks the treatments	No	No	Yes
Includes co-variables in the analysis	No	No	No

Choosing the appropriate method is dependent on the type of NI trial, the primary endpoints and the availability of historical trials and a different comparator. Chapter 7 of this thesis will review the available methods for setting an adjusted NI margin. In Chapter 8, the chosen methods will be applied to two case studies of setting the non-inferiority margin. The next chapter will review the available regulations regarding conducting, analysing, and reporting the NI trials.

Chapter 3 Guidelines for Non-inferiority Trials

3.1 Introduction

The concepts of active-controlled trials and non-inferiority (NI) trials have become more popular since the 1990s (Figure 1.1). As mentioned in Chapter 1, NI trials are conducted to test and market a new treatment, to find an alternative or second-line treatment or to prove the efficacy of an existent treatment when a placebo-controlled trial is not possible. Different sponsors are involved in the funding of NI trials, including pharmaceutical companies, public health institutes, and educational institutes. Moreover, NI trials have several methodological challenges, especially in choosing the appropriate active comparator and determining the NI margin.

These factors raise the need for guidelines on the conducting and reporting of active control and NI trials. In 1998, the international conference on harmonisation (ICH-E9) published the guidelines for conducting randomised control trials in general (ICH, 1998). Today there are several guidelines for the appropriate conduct of active-control and NI trials. These guidelines include ICH-E10, issued in 2001, regarding the choice of control group (ICH, 2001), Committee for Medicinal Products for Human Use (CHMP), issued in 2005, that discussed the choice of NI margin (CHMP, 2005), and the Food and Drug Administration (FDA) draft guidelines for NI trials, issued in 2010 (FDA, 2010), and its final version, issued in 2016 (FDA, 2016), which concentrate on the design and setting and analysis of NI trials.

In 2006 the Consolidated Standards of Reporting Trials (CONSORT) organisation released the CONSORT statement on the appropriate reporting of NI trials in medical journals, which was updated in 2012 (Piaggio et al., 2012). However, none of these guidelines establishes hard definitions; they provide guidance and recommendations only (FDA, 2016). Moreover, most of these guidelines concentrate on drug trials.

In this chapter, the most important American and European regulatory guidelines for non-inferiority trials will be summarised in Section (3.2), the main topics discussed in these regulations will be explained in Section (3.3), and finally, the differences between the European and American guidelines will be identified.

3.2 Regulatory Guidelines

3.2.1 ICH-E9: a statistical principle for clinical trials

ICH-E9 could be considered as the bedrock for the statistical, regulatory guidelines for clinical trials. It gives a broad description of the design of all types of clinical trials, including non-inferiority trials (ICH, 1998). No methods for determining the NI margin were discussed. However, it specified that the margin should be justified clinically. Concerning the confidence interval, it recommends that the use of a one-sided interval and a type I error should be separate from the use of a one-sided or two-sided test (ICH, 1998). Furthermore, it highlighted the use of the full set analysis (all patients randomly assigned to a treatment group having at least one efficacy assessment after randomisation) as being non-conservative (concluding the non-inferiority of an inferior treatment) in NI trials.

A draft for the ICH-E9 addendum on estimands and sensitivity analysis was published in August 2017 (ICH, 2017). The addendum defined the estimands for a confirmatory clinical trial as

"*The target of estimation to address the scientific question of interest posed by the trial objective.*"

The document describes NI trials as non-conservative (concluding the non-inferiority of an inferior treatment) trials, and, because of that, the choice of estimand should be aimed to minimise the number of protocol violations and non-adherence and withdrawals (ICH, 2017).

3.2.2 ICH-E10: choice of the control group in a clinical trial and related issues

ICH-E10 deals mainly with the choice of an appropriate control group in any clinical trial (superiority or NI trials). Different types of control groups are discussed in detail. It concentrates on different purposes of clinical trials and distinguishes between active control and placebo trials, and the concept of indirect comparison. However, no specific guidance was given on the choice of the non-inferiority margin. In these guidelines, the concept of assay sensitivity was highlighted and discussed. The constancy assumption was also discussed, even though the term constancy was never used (ICH, 2001).

3.2.3 European Medicines Agency (EMA) guidelines on choice of non-inferiority margin

Adopted in July 2005 by the Committee for Medical Products for Human Use (CHMP), the guidelines focused on the actual choice of the non-inferiority margin and described the different situations where it is appropriate to conduct the non-inferiority trial. These guidelines are concerned with both the absolute efficacy of the test treatment compared to the placebo and the relative efficacy of the test treatment to the active control. Besides, they discuss the choice of NI margin in more detail compared to ICH-9 and ICH-10. CHMP recommends that the decision to perform an NI trial and choice of a specific NI margin should be justified in the protocol and should be based on both statistical reasoning and clinical judgement.

According to CHMP guidelines, a three arm non-inferiority trial that includes placebo, active treatment and test treatment is the recommended design whenever possible since this kind of design will allow within-trial validation of the choice of NI margin. Choosing the appropriate margin will assure that the test treatment is clinically superior to the placebo. The primary focus of NI trials is on the relative efficacy of the test treatment and active control (CHMP, 2005). CHMP stated that it is not possible to perform NI trials in all situations. The decision on choosing NI design should be justified in the protocol, considering both the therapeutic area and the active control (CHMP, 2005). Using a wider NI margin is possible according to CHMP if the trial is an efficacy trial, and the test treatment has advantages in other aspects. Still, the superiority to placebo should be confirmed. Finally, in extreme situations, it could be acceptable to run a superior trial with a level of two-sided significance greater than 0.05 as an alternative to the NI trial (CHMP, 2005).

3.2.4 The extension of the Consolidated Standards of Reporting Trial (CONSORT)

The CONSORT statement for NI trials was published in 2006, with its extension published in 2012. It describes the publication of NI trials and the quality of published NI trials before and after 2006 and updates the recommendations to authors on how to report the design, conduct, and results of non-inferiority trials. It includes a checklist and flow diagram to help authors improve their reporting of NI trials (Piaggio et al., 2012). According to the

checklist, the authors should provide a rationale for choosing a non-inferiority design, providing the results from trials used to base the active control effect (Piaggio et al., 2012).

3.2.5 Standard Protocol Items: Recommendation for Interventional Trials (SPIRIT)

This statement guides the elements that should be included in a clinical trials protocol, including the scientific, ethical, administrative elements. It provides a checklist of 33 items that applies to protocols for all clinical trials, concentrating on content, not the formatting (Chan et al., 2013). Regarding the NI trials, it recommends the use of sensitivity analysis to assess the robustness of trial results and to handle missing data. It recommends the use of both Per Protocol (PP) and Intent to Treat (ITT) analyses (Chan et al., 2013).

3.2.6 The Food and Drug Administration (FDA) guidelines on non-inferiority clinical trials

In 2016, the FDA finalised the draft guidelines published in 2010 regarding conducting and interpretation of non-inferiority trials. The document is the most detailed regulatory guidelines document for pharmaceutical and biotechnological companies wanting to use the NI design to test new treatment efficacy and for treatment approval.

The guidelines are in four sections. Section 1 provides a general discussion of the main concept, design, and statistical analysis of non-inferiority trials. Section 2 provides details on different approaches used to determine the non-inferiority margin. Section 3 answers the most commonly asked questions about NI studies. Section 4 presents four examples for successful and unsuccessful NI trials to explain the different possible challenges that could arise during the process of designing, conducting, analysis, and the interpretation of non-inferiority trials (FDA, 2016).

This regulatory document highlighted the main issues with the NI trials and the difference between the superiority trial and NI trials in establishing the effectiveness, as well as reporting the main reasons for conducting NI trials and the number of NI trials needed for drug approval. FDA described the use of the fixed margin methods for determining the NI margin as the most conservative (chance of concluding the non-inferiority of an inferior treatment is low) method and the recommended method for setting the NI margin (FDA,

2016). The FDA regulations identified three alternative designs that can replace NI trials. These include an Add-on study, identifying a population not known to benefit from available treatment where the placebo-controlled trial is ethically acceptable and finally early escape, rescue treatment, randomised withdrawal for patients in placebo-controlled trials (FDA, 2016).

3.3 Main regulatory points regarding NI trials

3.3.1 Situations where NI trials could be used

According to FDA and ICH-10 guidelines, the non-inferiority active-controlled design should be used instead of superiority design if the use of the placebo arm in the trial is unethical or if there is an interest in comparing the effectiveness or assessing the sensitivity of a placebo-controlled trial (FDA, 2016; ICH, 2001). CHMP presented different situations where NI trials could be conducted. These include the situation where the experimental treatment has a considerably better safety profile than the active comparator; areas where bioequivalence trials are not possible; cases where no important loss of efficacy compared to the active control would be acceptable; disease areas where the use of a placebo arm is not possible and an active control trial is used to demonstrate the efficacy of the test product (CHMP, 2005).

3.3.2 Blinding

All regulatory guidelines state that blinding is a necessary process to minimise bias. An open-label design can be adopted if blinding is not possible (FDA, 2016; ICH, 2001).

3.3.3 Non-inferiority margin

All regulatory guidelines recommend that an acceptable pre-specified non-inferiority margin should be pre-defined, and should not be larger than the presumed entire effect of the active control in the NI trial. The determination of the margin in the non-inferiority trial is based on both statistical reasoning and clinical judgement. According to the FDA regulation, two margins (defined previously in Section (2.4.2)) should be identified: the statistical margin (M1) and the clinical margin (clinical judgement) (M2).

M1 should be identified based on previous experience in properly designed placebo-controlled trials sharing similar conditions to those planned for the NI trial and could be supported by dose-response or active control studies.

M2 should be based on clinical judgement and should be a fraction of M1 (FDA, 2016). In the anti-infective therapeutic field, FDA regulations recommend the use of 10% of the M1 margin as preferred NI margin (FDA, 2016). For cardiology, FDA regulations recommend the use of 50% of the statistically calculated M1 as the NI margin (M2) (FDA, 2016). However, in other fields, such as oncology, no specific percentage is recommended. The fixed margin approach has also been recommended as an approach for analysing NI trials (FDA, 2016). The fixed margin approach was described in detail in sections (2.4.2.1 and 2.5.1). According to the extension of the CONSORT statement, the margin should be specified in the publication (Piaggio et al., 2012).

3.3.4 Sample size estimation

The calculation of sample size for NI trials was described in detail in the FDA guidelines (FDA, 2016). The sample size of a non--inferiority trial should be based on a fixed margin approach and based on the need to rule out inferiority greater than M2, and this should be clearly stated in the protocol in the planning stage (FDA, 2016).

Both the clinically estimated margin (M2) and the estimated variance of treatment effect will affect the sample size calculation (FDA, 2016). An increase in M2 will lead to a decrease in the required sample size to conclude the non-inferiority. NI trials typically have larger sample sizes compared with superiority trials. However, Fleming argues that the need for a larger sample size for trials that have a rigorous margin is a myth that was introduced by some industrial representatives at the meeting of the FDA Anti-infective Drugs Advisory Committee in 2002 (Fleming, 2008). The effectiveness of the test treatment compared to the active control plays a vital role in determining the sample size. If the test treatment is more effective than the control, a smaller sample size could rule out any given non-inferiority margin. However, in less effective or inferior test treatment, the larger sample size is needed to rule out the non-inferiority (FDA, 2016).

The extension of the CONSORT statement required reporting of justification of sample size (Piaggio et al., 2012). The appropriate choice of NI margin will lead to appropriate sample size calculation in the case of the NI trial.

3.3.5 Analysis population

NI trials in terms of analysis populations are different to superiority trials. In intent to treat (ITT) analysis, participants are compared in terms of their results within the groups to which they were originally randomised, regardless of receiving the treatment, having dropped out or violated the protocol. ITT is the preferred analysis in superiority trials since it protects the trial from serious errors associated with selection bias, protocol violations, and loss for follow up. The situation is different in NI trials. ITT analysis alone is not preferred since it could lead to a false conclusion of non-inferiority for a less effective test treatment (FDA, 2016).

The alternative option is per protocol (PP) where only subjects meeting the inclusion criteria, receiving the treatment, and continuing until the end of the trial are considered in the final analysis. PP analysis is considered a conservative approach (chance of concluding the non-inferiority of an inferior treatment is low) in NI trials as it maximises estimates of the treatment difference. Hence PP analysis is the preferred primary analysis for NI trials, although the use of the PP analysis will lead to excluding patients from the analysis and could lead to an imbalance of the number of patients in each treatment arm, which will lead to bias in both directions (Rehal, Morris, Fielding, Carpenter, & Phillips, 2016).

All of the regulatory guidelines recommend the use of both ITT and PP analyses, with both of these analyses having equal importance and being reported in NI trials (CHMP, 2005; CPMP, 2000; FDA, 2016; ICH, 2001). However, the definitions of PP and ITT populations were not the same and obscure between the guidelines. The CONSORT statement described the PP analysis as excluding patients who did not take the treatment or were not protocol adherent. While ICH-E9 described the analysis population as asubset of patients who complied with the protocol, adding to that the use of "as treated analysis" or modified ITT analysis.

If differences emerge in the results of the two analyses (ITT and PP), further examination should be done (FDA, 2016).

3.3.6 Switching between non-inferiority and superiority

According to FDA guidelines, a planned NI trial can be tested for superiority without the need for adjusting for type I error. However, the conclusion of non-inferiority after a failed superiority trial gives uncertain results, and such a trial should be considered as a failed superiority trial (since the NI margin needs to be pre-specified before, not after, conducting the trial) (FDA, 2016).

Switching between superiority and non-inferiority could be possible in some situations. In 2000, The Committee for Proprietary Medicinal Products (CPMP) published a document setting out points to consider when switching between superiority and non-inferiority (CPMP, 2000). According to this document, switching could be feasible if:

- The non-inferiority margin was predefined and well justified
- Both the ITT and PP analysis were similar
- The trial was adequately designed and conducted according to the regulatory guidelines
- The trial has high sensitivity that is capable of detecting relevant existingdifferences
- The efficacy of the control treatment is shown by either direct or indirect evidence.

3.4 Differences between EMA and FDA guidelines

The growing interest in NI trials has led to the development of several regulatory guidelines from both EMA and FDA. All of these guidelines aimed to specify and regulate the conducting of NI trials. The guidelines from both EMA and FDA are conceptually similar. However, the terminology used was different: margin M1 in FDA guidelines corresponds to demonstrating efficacy in EMA guidelines. Moreover, M2 in FDA guidelines corresponds to establishing acceptable relative efficacy to active control in EMA (CHMP, 2005; FDA, 2016; ICH, 2001)

FDA guidelines stated clearly that the fixed margin method is the recommended method for analysing NI trials (FDA, 2016). By comparison, EMA recommended the use of both

statistical and clinical judgement and did not specify a specific method for selecting the margin (CHMP, 2005; ICH, 2001).

3.5 Summary

In this chapter, different regulatory guidelines were summarised and presented. All of the guidelines set recommendations on the appropriate designing and conducting of non-inferiority trials. However, none of them gives firm rules, and there is an apparent inconsistency between the guidelines that could negatively affect the quality and reporting of NI trials.

Despite the availability of these regulatory documents, the conducting of NI trials is still a challenge. Given that NI trials are conducted by both public and pharmaceutical industry bodies and can be used for the drug approval process (FDA, 2016), it was important to review these regulations in this chapter since they are considered as an important aspect of the designing, conducting and reporting of NI trials.

In the next chapter, a review of the NI trials that have been published in high impact journals will be presented; the review will aim to investigate whether these trials were conducted in accordance to the regulatory guidelines presented in this chapter. By the end of Chapter 4, a complete picture of the designing, conducting and reporting of NI trials will be formulated to meet the first objective of this thesis.

Chapter 4 Review of NI Trials Published in JAMA, Lancet, BMJ and NEJM in 2015

4.1 Introduction

The main challenges, assumption, and methods used for setting the NI margin were reviewed in Chapter 2 of this thesis and the main published regulatory guidelines regarding NI trials were reviewed in Chapter 3.

This chapter aims to investigate the design, analysis, interpretation and reporting of NI trials in four top medical journals. Moreover, it will investigate the use of historical information to set the NI margin. The main concentration will be on the setting and reporting of NI margin and how it follows the regulatory guidelines; how the sample size of the NI trial was reported. The secondary objective is to compare the trials according to the source of funding.

This chapter will start with the methodological section (4.2), where the methods of extraction and analysis will be presented. This will be followed by the results section (4.3), which will include the general characteristics of the trials, the NI margin setting and reporting and differences between the publicly and pharmaceutical company funded trials. A detailed discussion will be presented in Section 4.4, followed by a summary of the findings of this review in Section 4.5.

4.2 Methods

The Lancet, British Medical Journal (BMJ), Journal of American Medical Association (JAMA) and the New England Journal of Medicine (NEJM) were the medical journals chosen for inclusion in this review because they are considered as thehighest quality medical journals with robust publication standards and more likelihood of having a major influence on clinical practice. NI trials published in these journals will represent high standards of publication.

A search for NI trials published in these four journals between 1/1/2015 and 31/12/2015 in the PubMed database was performed by one reviewer (E.Duro). The original aim was to review trials conducted from 2010/2017. However, the advice from the confirmation review committee in August/ 2016 was that one year was sufficient to address the aims of this chapter and the thesis as a whole and for this reason only the year 2015 was used in the review.

The inclusion criteria were NI trials that were randomised clinical trials, done on adult humans, published in English and with the full text available. A standardised data extraction form was created (Appendix A). Data extracted using the form included general information on the journal's name, registration number on registry database, type of treatment (cardiovascular, anti-infective, surgical, gynaecology, and others), the phase of the trial (II, III, or IV), single centre or multicentre, and source of funding (public, private or both).

Information about blinding was obtained in accordance to the manuscript (open-label: no blinding, single: only the participants are blinded, double-blinding: both the researcher and the participants are blinded, or ambiguously stated: not clearly stated in the manuscript); type of statistical analysis used: ITT or PP or both (the definitions of ITT and PP were presented in section 3.3.6), primary endpoints (efficacy, safety, or both), sample size and power (calculation and justification), presence of placebo arm, study design (double arm, triple or four arm and parallel or crossover design) and conclusion (non-inferiority concluded or not, or if superiority was also concluded beside the non-inferiority), and the presence of ethical committee were retrieved.

Specific information on the setting of the NI margin was also extracted, including how it was calculated; if based on statistical consideration or clinical judgement or both; and how it was interpreted, if the primary analysis was similar to the one stated in the protocol or not. Clinical trials registries were used to fill out any missing information regarding the protocol, type of population used in the analysis and NI margin. Data were summarised, and descriptive statistics were obtained using SPSS 22 (SPSS Inc, USA; www.spss.com).

4.3 Results

In total, 387 articles were retrieved. Only 45 articles were published in the Lancet, BMJ, JAMA and NEJM, of which 37 were analysed, six articles were excluded because they were published in 2016, and two articles were excluded because they were review articles (Figure 4.1).

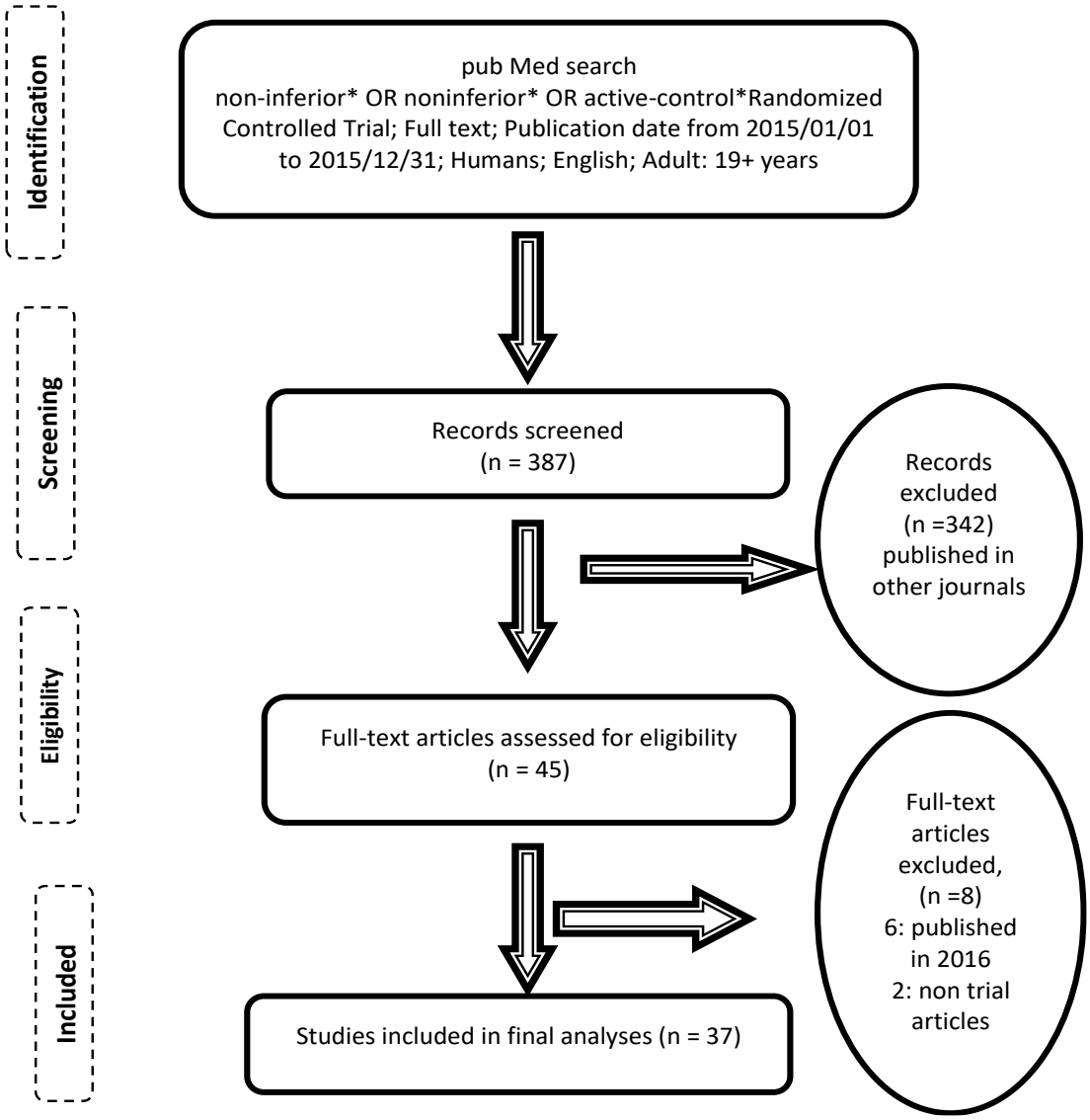

Figure 4-1 Flow chart for the trials extraction process

4.3.1 General Characteristics

Table 4.1 provides the general characteristics of the included trials. All of the 37 trials included in this review were randomised, multicentre trials. The study protocol was available online for most of the included trials. Out of these 37 articles, 15 were published in the Lancet, 12 in NEJM, five in BMJ and five in JAMA. Regarding the type of treatment, 12 (32.4%) of the studies were cardiovascular, and homoeostasis studies, six (16.2%) were anti-infective, four (10.8%) were oncology studies, four (10.8%) were trials on gynaecology and obstetrics. Additionally, three (8.1%) related to surgical procedures, three (8.1%) to each of autoimmune disease and rheumatology, and four (10.8%) to each of dermatology, diabetes mellitus, ophthalmology, vaccines and respiratory disease. A full list of included trials is presented in Appendix B. According to the source of funding, 19 (51.4%) of the trials were funded publicly, 15 (40.5%) were funded by pharmaceutical companies, and in three (8.1%) trials, the funding was provided by a combination of public and private sector organisations. All of the trials were multicentre trials with a median sample size of 571, minimum of 106 and maximum of 14215 patients.

Regarding the blinding, 25 (67.6%) of the studies were open-label studies (no blinding); among these open-label trials, in 15 (60%) blinding was not possible, with no specific reason given for the non-blinding in the other ten (40%) trials. Blinding is considered an essential part of any randomised controlled trial, either superior or NI trials. Although all the regulatory guidelines (FDA, 2016; ICH, 2001) recommended the use of double blinding to reduce the chances of bias in randomised clinical trials, only eight trials in this review were double-blinded, and 25 trials were open-label. Most of the open-label trials (60%) justified their use of open-label design by stating that blinding was not possible, while the remaining 40% did not justify the use of open-label design.

The phase of the trial was not reported in 27 (73%) trials, eight trials (21.6%) were described as phase III trials, and the other two (5.4%) were described as phase II trials. The primary endpoint was efficacy endpoint in 30 (81.1%) trials, five (13.5%) trials had two primary endpoints for efficacy, as an NI trial and safety as a superiority trial. One trial

tested both the efficacy and safety as an NI design and the other one tested safety as NI and efficacy as superiority.

Table 4.1 The characteristics of included NI trials

Category	Number of trials (%)
Type of Treatment	
Anti-infective	6 (16.2%)
Cardiology	12 (32.4%)
Oncology	4 (10.8%)
Others	15 (40.6 %)
Funding	
Public	19 (51.4%)
Private	15 (40.5%)
Both	3 (8.1%)
Blinding	
Open Label	25 (67.6%)
Single Blinded	4 (10.8%)
Double Blinded	8 (21.6%)
Conclusion	
NI concluded	25 (67.6%)
NI not concluded	8 (21.6%)
Both NI and Superiority concluded	4 (10.8%)
Analysis	
Intent to Treat (ITT)	9 (24.3%)
Per Protocol (PP)	5 (13.5%)
Primary ITT and PP as sensitivity	18 (48.6%)
Primary PP and ITT as sensitivity	5 (13.5%)

ITT; Intent to treat, PP; per protocol

Regarding the study design, 30 (81.1%) trials had two arms parallel design. Four trials had three arms, one trial had four arms, one had a 2X2 design, and one had crossed over design. Ten (27%) of the trials conducted an interim analysis. Sensitivity analysis was conducted in 28 (75.7%) of the trials. Finally, 23 (62.16%) of the studies used both ITT and PP analyses.

Most of the manuscripts reported that the reason for choosing the NI trial instead of the superiority trials was the presence of standard active treatment instead of placebo. The choice of the active comparator (the active control) was justified in all of the trials as being the most used or recommended treatment by the regulations.

Given that the classical NI trial is considered to be the trial where two active treatments are compared to conclude the non-inferiority of the test treatment to the active control and indirectly superiority of test treatment to placebo, the design of the reviewed trials was not the classical active-controlled NI form. Only 21 (56.8%) of the trials compared two treatments (active versus test treatment) and six (16.2%) trials compared two surgical procedures. The NI design was used to compare the duration for the same treatment (Barone et al., 2015; Bernard et al., 2015), route of administration (Cox et al., 2015; Le Page et al., 2015), surgical versus medical treatment (Kehoe et al., 2015; Salminen et al., 2015).

Usually, NI trials do not include a placebo arm since the presence of a placebo arm is considered unethical or not good practice. However, in this review, four trials included a placebo arm. In one trial (Bachelez et al., 2015) the placebo was used to test the superiority of the test treatment (tofacitinib) versus placebo and to test the non-inferiority of tofacitinib compared to the active treatment etanercept. This trial was the only trial that used a placebo in this way and concluded the superiority of the test treatment to the placebo in addition to the non-inferiority of the test treatment to the active control directly. In the ELIXA trial (Pfeffer et al., 2015) the aim was to conclude that the test treatment (lixisenatide) was not inferior to the placebo in regard to cardiovascular outcome in diabetic patients (since the efficacy of this treatment to improve the glycaemic control and weight reduction in the diabetic patient was already known). However, it was not understood why the author used the NI design instead of safety superiority design (the study was funded by the company who manufactured this medicine). This study concluded that the lixisenatide was not inferior to placebo in the reduction of cardiac events in patients with type II diabetes.

Two other trials used placebo as test treatment and aimed to show the standard treatment was no better than placebo. The BRIDGE study (Douketis et al., 2015) aimed to prove no perioperative anticoagulant bridging was not inferior to bridging with low molecular weight heparin in patients with atrial fibrillation who would be having a surgical procedure. The RAPID trial (Radford et al., 2015) was an oncological trial that aimed to prove no further

radiotherapy was not inferior to further radiotherapy in patients with Hodgkin lymphoma. In both studies, the investigators claimed that the evidence of the effectiveness of the active control was weak and based on observational studies. All these trials contain a placebo arm in their design, but still, it is used as an experimental arm in the trial, which is unusual for NI trials. The ASPECT-cUTI study (Wagenlehner, Umeh, Steenbergen, Yuan, & Darouiche, 2015) was the only study that mentioned the possibility of placebo creep in the chosen dose of active control.

Regarding the population included in the analysis, the regulatory guidelines recommend the use of both ITT and PP design, with PP as the primary analysis if the conclusion was different between the two populations (FDA, 2016). The combination of ITT and PP analysis was the most common type of analysis in 23 trials (62.12%). However, the definitions of the ITT and PP population used were not the same across the trials. The use of modified ITT (mITT) population was evident in most of the trials, instead of the classical ITT population. Besides, the use of the words "as a treated population" to reflect the modified PP was popular too. One trial described four different populations (Goldstein et al., 2015); modified ITT(mITT), ITT efficacy (ITT E), ITT safety (ITT S), and per protocol (PP) population. The disagreement between the different regulatory guidelines in the definition of the PP versus ITT populations is the main reason for these different definitions (Rehal et al., 2016). Both ITT and PP designs have their pros and cons and neither of them is considered as the gold standard for NI trials (Rehal et al., 2016). In the case of NI trials using PP, analysis will exclude patients with missing data and will give results that are more conservative (chance of concluding the non-inferiority of an inferior treatment is low) compared to the ITT design (FDA, 2016). The population included in the primary analysis should be predefined in the protocol. Moreover, the methods used for handling missing data should be specified in the protocol. In this review, most of the trials that used ITT analysis used multiple amputations to handle the missing data.

4.3.2 NI margin setting

Table 4.2 presents the characteristics of the NI margin in the reviewed trials. All of the reviewed trials reported their NI margin. The methods for determining NI margin were not evident in nine (24.3%) trials. In ten (27%) trials, the margin was calculated based on previous studies only. Clinical judgement alone was used in six (16.2%) trials. The NI margins were justified

based on both clinical judgement and historical trials in nine (24.3%) trials and based on the regulatory guidelines in only three (8.1%) trials.

There was no explicit calculation formula; none of the trials mentioned fixed margin method or synthesis method or any other specific methods. The trials aimed to test the non-inferiority only in 27 (73%) trials and aimed to test both the non-inferiority and superiority in ten (27%) of the trials.

Sensitivity analysis was reported in 28 (75.7%) of the trials, and usually this means the use of PP or ITT analysis as a second analysis. The interim analysis was reported only in ten (27%) of the trials. The risk difference was used as a measure of effect for the NI margin in 31 (83.3%) trials, the hazard ratio was used in five trials (13.5%) and one trial (2.7%) used the relative risk as a measure of effect. Type I error was determined to be 0.05 in 17 (45.9%) of the studies and the power of 80% in 17 trials (45.9%). All trials except one justified the sample size calculation (Behringer et al., 2015).

Additionally, 27 (72.9%) trials reported the use of two-sided 95% CI. Among these, 11 trials reported type I error of 0.025 and 12 (32.4%) reported the type I error as 0.05, while six trials did not report the type I error. Non-inferiority was concluded in 29 (78.4%) of the trials, among which four trials concluded the superiority of the test treatment over the active control in addition to the non-inferiority.

Table 4.2 The characteristics of the NI margin

Category	Number of trials (%)
NI margin Justification	
Based on historical data	10 (27.0%)
Based on clinical judgement	7 (18.9%)
Based on both historical data and clinical judgement	8 (21.6%)
Based on regulation	3 (8.1%)
Not stated	9 (24.3%)
Confidence Interval (CI)	
Two-sided 95% CI	27 (73.0 %)
One-sided 95% CI	4 (10.8%)
Two-sided 90% CI	3 (8.1%)
One-sided 90% CI	2 (5.4%)
One sided 97.5% CI	1 (2.7%)
Type I error	
0.025	12 (32.4%)
0.05	17 (45.9%)
0.1	2 (5.4%)
Not stated	6 (16.2%)

4.3.3 Public versus private funding

Table 4.3 presents the differences between the public and private funded trials. Regarding these differences, nine of the trials funded by pharmaceutical companies were cardiovascular trials, two of them were anti-infective, others included vaccines, diabetes, rheumatology and dermatology. There was a statistically significant difference between the trials funded publicly and trials funded by pharmaceutical companies concerning the conclusion. Among the privately funded trials, out of 15 trials, ten (66.7%) of them concluded non-inferiority, four (26.7%) concluded superiority and only one trial (6.7%) failed to conclude non-inferiority. Among the publicly funded trials, 13 out of 18 concluded non-inferiority, six failed to establish non-inferiority, and no superiority was concluded (p-value =0.02).

Regarding the blinding process, there was a statistically significant difference between the publicly funded trials and trials funded by pharmaceutical companies. Most of the publicly funded trials were open label 18 (94.7%), and only one trial was double-blinded. In contrast, 40% of private funding trials were open-label trials, 20% were single-blinded, and 40% were double-blinded (P-value < 0.01). There were no statistically significant differences between the public trials and trials funded by pharmaceutical companies concerning the NI methods or the primary analysis.

Table 4.3 Differences between public and private funded trials

	Public trials, N (%)	Private trials, N (%)	Both N (%)
Type of Trial			
Anti-infective	4 (21.1%)	2.0 (13.3%)	0.0 (0.0%)
Cardiovascular and haemostasis	2 (10.5%)	9.0 (60.0%)	1.0 (33.3%)
Oncology	2 (10.5%)	0.0 (0.0%)	1.0 (33.3%)
Gynaecology	4 (21.1%)	0.0 (0.0%)	0.0 (0.0%)
Others	7 (36.8%)	4.0 (26.6%)	1.0 (33.3%)
Methods for determining NI margin			
Historical data	3 (15.8%)	5.0 (33.3%)	2.0 (66.7%)
Clinical Judgement	4 (21.2%)	2.0 (13.3%)	0.0 (0.0%)
Both historical data and clinical judgement	8 (42.1%)	1.0 (6.7%)	0.0 (0.0%)
Based on regulation	1 (5.3%)	2.0 (13.3%)	0.0 (0.0%)
Not stated	3 (15.8%)	5.0 (33.3%)	1.0 (33.3%)
Blinding			
Open label	18.0 (94.7%)	6.0 (40.0%)	1.0 (33.3%)
Single blinding	0.0 (0.0%)	3.0 (20.0%)	1.0 (33.3%)
Double blinding	1.0 (5.3%)	6.0 (40.0%)	1.0 (33.3%)
Conclusion			
NI established	13.0 (68.4%)	10.0 (66.7%)	2.0 (66.7%)
Superiority established	0.0 (0.0%)	4.0 (26.7%)	0.0 (0.0%)
NI not established	6.0 (31.6%)	1.0 (6.7%)	1.0 (33.3%)
Type of primary analysis			
ITT	4.0 (21.1%)	5.0 (33.3%)	0.0 (0.0%)
PP	3.0 (15.8%)	2.0 (13.3%)	0.0 (0.0%)
Primary ITT and sensitivity as PP	11.0 (57.9%)	5.0 (33.3%)	2.0 (66.7%)
Primary as PP and sensitivity as ITT	1.0 (5.3%)	3.0 (20.0%)	1.0 (33.3%)

ITT; intent to treat, PP per protocol

4.4 Discussion

4.4.1 General Characteristics and NI margin

The conducting and reporting of NI trials are associated with regulatory and statistical challenges and usually this kind of trial is poorly conducted (Rehal et al., 2016), mostly due to disagreement between the regulatory guidelines. In this review, the NI trials published in 2015 in the top four medical journals were used to assess the quality of the published NI trials in general and to investigate how the challenges and assumptions of NI trials (the assay sensitivity, bias minimising and constancy assumption) were presented in the practical field of setting and conducting the NI trials.

All regulatory guidelines demanded reporting of the NI margin and its justification by statistical and medical judgement. All reviewed trials specified the NI margin used (100%) and the chosen margin was justified in 28 trials (75.6%). Only nine trials used both statistical relevance and medical judgement to establish the NI margin as recommended by the regulatory guidelines. Moreover, even when they did, the justification was ambiguous with little detail and usually referred to references that do not clearly state how the margin was chosen.

Moreover, most of the trials reflected the choice of the margin as based on "investigator assumptions" without any further explanations. Two trials justified the use of a wide NI margin that was determined by clinical judgement only on the basis of there being no historical data available (Bensdorp et al., 2015; Cooper et al., 2015). The subjectivity in the setting of NI margin was due to the use of clinical judgement that depended on the clinicians' opinion rather than being evidence-based. All trials that depended on clinical judgement only concluded the NI, while only two thirds of the trials that used both statistical and medical judgement concluded the non-inferiority. None of the reviewed trials explained or mentioned the method used to set the NI margin even in the study protocol that was published online. The NI margin is usually presented as a percentage or number and justified either on a clinical basis or on historical evidence.

The changes in the active control efficacy over time were not reported in most of the trials. However, in a trial that assessed the efficacy and safety of FXI-ASO compared to enoxaparin as prophylaxis for venous thrombosis after total knee arthroplasty, the investigators considered changes in the active control efficacy compared to the placebo and reduced the selected NI margin by 50% to account for the retention of enoxaparin effect compared to placebo (Büller et al., 2015).

There was an inconsistency between the type I error used for sample size calculation and the confidence interval used for the conclusion. Most of the regulatory guidelines recommend the use of two-sided 95% CI with a corresponding one-sided type I error of 0.025 (CHMP, 2005; FDA, 2016; Piaggio et al., 2012). In this review, 23 (62.16%) of the trials used the two-sided 95% CI, 11 of them used the type I error of 0.025 and the other 12 used one-sided type I error of 0.05. Both type I error and power of the study should be reported and whether the type I error is one-sided or two-sided. Use of two-sided 90% CI is also acceptable if the type I error is stated to be one-sided 0.05, as was the case with the CAP-START trial which used two-sided 90% CI with corresponding one-sided 0.05 type I error (Postma et al., 2015).

The non-inferiority was established in 29 (78.4%) of the trials. These results could be influenced by the fact that trials with positive results are more likely to be published, regardless of the study design (superiority or non-inferiority)(Lee, Bacchetti, & Sim, 2008). Out of nine trials that tested both superiority and non-inferiority, only four established the superiority of the test treatment compared to the active control. Switching from non-inferiority to superiority is acceptable by all of the regulatory guidelines and no adjustments are needed for type I error (FDA, 2016; Lewis, 2001).

4.4.2 Public versus private funded trials

Around half of the reviewed trials were publicly funded trials. Compared to the pharmaceutical companies, public funded trials tend to use a more conservative (more controlling for type I error) margin and usually compare already existing treatments, not a new treatment, or compare two well-known regimens.

More details regarding sample size, randomisation and blinding can be seen in private funded trials since most of the time these trials will be used for drug approval application. However,

the way the NI margin was chosen was not clear in two thirds of privately funded trials and only one private funded trial used both statistical and clinical judgement to justify the chosen NI margin. Fifty-three per cent of the privately funded trials used both ITT and PP analysis compared to 63.15% of the publicly funded trials. Only one privately funded trial failed to declare the non-inferiority compared to six publicly funded trials. The four trials that concluded superiority were privately funded trials.

One of the reviewed trials that can illustrate the manipulation of NI margin and the conclusion is the PROCEED II trials (Ardehali et al., 2015). This trial was a prospective, open-label, multicentre, randomised non-inferiority trial, funded by Trans Medics (the manufacturing company for Organ Care system). The trial aimed to assess the "clinical outcomes" of the Organ Care System compared to the standard cold storage of human donor hearts for transplantation. The test treatment was the Organ Care System, and the active control was the standard cold storage of human donor hearts. Both efficacy and safety endpoints were determined. The primary endpoint was 30-day graft survival with 10% difference as non-inferiority margin. Both ITT as primary analysis and PP analysis were performed. The investigators justified the use of a cold storage system as the standard care. Sample size calculation was based on an NI margin of 10% difference, with the use of a normal approximation test, and a one-sided α level of 0.025; the inclusion of 54 patients per treatment group would provide 80% power. The final sample size was 64 patients in each arm.

Regarding the role of funding the investigators stated (Ardehali et al., 2015):

"The funder of the study had a role in study design and data collection. The authors were responsible for data interpretation, data analysis, and writing of the report."

The results were for the ITT analysis (128 patients were included) of the 30-day patients, and graft survival rates were 94% (n = 63) in the Organ Care System group (T) and 97% (n = 61) in the standard cold storage group (C), (difference 2·8%, one-sided 95% upper confidence bound 8.8; P=0·45) (Ardehali et al., 2015).

For the PP analysis 121 patients were included in the final analysis; the 30-day patients and graft survival accounted for 93% in the Organ Care System group (T) and 97% in the cold storage group (difference 3.4%, one-sided 95% upper confidence bound 9.9; P-value = 0·39).

The incidence of severe rejection in the Organ Care System was 11 (18%) compared to nine (14%) in the standard group (between-group difference was four, 95% CI (-8; 17, p-value = 0.52). In their conclusion, the authors (Ardehali et al., 2015) stated:

"*In conclusion, our findings show that the clinical outcomes of donor hearts adequately preserved with the Organ Care System platform are non-inferior to the outcomes of those preserved with standard cold storage. Evaluation of the metabolic assessment capability of the Organ Care System requires further study*."

There was a discrepancy between the protocol and the actual conducting of the study; in their protocol, the investigators stated that the PP analysis would be the primary analysis for this study (TransMedics, 2008). However, ITT analysis was the actual primary analysis for this trial without any protocol amendments. In PP analysis, the upper limit of 95% CI was 9.9, and the cut-off point was 10% (it is not clear if it is considered clinically acceptable to declare non-inferiority of the Organ Care System); the P-value for non-inferiority was not statistically significant. Moreover, for the other secondary points, there were no statistically significant differences between the groups, i.e. no additional benefits for the Organ Care System over the standard treatment were reported. Even with all these serious considerations, the authors still concluded the non-inferiority of the test treatment compared to the standard treatment.

As mentioned earlier, publicly funded trials tend to be more conservative (more controlling for type I error) in choosing the NI margin. For example, in a publicly funded trial that aimed to compare azithromycin (T) versus doxycycline (C) for urogenital chlamydia trachomatis infection (Geisler et al., 2015), the aim was to conclude the non-inferiority of azithromycin (1 g in one dose) to doxycycline (100 mg twice daily for seven days). The primary endpoint was treatment failure (efficacy). The study was a two arms parallel study with a sample size of 567 patients, as an open-label study. Both treatments had already been tested and recommended by the Centre for Disease Control in the USA (CDC). Previous studies had been conducted to investigate the efficacy of both treatments. An interim analysis was used to recalculate the sample size. The primary analysis was per protocol analysis. A non-inferiority margin of 5% difference was chosen. Regarding the NI margin setting the investigators stated that

"*This non-inferiority study was designed to test the null hypothesis that the absolute rate of azithromycin treatment failure would be at least 5 percentage points higher than the absolute*

rate of doxycycline treatment failure against the alternative hypothesis that there would be no difference between regimens, with a failure rate of 3% for both (a rate that was based on the results of the meta-analysis).”

Regarding their justification of using NI margin of 5%, the author stated:

“The decision to use the difference cut off of 5 percentage points was based on the reported high cure rates for both treatments.”

In terms of clinical judgement, the investigator stated:

“this difference was considered by the investigative team to be an appropriate cut-off to establish the clinical non-inferiority of azithromycin to doxycycline.”

In the results, the doxycycline group had no treatment failure compared to five (3.2%; 95% CI, 0.4 to 7.4%) patients in the azithromycin group. The observed difference in failure rates was 3.2%, with an upper boundary of the 90% CI of 5.9 percentage points, which exceeded the pre-specified absolute 5-percentage cut-off point for establishing the non-inferiority of azithromycin. The non-inferiority margin of 5% is considered conservative (chance of concluding the non-inferiority of an inferior treatment is low) compared with the 10% margin difference recommended by FDA regulations (FDA, 2016).

The chief investigator was contacted by email and asked for the reason for using this conservative margin (Geisler, 2016), and the answer was:

“*The reason for using the smaller non-inferiority margin of 5% was because the anticipated treatment failure rate for both regimens was only 3 %. Also, typically for a drug to be recommended first-line by CDC, it should have an efficacy of 95% or higher*.”

The evidence from the trial is that azithromycin is not a newly introduced treatment; instead it is a well-established treatment for urogenital tract infections, and the trial was not funded by the manufacturing company. Moreover, for other benefits of using azithromycin rather than doxycycline, the compliance rate was higher for the azithromycin group compared to doxycycline. For adverse events, 23% of participants in the azithromycin group had adverse events compared to 27% in the doxycycline group. The use of a conservative margin (more

control for type I error) could be the reason for the failure to establish the non-inferiority of azithromycin compared to doxycycline. The use of only per protocol for analysis reduced the chance of establishing the non-inferiority.

4.4.3 Comparison with other reviews

All included trials reported the NI margin. The rate of justification was higher in this review (75.4%) compared to the other similar reviews; in systematic reviews of NI trials that were published in high impact journals between 2010 and 2015 (Rehal et al., 2016) the rate of reporting the NI margin was 98% and the rate of justification was 45%.

Wangg et al. stated that 97.8% of the trials reported the NI margin and only 45% justified the NI margin used (Wangge et al., 2010), while Schiller et al. stated that 94% of the reviewed trials reported the NI margin and only 23% justified the used NI margin (Schiller, Burchardi, Niestroj, & Kieser, 2012). However, the Schiller review included all NI trials published in 2009 regardless of the quality of the published journal. The rate of justification was even lower (20%) in Henanff et al. (Le Henanff, Giraudeau, Baron, & Ravaud, 2006) who reviewed the NI trials in 2003, 2004 before the publication of the CONSORT statement in 2006 (Piaggio, Elbourne, & Altman, 2006). The reason for improvement could be the fact that the included journals in this review were the top four medical journals, which had stringent guidelines for publication.

Concerning the population included in the analysis, 23 (62.16%) of the trials in this review used both PP and ITT analysis, which was higher than Wangge et al. (Wangge et al., 2010) who reported that 42% used both analyses. Also, it is higher than Schiller et al.'s (Schiller et al., 2012) finding that 42% of the trials used both ITT and PP. The reason for this high rate could be that this review is the most recent one.

Most of the trials in this review used two-sided 95% CI, which was a similar finding to previous reviews (Le Henanff et al., 2006; Rehal et al., 2016; Schiller et al., 2012; Wangge et al., 2010).

4.4.4 Limitations

This review reflects the publication of NI trials in high impact journals only and the results of this review cannot be generalised to other low impact journals. Second, only one reviewer extracted and reported the results of this review, which increased the chance of bias due to subjectivity, especially in regard to the methods used for NI margin justification. However, using the top four medical journals and the most recent year (at the time of review) will present a good picture of the quality of published NI trials.

4.5 Summary

Comparing with previous reviews, there was an improvement in the reporting within published NI trials (Le Henanff et al., 2006; Rehal et al., 2016; Schiller et al., 2012; Wangge et al., 2010). This improvement can be seen in the percentage of reporting of the NI margin and reporting of methods for setting the margin. However, the reporting of NI trials in the top medical journals is still not compatible with the regulatory guidelines, especially in terms of blinding, the population included in the analysis, and reporting and justification of the NI margin used.

This chapter found that around sixty per cent of the trials that justified the NI margin used historical evidence in their justification of the NI margin, which reflects the importance of historical information in setting the NI margin. Most of them justified the use of active control. However, the subjectivity of using clinical judgement only was high since the medical judgement for setting the NI margin was hard to investigate due to its subjectivity.

Up to this point, this thesis has investigated the importance of historical data in the designing phase of the NI trial and choosing the appropriate active control in Chapter 2. Chapter 3 reviewed the available regulations regarding the NI trials and concluded that the only method recommended by the regulation is the fixed margin method, which compares a confidence interval from historical placebo-controlled trials with the confidence interval from the NI trial. In the current chapter, it was explained that around sixty per cent of the included NI trials that justified their NI margin based this on the historical data either alone or in combination with clinical judgement.

The conclusion from all these chapters is that the historical placebo-controlled trials have a critical role in designing, setting, and analysis of NI trials, which was evident from the available

literature (Chapter 2), from regulatory guidelines (Chapter 3) and in practice (Chapter 4). Moreover, it was concluded that any changes in the treatment difference between the placebo and active control in the placebo-controlled trials would lead to biased estimation and false conclusion of non-inferiority of an inferior experimental treatment (Chapter 2).

The next chapter of this thesis will investigate the changes in the treatment differences between the placebo and the active control over time (the constancy assumption). The correlations between the year of publication and the treatment effect will be measured in Chapter 5, followed by a regression analysis of these changes to build a predictive model to estimate the treatment effect based on the year of publication in Chapter 6. Chapter 7 will review the available methods to adjust the NI margin, then Chapter 8 will apply both adjusted and non-adjusted methods for setting the NI margin in an NI trial. Chapter 9 will present the final discussion and conclusion.

Chapter 5 Changes in Treatment Response over Time

5.1 Introduction

The setting of NI margin depends heavily on an indirect comparison between the test treatment and the placebo from historical placebo-controlled trials that compare the placebo with the active control used in the NI trial. The primary assumption regarding the NI margin setting is the constancy of the active control effect over time compared to placebo and the assay sensitivity of the active control. Moreover, the main problems with indirect comparison are the presence of bio-creep and placebo creep, as discussed in Chapter 2.

As mentioned in Chapter 2 (section 2.3.5), there is an argument regarding the constancy assumption and changes in the active treatment and placebo over time. Many studies have demonstrated improvement of the placebo response over time in different therapeutic areas (Julious & Wang, 2008; Walsh, Seidman, Sysko, & Gould, 2002) and they have argued that could be the reason for the violation of the constancy assumption and the presence of placebo creep.

The changes in the treatment effect can be seen in the effect of aspirin as a painkiller. Aspirin has been used as a painkiller for more than a hundred years. If a clinical trial conducted in 1950 that compared aspirin to a placebo concluded the efficacy of aspirin as a painkiller and the same trial was repeated in 2018, would the difference between aspirin and placebo stay the same? In addition, if there was a difference, would it be due to the improvement of the placebo effect (due to the improvement of the general care or other circumstances) or due to a reduction in the efficacy of the aspirin as a painkiller? Moreover, what if the aspirin was used as an active control in an NI trial in 2020? Would it be valid to assume that the efficacy of aspirin in 1950 was the same as in 2020? For the 2020 NI trial would the constancy assumption and assay sensitivity both hold? Alternatively, would they change over time?

Assuming that the effect of placebo improves over time and at the same time the effect of active control decreases over time, based on that assumption, the use of historical placebo-controlled trials to estimate the effect of placebo in the present time will be biased toward non-inferiority if no adjustment for the time is made. This chapter will investigate if there any changes in the

placebo and active control effects over time using the Cochrane reviews of placebo-controlled trials published in 2015/2016.

The review will investigate the effect of year of publication (as a proxy to the time of trial conducting) on the difference between the active treatment and the placebo (effect size) over time by measuring the weighted correlation between standardised mean difference and year of publication in different therapeutic areas. In addition, it will investigate the effect of year of publication on placebo and active treatment responses after controlling for sample size.

First, the methods used for reviewing the Cochrane reviews will be explained in Section 5.2. The results will be presented in Section 5.3. Three examples from the included reviews will be presented in Section 5.4 to illustrate the effect of year of publication on the treatment effect. Finally, discussion and conclusions will be presented on the main findings from this review.

5.2 Methods

5.2.1 Study design and data collection

This study is an overview of the Cochrane systematic reviews of placebo-controlled trials published in the Cochrane database from January /2015 to December /2016.
The inclusion criteria for selecting the relevant systematic reviews are:

- Cochrane reviews of placebo-controlled trials
- Defined as placebo-controlled trials by the review's author regardless of the type of control group used (placebo, no treatment, usual care)
- Meta-analysis was performed.
- The meta-analysis included at least four placebo-controlled trials.
- Meta-analyses published in 2015-2016

The exclusion criteria are:

- Reviews that were withdrawn from publication
- Over reviews or reviews that included active-controlled trials
- Reviews containing three or fewer trials
- Reviews where meta-analysis was not performed
- Reviews where all trials were conducted in the same year

The main reason for choosing reviews published between 2015 and 2016 was that these reviews would be the most recent ones and include all recent updates. Already conducted meta-analyses of the published systematic reviews was chosen to ensure similarity between the trials in the treatments used and the measure of effect, thereby ensuring that these trials can be compared to each other (this is usually the same case for the historical trials that were used in setting the NI margin). The chosen number of trials included was four or more trials because the aim was to exclude the last trial and predict its effect estimate from the remaining trials in the meta-analysis using the included reviews (Chapter 6). Moreover, including less than four trials could lead to more extreme results, especially when measuring the partial correlations.

Rayyan (Ouzzani, Hammady, & Fedorowics, Zbys, Elmagarmid, 2016), a web-based systematic review manager, was used to conduct the systematic review and retrieve the needed information. The keyword used in the primary search in the title and abstract was "placebo", the abstracts were reviewed, and the inclusion and the exclusion criteria were applied.

From each included review, information regarding Cochrane ID, publication year, Cochrane group, and medical speciality was retrieved. From each review, the first meta-analysis was chosen as the data point in the analysis unless it had no meta-analysis or had less than four trials. If the first meta-analysis could not be chosen, the next meta-analysis with more trials was chosen. If in any meta-analysis, the subgroups shared the weight in the study, they were included as one analysis. If the weight of the study was not shared between the subgroups, only the subgroup with the largest number of trials was included.

From each included meta-analysis, information regarding the year of publication, number of trials included, type of control group used, the active treatment used, total number of patients, weight of each trial, total number of patients in the placebo and active treatment group, placebo and active treatment effect, measure of effect used, type of analysis (fixed or random), and heterogeneity were retrieved. Besides, the main estimate and 95% CI, and the last trial's main estimate and 95% CI were retrieved, as well as information regarding the risk of bias and the quality of evidence.

5.2.2 Standardising the difference

There was a need to obtain a standardised measure of effect to compare both the binary data and numerical data. For the binary data, the measure of effect was transferred to the odds ratio and then the standardised mean difference (SMD) was calculated from the odds ratio (Borenstein, Hedges, Higgins, & Rothstein, 2009) using the formulae below to convert the effect sizes to SMD

$$SMD = Log\ OddsRatio \times \frac{\sqrt{3}}{\pi} \quad (5.1)$$

$$V_{smd} = V_{logoddsratio} \times \frac{3}{\pi^2} \quad (5.2)$$

$$V_{logoddsratio} = \frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}, \quad (5.3)$$

where V denotes the variance of the log odds ratio, A is the number of events in the treatment group, B is the number of no events in the treatment group, C is the number of events in the control group and D is the number of no events in the control group.

For continuous data, the measure of effect was the mean difference that was transformed into the SMD using Borenstein et al. (2009)

$$SMD = \frac{\overline{X1} - \overline{X2}}{S_{within}} \quad (5.4)$$

$$S_{within} = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \quad (5.5)$$

$$V_{smd} = \frac{n_1+n_2}{n_1 n_2} + \frac{d^2}{2(n_1+n_2)}, \quad (5.6)$$

Where S_{within} is the within-groups standard deviation, pooled across groups, S1 is the standard deviation of the placebo group, S2 is the standard deviation of the control group, n_1 is the sample size of the control group, n_2 is the sample size of the active treatment group, and V_{smd} is the variance of the SMD.

5.2.3 Statistical analysis

The correlation coefficient was used in previous studies to measure the association between the year of publication and the effect size (Julious & Wang, 2008; Walsh et al., 2002). In this analysis, the correlations (both for all trials in general and by the meta-analysis) between the year of publication and the SMD, placebo and active treatment and sample size were measured. Moreover, partial correlations were adjusted for the sample size between the year of publication and SMD, placebo, and active treatment was measured to assess the relationship between the year of publication as a proxy to the time of trial conducting and the effect size and the response of placebo and the active treatment, all in relation to the sample size. The year of publication was used as a proxy for the year of trial conducting, since this was the most appropriate and available information from all trials and was used in most of the literature as a proxy for year of conducting the trial (Agid et al., 2013; Julious & Wang, 2008; Walsh et al., 2002).

A parametric (Pearson) correlation and non-parametric (Spearman) correlation between the SMD and year of the trial's publication was calculated. The reviews included both reviews with positive (healing, improvement, etc.) and adverse outcomes (death, relapse, pain intensity). To perform one scale of measure, the SMD for the reviews of negative outcomes was transformed into a positive outcome, and then the correlations were calculated. All reviews presented in this chapter reported a positive outcome. As a sensitivity analysis, the absolute SMD was used too instead of transformation (the results from the absolute SMD are presented in Appendix C).

The correlation is considered weak if the correlation coefficient is from [0 , 0.3], moderate if the correlation coefficient is [0.3, 0.5], and a correlation coefficient of more than 0.5 is considered a strong correlation (Burns & Grove, 2007). The aim of subdividing the correlations into weak, moderate and strong correlations was to demonstrate the percentage of correlations that fell into these two categories regardless of the sign of the correlation.

For the illustrated examples presented in this chapter, pairwise meta-regression with the year of publication as a covariate was used to assess the effect of year of publication on the main estimate of treatment difference between the placebo and the active treatment using the bubble plot. Pairwise meta-regression is a pairwise meta-analysis that can be adjusted for covariates. It was used in the literature to assess the effect of different covariates in the changes of placebo

effect over time in different therapeutic areas (Agid et al., 2013; Khan et al., 2017, 2018b, 2018c); more details about the pairwise meta-regression will be presented in Chapter 7.

SPSS version 24 (IBM Corp, 2016) was used to collect and analyse the data. Each meta-analysis was treated as a separate SPSS data file then all these databases were aggregated in one SPSS file and organised by Cochrane ID. For the meta-regression, the R meta-package was used for the analysis (Schwarzer, 2007).

5.3 Results

5.3.1 Data extraction

Following systematic reviews conducted in the Cochrane database, 684 titles were identified to have a placebo term in the abstract or the title. Of these, 289 titles were excluded after reviewing the abstract, and 98 titles were excluded after a secondary assessment of the review (reviewing the manuscript). The final sample included 236 reviews for analysis. Figure 5.1 represents the flow diagram for the data extraction process.
The main reasons for exclusion in the full-text article assessment were: three or fewer trials in the review (238 reviews); 53 reviews had no trials; 56 reviews had one trial; 69 had only two trials in the review; and 60 reviews had three trials. In 138 reviews, data could not be pooled for meta-analysis.

Additionally, 59 reviews had the wrong study design: 52 were for active-controlled trials; two reviews were NMA; and five reviews were overviews of Cochrane reviews. Four reviews used non-medical treatment; four reviews were withdrawn from publication; three reviews had missing information; and in one review all trials were conducted in the same year.

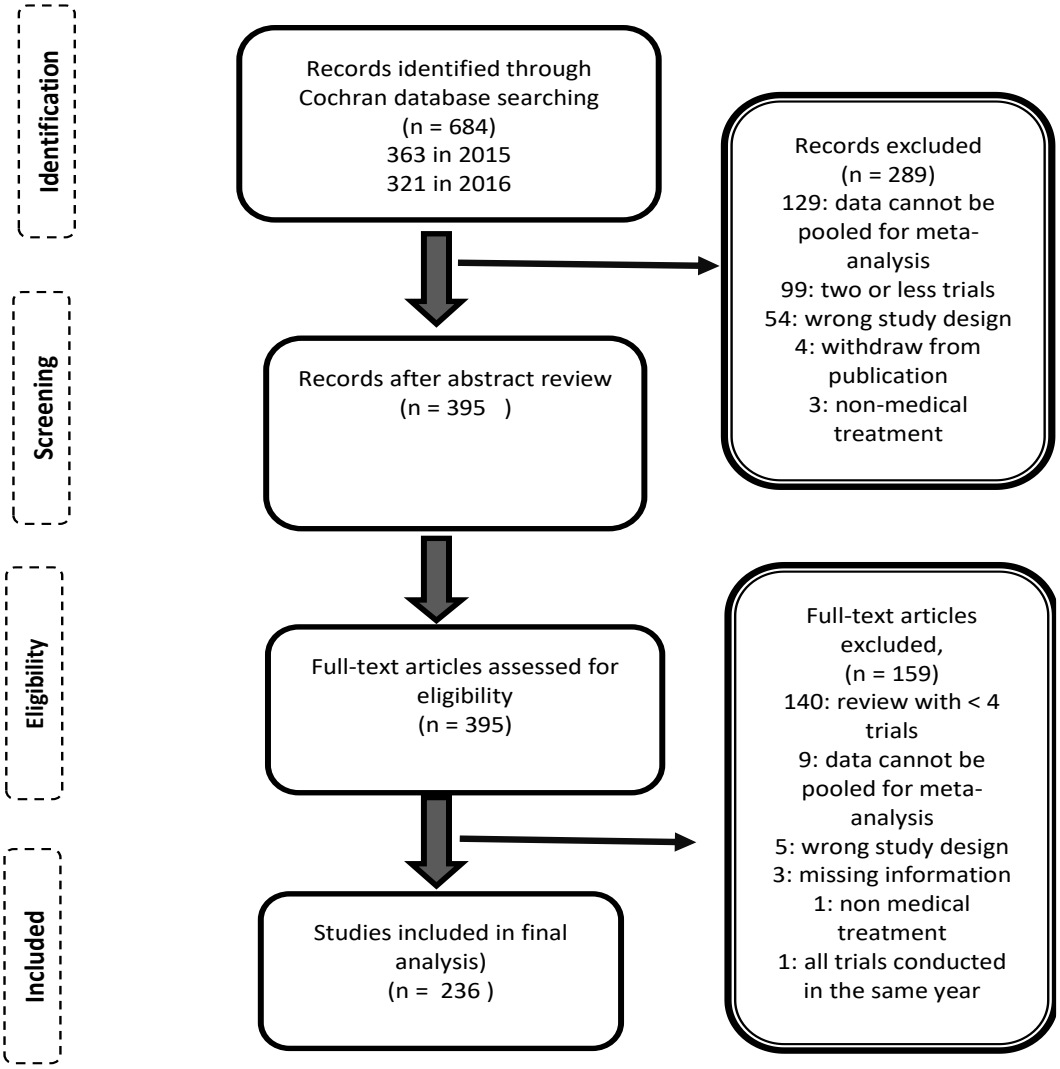

Figure 5-1 Flowchart for the process of data extraction

5.3.2 Characteristics of the included reviews

In total, 2489 placebo-controlled trials from 236 meta-analyses were included in the final analysis. Among the meta-analyses, 155 (65.4%) measured negative outcomes, and 82 (34.6%) measured positive outcomes. Primary meta-analysis was used in 152 reviews (64.1%). The median number of trials was seven trials, and the mean was 9.9 trials, with a minimum number of four trials and a maximum of 51 trials.

The years of trial conducting ranged from 1931 to 2016. The year difference ranged from one year to 80 years. Among the included meta-analyses, 76 (32.1%) used mean difference as the measure of effect. The risk ratio was used in 131 (55.3%), the odds ratio in 27 (11.4%) of the meta-analyses, and the risk difference in only three meta-analyses (1.3%). The most common outcome measured was pain, 30 (12.17%), followed by death, 26 (11%), in the included meta-analyses. The median sample size was 1160 participants with IQR (interquartile range) (494 - 2229), the minimum sample size was 105 and the maximum was 43290 participants. Additionally, 103 (43.5%) meta-analyses used the fixed effect model and 134 (56.5%) used the random effects model.

Figure 5.2 demonstrates the relation between the number of trials in the meta-analyses and the used model. The mean number of trials for the fixed effect model meta-analyses was 9.8 compared to 11 trials in the random effects meta-analyses. However, the difference was not statistically significant.

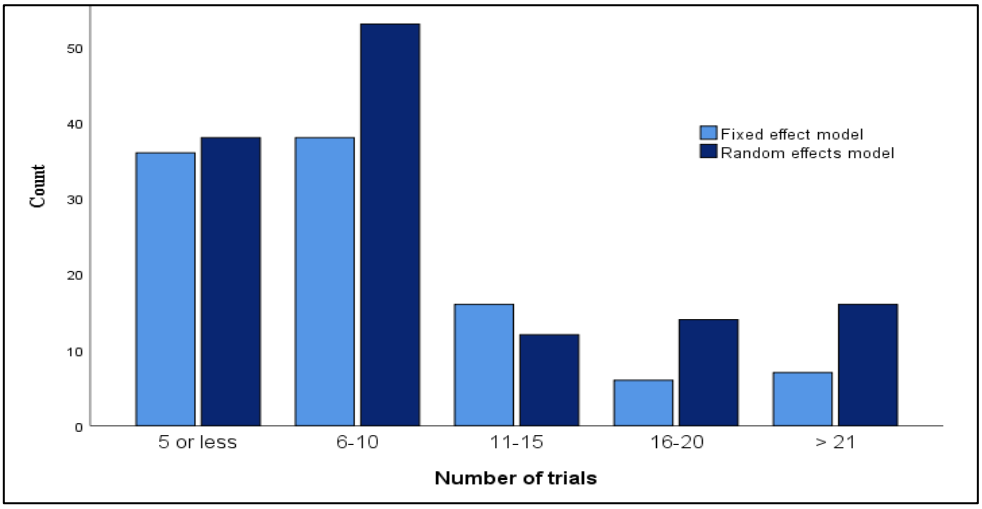

Figure 5-2. Type of model used per number of trials

Increasing the sample size was associated with more positive results. Specifically, 170 (72%) of the included meta-analyses concluded statistically significant results and only 66 (28%) failed to reject the null hypothesis. Among meta-analyses with 4-7 trials, 66.4% had statistically significant results compared to 78.1% of the meta-analyses containing more than seven trials. The mean number of trials in the meta-analysis with statistically significant results was 11.34 compared to eight trials for the meta-analyses with non- statistically significant results (p-value = 0.012).

Regarding the different types of control group included in the analysed meta-analyses (Figure 5.3), 42.4% of the included meta-analysis defined the control group as a placebo, placebo or no treatment was used in 24.2% of the reviews, and 17.8% defined control group as placebo or no treatment or usual care.

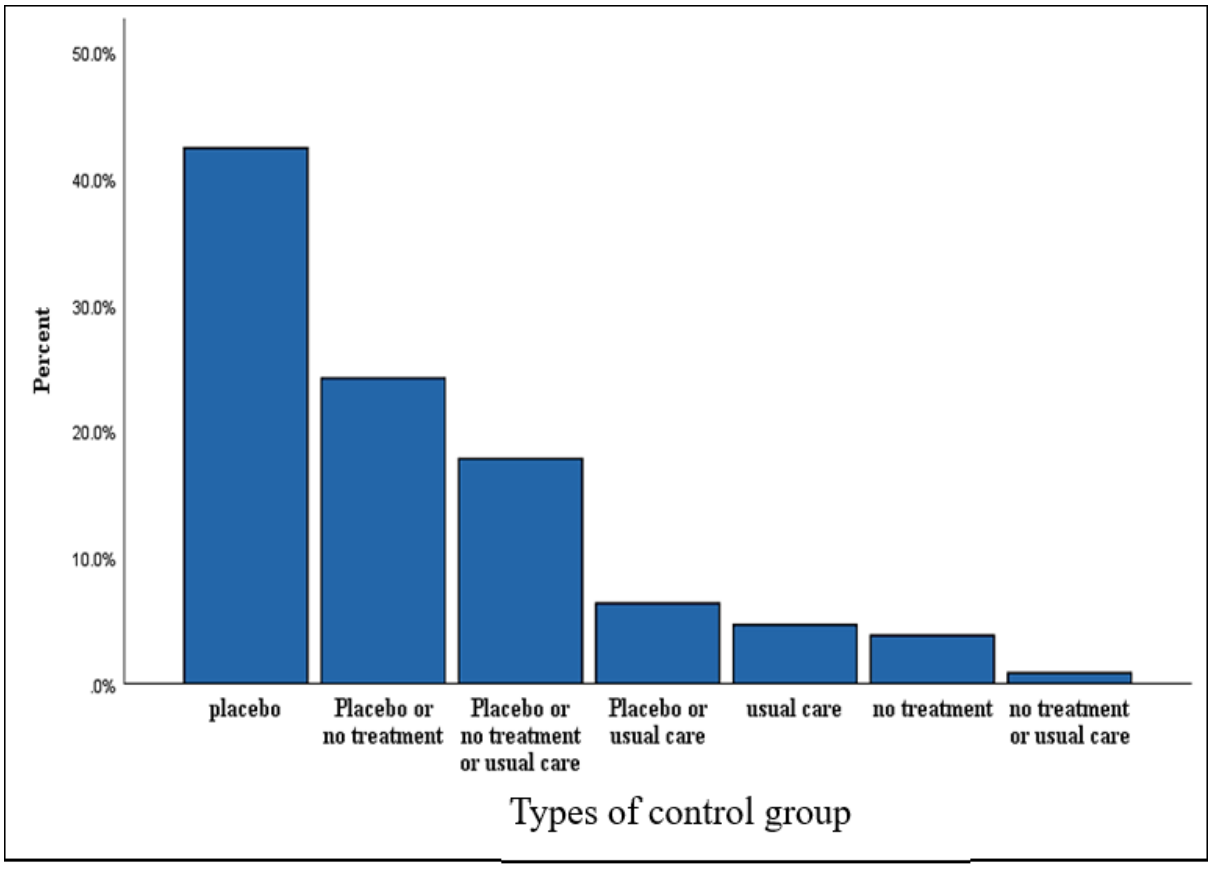

Figure 5-3 Different definitions of control group

Table 5.1 explains the different Cochrane groups included in the review. There were 17 different therapeutic areas; the most frequently occurring was Gynaecology and Obstetrics with 37 (15.6%) reviews.

Table 5.1 Distribution of the reviews by Cochrane groups

Cochrane group	Frequency
Pain, Palliative and Supportive Care Group	22.0 (9.3%)
Pregnancy and Childbirth Group	20.0 (8.5%)
Gynaecology and fertility group	15.0 (6.4%)
Heart Group	13.0 (5.5%)
Anaesthesia, Critical and Emergency Care Group	11.0 (4.7%)
IBD Group	11.0 (4.7%)
Musculoskeletal Group	10.0 (4.2%)
Stroke Group	9.0 (3.8%)
Kidney and Transplant Group	9.0 (3.8%)
Airway group	8.0 (3.4%)
Hypertension Group	8.0 (3.4%)
Acute Respiratory Infections Group	7.0 (3.0%)
Infectious Disease Group	7.0 (3.0%)
Vascular Group	7.0 (3.0%)
Common Mental Disorders Group	6.0 (2.5%)
Drugs and Alcohol Group	5.0 (2.1%)
ENT Group	5.0 (2.1%)
Neonatal Group	5.0 (2.1%)
Neuromuscular group	5.0 (2.1%)
Schizophrenia Group	5.0 (2.1%)
Skin Group	5.0 (2.1%)
Upper GI and Pancreatic Diseases Group	5.0 (2.1%)
Developmental, Psychosocial and Learning Problems Group	3.0 (1.3%)
Epilepsy Group	3.0 (1.3%)
Wounds Group	3.0 (1.3%)
Hepato-Biliary Group	3.0 (1.3%)
Tobacco Addiction Group	3.0 (1.3%)
Bone, Joint and Muscle Trauma Group	2.0 (0.8%)
Cystic Fibrosis and Genetic Disorders Group	2.0 (0.8%)
Dementia and Cognitive Improvement Group	2.0(0.8%)
Eye and Vision Group	2.0 (0.8%)
Haematological Malignancies Group	2.0 (0.8%)
Incontinence Group	20 (0.8%)
Metabolic and Endocrine Disorders Group	2.0 (0.8%)
Movement Disorders Group	2.0 (0.8%)
Other groups	7.0 (3.0%)
Total	236.0 (100.0%)

IBD; inflammatory bowel disease

All included reviews included the risk of bias assessment following the author view. Randomisation, blinding, attrition, and reporting were all included in the assessment of the risk of bias. Regarding bias, 20.3% of the reviews were considered at high risk, 55.7% had a moderate risk of bias, in 9.7% of the reviews the risk of bias was hard to assess, and only 14.3% were considered to have a low risk of bias. The quality of evidence was reported for most of the reviews in the summary of findings box. The quality of evidence reflects the author's confidence in the estimate of effect. Only 19.8% of the reviews had high quality of evidence, 36.5% had moderate quality, while 43.5% of the reviews had low to very low quality of evidence (Table 5.2).

Table 5.2 Risk of bias and quality of evidence

Risk of Bias	Frequency, N (%)	Quality of Evidence	Frequency, N (%)
High	48.0 (20.3%)	Very Low	19.0 (8.1%)
Moderate	131.0 (55.5%)	Low	84.0 (35.2%)
Low	34.0 (14.4%)	Moderate	87.0 (36.9%)
Unclear	23.0 (9.7%)	High	47.0 (19.9%)
Total	236.0 (100.0%)	Total	236.0 (100.0%)

The heterogeneity measured in $I^2$was reported for all reviews and ranged from 0 per cent to 99 per cent. The heterogeneity was defined as mild if $I^2 < 30\%$, moderate if, $30\% \leq I^2 \leq 50\%$ and considerably high if $I^2 > 50\%$ (Whitehead, 2002). Half of the meta-analyses had mild heterogeneity, with 74 meta-analyses having $I^2 = 0\%$; 32.2% of them had considerably high heterogeneity; and only 17.8% had moderate heterogeneity. The heterogeneity was higher, with statistical significance, for meta-analyses that used a random effects model and for meta-analyses with a large number of trials.

5.3.3 Results of Correlations

Correlations between the year of publication and sample size, placebo effect, active treatment effect, SMD were obtained. Partial correlations between the year of publication and placebo effect, active treatment effect, SMD after controlling for the sample size were obtained. Both Pearson and Spearman correlations were obtained. Correlations were measured for the 2489 trials in general and then individually for each meta-analysis. The results for parametric and non-parametric correlations were similar. The results regarding Spearman correlation are presented in Appendix C.

5.3.3.1 Correlations between total sample size and year of trial publication

For all included trials the correlation between the sample size in a trial and the year of publication was positively correlated with the Pearson correlation, 0.038, 95% CI [0.006; 0.086], and the Spearman correlation was 0.15, 95% CI [0.0116; 0.194]. Regarding the meta-analyses included, in 179 (75.5 %) meta-analyses the correlation between the year of publication and the sample size was a positive correlation with the median correlation = 0.2. In 72.5% of the included meta-analyses, the Spearman correlation between the sample size and the year of publication was a positive correlation with mean correlation = 0.28 (Figure 5. 4).

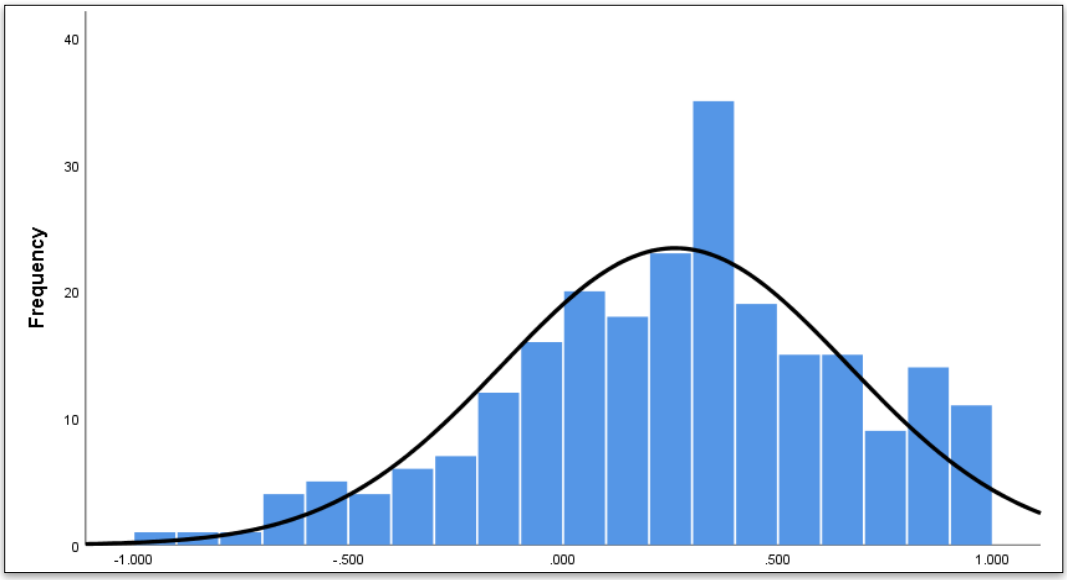

Figure 5-4: Histogram for the correlation between sample size and year of publication

5.3.3.2 Correlation between the SMD and the year of publication

The correlations for the reviews with negative outcome were transformed to positive outcomes to present one scale of measure. Regarding all included trials in general, the year of publication was negatively correlated with SMD, with Pearson correlation of - 0.013, 95% CI [-.055; 0.03] and Spearman correlation of - 0.048, 95% CI[-.085; -0.007]. Regarding meta-analyses, 58.2% of the meta-analyses had a negative correlation between the standardised mean difference and the year of publication. The median correlation was - 0.12, mean was - 0.083 and the standard deviation (SD) = 0.43. For the Spearman correlation, the median was - 0.11, mean was - 0.087 and SD = 0.43. (Figure 5.5)

Regarding the partial Pearson correlation after controlling for the sample size, 56.8% of the reviews had a negative correlation between the year of publication and the SMD after controlling for the sample size. The median correlation was - 0.093, mean was - 0.059, and SD was 0.48 (Table 5.3 and Figure 5.6).

Table 5.3 Correlation and the partial correlation between SMD and the year of publication

Correlation	Pearson correlation	Partial correlation
Strong Negative	47.00 (19.90%)	48.00 (20.30%)
Moderate Negative	38.00 (16.10%)	26.00 (11.00%)
Weak Negative	53.00 (22.50%)	60.00 (25.40%)
Weak Positive	45.00 (19.10%)	44.00 (18.60%)
Moderate Positive	32.00 (13.60%)	25.00 (10.60%)
Strong Positive	21.00 (8.9 0%)	33.00 (14.00%)
Total	236.00 (100.00%)	236.00 (100.00%)

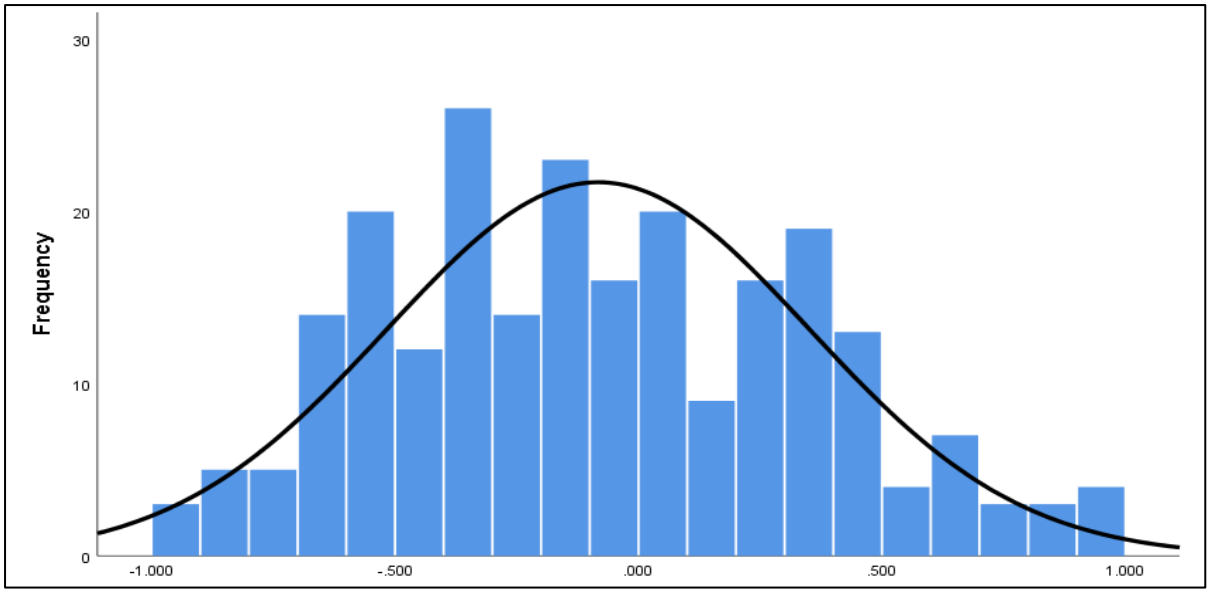

Figure 5-5 Pearson Correlation between standardised mean difference and the year of publication

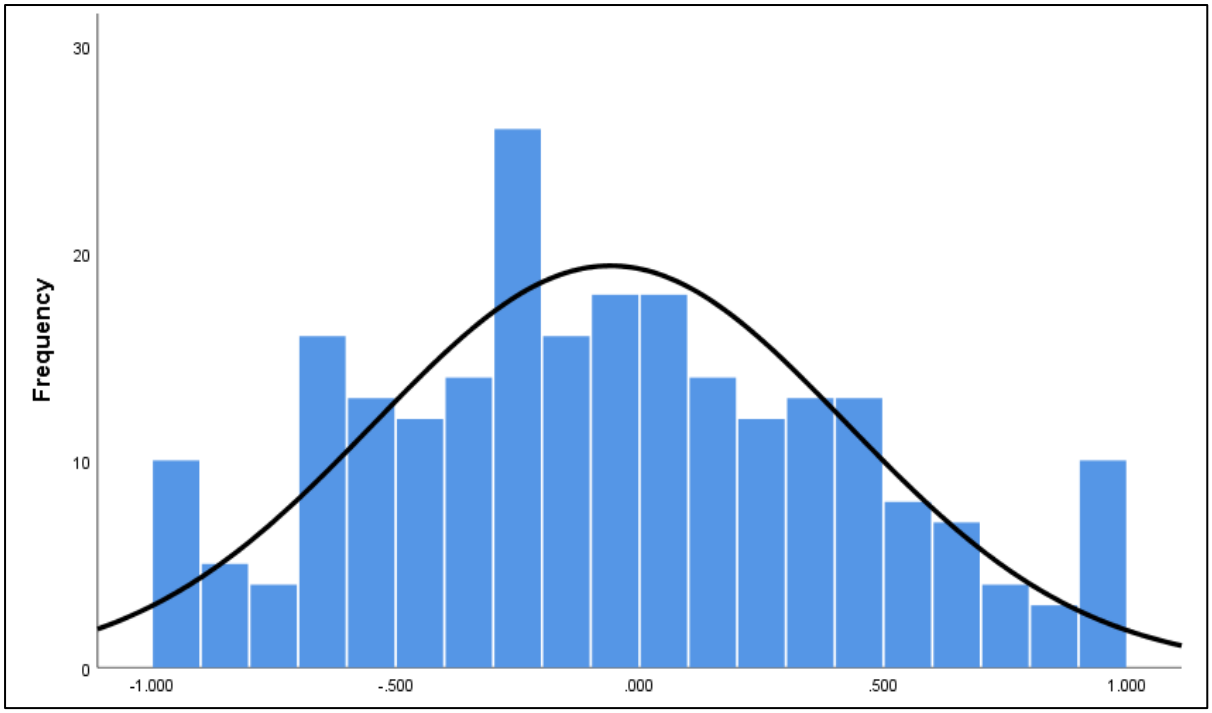

Figure 5-6 Partial Correlation between standardised mean difference and the year of publication

5.3.3.3 Correlation between the placebo response and the year of publication

As mentioned earlier, the correlations for the reviews with negative outcome were transformed into positive outcomes. In 58.6% of the reviews there was a positive correlation between the placebo response and the year of publication (Table 5.4). The median correlation was 0.09, mean was 0.07, and SD was 0.44. Regarding the Spearman correlation, the median was 0.06, mean was 0.05, and the SD was 0.44 (Figure 5.7).

Table 5.3 Correlation between the placebo response and the year of publication

Correlation	Pearson correlation, N (%)	Partial correlation, N (%)
Strong Negative	29.0 (12.9%)	34.0 (15.1%)
Moderate Negative	16.0 (7.1%)	23.0 (10.2%)
Weak Negative	47.0 (20.9%)	50.0 (22.2%)
Weak Positive	60.0 (26.7%)	47.0 (20.9%)
Moderate Positive	35.0 (15.6%)	28.0 (12.4%)
Strong Positive	38.0 (16.9%)	43.0 (19.1%)
Total	226.0 (100.0%)	225.0 (100.0%)

A partial correlation after controlling for the sample size was obtained in 226 reviews. In 52.2% of the reviews there was a positive correlation between the year of publication and the placebo response after controlling for the sample size. The median correlation was 0.05, mean was 0.04, and SD was 0.44. These results mean that increasing the year of publication will increase the placebo response, i.e. placebo response improved over time (Figure 5.8). There were no differences between the Pearson and Spearman correlations (Appendix C). Reviews in which the placebo response was missing were excluded from the analysis.

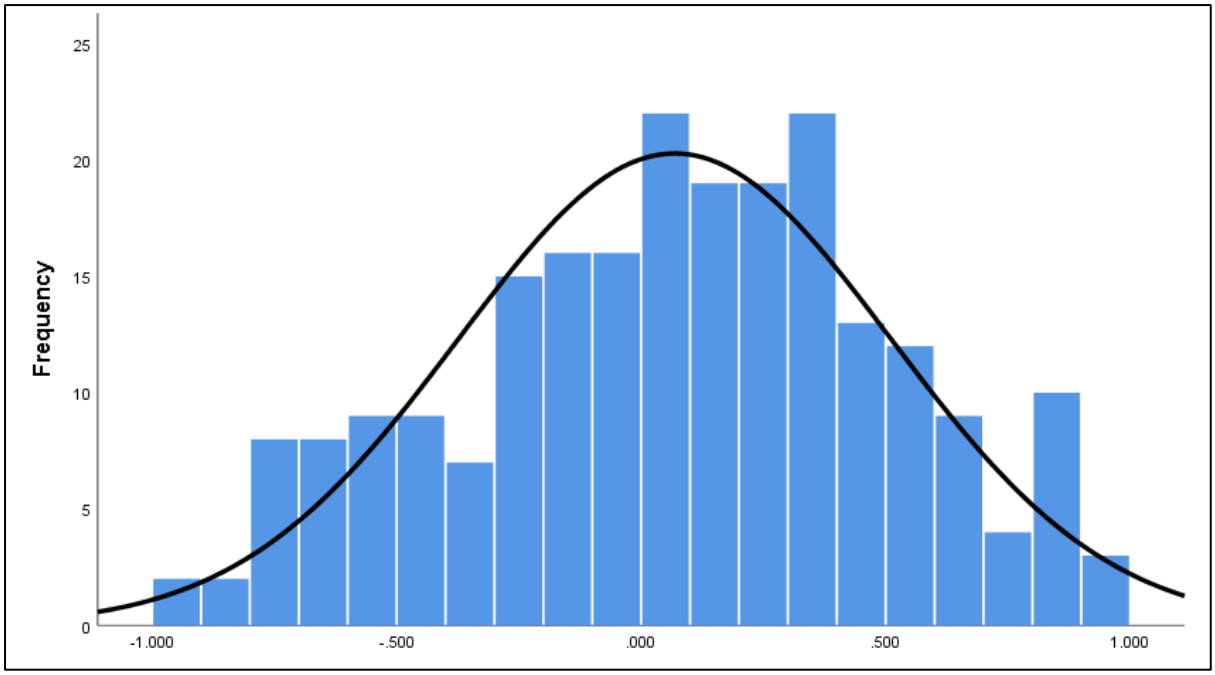

Figure 5-7 Pearson Correlation between placebo and year of publication

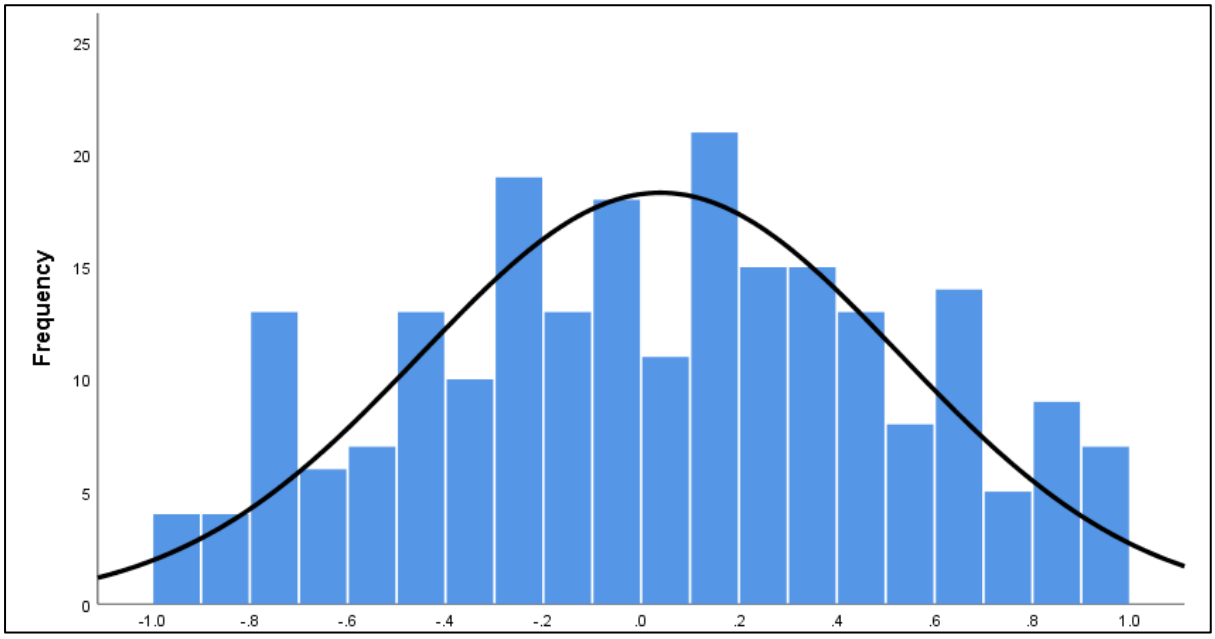

Figure 5-8 Partial Correlation between placebo and year of publication

5.3.3.4 Correlation between the active treatment response and the year of publication

The correlations for the reviews with negative outcome were transformed to positive outcomes to present one scale of measure between the negative and positive reviews. In 52% of the reviews there was a negative correlation between the active treatment response and the year of publication (Table 5.6). The median correlation was - 0.04, mean was - 0.02, and SD was 0.43. The median Spearman was - 0.10, mean was - 0.05, and SD was 0.43 (Figure 5.9).

The partial Pearson correlation after controlling for the sample size was obtained in 226 reviews. In 51.7% of the reviews there was a negative correlation between the year of publication and the active treatment response after controlling for the sample size. The median correlation was - 0.03, mean was - 0.02, and SD was 0.43 (Figure 5.10). There were no differences between the Spearman and the Pearson correlations (appendix C).

Table 5.4 Correlations between the active treatment and the year of publication

Correlation	Pearson correlation	Partial correlation
Strong Negative	34.00 (15.10%)	43.00 (19.10%)
Moderate Negative	31.00 (13.80%)	35.00 (15.60%)
Weak Negative	51.00 (22.70%)	38.00 (16.90%)
Weak Positive	58.00 (25.80%)	51.00(22.70%)
Moderate Positive	18.00 (8.00%)	18.00 (8.0%)
Strong Positive	33.00 (14.70%)	40.00 (17.80%)
Total	226.00 (100.00%)	225.00 (100.00%)

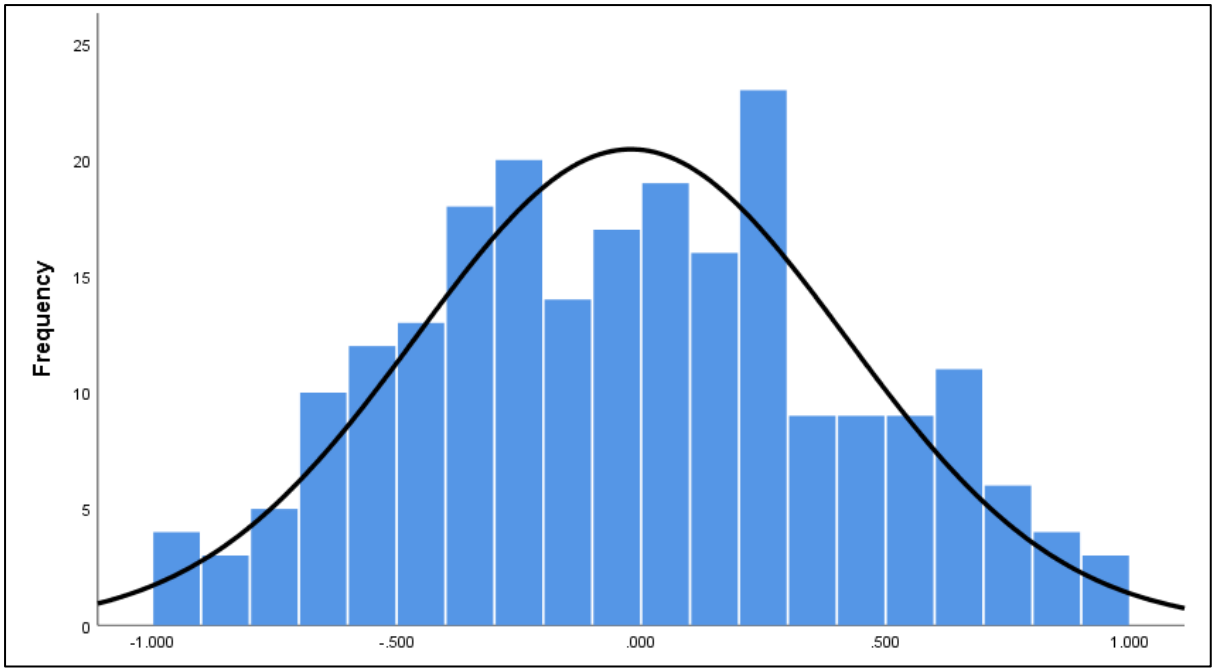

Figure 5-9 Pearson Correlation between the active treatment and the year of publication

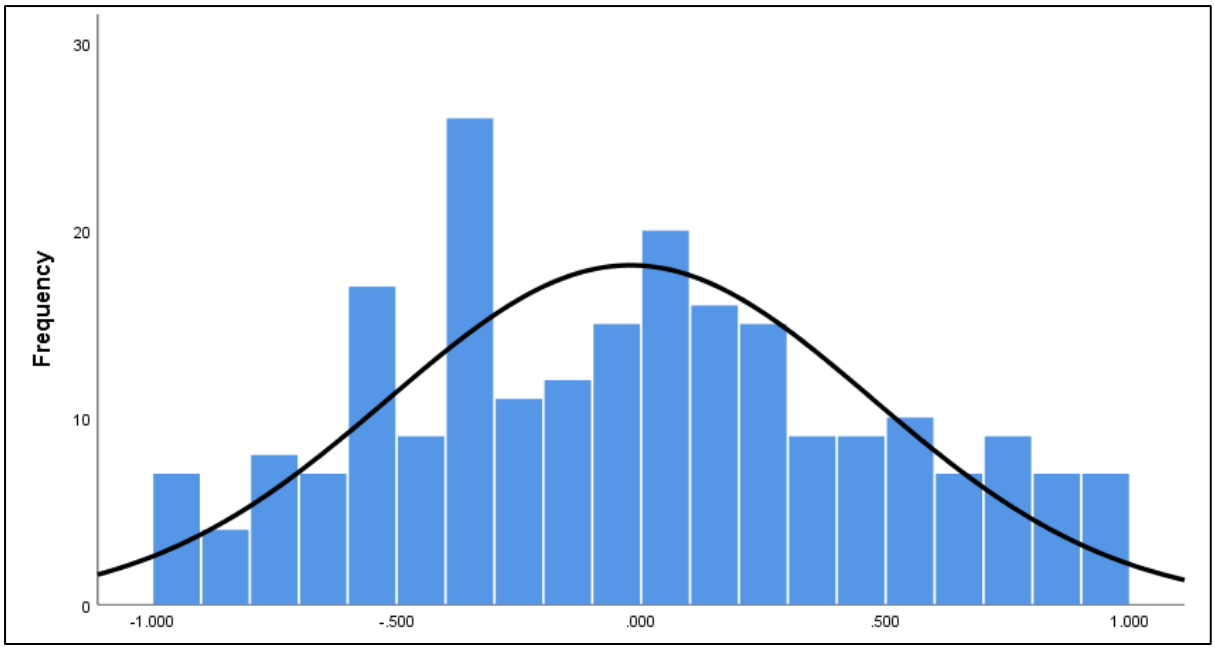

Figure 5-10 Partial Correlation between the active treatment and the year of publication

The obtained correlations and partial correlations were affected only by the number of trials included in the analysis; meta-analyses with a smaller number of trials had the strongest correlations (SMD, placebo and active treatment) on both sides (positive and negative). The relation between the SMD correlations and the placebo correlations was in the opposite direction, while the relation between the active treatment correlations and the SMD correlations was in the same direction for both the Pearson and partial correlations (Figure 5.5 and Figure 5.6). These results indicate that the active treatment, not the placebo, had the main impact on the changes of the SMD response over time.

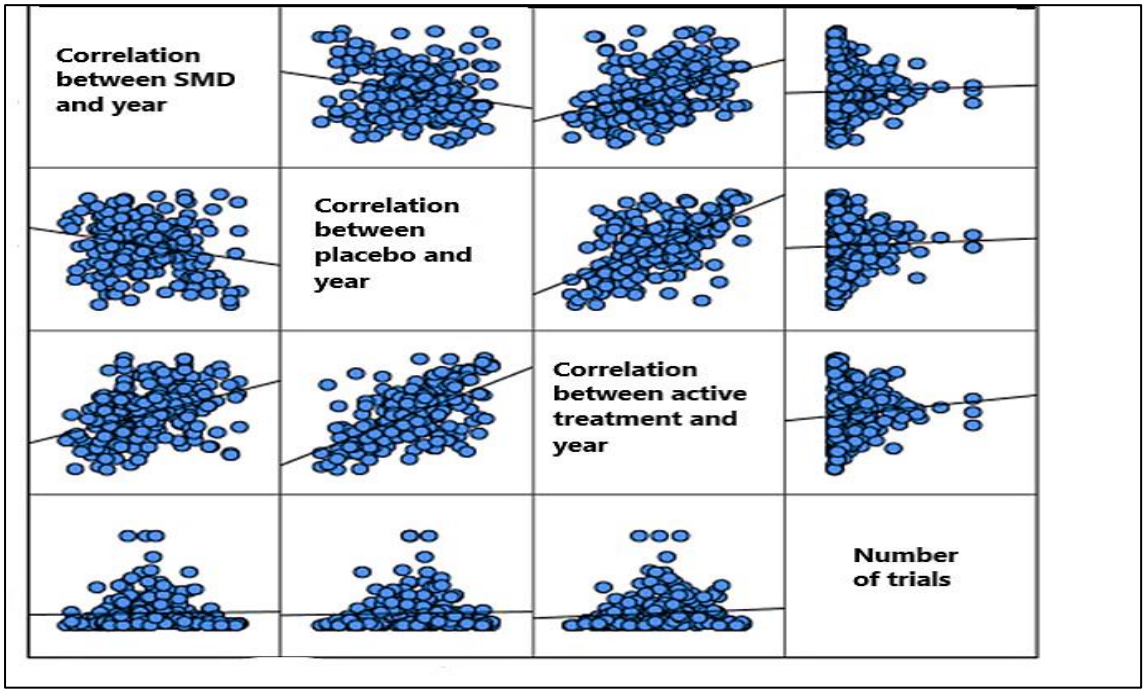

Figure 5-11 Matrix scatter plot showing the correlations between placebo, active treatment, SMD and the number of trials in the meta-analysis

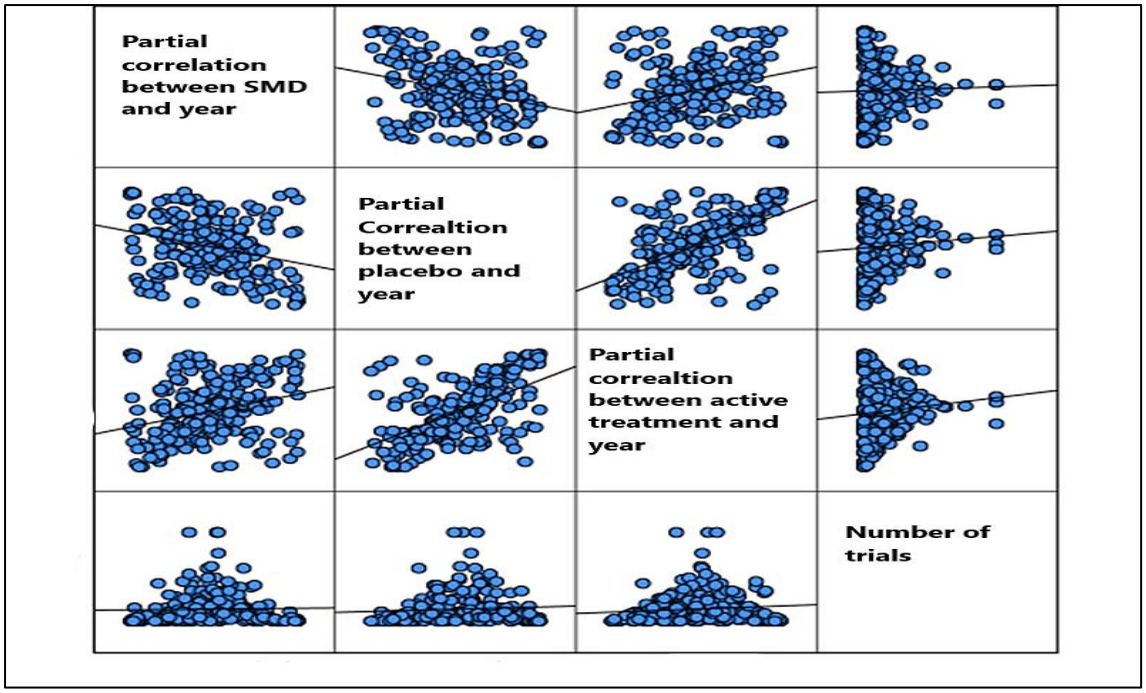

Figure 5-12 Matrix scatter plot showing the partial correlations between placebos, treatment, SMD and the number of trials in the meta-analysis

5.4 Illustrated examples

Three meta-analyses (from the included 236 meta-analyses) were selected to illustrate the changes in the treatment response and placebo response over time in this section. The first meta-analysis (Adams, Sekhon, & Wright, 2015) was selected because it was from the cardiovascular therapeutic area, which was the most common in this review and the least studied in the literature regarding the changes in placebo effect over time. Additionally, it used an objective outcome measure, depending on the changes of total cholesterol in the blood, and finally, it included a large number of trials with a 20-year difference in the publication years. The second review (Enthoven, Roelofs, Deyo, Van Tulder, & Koes, 2016) was chosen because it used a subjective outcome measure (pain) with a smaller number of trials included and range of publication years of over 20 years. The third review (Se et al., 2016) related to the prevention of depression, and both the therapy and the outcome were measured subjectively.

5.4.1 Atorvastatin for lowering lipids

This review was published in 2015 and aimed to assess the effects of various doses of atorvastatin on body lipids (total serum cholesterol, low-density lipoprotein (LDL), high-density lipoprotein (HDL) and triglycerides) in individuals with and without evidence of cardiovascular disease. It included 296 trials in total (242 before and after trials and 54 placebo-controlled trials), with 38,817 patients in total. The main conclusion was that atorvastatin decreases total blood cholesterol and LDL-cholesterol in a linear dose-related manner. In general, the evidence from this review is considered as high-quality evidence and the risk of bias is considered as moderate (Adams et al., 2015).

The meta-analysis used in this example contains 24 placebo-controlled trials published from 1995 to 2014. The total sample size was 1902 participants. The active treatment was atorvastatin 10 mg. The mean difference of cholesterol reduction was the outcome measurement (negative outcome), and the fixed effect meta-analysis model was used to calculate the estimate. The final estimate was a statistically significant difference between the atorvastatin and the placebo in reduction of total cholesterol (mean difference (MD) = -25.44, 95% CI [-26.38; -24.5]) (Figure 5.7).

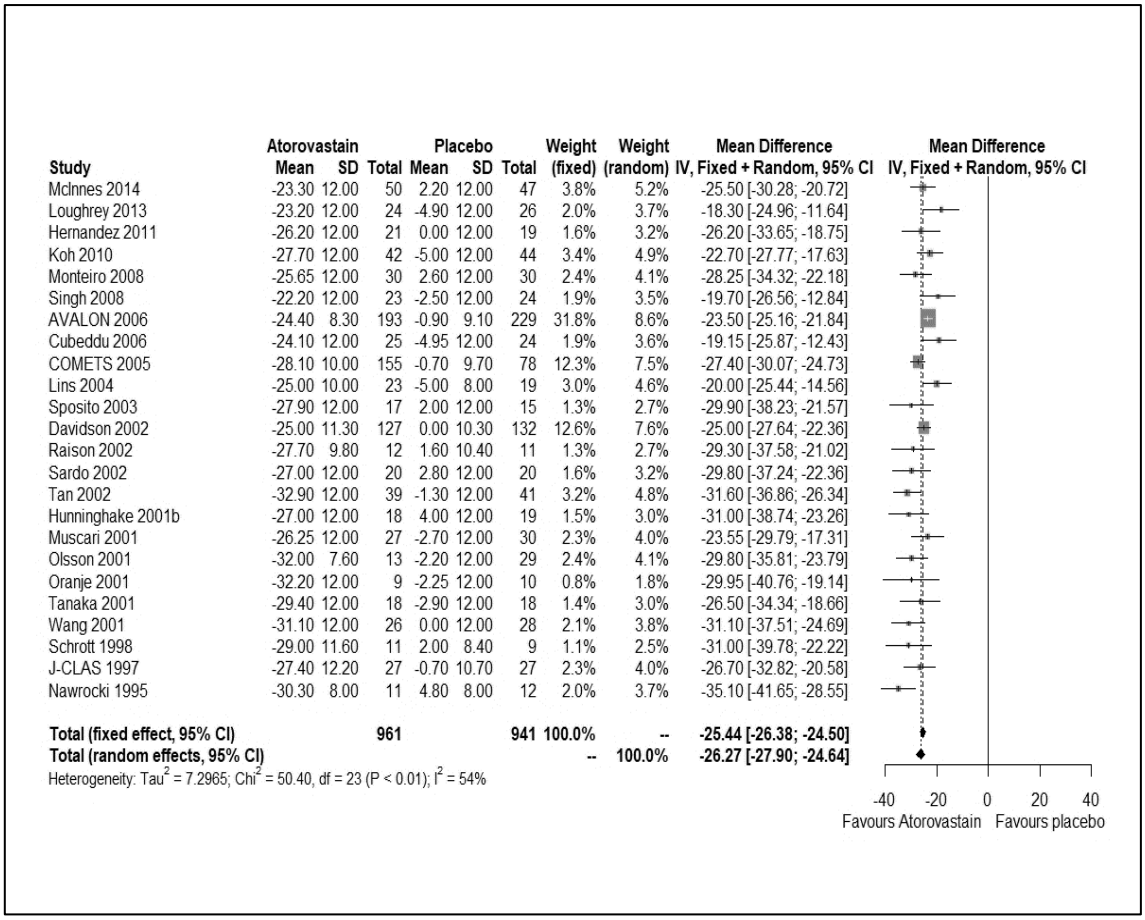

Figure 5-13 Forest plot of the effect of atorvastatin on lowering blood cholesterol

After transformation from negative to a positive outcome, there was a strong positive correlation between the sample size and the year of publication (r = 0.79, p-value = 0.03). This means that recent trials have a larger sample size than old ones. The correlation between the placebo and year of publication was a non- statistically significant moderate positive correlation (r = 0.33, p-value = 0.10). This correlation did not change after controlling for the sample size, the partial correlation was moderate positive correlation (r = 0.33, p-value = 0.13). The effect of placebo on lowering the cholesterol level was improved over time.

The correlation between the atorvastatin and the year of publication was a statistically significant strong negative correlation (r = - 0.63, p-value < 0.01). The partial correlation after controlling for the sample size decreased but was still strong negative correlation (r = - 0.612, P-value < 0.01).

In 1995, the atorvastatin decreased the cholesterol level by 30.30 mg/dl compared to 23.30 mg/dl in 2014. The correlation between the SMD and year of publication was a strong negative correlation (r = - 0.65, p-value < 0.01). After controlling for the sample size, the partial correlation was also strong negative correlation (r = - 0.66, p-value <0.01). The difference between atorvastatin and placebo was decreased over time, even after controlling for the sample size (Figure 5.8).

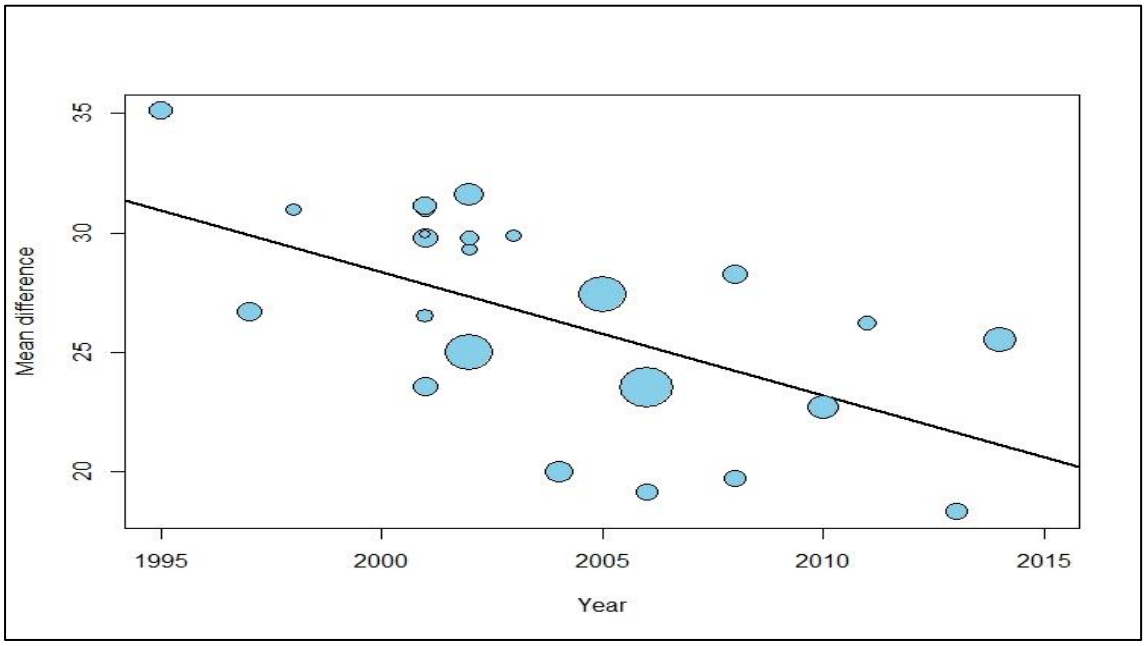

Figure 5-14 Bubble plot of the estimate mean difference between the atorvastatin and placebo by year of publication

5.4.2 Non-steroidal anti-inflammatory drugs for chronic low back pain (2016)

This review aimed to assess the effects of non-steroidal anti-inflammatory drugs (NSAID) among people with chronic back pain. It includes 13 trials in total (six trials are placebo-controlled trials; the other seven trials are active-controlled trials). Included in total were 1354 participants with follow up between nine days and 16 weeks. The main conclusion was that NSAID effectively reduced pain and disability associated with low back pain compared to placebo. In general, the evidence from this review is considered as low-quality evidence and the risk of bias is considered as a moderate risk (Enthoven et al., 2016).

The meta-analysis used in this example contains six placebo-controlled trials published from 1982 to 2013. The total sample size was 1354 participants. The active treatment was different types of NSAID, and the outcome of interest was a reduction in the pain intensity from the baseline (negative outcome). The mean difference was the outcome measurement, and the random effect model was used to calculate the estimate. The final estimate was a statistically significant difference between the NSAID and the placebo in reduction of pain intensity (mean difference (MD) = - 6.97, 95% CI [-10.74; -3.19]) (Figure 5.9).

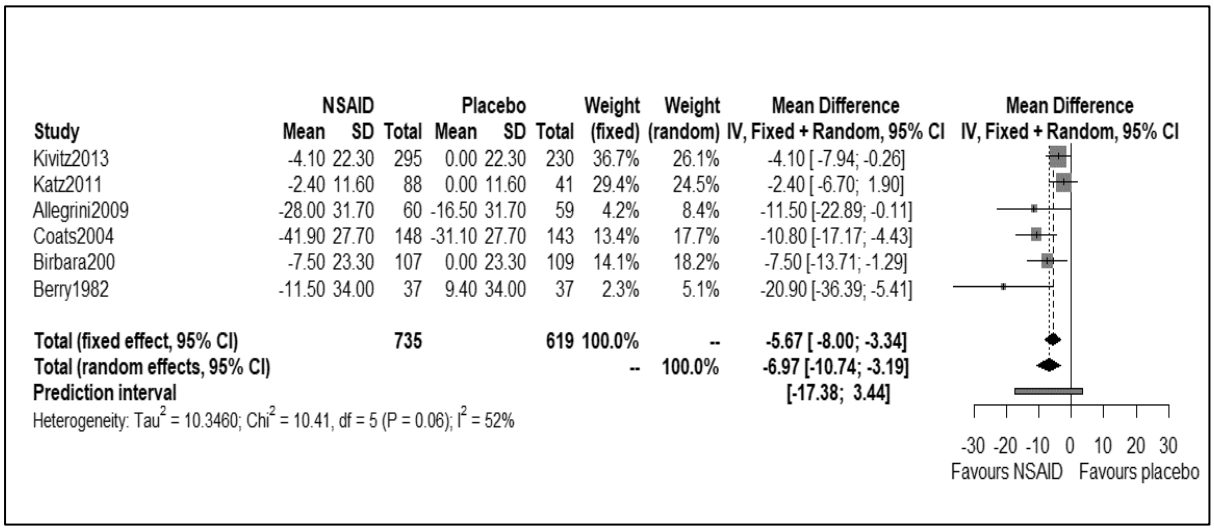

Figure 5-15 Forest plot of the effect of NSAID on the reduction of pain intensity compared to placebo (*size of the bubble reflects the sample size***)**

There was a strong non- statistically significant positive correlation between the sample size and the year of publication (r = 0.52, p-value = 0.26). That means recent trials have a larger sample size than old ones. The Spearman correlation was the same (r = 0.54, p-value = 0.26).

The correlation between the placebo and year of publication was a non- statistically significant moderate positive correlation (r = 0.36, p-value = 0.48). That means the efficacy of placebo to reduce the pain was increased over time; this correlation did not change after controlling for the sample size, the partial correlation was a moderate positive correlation (r = 0.34, p-value = 0.57). The correlation between the NSAID and the year of publication was a weak negative correlation (r = - 0.04, p-value = 0.90). The Spearman correlation was a moderate negative correlation (r = -0.45, p-value = 0.33). These correlations indicate that the efficacy of NSAID decreased with time. The partial correlation after controlling for the sample size changed to a weak positive correlation (r = 0.003, p-value= 0.9).

The correlation between the SMD and year of publication was a strong negative correlation (r = - 0.83, p-value = 0.04). After controlling for the sample size, the partial correlation was also a strong negative correlation (r = - 0.92, p-value = 0.03). The difference between the NSAID and placebo was decreased over time, even after controlling for the sample size. The difference between NSAID and placebo was decreased over time, even after controlling for the sample size (Figure 5.10).

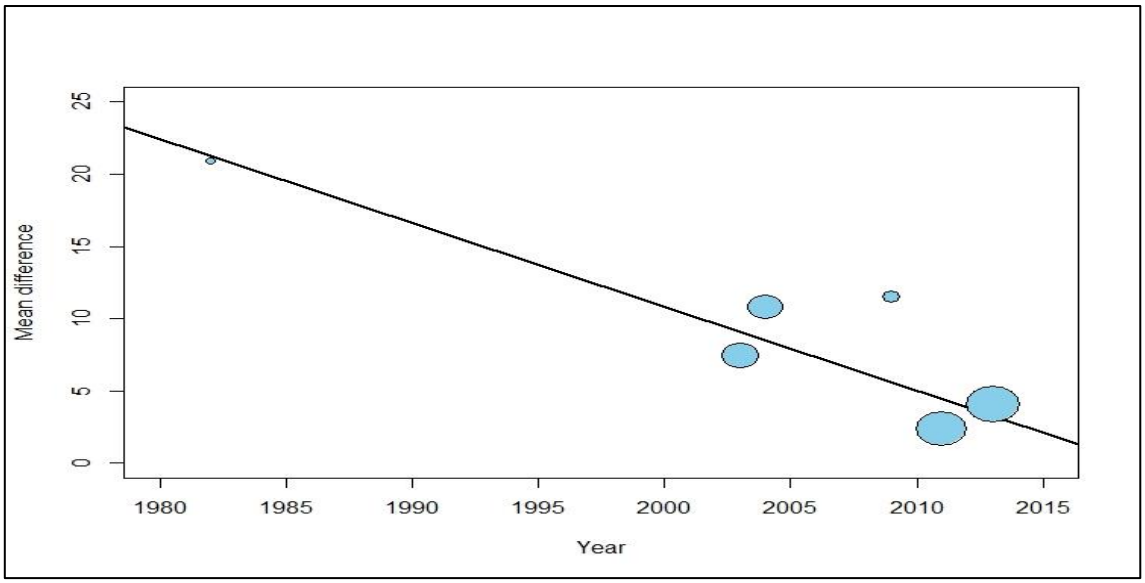

Figure 5-16 Bubble plot of the estimate mean difference by year of publication (*size of the bubble reflects the sample size*)

5.4.3 Cognitive behavioural therapy (CBT), third-wave CBT and interpersonal therapy (IPT) based interventions for preventing depression in children and adolescents

This review was published in 2016 and aimed to investigate the effectiveness of evidence-based psychological interventions (including cognitive behavioural therapy (CBT), interpersonal therapy (IPT) and third wave (CBT)) in preventing the onset of the depressive disorder in children and adolescents (Se et al., 2016).

The primary outcome was depression diagnosis at medium-term follow up (up to 12 months), based on 32 trials with 5965 participants. The treatment used was a psychological, behavioural interventional therapy, not a physical drug therapy. The risk difference was the measure of treatment effect used with the random effects model, which was subjectively measured. Two subgroups were included in the analysis (Targeted and Universal).

The result was a statistically significant reduction of the risk of having a diagnosis of depression for participants receiving an intervention compared to those receiving no intervention (risk difference (RD) - 0.03, 95% CI [-0.05; -0.01], P-value = 0.01) (Figure 5-17). The year difference was 21 years, from 1993 to 2014 — the outcome measure in this study.

Both treatment groups were healthy at the beginning of the study. The aim was to assess the efficacy of this treatment in preventing depression but not in treating patients who had already been diagnosed with depression.

There was a moderate non- statistically significant positive correlation between the sample size and the year of publication (r = 0.25, p-value = 0.17). That means recent trials have a larger sample size than old ones. The Spearman correlation was the same (r = 0.17, p-value = 0.36).

The correlation between the placebo and year of publication was a non- statistically significant weak positive correlation (r = 0.24, p-value = 0.2). That means the efficacy of the placebo was increased over time. This correlation did not change after controlling for the sample size, the partial correlation was weak positive correlation (r = 0.26, p-value = 0.12).

The correlation between the CBT (active treatment) and the year of publication was weak positive correlation (r = 0.06, p-value = 0.13). The Spearman correlation was weak positive correlation (r = 0.18, p-value = 0.33). The partial correlation after controlling for the sample size changed to weak positive correlation (r = 0.15, p-value = 0.43).

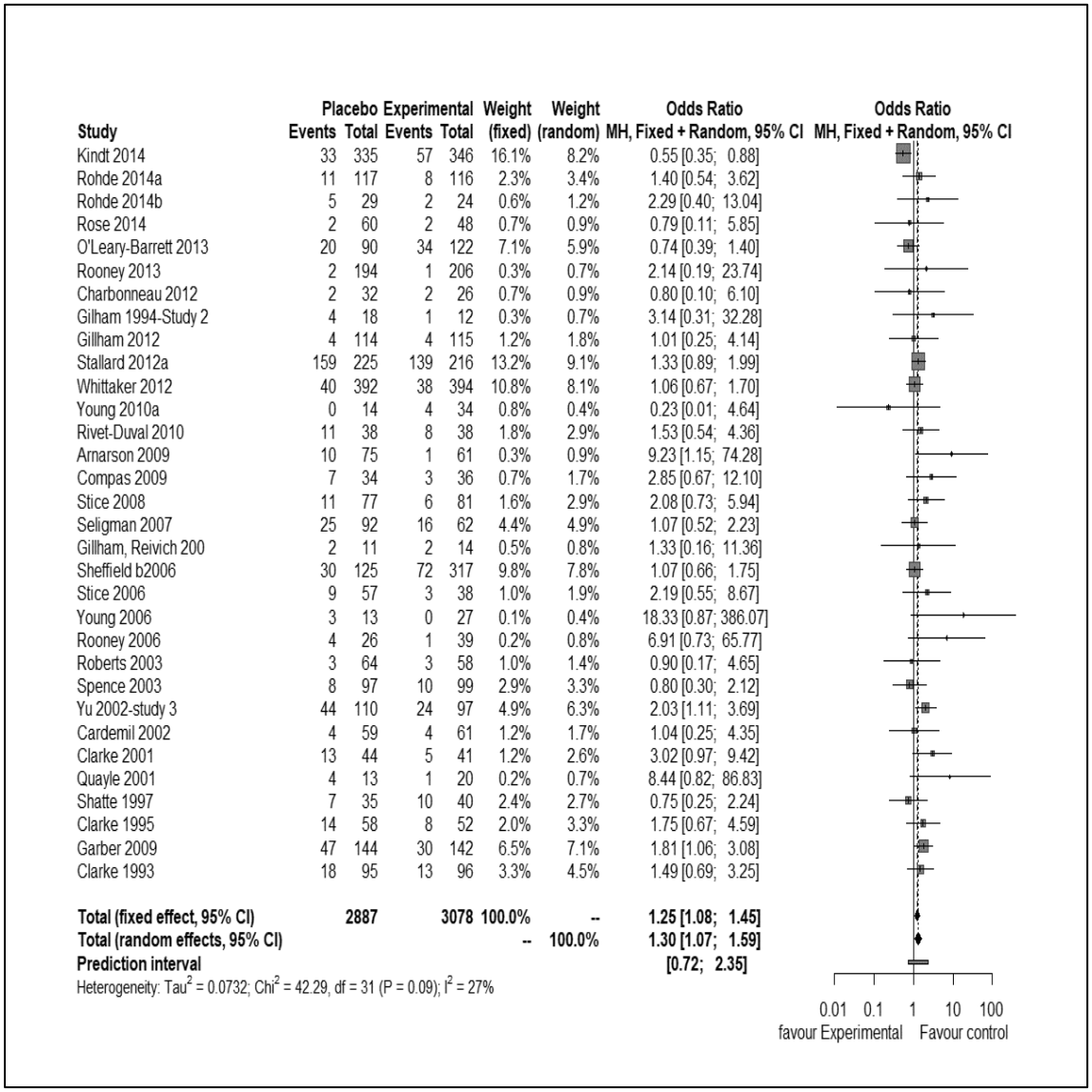

Figure 5-17 Forest plot of comparison of psychological intervention versus no intervention

The correlation between the SMD and year of publication was a weak negative correlation (r = - 0.15, p-value = 0.42). After controlling for the sample size, the partial correlation was also weak negative correlation (r = - 0.08, p-value = 0.6). The difference between the CBT and placebo was decreased over time even after controlling for the sample size (Figure 5-18).

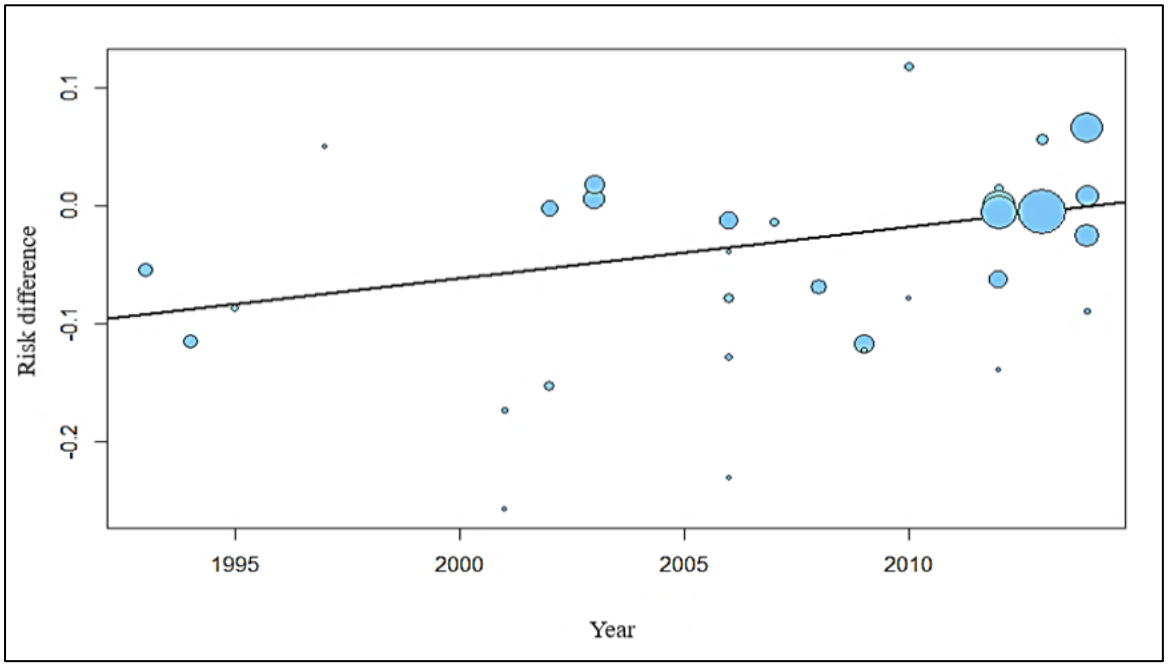

Figure 5-18 Bubble plot for the changes in the risk difference by year of publication (*size of the bubble reflects the sample size*)

5.5 Discussion and conclusion

A review of the Cochrane reviews of placebo-controlled trials was performed in this chapter. The aim was to investigate the effect of changes over time (year of publication) on the difference between the active and control treatment (placebo) by measuring the correlation between the SMD and the year of publication.

The correlations of the SMD varied from strong positive correlations in 21(8.9%) reviews to strong negative correlations in 47(19.9%) reviews. The median correlation between SMD and year of publication was skewed toward the negative, with a weak negative correlation of -0.1. Even though the correlation is considered weak, its negative sign refers to the inverse

relationship between the SMD and the year of publication. That means the difference between the active treatment and placebo was larger in older trials compared to the most recent ones.

There was one review study that also investigated this issue. Rattehalli et al. assessed the effect of antipsychotics to treat schizophrenia (the rate of drop out) using 12 clinical trials from 1992 to 2014. The correlation between the standardised mean difference and year of publication was -0.35, which means that the difference between the placebo and active treatment decreased over time.

Most of the published studies concentrate on the placebo response over time. The results of this review were similar to Agid et al. (Agid et al., 2013), We et al. (We et al., 2012) , Nielsen (Nielsen, 2016), Linde et al. (Linde et al., 2016) and Hm et al. (Hm et al., 2009) in assessing the placebo effect over time. The current review showed that the correlation between the placebo and the year of publication was a positive correlation of 0.07, which reflects the positive relationship between the placebo and the year of publication.

The strongest correlations were between the sample size and year of publication, with a mean correlation of 0.2. That means the sample size in the meta-analysis increased with increase in the publication years. For the effect size of the active treatment itself, the correlation was -0.05. These results reflect a decrease in the effect of the active treatment over time, which was larger in older studies than the recent ones.

Putting all of this together, the results from this chapter indicate a decrease in the effect size of the active treatment and increase in the effect size of the placebo that led to decrease in the difference (SMD) between the active treatment and placebo. These results were illustrated by the three included examples that explain the changes in the treatment differences between the active treatment and the placebo. These correlations varied from strongly positive to strongly negative. The only factor affecting the results was meta-analyses with a smaller number of trials that had extreme correlations.

The three included examples concluded that the changes in the SMD were due to the changes in both the placebo (improvement) and the active control (reduction), with the changes being more apparent in the active control than the placebo.

There has always been an argument that the improvement in the placebo effect group is due to changes in the population and the standard treatment (Kamper & Williams, 2013). However, in this review the changes were noticed even for therapeutic areas that used objective outcomes.

In the first illustrated example, the outcome measure was the total cholesterol level in the blood, and the improvement in the placebo group over time was quite clear in comparison to the atorvastatin group. That means the improvement was not due to the placebo effect only; instead it was due to changes in the adjuvant treatment and due to the regression to the mean phenome. The type of placebo included did not affect the results of the correlations.

The changes in the sample size over time were clear too. Regarding the sample size, 75.8% of the included reviews had a positive correlation between the year of publication and sample size, which means that recent studies tend to have a larger sample size than the oldest ones. These results are considered as a proof that the changes in the placebo and active treatment responses were due to regression to the mean, changes in the population and the improvement in the adjuvant treatment, but not due to the effect of the placebo itself.

The fact that larger studies are the most recent ones raises another argument regarding the type of model that should be used in a meta-analysis. A fixed effect model depends on the sample size and gives more weights for larger studies, which are usually the recent ones. While a random effects model gives more weights for the smaller (older) studies to account for any possible heterogeneity. In my opinion, in the case of NI trials, where the boundaries of the confidence interval are more important than the point estimate, using a fixed effect model may be more preferred than the random effect model since it will give more weight to the most recent studies (the larger ones). This hypothesis will be investigated in Chapter 6.

This review used published data only, which could be considered as a limitation for the generalisability of the results since published trials are usually trials with positive results (H. Rothstein et al., 2005). Moreover, in this review, 72% of the included meta-analyses have statistically significant positive results. This could increase the possibility of publication bias. However, this kind of data (published trials) is usually used in the indirect comparison situations, either in general or in the estimation of the NI margin from the historical data. That means this review is very relevant to the real situation in NI trials.

The use of the year of publication as a surrogate for the year of trial conducting could be considered as a limitation for this review. That is because the meta-analyses used hadalready been published and it was difficult to find the actual year of trial conducting, especially for the older trials. Another limitation was that some of the included meta-analyses had the smaller sample size of four trials. This could affect the reliability of the results. However, the parametric and non-parametric results were both similar.

Subdividing the correlations to strong, moderate and weak correlations could also be considered as a limitation of this review. However, the aim of this categorisation was to demonstrate the strength of the correlations regardless of their direction, since the median and mean correlations were weak in general.

Different therapeutic areas, different treatments, and different types of placebo groups were included in this review, which is considered as a strength for this review. Overall, regarding the treatment difference between the active control and placebo, 58.5% of the included meta-analyses had either a moderate or strong correlation with time. That means the constancy could be assumed in only 41.5% of the included meta-analyses. Adjusting to the sample size, improved the percentage of constancy to 44%, but constancy was still lacking in more than half of the included meta-analyses. For the placebo response, the constancy assumption held in 47.5% and this was reduced to 43.2% after controlling for the sample size. The active control effect was constant in 48.4% of the included meta-analyses and this proportion was reduced to 40% after controlling for the sample size, which means the sensitivity of the active control was not constant.

These results indicate that assuming constancy of the treatment difference between the active control and placebo and assuming that the sensitivity of the active control will not change over time will lead to a biased estimate of the treatment effect. In the case of the non-inferiority the trial, this will lead to the conclusion of the non-inferiority of an inferior test treatment.

These results highlight the importance of time changes in the case of indirect comparisons between different treatments, especially in the case of NI trials, which depend heavily on the indirect comparison between the placebo (P), and the experimental treatment (T) via the active treatment (C) assuming the constancy. In the next chapter, the magnitude of the changes in the

treatment difference will be studied using a regression model and aiming to predict the treatment difference using the available historical trials.

Chapter 6 Incorporating Time in the Estimation of the Treatment Effect Based on Historical Trials

6.1 Introduction

The main aim of the thesis is to quantify the non-inferiority margins when using retrospective data to inform the decision, and their effect on the analysis of NI trials. Chapter 5 found that the changes (reduction) in the treatment difference between the placebo and active treatment were due to improvement in the placebo response and decrease in the active treatment response over time. This chapter will investigate factors that affect the prediction of a future trial based on the available historical information using a weighted linear regression model.

The detailed objectives will be presented in Section 6.2, followed by the methods used to formulate the dataset, build and validate the regression model in Section 6.3. The results will be presented in Section 6.4, followed by the discussion and conclusion in Section 6.5.

6.2 Aim and objectives

This chapter aims to investigate factors that affect the estimate of a future trial based on the available historical trials using the weighted linear regression to predict the standardised mean difference in a trial based on the standardised mean difference from a meta-analysis of previous trials.

The objectives are

- To compare the results of the point estimate using both fixed and random models
- To assess the relationship between the point estimate (SMD) of a future trial (SMDlt) and the point estimate (SMD) of a meta-analysis of retrospective trials (SMDdl)
- To assess the relationship between the point estimate (SMD) of a future trial (SMDlt) and the characteristics of the meta-analysis of previous trials
- To build a regression model of prediction using SMDlt as the response variable
- To validate the developed model

6.3 Methods

6.3.1 Formulating the dataset

A dataset was formulated from the previously collected reviews (236 Cochrane reviews (meta-analyses) used in Chapter 5) to form a database for the analysis in this chapter. The included reviews are those with more than three trials conducted in different years after deleting the last trial(s). The excluded reviews were the reviews with only two trials remaining after removing the most recent trial (last trials), and reviews where all trials were conducted in the same year.

The database contains the original estimate of the treatment effect from this meta-analysis and 95% CI and the significance level, the SMD and its 95% CI for all trials in the meta-analysis, the calculated SMD after deleting the last trial(s) and its 95% CI, and the SMD for the last trial (s). It also includes the number of trials included in each meta-analysis, year difference between the last trials and first trials, and year difference between last trial and most recent trial after deleting the last trial. Other general information regarding the therapeutic area, active treatment, the original measure of effect used, heterogeneity, risk of bias, level of evidence, type of placebo and number of patients is included in the analysis. The unit of analysis is the meta-analysis not the trials.

For each meta-analysis in our database, three standardised mean differences were calculated:

- The SMD for all trials included in the original meta-analysis (SMD)
- The SMD for all trials included in the original meta-analysis, excluding the most recent (last) trial(s) (*SMD*dl)
- The SMD for the most recent (future) trial(s) (*SMD*lt)

Some of the included reviews had more than one last trial (most recent). There were two possible approaches to overcome this problem. The first approach was to use the most recent trial by its month of publication to determine the last trial. Other trials that were published in the most recent year but earlier in that year were included with the meta-analysis after deleting the last trials (*SMD*dl). Even with this approach, there were trials conducted in the same month or the same trials were used twice in the meta-analysis. For these reviews, a meta-analysis of these last trials was conducted, and the pooled estimate of the standardised mean differences

from all these trials was used. Using this approach, the changes in the same year could be measured.

The second approach was to calculate the SMDlt when there was more than one last trial in a review; a meta-analysis for all last trials published in the same year (the type of model used (FE or RE) was applied in accordance to the original model used in the meta-analysis of all trials). The pooled estimate of the standardised mean differences from these trials was used as the point estimate of the last trial (SMDlt).

To investigate the effect of using the random and fixed models in the estimate of the meta-analysis and its 95% CI, both random and fixed effect models were applied in addition to the main model used in the meta-analysis. According to the model used, there will be a dataset for the original model, a dataset for a fixed model and a dataset for a random model.

Six different datasets were formulated based on the analysis approach for dealing with multiple last trials and model used in the meta-analysis (Table 6.1).

Dataset 1: This dataset had SMDdl from the previous trials and for SMDlt (meta-analysis for the reviews with more than one trial) + the original model used (fixed or random).

Dataset 2 : Based on the analysis, this dataset had SMDdl from the previous trials and for SMDlt (meta-analysis for the reviews with more than one trial) + fixed effect model.

Dataset 3: Based on the analysis, this dataset had SMDdl from the previous trials and for SMDlt (meta-analysis for the reviews with more than one trial) + random effect model.

Dataset 4: This dataset had SMDdl from the previous trials and for SMDlt from only one last trial (the most recent one by month) + the original model used (fixed or random).

Dataset 5: Based on the analysis, this dataset had SMDdl from the previous trials and for SMDlt from only one last trial (the most recent one by month) + fixed effect model.

Dataset 6: Based on the analysis, this dataset had SMDdl from the previous trials and for SMDlt from only one last trial (the most recent one by month) + random effect model.

Table 6.1 Different used datasets

Data	Model Used	Type of last trial included
Dataset 1	Original Model used in the review	**SMDdl**: all trials in the meta-analysis excluding all last trials **SMDlt**: an estimate from a meta-analysis of the last trials (if more than one last trial included) and the estimate of the last trial if there is only one trial
Dataset 2	Fixed Model	
Dataset 3	Random Model	
Dataset 4	Original Model used in the review	**SMDdl**: all trials in the meta-analysis excluding only the most recent last trial **SMDlt**: an estimate from the most recent last trial
Dataset 5	Fixed Model	
Dataset 6	Random Model	

6.3.2 The effect of the model used

As mentioned in earlier chapters (Sections 2.4, 5.6) the choice between the fixed and random effects model could influence the setting of the NI margin. Although the random effects model accounts for heterogeneity, it gives more weight for smaller "older trials" compared to the fixed effect model.

To investigate the effect of using the random and fixed models on the estimate of meta-analysis and its 95% CI, both random and fixed effect models were applied in addition to the main model used in the meta-analysis. The differences between the fixed or random effects models' datasets were investigated using the Bland Altman plots as a measure of agreement between the SMD for all trials from the fixed and random datasets (Bland & Altman, 1999). The Bland Altman is the recommended measure for the comparison between two different methods (Machin, Campbell, & Walters, 2008).

6.3.3 Building the weighted regression model

For the predictive model, Dataset 1 was the most realistic dataset. In dataset one the model was the original model in the review, and the meta-analysis with more than one last trial was treated equally by using a meta-analysis to estimate the SMDlt compared to dataset four which used the month of publication as a surrogate for the chosen trials, and this is usually inaccurate. Datasets 2, 3, 5, 6 used either fixed or random models, which reduced the chance of generalisability of the predictive model. For all these reasons, dataset 1 was chosen to be the dataset for the development of the predictive model.

The main aim of this chapter is to investigate if it is possible to predict the estimate of a trial based on a meta-analysis of previous similar trials using the regression model. Regression is considered as the most frequently used method for prediction. It is considered a powerful and more flexible method (Kutner, Nachtsheim, Neter, & Li, 2005). Due to the nature of the outcome variable available from the constructed dataset, a multiple linear regression model will be the appropriate model to use to construct the predictive model.

Dataset 1 will be divided randomly using R into 75% training dataset that includes 168 meta-analyses to build a regression model and 25% test dataset that includes 56 meta-analyses to test and validate the model. A model will be developed to predict the values of SMDlt (dependent variable), using the SMDdl as the independent variable and the year of the predicted trial (Ylt), the year of last trial publication (Ydl) and year of first trial (Y1) as co-variables in the model. In addition, the year differences between the first and last year in the meta-analysis of historical trials and the year of the predicted trials will be tested as possible co-variables in the model.

The independent variable (SMDdl) used in the model is constructed from a meta-analysis of several trials, and because of that, each case in the dataset will have a different weight according to the sample size of the meta-analysis. For this reason, using weighted multiple regression (WLS) will be more appropriate than using multiple linear regression (Solon, Haider, & Wooldridge, 2015). Weighted regression will give each meta-analysis its proper amount of influence over the parameter estimate. Based on the fact that the sample size is increasing by time with the median correlation between sample size and year of publication of 0.2 (Section 5.5.4.1), the model will be weighted for the total sample size of the historical trials (Ndl).

The multiple linear regression model used is represented in the equation (6.1) (Kutner et al., 2005)

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{p-1} x_{i,p-1} + \varepsilon_i \qquad (6.1)$$

where Y is the response variable, $\beta_0, \beta_1, \beta_2 \ldots \ldots + \beta_{p-1}$ are the parameters, $x_{i+}x_{i2}, \ldots \ldots . x_{i,p-1}$ are the predictors, ε_i is the measurement error N(0. σ^2) and i= 1… n.

For the weighted least square, the coefficients of estimates can be calculated using equations (6.2) and equation (6.3)

$$\beta_i = \frac{\sum w_i \sum w_i x_i Y - \sum w_i x_i \sum w_i Y}{\sum w_i \sum w_i x_i^2 - (\sum w_i x_i)^2} \qquad (6.2)$$

$$\beta_0 = \frac{\sum w_i Y - b \sum w_i x_i}{\sum w_i} \qquad (6.3),$$

where w_i is the weight for each case (meta-analysis) in the model. The model will be weighted for the total sample size of the historical trials (Ndl).

The model adequacy will be checked by checking the assumptions of multiple regression model (Montgomery, Peck, & Vining, 2006): (1) the relation between the dependent variable and the independent variable is linear; (2) the error (ε)has zero mean with constant variance; (3) the errors (ε) are uncorrelated and normally distributed; (4) for the weighted regression, the weights must be known (Ndl). The model adequacy check will include residual analysis, a test of lack of fit, looking for high leverage and influence observation, and checking for outliers.

As a secondary objective for building the regression model, the agreement between the predictors and the observed values of the last trial estimate will be measured using a Bland-Altman plot (Bland & Altman, 1999). The results are presented in Appendix D.

All analyses were done in both SPSS (IBM Corp, 2016) and R (R Development Core Team, 2008).

6.3.4 Validation of the regression model

The model validation is necessary to check if the model will work successfully in the real working environment (Montgomery et al., 2006). A proper validation should include checking if the regression coefficients' signs and magnitude are reasonable. The stability of the regression coefficients should be investigated. Also, the prediction performance of the model should be checked (Montgomery et al., 2006). Different methods could be used for validation of the regression models, and these include bootstrapping and cross-validation.

Bootstrapping is one of the most common methods used to provide an accurate estimate, especially when the size of the sample data is considered small (Kutner et al., 2005). In this chapter, bootstrapping will be used to validate the weighted linear regression built from the training data set.

Cross-validation is one of the most common methods for validation of regression models (Kutner et al., 2005). Two approaches are available depending on the sample size of the original data. The first approach is used when rich datasets are available and involves dividing the dataset into 3 parts: training (50%), validation (25%) and test sets (25%). The model will be fitted using the training set. The validation set will be used to assess the prediction error rate. The test set is used to assess the general error of the final model. The other approach, which is used in case where the available dataset is not too large (as in the current case), is to divide the data into a training set (75%) and test set (25%), then the model will be built using the training set and cross-validated and tested using the test dataset.

There are other methods for cross-validation, including leave one out cross validation (LOOCV), K-fold cross-validation or the repeated K-fold cross-validation. However, interpretation of the results should be done with caution since the root mean squared error (RMSE) calculated from these methods tends to be higher in the case of weighted regression (Kutner et al., 2005). The dataset used for model prediction comprises only 224 meta-analyses. On that basis, the decision was made to cross-validate using a training and test dataset and the bootstrapping only. The other methods for cross-validation are presented in Appendix D.

6.4 Results

6.4.1 Characteristics of the included meta-analyses (whole dataset)

Out of the 236 meta-analyses included in the previous analysis, only 224 were included in the final analysis. Twelve meta-analyses were excluded from the analysis.

- Six were excluded because the remaining trials after deleting the last were conducted in the same year, and that meant there would be no year difference between the trials.
- In five meta-analyses only two trials remained after deleting last trials, and since the aim was to include at least 3 trials in the meta-analysis, it was decided to remove these trials.
- In one meta-analysis, the three remaining trials were conducted in the same year, so there was no treatment difference between the trials.

It was identified that 172 (76.8%) meta-analyses had only one trial as last trials; 34 (15.2%) had two trials as last trials; ten (4.5%) had three last trials; six (2.7%) had four trials as last trials; and two meta-analyses had six trials as last trials. After choosing the most recent trial from the last trials, 218 meta-analyses had one last trial and six reviews had two trials described as last trials (Figure 6.1).

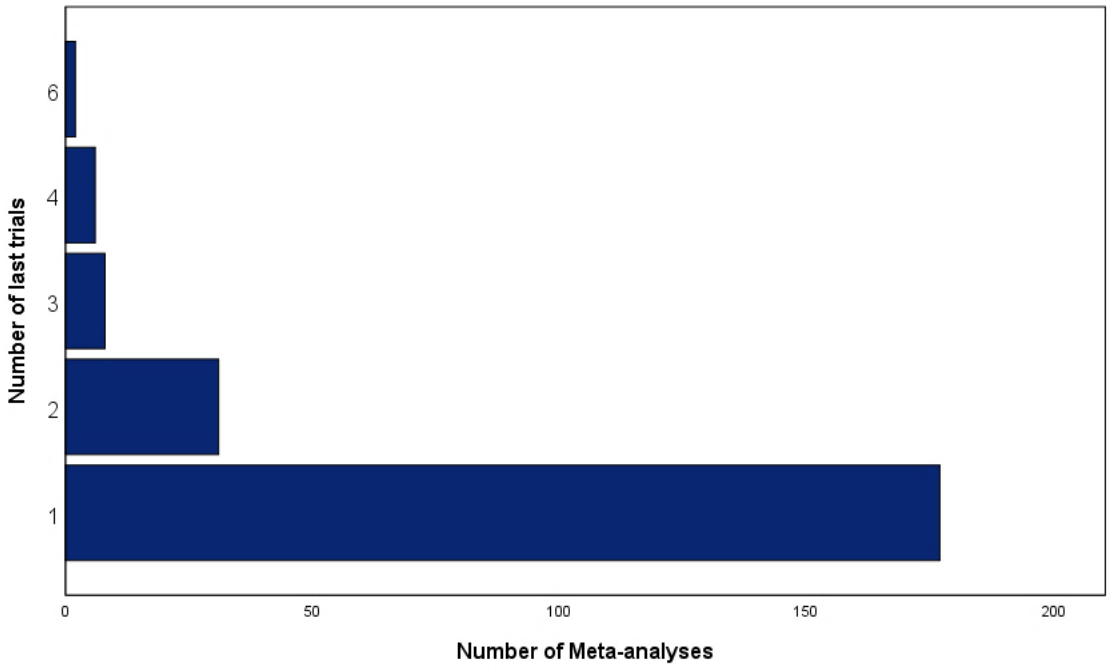

Figure 6-1 Number of trials

From 224 reviews included, 111 reviews were published in 2016, 113 reviews in 2015. The total number of trials included in each of the meta-analyses ranged between 4 and 51 trials, with a mean number of 10.31 trials, SD = 7.5, and a median of 8 trials. The total sample size ranged from 105 to 43290 patients, the median being 1244 patients and IQR (526-2251).

The year of publication ranged from 1931 to 2016 with the year difference between the oldest and the most recent trials varying from two years' difference to 80 years' difference. The difference between the last trial and trial before it ranged from one to 24 years.

Risk ratio was the measure of effect in 125 (55.8%) of the reviews, the mean difference was used as a measure of effect in 72 (32.1%) of the reviews, 24 (10.7%) reviews used the odds ratio and four (1.3%) reviews used risk difference as the measure of effect. A fixed effect model was used in 96 reviews and random effect model in 128 reviews (Figure 6.2). Heterogeneity was statistically significantly higher when the random effect model was used, with mean I^2= 40.48% compared to 23.7% for fixed effect models. There was no difference in the number of trials included in the analysis between fixed and random models, with the mean number of trials 10 and 11 trials, respectively.

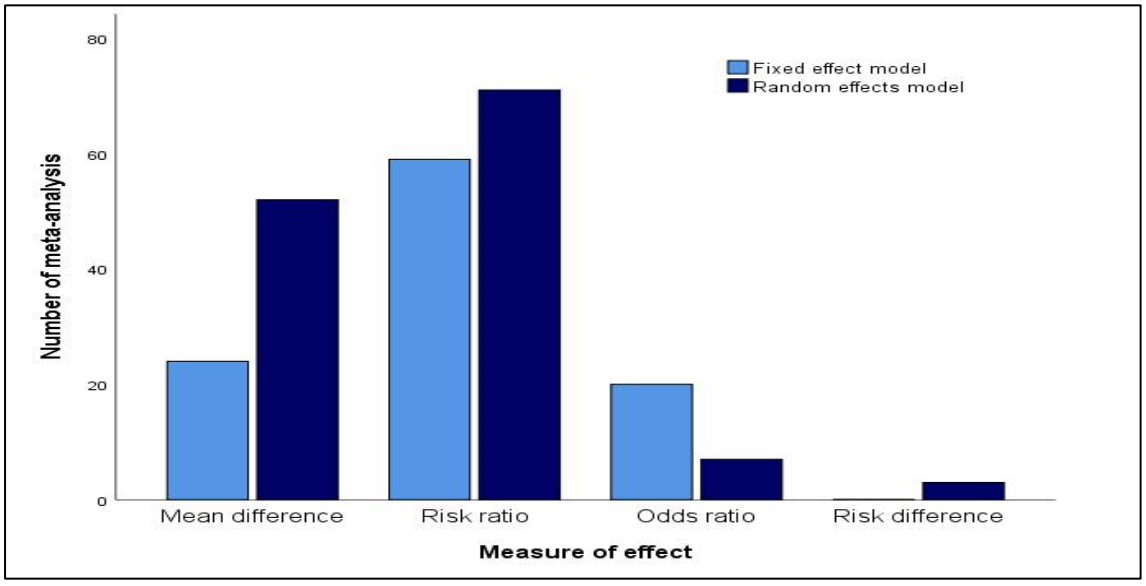

Figure 6-2 Type of model used according to the measure of effect

Regarding the placebo type, 92 (41.1%) of the reviews defined the control group as (placebo) only, while 56 (25%) reviews used (Npno treatment or placebo) as the control group. Forty-two (18.8%) of the reviews defined the control group as (placebo, usual care or no treatment) and 15 (6.7%) defined it as (usual care or placebo), nine (4%) reviews as usual care, eight (3.6%) as no treatment, and in two reviews no treatment or usual care was used.

Regarding the risk of bias, 32 (14.3%) of the reviews had low risk of bias, 125 (55.8%) had moderate risk of bias, 47 (21%) had high risk of bias and in 20 (8.9%) of the reviews the risk of bias was described as unclear. Regarding the quality of evidence, the evidence was very low quality in 19 (8.5%), low in 80 (35.7%), moderate in 81 (36.2%) and high quality in 44 (19.6%) (Figure 6.3).

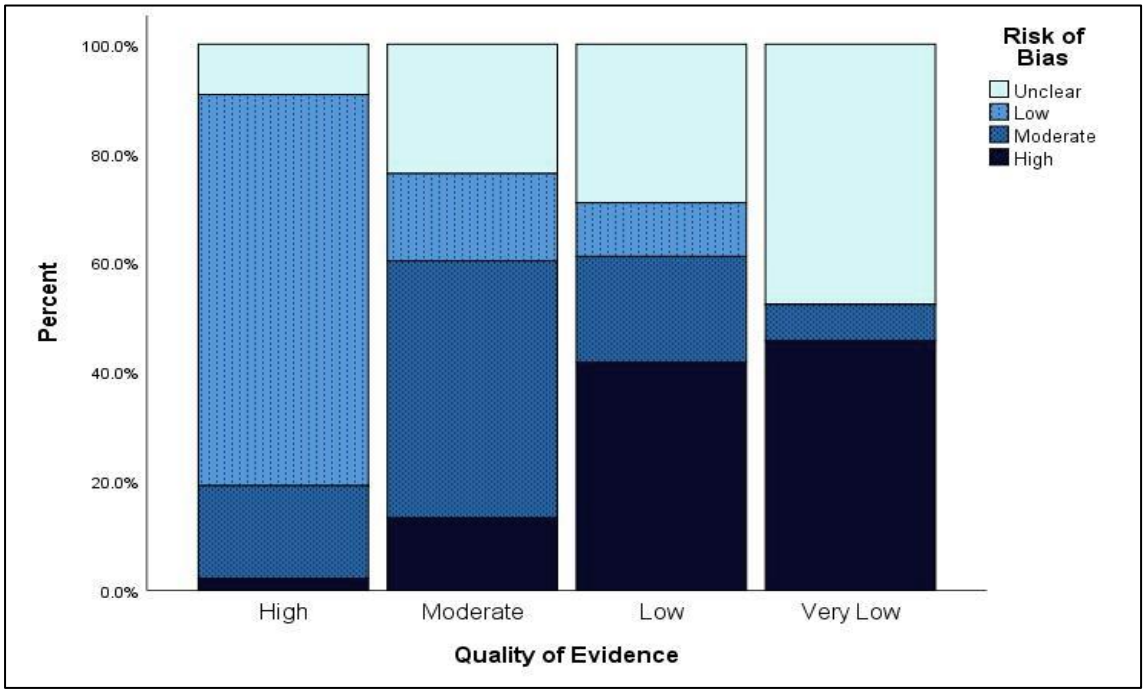

Figure 6-3 Quality of Evidence and Risk of Bias

Table 6.2 Mean and Standard deviation for SMD, SMDdl, SMDlt for six different datasets

	SMD (mean, SD)	SMDdl (mean, SD)	SMDlt (mean, SD)
Dataset 1	-0.059 (0.580)	-0.062 (0.600)	-0.078 (0.770)
Dataset 2(fixed)	-0.052 (0.550)	-0.055 (0.590)	-0.072 (0.750)
Dataset 3 (random)	-0.063 (0.600)	-0.06 (0.620)	-0.067 (0.820)
Dataset 4	-0.059 (0.580)	-0.061 (0.60)	-0.067 (0.820)
Dataset 5(fixed)	-0.052 (0.550)	-0.053 (0.580)	-0.067 (0.820)
Dataset 6(random)	-0.063 (0.600)	-0.063 (0.620)	-0.067 (0.820)

SD= standard deviation, SMD: the standardised mean difference of all trials, SMDdl: a standardised mean difference of all trials after deleting the last, SMDlt: a standardised mean difference of the last trial (s).

Regarding the differences in the estimate from the different datasets extracted, there was no statistically significant difference between these datasets (Table 6.2). The mean SMD for the total trials using method one or method two for data extraction (mentioned earlier in 6.4.1) was similar for the original model and using either fixed or random model. For the comparison between the different models, (fixed and random) datasets one, two, and three will be used. For the model prediction, dataset one will be used. From Table (6.2), the estimate from the last trial is smaller than the estimate of the previous trials regardless of the type of model used or the type of last trial extraction. Moreover, the point estimates were higher than both the SMDdl and SMDlt. The estimate from the fixed models was higher (closer to zero) than the estimate from the random models for SMD, SMDdl and SMDlt.

Regarding the Cochrane group, in total, 42 different groups were included in the analysis. Twenty meta-analyses (8.9%) were from the pregnancy and childbirth group, 19 (8.4) were from the pain and palliative supportive group, 14 (6.2%) from the gynaecology group, 13(5.8 %) from the heart group, while the other groups varied between 11 and one meta-analyses. Concerning the therapeutic area, 33 (14.7%) meta-analyses were cardiovascular, neurology and obstetrics and gynaecology accounted for 26 (11.6%) each, 21 (9.3%) were infections, 20 (8.9%) were psychiatry, 19 (8.4%) were gastroenterology. Additionally, anaesthesia accounted for 18(8%) meta-analyses, nutrition accounted for 12 (5.3%), respiratory and urology accounted for 10 (4.4%) meta-analyses each, 8 (3.6%) were orthopaedics and 22 (9.8%) of the meta-analyses were distributed in other therapeutic areas.

6.4.2 The effect of type of model on the estimate of SMD

Paired sample t-test was used to test if the differences between these the two models were statistically significant or not. In general, there was no statistically significant difference in the SMD estimated from the fixed or random model with the mean SMD for the fixed models -0.05, 95% CI [-0.13; 0.021] compared to -0.06, 95% CI [-0.14; 0.017] for the random models with mean difference between the two models -0.01 and 95% CI [-0.044; 0.025]. However, there was a statistically significant difference in the length of the 95% CI, with mean width for the fixed model =0.89, 95% CI [0.79; 0.98] and the random model = 0.98 with 95% CI [0.88; 1.08], the mean difference in the length between the random and fixed model was 0.09, 95% CI [0.064; 0.14].

Figure (6.4) illustrates the differences in frequency distribution for SMD between the fixed and random model. The distribution of the random effects model was wider than that of the fixed effect model; this reflects the wider confidence interval for the random model compared to the fixed model. The point estimate from both models was similar.

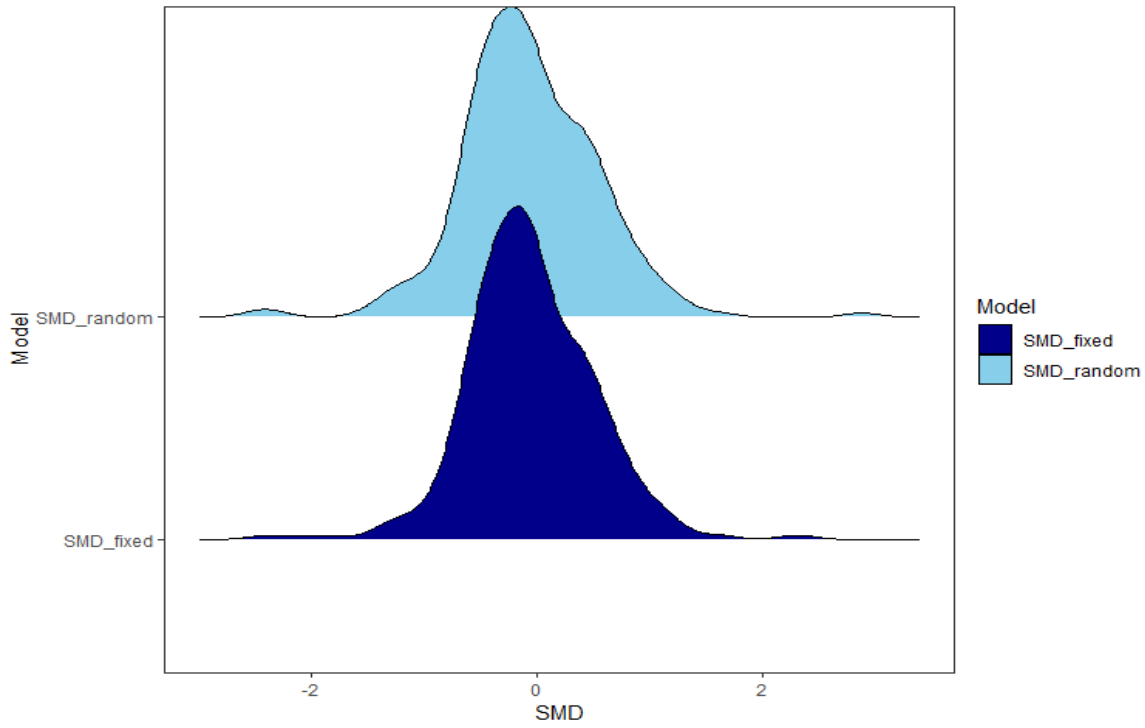

Figure 6-4 Comparison between the frequency distribution for SMD using fixed and random models

The Bland-Altman plot was used to assess the agreement for the point estimates and the 95% CI boundaries to test the agreement between the fixed and random model. Figure (6.5) represents the Bland Altman plot for the agreement between the fixed and random models for the point estimate, the width of the 95 % CI and the upper and the lower boundaries of the 95 % CI in accordance to the heterogeneity and the sample size (N). Figure (6.6) represents the Bland Altman plot for the agreement between the fixed and random models for the point estimate, the width of the 95 % CI and the upper and the lower boundaries of the 95 % CI in accordance to the heterogeneity and the total number of trials included.

For the point estimate, the average of difference (bias) = -0.01, SD =0.1 and the limit of agreement (LOA)was- 0.23; 0.21. For the width of the 95% CI the absolute distance between the lower and upper limits of the confidence interval was used to measure the width of the 95% CI From the graph, , the average of difference (bias) = -0.09, SD =0.2 and the limit of agreement was - 0.5; 0.31.

For the agreement between the 95% CI boundaries of the fixed and random model, regarding the upper boundaries of the 95% CI the average difference (bias) = -0.07, SD =0.17 and the limit of agreement was - 0.41; 0.25. For the lower limits of the 95% CI, the average difference (bias) = 0.09, SD =0.17 and the limit of agreement was - 0.24; 0.43.

From the figures (6.5 and 6.6), there was good agreement between the point estimates of the fixed and random model, with the smaller average difference of -0.01 and narrower limit of agreement of (- 0.23; 0.21). The agreement was less for the boundaries of the 95 % CI and the width of the 95% CI, with the averages of the difference of the upper, lower and width being 0.07, 0.09 and -0.09 respectively and the LOA being wider.

For the point estimate, there is a clear agreement between the fixed or random model in the case of the meta-analysis with mild heterogeneity, larger sample size, and a larger number of trials, this conclusion based on the graphs only. When the heterogeneity of the meta-analysis increases, the agreement is less apparent, and the difference between the two models is statistically significant. These results support the finding from the literature that the random effect model accounts for heterogeneity but cannot explain it, while the fixed effect model does not account for heterogeneity (Cooper, Sutton, Morris, Ades, & Welton, 2009; DerSimonian & Kacker, 2007).

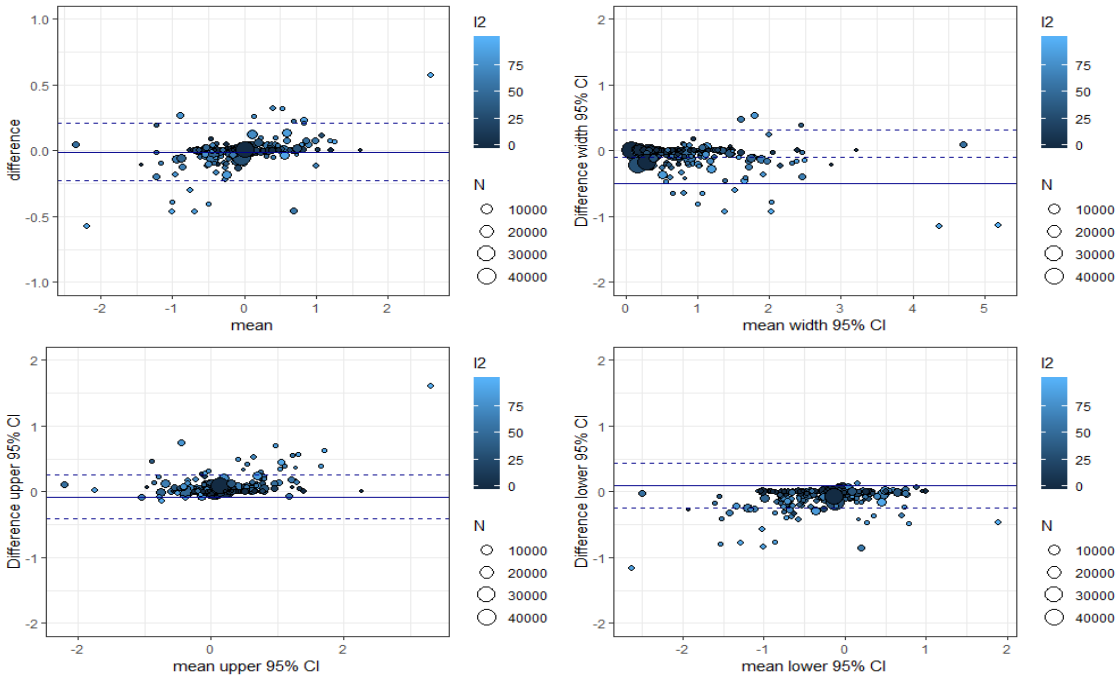

Figure 6-5 Bland-Altman graph for agreement between random and fixed models
(*According to the heterogeneity of the model, the size of the bubble represents the sample size in the meta-analysis)*

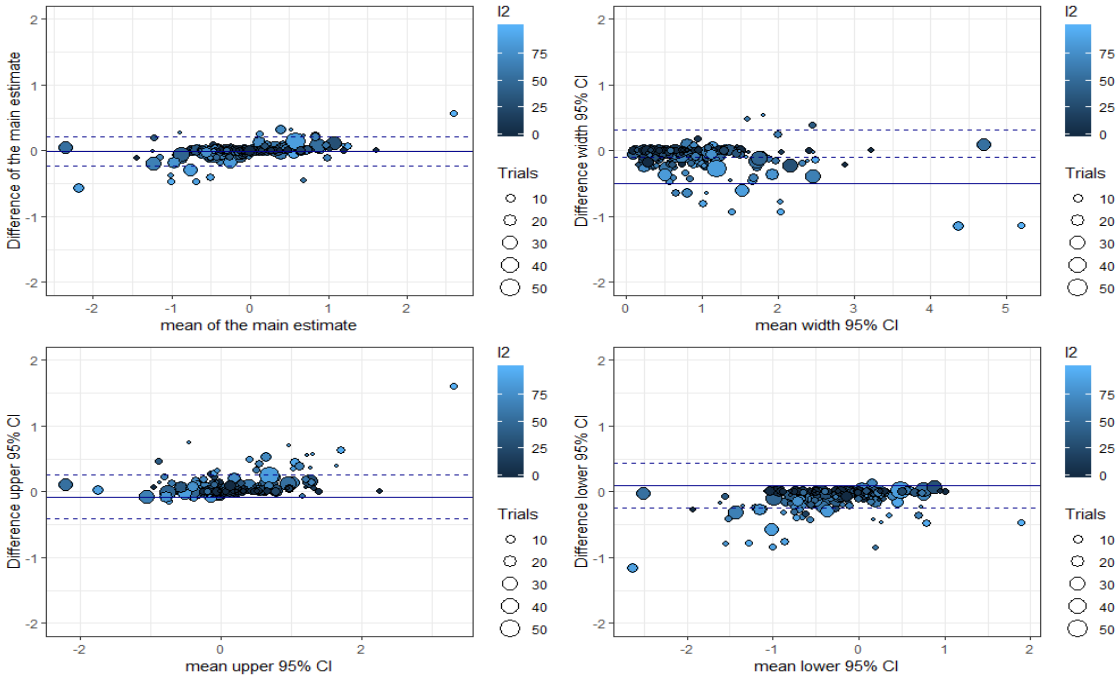

Figure 6-6 Bland-Altman plot for agreement between random and fixed model
According to the heterogeneity, the size of the bubble reflects the number of trials in the meta-analysis

6.4.3 Prediction of the SMD of the last trial from a meta-analysis of previous trials

6.4.3.1 Description of data used for analysis (training data set)

The dataset of 224 meta-analyses was randomly divided into two datasets using R sampling process: the training dataset that included 75% of the whole dataset (168 meta-analyses) and the test dataset that contained 56 meta-analyses. Regarding the outcome variable, the SMD of the predicted trial (SMDlt), the mean (SD) = -.11 (0.75), the Min, Max = (-3.61; 2.28). Regarding the independent variable, the SMD from the historical trials (SMDdl), the mean (SD) = -0.07 (056), Min, Max= (-2.39; 1.3).

Table (6.3) represents the differences between the three datasets in regard to the variables that will be used in building and testing the regression model.

Table 6.3 Differences between the whole, training and test datasets used for analysis

Comparison		Whole dataset	Training data	Test data
Number of reviews		224.00	168.00	56.00
SMDdl	Mean (SD)	-0.06 (0.60)	-0.07 (.56)	-0.03 (0.70)
	Median	-0.11	-0.16	-0.04
	Min; Max	-2.39; 3.25	-2.39; 1.30	- 1.31; 3.25
SMDlt	Mean (SD)	- 0.07 (0.77)	- 0.11 (0.75)	0.04 (0.80)
	Median	- 0.05	- 0.08	0.04
	Min; Max	- 3.61; 2.46	- 3.61; 2.28	-1.31; 3.25
Sample size (Ndl)	Mean (SD)	2389.00(4879.00)	2166.00 (4289)	3057.00(6330.00)
	Median	934.00	1002.00	842.00
	Min; Max	67.00; 38862.00	67.00; 34996.00	111.00; 3886.00
Number of trials (Kdl)	Mean (SD)	9.46 (8.10)	9.50 (8.63)	9.16 (6.40)
	Median	7.00	7.00	6.05
	Min; Max	3.00; 50.00	3.00; 50.00	3.00;28.00
Y meta	Mean (SD)	15.53 (11.30)	15.6 (11.90)	15.25 (9.30)
	Median	13.00	12.50	13.00
	Min; Max	1.00; 75.00	1.00; 75.00	3.00; 46.00
Y2	Mean (SD)	19.10 (11.50)	19.13 (12.13)	19 (9.70)
	Median	16.00	15.00	17.00
	Min; Max	1.00; 75.00	2.00; 80.00	5.00; 48.00
Y3	Mean (SD)	3.57 (3.50)	3.50 (3.75)	3.80(9.30)
	Median	2.00	2.00	3.00
	Min; Max	1.00; 24.00	1.00; 24.00	1.00; 3.00
First year	Min; Max	1931; 2013	1931; 2013	1966; 2007
Recent year	Min; Max	1965; 2015	1965; 2015	1982; 2015
Predicted year	Min; Max	1977;2016	1977; 2016	1990; 2016

***SMDdl**: SMD from the historical trials, **SMDlt**: SMD for the predicted trial, **Ndl**: sample size of the meta-analysis, **Kdl**: number of trials included in meta-analysis, **Ymeta**: year difference in meta-analysis, **Y2**: year difference between the first trial and the predicted trial, **Y3**: year difference between last trial in the meta-analysis and the year of predicted trial*

6.4.3.2 Building the regression model

The sample size of 224 was considered small. For that reason, to obtain a valid model across validation was used to develop and test the model. The developed model was based on 75% of the data as a training dataset and test and validated on the remaining 25% of the data. The results of the final training model included 168 meta-analyses that were randomly selected, and the final tested model contained the remaining 56 meta-analyses (Table 6.3).

The mean (SD) year difference in the historical meta-analysis was 15.6 (12.5) years with a minimum of one and maximum of 75 years. The mean (SD) number of trials in the meta-analysis was 9.5 (8.6) and the median was seven trials with a minimum of three trials and maximum of 50 trials. Regarding the sample size for the meta-analysis, the mean number of participants was 2,166 (4,289), the median was 1,002, the minimum number was 67 and the maximum number was 34,996.

A weighted multiple regression model was built using the 75% training dataset that was constructed randomly from the full dataset to test if the SMD of the last trial (SMDlt) could be predicted from the SMD of the previous meta-analysis (SMDdl) and what changes there would be in the SMDdl. The model included the SMD of the last trial as the dependent variable (SMDlt), SMD from the previous meta-analysis as the independent (predictor) variable (SMDdl). The covariates tested in the model were the year difference between the last trial and the oldest trial in the meta-analysis of previous trials (Y1), year difference between predicted trial and the trials before (Y2), and the year difference in the meta-analysis of the previous trials (Y3). Additionally, the year of publication of predicted trial (Ylt), the year of the first and last trial in the meta-analysis of the previous historical trials (Y1st, Ydl) and the number of trials in the previous meta-analysis (Kdl). The model was weighted by the sample size of the previous meta-analysis (Ndl). Stepwise regression was used, and only variables that statistically significantly affected the SMDlt were presented in the final model. Table 6.4 illustrates the results of the weighted regression model using the training dataset.

The final fitted regression model was

$$Y(SMDlt) = 30.32 + 0.92 \times (SMDdl) - 0.005 \times (Ymeta) - 0.015 \times (Ylt) \qquad (6.4)$$

Table 6.4 Summary of the results of the regression model

Model	B	Std. Error	Beta	t	Significance	95% CI of β
(Constant)	30.320	10.689		2.830	0.005	(9.210; 51.420
SMDdl	0.920	0.065	0.747	14.236	< 0.001	(0.792; 1.047)
Ymeta	-0.005	0.002	-0.116	-2.223	0.028	(-0.009 ;-0.001)
Ylt	-0.015	0.005	-0.149	-2.833	0.005	(- 0.026; - 0.005)

Weighted Least Squares Regression weighted by the sample size of the historical meta-analyses. Dependent Variable: ***SMDlt;*** *Standardised mean difference of predicted trial,* ***Ymeta****: year difference between first and last trials in the meta-analysis,* ***SMDdl****= Standardised mean difference of historical trials,* ***Ylt****= year of publication of the predicted trial*

The model indicated that the SMD from the meta-analysis of the previous trials (SMDdl), year difference in the historical meta-analysis and the year of the predicted trial (Ylt) explains only 55.1% of the variance in the model (Adjusted R^2= 0.551, F statistics for overall significance = 69.175, p-value <0.0001). The point estimate of the meta-analysis of historical trials (SMDdl) statistically significantly predicted the SMD of the future trial (β = 0.92, 95% CI (0.79; 1.05)), for each unit increase in SMDdl the SMDlt increased by 0.92 units (i.e. the SMD of any future trial will be 0.92 of the point estimate of previous trials after controlling for the other variables). The year difference between the oldest and most recent trial in the meta-analysis of previously conducted trials affected the SMD of the future trial. For every one year increase in the difference in the meta-analysis, the SMD of the future trial will be decreased by -0.005 (β = -0.005, P-value = 0.028). The year of the predicted trial (Ylt) statistically significantly affected the estimate of the future trial (SMDlt) (β= - 0.015, P-value = 0.005). For each year increase in the future trial, SMD will be reduced by 0.015.

A similar model was built using the whole dataset (for comparison). The results of the whole dataset can be presented as

$$Y(SMDlt) = 36.140 + 0.881 \times (SMDdl) - 0.008 \times (Ymeta) - 0.018 \times (Ylt) \quad (6.5)$$

The detailed model is presented in Appendix D.

6.4.3.3 Checking the regression model adequacy

Figure 6.7 and Figure 6.8 illustrate the diagnostic plot for the regression model. There were no possible outliers in the model. The residuals were normally distributed. Regarding the collinearity, both the variance inflation factor and the tolerance level were low, and that indicates multicollinearity was not a concern in the model. The assumption of independent errors was met when the Durbin-Watson value was 2.2 (Kutner et al., 2005).

The scatter plot of predicted values against the residuals showed that the data met the assumptions of homogeneity of variance and linearity. There was a random pattern in the plot for the predicted and the residual values.

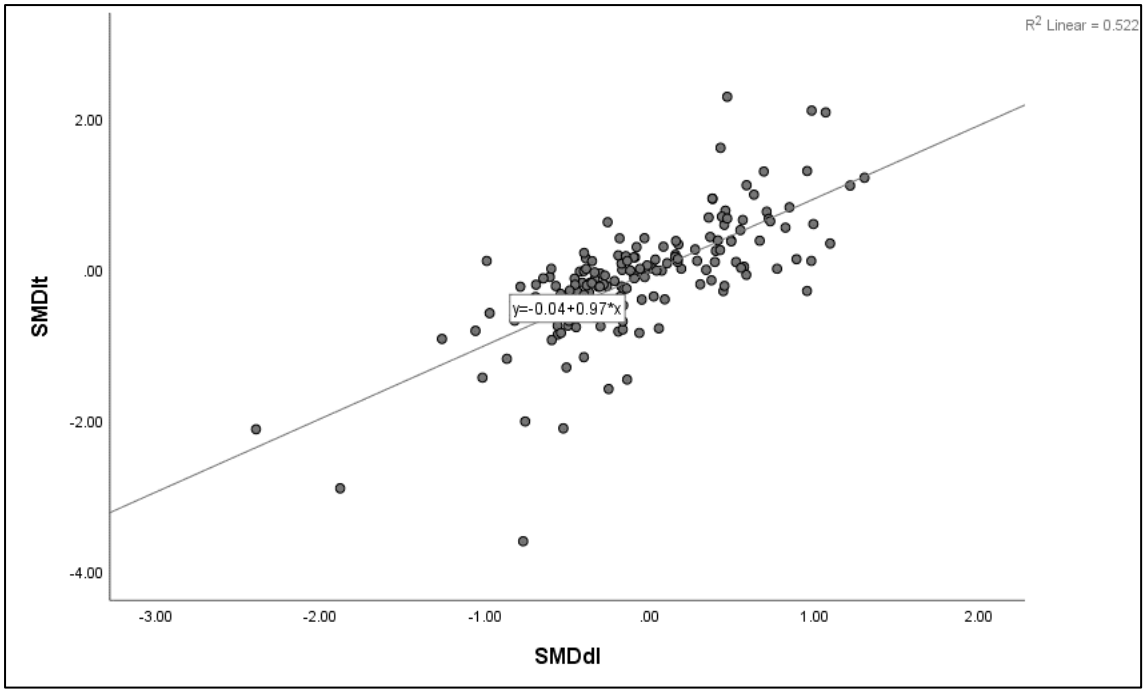

Figure 6-7 Scatter plot between the independent variable (SMDdl) and the predicted variable (SMDlt)

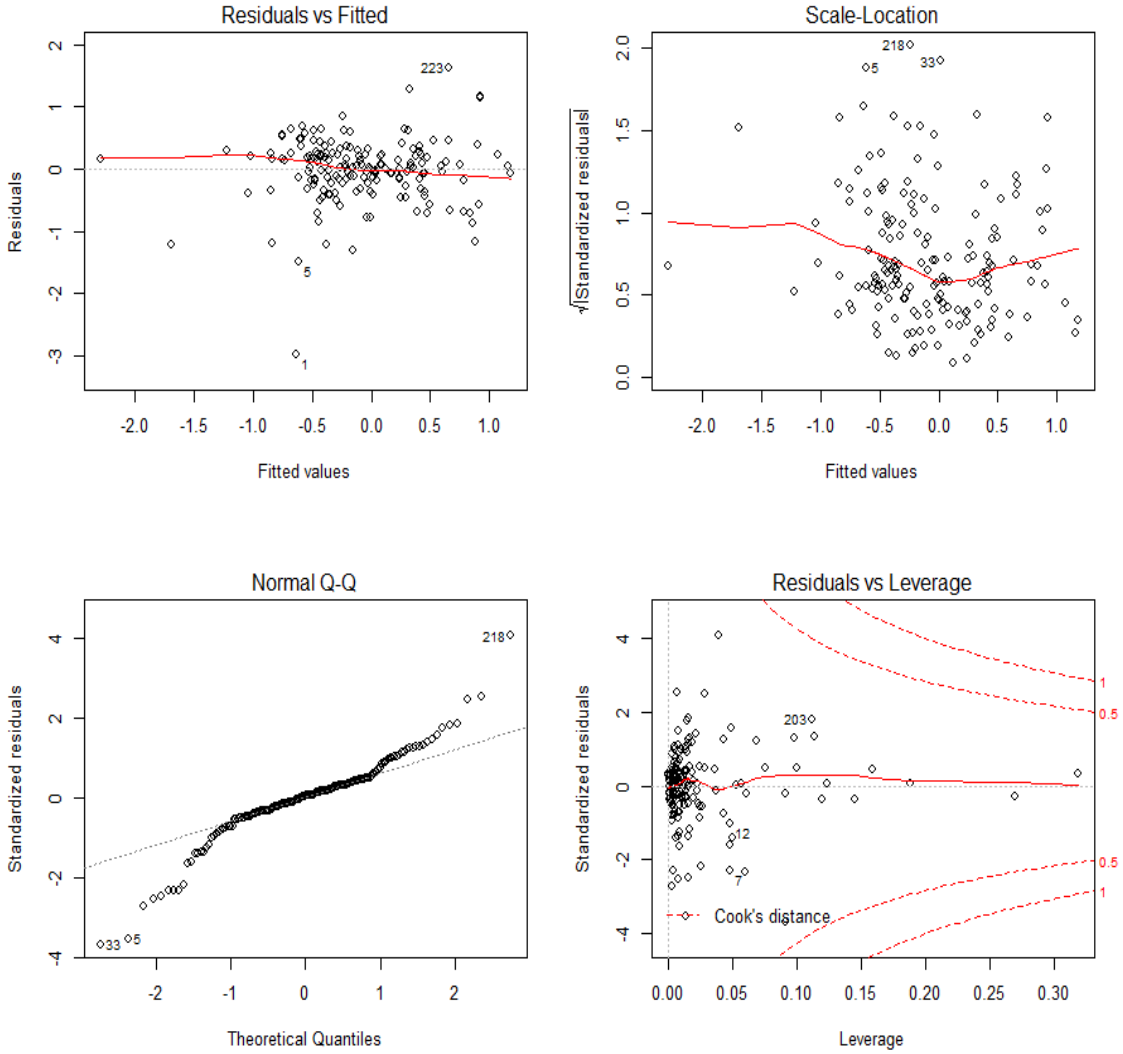

Figure 6-8 Diagnostic plots for the fitted regression model using the training dataset

6.4.3.4 Regression model validation

6.4.3.5 Bootstrapping of the regression model

Bootstrapping was used as a method of validation in the training dataset. From the sample data 1000 repeated samples were drawn, and the results of this bootstrapping are given in Table (6.5) According to the table below, the level of bias in the included variables was very low. The biased corrected 95% CIs were similar to the original model except for the year difference that had a borderline p-value of 0.054.

Table 6.5 Bootstrapping for the regression model of the training dataset

	Bootstrapping				
					BCa 95% CI
	B	Bias	Std. Error	Sig.(2-tailed)	
Constant	30.316	-1.066	11.911	.005	(7.988; 49.323)
SMDdl	.920	.003	.077	.001	(0.760; 1.075)
Ylt	-.015	.001	.006	.005	(-0.028; -0.001)
Ymeta	-.005	.000	.003	.054	(-0.011; -0.001)

*Bootstrap results are based on 1000 bootstrap samples, **SMDdl**: Mean standardised difference from the meta-analysis, **Yet**: the year of predicted trial, **Ymeta**: year difference in the meta-analysis, **BCa 95% CI**: bias-corrected and accelerated 95% CI*

6.4.3.6 Regression model validation

As mentioned earlier, 75% of the data was used to build the model (training data). The next step will be to test and validate the model on the remaining 25% of the data (test data). The developed model from the training set was used to predict the SMDlt in the test data set. The predictive model from the training set was given in equation (6.4)

This model (Equation 6.4) was used to predict SMDlt from the test dataset. The correlation between the predicted values and the observed values of the SMDlt in the test dataset was 0.7, which is a strong positive correlation (Figure 6.9). $R^2 = 0.4$, which means that the model explained 40% of the variation in the test data. The Root Mean Squared Error (RMSE) was 0.6, and the Mean Absolute Error (MAE) was 0.4. Both of these are considered low, and that means the performance of the model is adequate.

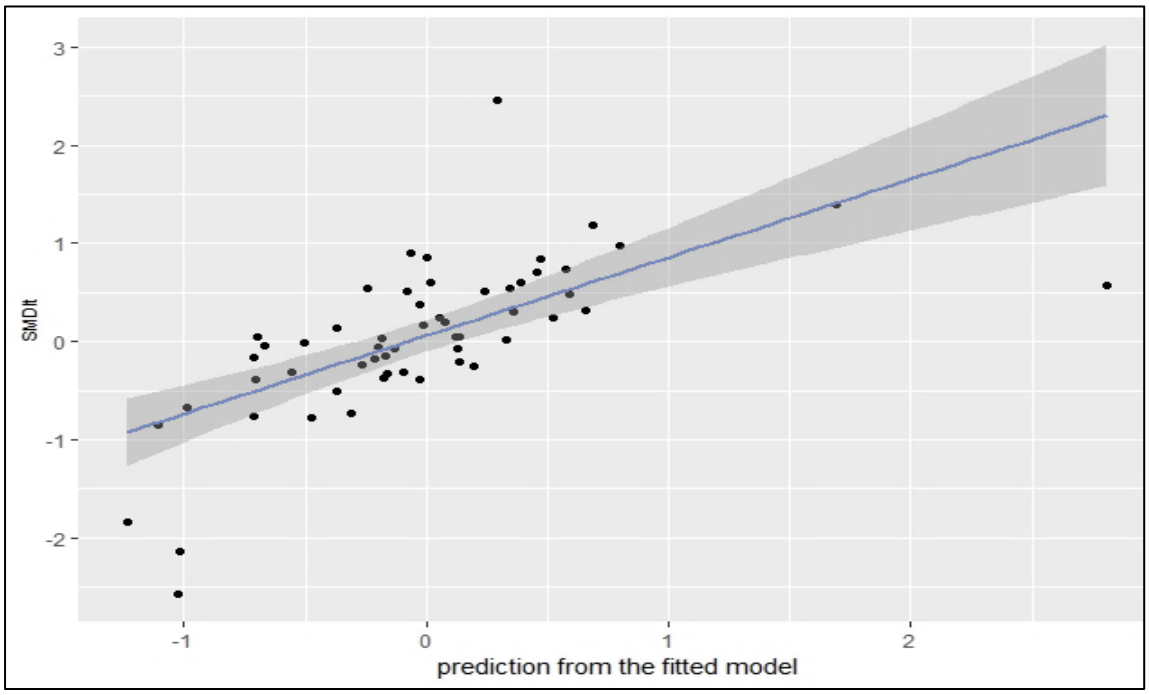

Figure 6-9 Scatter plot between the predicted values using the training model and SMDlt from testing dataset

6.5 Summary and discussion

6.5.1 Main findings and interpretation

This chapter investigated the effect of using either a fixed or random model in the final estimate of the meta-analyses. Using the data set from Chapter 5, the meta-analyses were examined using both random and fixed models to compare the results. There was no statistically significant difference between the point estimates in the meta-analyses using a fixed or random model; however, the 95% CI was statistically significantly wider using random effect models compared to the fixed effect model. Using the Bland Altman plot to measure the agreement between the two models obtained a high degree of agreement between the two models in regard to the point estimate; this agreement was less in the case of the 95% CI boundaries and the width of the 95% CI.

Heterogeneity with the meta analysis, the number of trials and the total sample size of the meta-analysis all had effects on the agreement between the fixed and random model; meta-analyses with higher heterogeneity, small overall sample size and a small number of trials showed less agreement between the estimates from the two models. This may be due to smaller studies having higher heterogeneity when compared to larger ones (IntHout, Ioannidis, Borm, & Goeman, 2015).

In NI trials, the NI margin was formulated using the boundaries, not the point estimate. Thus, setting the NI margin based on the fixed or random model could lead to different margins, especially for a meta-analysis with a smaller number of trials and smaller sample size. The random effects model will use a wider 95% CI with extreme boundaries, which could lead to the conclusion of non-inferiority of an inferior treatment. Moreover, the results in Chapter 5 of this thesis indicated that smaller studies tend to be earlier studies. Thus, in the context of this thesis with respect to NI trials, it can be concluded that using a random-effects model to estimate the effect of active control compared to placebo will give more weight to the older heterogeneous small studies that could lead to a biased estimate of the effect of the active control over placebo. For both these reasons, the recommendation is to use a fixed effect model for the situation where this is going to inform a determination of non-inferiority limit consequence by indirect comparison.

The primary aim of this chapter was to build a model to predict the point estimate of a trial based on previously available trials. The model was built using data from 2310 trials from 224 meta-analyses of placebo-controlled trials from the different medical fields and multiple linear regression that was weighted for the sample size of the meta-analyses. The main predictor for the point estimate of a trial was the point estimate of previous trials; year differences in the meta-analysis and the year of the predicted trial were the other co-variables in the model.

The three main variables that affect the estimate of any future trial were the estimate from the meta-analysis of previous trials, the year difference in the meta-analysis and the year of the predicted trial. Increase of one unit in the point estimate of the historical meta-analysis will lead to an increase in the predicted estimate of the future trial by 0.92. For the year difference in the meta-analysis, for each year increase in difference, the predicated estimate will be reduced by 0.005. For the year of prediction, for each year increase in the prediction, the predicted estimate will be reduced by 0.015. The model created in this chapter takes into

consideration both the estimate from previous trials and the year differences between the trials, and that will lead to a more accurate estimate than using the results from the regular meta-analysis.

The results from this model support the results from Chapter 5 and other literature (Ioannidis & Lau, 2001) that concludes the treatment effect is not constant over time. These results highlight an important issue of the bias that could arise from using the estimate of historical meta-analysis for indirect comparison without any further adjustment. It also highlights the need for the most appropriate estimate of effect and raises the question of whether using the results from the most recent, more extensive, trials would be better for estimating the real treatment effect than the overall results, especially in the context of this thesis (Borzak & Ridker, 1995).

In summary, the estimate from the meta-analysis of the historical trial can explain only 55% of the estimate of the future even after adjusting for the time and the year difference in the meta-analysis. However, it is an excellent predictor of the estimate of any future trial. Using meta-analysis of retrospective trials to predict the estimate of future trial or using it in indirect comparison will introduce bias since, as mentioned earlier, its predictive power is only 55%. According to our results, the difference between the estimate from a meta-analysis of historical trials and the predicted trial will be in the range of (0.79; 1.047). This difference is affected by the year of publication of the future trial and the year difference in the meta-analysis of historical trials. For each year increase, the prediction of the estimate will be reduced by 0.015 and for each year increase in the difference between the years of the first and last trials in the meta-analysis the prediction will be decreased by 0.005. These differences could be seen as small differences even though they arestatistically significant. However, the fact that a standardised difference was used in the range from (-3, 3) indicates that these differences are considered moderate changes and should be considered.

In the case of NI trials, usually the estimate from the historical placebo-controlled trials is used as the estimate of the treatment difference between the putative placebo and the active control in the NI trial, with a degree of adjustment in the case of the fixed margin approach by using the upper or the lower limit of the 95% CI. However, according to the results of this chpater, the used estimate is a biased estimate and does not reflect the actual efficacy of the active control compared to the putative placebo in the NI trial, since the actual estimate will be

different by 0.92 (0.79; 1.047). Moreover, the conducting year of NI trials and the time difference in the meta-analysis will also affect the predictor of the estimate in the predicted trial.

6.5.2 Strengths and limitations

The main strength of the model developed in this chapter is that it comes from different therapeutic areas and could be viewed as a general model. Moreover, it takes into consideration the year difference between the oldest and the planned trial (where the placebo effect needs to be predicted) and the year of the planned trial.

The main limitation of the model is the use of transformed data. Transformation of the point estimate from odds or risk ratio to the standardised mean difference could lead to a reduction in the power of prediction for this model, even though our estimate will contribute to reducing the bias in the case of NI trials.

Splitting the dataset into training and test datasets was used for model validation. This method was chosen due to the relatively small sample size. However, other methods for validation, including leave one out cross validation and K-fold cross validation, were used and are presented in Appendix D2. Some points were considered as posssible outliers; however, removing these points from the analysis did not change the results of the regression model and for that reason the decision to keep these points was taken.

Trials with positive findings tend to be published more frequently than trials with no findings (Rothstein et al., 2006). For the dataset used in this chapter, the risk of publication bias was higher since only published data (meta-analyses) were used in the final model. In addition, trials with negative findings tend to be published later than trials with positive findings (Rothstein et al., 2006) and using the year of trial publication as a surrogate for the trial conduct year could impact the results.

6.5.3 Implications for the thesis aims and objectives

In NI trials, historical trials are usually used to set the NI margin, using the upper or lower part of the 95% CI in the case of a fixed margin effect. Two conclusions can be drawn from this chapter in regard to NI trials. First, the fixed effect model is the recommended model to use in

the case of NI trials since its 95% CI is narrower and it gives less weight for smaller older studies. Second, in this chapter, it was found that the treatment effect in a future trial would range from 0.79 to 1.047 of the treatment effect of historical trials. That means, in NI trials, the constancy should not be assumed; instead, it should be assessed and according to that the NI margin should be formed.

In the next chapter, a review of the possible methods that can be used to assess the constancy and adjust for a time in the setting of the NI margin will be conducted, followed by the application of the chosen methods on two case studies of NI trials.

Chapter 7 Methods for Adjusting for Time in Non-Inferiority Trials

7.1 Introduction

Reviewing the literature in Chapter 2 and the review of the NI trials in Chapter 4 resulted in the conclusion of the importance of using the historical information in setting the NI margin for indirect comparison, with more than 50% of reviewed trials depending on the historical trials in setting the NI margin.

Chapter 2 highlighted the assumptions regarding the NI trials that include the constancy assumption and assay sensitivity of the active control and controlling for the placebo effect. Violation of these assumptions will lead to a biased estimate from the NI trials that could lead to a conclusion of the non-inferiority of an inferior treatment. The changes in the effect sizes of the placebo and the active control and treatment difference were studied in Chapters 5 and 6. All of these results confirm the importance of incorporating time changes in setting the NI margin.

In the context of this thesis, the aim of this chapter is to develop criteria to select a method to set an adjusted NI margin for time changes. To review the available methods to set and adjust for a covariate (time) in the NI trials with choosing the appropriate methods for adjusting for the time in indirect comparisons.

Criteria for the best performance will be formulated based on the literature review in Chapters 2 and 3 and the results of Chapters 5 and 6. These criteria will be applied to the different methods. In the following chapter, the chosen method will be applied to data from the placebo-controlled reviews from Chapter 5 to set an adjusted NI margin.

7.2 Criteria for a good performance method

The main aim in the NI trial is to establish the non-inferiority of the tested treatment compared to active treatment by an indirect conclusion of the superiority of the tested treatment to the placebo. Any chosen method to set the adjusted NI margin should maintain the three main assumptions in the NI trial (Julious, 2011):

- Assay sensitivity: the chosen active control is the most efficient available treatment
- Bias minimising: the differences between the historical placebo-controlled trials and the NI trial are minimum (the same endpoint, similar population characteristics, etc...)
- Constancy assumption: the efficacy of the active control is the same in the placebo-controlled trials and the NI trial.

It is proposed that the criteria for the chosen method include adjusting for any possible covariates, including all possible active controls that can be used in the design and the analysis phase of the NI trial. These criteria are based on the assumptions and methodological needs concerning NI trials to produce accurate and reliable results that are adjusted for any possible biases (FDA, 2016), and on the findings from the literature review in Chapter 2, the regulations in Chapter 3 and the results of Chapters 5 and 6.

7.2.1 Adjusting for any possible covariates while setting the NI margin

From the results of Chapters 5 and 6, it was found that the time changes affect the estimate of any future trials based on the results from the meta-analysis of the historical trials. Adjusting for the time difference between the placebo-controlled trials and the NI trial will reduce the possibility of biases. Moreover, accounting for time differences between the NI trials and the historical trials is also essential for the constancy assumption.

7.2.2 Including all possible active controls

In most therapeutic areas, there is more than one possible active control (standard treatment). Comparing all possible active controls in the designing phase and while setting the NI margin is essential for the assay sensitivity of the active control and including all relevant trials in the analysis to compare these active controls together or against the placebo will ensure the sensitivity of the chosen active control.

7.2.3 The phases of the NI trial

Setting the NI margin in the design phase is one of the most important regulatory and methodological challenges concerning the NI trial. The chosen method should be used in the designing phase to choose the active control, to set the sample size of the NI trial and to set the NI margin. In the analysis phase, the same method should be used to compare the results from the current NI trial and historical trials, to estimate the efficacy of the tested treatment compared to placebo and to rank all possible treatments compared to placebo.

7.2.4 Type of data

Two types of data can be used in the case of NI trials to compare the results from the NI trial and the historical trials: individual patient-level data (IPD) and aggregated data (AD). IPD is preferred to the AD in the case of indirect comparison. However, IPD is not available most of the time, especially from the historical placebo-controlled trials. Any used method should be able to handle both the IPD and AD in the case of hard to reach IPD.

7.2.5 Computational flexibility

A method that can be used in both Frequentist and Bayesian approaches and that offers ease of use and coding is preferred.

Table 7.1 gives a summary of the developed criteria.

Table 7.1 Criteria for the best performance method

Criteria	Description
Adjusting for Co-variables	Adjusting for the differences between the including trials, to minimise the biases originated from the difference between the NI trial and the placebo-controlled trials. Adjusting for time differences to ensure the constancy assumption
Including more than one active control	To ensure the assay sensitivity assumption
Can be used both for the designing phase and the analysis phase of NI	For the designing phase to calculate the sample size, and set the NI margin. For the analysis phase to compare the results from the NI trial and other placebo-controlled trials and rank the included treatments.
Can handle both the IPD and AD data	In the case of historical trials, usually, it is difficult to have access to the patient data level
Computational flexibility	Using both the Frequentist and Bayesian approach and with availability of software to conduct the analysis

7.3 Methods for adjusted non-inferiority margin

7.3.1 Adjusted regulatory approaches

This approach was built according to the predictive model developed in Chapter 6. The predicted estimate for a future trial from previous historical trials depends on the point estimate from the meta-analysis, the year differences in the meta-analysis and the year of the future predictive trial.

$$Y(SMDlt) = 30.32 + 0.92 \times (SMDdl) - 0.005 \times (Ymeta) - 0.015 \times (Ylt)$$

The parameters in this model have min and max-predicted years of 1977 and 2016. The year difference in the meta-analysis ranged from one to 75 years, and the SMD from the historical trials ranged from (-2.39 to 1.3). Using this model, the estimate for any future trial can be predicted. This estimate will be adjusted for the time differences between the NI trial and the historical trials.

For the fixed margin approach, either the main predicted estimate (more reliable) or the worst limit of the predicted 95% CI) can be used. For the synthesis method, the adjusted predicted estimate and its 95% CI can be incorporated in the 95% CI from the NI trial.

The differences between the adjusted regulatory and the non-adjusted regulatory approaches (discussed in Chapter 2) are illustrated in Table 7.2. The main advantage for this method is that it adjusts for time and can predict the estimate of the future trial based on the historical trials, the year difference in the historical meta-analysis, in addition to the year of predicted trial. The main limitation for this proposed approach is that it cannot compare more than three treatments (only one active control will be included), cannot adjust for any other difference between the trials, is limited to the year 2016 and cannot be used beyond this limit. Moreover, the predictivity of this model was only 55.1%.

Table 7.2 Comparison between the adjusted and non-adjusted regulatory approaches based on the methodology criteria

Criteria	Non-adjusted regulatory approaches	Adjusted regulatory approaches
Adjusting for Co-variables	No	Adjust for time only
Including more than one active control	No	No
Used in the designing and analysis phase of NI	Yes	Yes
Can handle both the IPD and AD data	Yes	Yes
Computational flexibility	Yes	No
Can rank the included treatments	No	No

7.3.2 Pairwise meta-regression for adjusting for time

Meta-regression was introduced as a method to explain the heterogeneity in the pairwise meta-analysis (Thompson & Sharp, 1999). Usually, meta-regression compares two treatments (pairwise meta-regression) with adjusting for any possible covariates between the trials. The meta-regression model is a meta-analysis model that includes study level covariates to test the impact of covariates for statistical significance (Hoaglin et al., 2011). In a pairwise meta-analysis, including co-variables or effect modifiers could reduce biases introduced by heterogeneity or inconsistency between treatment comparisons (Hoaglin et al., 2011). Usually, meta-regression is not recommended if the number of studies included is less than ten (Higgins & Green, 2008), especially when there are multiple covariates in the model. If the number of studies is small, multiple covariates are not recommended in meta-regression (Borenstein, Hedges, 2009). This could be considered as a limitation for the meta-regression. However, it has been used for a smaller number of studies (Dranitsaris, Jelincic, & Choe, 2011) with caution to include only one covariate at the time.

The general principles of meta-regression are the same as those for the regression models. The effect estimate is the dependent (outcome) variable, and the characteristics of the study (the potential effect modifiers or covariates) are the independent variables. The difference between regular regression models and meta-regression models is that in the latter the included studies are weighted (Thompson & Sharp, 1999).

The obtained coefficient from the meta-regression model will affect the size of the main estimate of "the outcome variable" changes by the co-variables. The p-value from the coefficient of the covariate indicates whether the difference is statistically significant or not. Both fixed and random effects models can be used.

Pairwise meta-regression has been used in NI trials to include covariates in the analysis. Eckert & Falissard (2006) used a pairwise meta-regression to compare the direct and indirect estimate between escitalopram and venlafaxine anti-depressant treatments using six placebo-controlled trials that compared escitalopram versus placebo and four comparing venlafaxine and placebo to indirectly compare venlafaxine and escitalopram (to establish non-inferiority of escitalopram to venlafaxine). The covariates included in the model were age, gender repartition, and severity at baseline.

The results from the indirect comparison were compared to the results from the two non-inferiority trials that compared venlafaxine and escitalopram. Eckert et al. concluded that the direct and indirect comparison results were both the same; the type of model used was not stated clearly (Eckert & Falissard, 2006).

Dranitsaris et al. (2011) used meta-regression to compare dalteparin to enoxaparin indirectly with the presence of the common comparator (placebo). They compared the results from five placebo-enoxaparin trials and four dalteparin-placebo trials to estimate the indirect treatment effect of dalteparin and enoxaparin compared to placebo. They used the active treatment as the primary independent variable, and the estimate from the trials compared to placebo (relative risk) as the outcome variable. The type of model used was not stated clearly. Moreover, they were able to adjust for the duration of therapy, treatment schedule, geographical region and the year of randomisation (Dranitsaris et al., 2011). Dranitsaris et al. concluded that meta-regression is considered as an appropriate method for adjusted indirect comparison in the case of the presence of the common comparator (placebo) (Dranitsaris et al., 2011).

Witte et al. (2011) used pairwise meta-regression to design an NI trial for supportive treatments in kidney transplantation (Witte, Schmidli, O'Hagan, & Racine, 2011). They proposed a random effect meta-regression and used each treatment as covariates in the model. They suggested the use of fixed margin or synthesis approaches to set the NI margin. In this study, there were no other covariates in the model apart from the treatments.

Meta-regression has been used to measure the changes in the placebo response over time in multiple therapeutic areas (Furukawa et al., 2018; Khan et al., 2017, 2018a, 2018b). In all these studies, the changes in placebo response over time were confirmed, using the year of publication as a covariate in the regression model.

Based on these studies, and in the context of this thesis to adjust for a time while setting a NI margin from the indirect comparison, pairwise meta-regression could be used in the designing phase of the trial as an alternative to the unadjusted pairwise meta-analysis to test for the constancy first and then to predict the active treatment response compared to the placebo in the year of NI trial conducting. This method could lead to a more accurate NI margin that takes into consideration the changes in the efficacy of the active control compared to placebo over time. The steps for the proposed method are:

1- Meta-regression of the placebo-controlled trials that compare the active treatment to placebo is conducted with the year of publication as a covariate in the regression model (if the year of trial conducting is available for all trials it is better to use that instead of the year of publication).
2- The constancy assumption will be assessed based on the bubble plot of the estimated effect size.
3- If the constancy assumption holds, then the estimate and the 95% CI from the pairwise meta-analysis will not be different from the estimate from the meta-regression and can be used to set the NI margin.
4- If the constancy assumption does not hold, then the predicted estimate adjusted for the time of NI trial conducting and the 95% CI should be used to set the NI margin.

Table 7.3 The methodology selection criteria for pairwise meta-regression

Criteria	Meta-regression
Adjusting for co-variables	Yes
Including more than one active control	No
Used in the designing and analysis phase of NI	Limited (only in the designing phase)
Can handle both the IPD and AD data	Yes
Computational flexibility	Limited
Can rank the included treatments	No

This method will lead to assessment of the constancy assumption and to setting an adjusted NI margin that accounts for any possible changes in the treatment effect over time. The main limitation of this method would be that the number of included historical trials should be ten or more. Moreover, this method cannot be used if there is more than one active control, and it is limited to the designing phase of the NI trial and cannot be used in the analysis phase. Table 7.3 applies the methodological selection criteria for the pairwise meta-regression.

7.3.3 Network meta-regression

A network meta-regression approach could be used to adjust for covariates or effect modifiers in a network meta-analysis model. This approach became popular in recent years as a preferred method to adjust for covariates in indirect comparison (Cooper et al., 2009; Donegan, Welton, Tudur Smith, D'Alessandro, & Dias, 2017; Eckert & Lançon, 2006; Liang et al., 2014).

Network meta-regression is a newly developed approach used to explain and control for the heterogeneity and inconsistency in the network meta-analysis (mixed treatment comparison) by adjusting for possible effect modifiers in the network meta-analysis.

Until now, only Bayesian frames are available for network meta-regression. Different software is available for analysis, including WIN Bugs, SAS, STATA, and gemtc R package (Valkenhoef & Kuiper, 2016).

Similar to the pairwise meta-regression, network meta-regression is not recommended if the number of included studies is less than ten (Higgins & Green, 2008). This is considered as the most critical drawback regarding the network meta-regression model due to concern over low power of the analysis if the number of trials is small compared to the number of included comparisons (Cooper et al., 2009). The advantage of using the network meta-regression over the pairwise meta-regression is that more than one active treatment can be included in the comparison, while the pairwise meta-regression can handle only two treatments. Moreover, network meta-regression can be used in both the designing phase to set the NI margin based on different active controls and in the analysis phase by incorporating the NI trial in the network of analysis. Moreover, the network meta-regression can be used when there are no direct placebo-controlled trials that compare the active treatment and the placebo.

Cooper et al. (2009) introduced three possible mixed treatment comparison methods (network meta-regression models) with co-variables depending on the type of regression coefficient used in the model (Cooper et al., 2009).

Model 1: The regression coefficient for each treatment is different

This model assumes that each treatment by covariate interaction for the comparison between the active treatment and the control is different and unrelated. Equation 8.1 illustrates the framework of network meta-regression when the coefficient is independent

$$r_{jk} \sim Binomial\left(p_{jk}, n_{jk}\right) for \text{ trial } j, \text{treatment } k$$

$$logit\ (p_{jk} = \begin{cases} \mu_{jb} & b = A, B, C\ if\ k = b \\ \mu_{jb} + \delta_{jkb} & if\ k\ alphabetcally\ after\ b \end{cases}$$

$$\delta_{jbk} \sim Normal\left(d_{bk} + \beta_{bk} X_j, \sigma^2\right) \sim Normal\left(d_{Ak} - d_{Ab} + (\beta_{Ak} - \beta_{Ab}) X_j, \sigma^2\right), \quad d_{AA}, \beta_{AA} = 0 \qquad (8.1)$$

Where μ_{jb} is the log odds ratio in trial *j* compared to baseline treatment *b*, β_{bk} is the change in the log odds ratio of an event per unit change in covariate X_j for treatment k relative to control treatment b, δ_{jbk} is the trial log odds ratio of *k* compared to *b*, d_{bk} is the pooled log odds ratio for the treatment k relative to b if a covariate X_j =0, number of r of treatments in the network = *k* and the number of relative treatments is *k-1*.

Model 2: The regression coefficient is exchangeable: this model assumes all interactions of treatment by covariates are different but related, equation (8.2).

$$r_{jk} \sim Binomial\left(p_{jk}, n_{jk}\right) for \text{ trial } j, \text{treatment } k$$

$$logit\ (p_{jk} = \begin{cases} \mu_{jb} & b = A, B, C\ if\ k = b \\ \mu_{jb} + \delta_{jkb} & if\ k\ alphabetcally\ after\ b \end{cases}$$

$$\delta_{jbk} \sim Normal\left(d_{bk} + \beta_{bk} X_j, \sigma^2\right) \sim Normal\left(d_{Ak} - d_{Ab} + (\beta_{Ak} - \beta_{Ab}) X_j, \sigma^2\right)$$

Where $\beta_{Ak} \sim Normal(B, \sigma_B^2)$, $d_{AA}, \beta_{AA} = 0$ (8.2)

Model 3: The regression coefficient is shared (common)

$$r_{jk} \sim Binomial\left(p_{jk}, n_{jk}\right) for \text{ trial } j, \text{treatment } k$$

$$logit\ (p_{jk} = \begin{cases} \mu_{jb} & b = A, B, C\ if\ k = b \\ \mu_{jb} + \delta_{jkb} & if\ k\ alphabetcally\ after\ b \end{cases}$$

$$\delta_{jbk} \sim \begin{cases} Normal\left(d_{Ak} + \beta\ \ X_j, \sigma^2\right) \sim Normal\left(d_{Ak} - d_{AA} + \beta X_j, \sigma^2\right) if\ b = A \\ Normal(d_{bk}, \sigma^2) \sim Normal(d_{Ak} - d_{Ab}, \sigma^2) \qquad if\ b \neq A \end{cases}$$

where $d_{AA}, \beta_{AA} = 0$ (8.3)

The way of choosing between these three models depends on the nature of covariates and the number of trials available per comparison, since in the case of model one the number of comparisons will be higher compared to model 3 (Cooper et al., 2009). The model's goodness of fit should decide which coefficient to use by comparing the deviance information criterion (DIC) with low DIC preferred. Moreover, the interpretation and the usefulness of the model to the clinician should be taken into consideration (Cooper et al., 2009). For the purpose of this thesis, the shared model will be used.

There was not enough literature regarding the use of network meta-regression in general or in the case of NI trials. A search of the Web of Science for publications with network meta-regression in their titles returned only 15 publications, starting from the year 2012. The peak was in 2018 with seven publications. Moreover, there was only one study regarding the conducting of network meta-regression in NI trials (Kent et al., 2018).

Kent et al. used network meta-regression to choose the appropriate active control, to set the non-inferiority margin and to calculate sample size for the future NI trial based on the network meta-regression on the available data for endovascular abdominal aortic aneurysm (Kent et al., 2018). They used the follow-up time as a co-variable in the model and they adjusted for what were considered as confounders (age, gender and the mean aneurysm diameter) (Kent et al., 2018).

The limitations regarding this method are the reducing power of analysis when the number of covariates increases. Even though there were no minimum limits of the number of trials that should be included in the analysis, the power of the analysis for the network meta-regression reduced as the number of studies decreased (Cooper et al., 2009). Most published studies include only one covariate at a time in the analysis, and only the Bayesian approach is available.

Table 7.4 summarises the methodological criteria for the network meta-regression

Table 7.4 Methodological criteria for the network meta-regression

Criteria	Network Meta-regression
Adjusting for Co-variables	Yes
Including more than one active control	Yes
Used in the designing and analysis phase of NI	Yes
Can handle both the IPD and AD data	Yes
Computational flexibility	Limited (no Frequentist approach)
Can rank the included treatments	Yes

7.3.4 Methods depending on IPD data

Different methods were introduced to adjust for covariates in indirect comparison in general and in the case of NI trials; some of these methods depend on the use of individual patient data (IPD) from the available trials to adequately adjust for any possible effect modifiers like age and gender (Phillippo et al., 2018). The most commonly used methods are the Matching Adjusted Indirect Comparison (MAIC) (Ishak, Proskorovsky, & Benedict, 2015; Signorovitch et al., 2012, 2010) and Simulated Treatment Comparison (STC) (Caro & Ishak, 2010; Ishak et al., 2015). Both of these methods depend on the availability of the individual patients' data and the presence of a common comparator to generate an indirect comparison that is adjusted for any possible effect modifiers. The calculation is done either by reweighting the treatment effect of the AB trial and applying it in the AC trial in the MAIC or to simulate the AB treatment effect in the AC trial in the STC method (Phillippo et al., 2018).

Both methods could be applied in the case of NI trials since in these trials the active control is the common comparator between the placebo and the test treatment. However, both methods adjust for only patients-based covariates, not for trials-based covariates. In this case, they cannot adjust for time differences (year of publication) on the patient level. The availability of IPD data is another consideration because IPD is not usually available from all historical data.

Nie & Soon (2010) presented a covariate-adjusted regression model to assess the constancy assumption in NI trials and set a justified margin when the constancy assumption was violated (Nie & Soon, 2010). However, their model depends on the availability of IPD data and cannot be applied to aggregated population data (AD). Moreover, it adjusts only if the constancy assumption is violated and depends mainly on the historical data to test the constancy assumption (Xu, Barker, Menon, & D'Agostino, 2014). This model was modified by Xu et al. (2014) to modify the covariate adjustment using both fixed and synthesis methods in one and two stage approaches using the IPD, not the aggregated population data (AD) (Xu et al., 2014). Table 7.5 summarises the methodological criteria for the IPD methods.

Table 7.5 Applying the selection criteria for the IPD based methods

Criteria	IPD
Adjusting for Co-variables	Limited (patients based only)
Including more than one active control	Limited
Used in the designing and analysis phase of NI	Yes
Can handle both the IPD and AD data	No
Computational flexibility	Yes
Can rank the included treatments	Yes

7.4 Summary

When setting NI margins using the regulatory approaches, there is no adjustment for the changes in the placebo and active treatment effect over time. However, using a fraction of the effect size of the active control fixed margin could be considered as an approach for adjustment.

Adjusting for covariates in individual patient data (IPD) level is another approach to adjustment for covariates in indirect comparison using MAIC or STC methods. However, the IPD usually is not available for all trials, especially the older trials in the case of NI trials. Regarding the aggregated data, both pairwise meta-regression and network meta-regression could be used to adjust for indirect comparison in NI trials. However, pairwise meta-regression can be used only in the designing phase of the trial and can include only one active control.

The use of network meta-regression could be promising in the case of NI trials for many reasons. First, it could adjust for the time (year of publication) or any other possible effect modifiers or confounders between the historical trials and NI trials. Second, it can include all possible active controls in the network to compare the efficacy of the test treatment with different available treatments and placebo. Third, network meta-regression could rank the treatment according to best treatment and can compare the estimate from both the direct and indirect comparison in the same networks. The drawbacks of network meta-regression are the need for a large number of trials to include more comparisons and its susceptibility to ecological bias, as well as its lower power to detect differences (Phillippo et al., 2018). Moreover, only the Bayesian approach is available until now. Table 7.6 reflects the differences between the different methods for setting the NI margin.

In summary, for the purpose of this thesis (adjusting for the time in case of indirect comparison), network meta-regression is considered the best method, followed by the pairwise meta-regression. Network meta-regression and pairwise meta-regression will be applied and compared using the available data from the review in Chapters 5 and 6.

As mentioned earlier, IPD methods cannot be applied for adjusting for time. However, it could be applied for adjusting for the patients level data, which indirectly could account for the differences in time between the trials. Moreover, the major limitation for the IPD methods is the shortage of patients level data, especially for the older placebo-controlled trials. The

proposed adjusted regulatory approach and the IPD methods cannot be used due to the limitations mentioned earlier.

In the next chapter, two reviews from the Cochrane reviews of placebo-controlled trials that were discussed in Chapters 5 and 6 will be used. Pairwise meta-regression will be used to assess the constancy by including the year of publication as a covariate in the model. then an adjusted NI margin will be calculated from the output of the pair-wise meta-regression. Different M2s will be constructed based on different percentages of M. The sample size of the hypothetical NI trial will be calculated based on the selected NI margin (M2). Network meta-regression will be used for the analysis of the hypothetical NI trial based on the different NI margins proposed.

Table 7.6 Characteristics of methods used to set NI Margin

Criteria	Pairwise Meta-regression	Network meta-regression	Adjusted regulatory approaches	IPD methods
Adjusting for Co-variables	Limited	Yes	Limited	Limited
Including more than one active control	No	Yes	No	Limited
Used in the designing and analysis phase of NI	Limited (Design)	Yes	Yes	Yes
Can handle both the IPD and AD data	Yes	Yes	Yes	No
Computational flexibility	Limited	No	Limited	Yes
Can rank the included treatments	No	Yes	No	No

Chapter 8 Applying the Proposed Adjusted Method for Setting and Analysis of NI Trials

8.1 Introduction

Chapter 7 reviewed the possible methods that can be used to adjust for a time in the case of NI trials. Criteria for methods that can be used to adjust for a time in the case of indirect comparison were also developed based on the assumptions of NI trials (assay sensitivity, bias minimising and constancy assumption), and the ability to adjust for time and the ease of use. Th conclusion was that pairwise meta-regression in the designing phase and network meta-regression in the analysis phase were the recommended methods to use for adjusting for a time in indirect comparison.

This chapter aims to propose a new method of using the pairwise meta-regression for assessing the constancy assumption and setting the NI margin in the designing phase (using adjusted fixed margin approach), and the network meta-regression to indirectly compare the placebo response to the test treatment with the year of conducting as a covariate in the analysis phase. Moreover, the changes in the used percentage of M1 to form M2 and the effect of year of conducting on the 50% of M1 will be discussed.

The chapter is structured as follows. In Section 8.2, the methods for the setting and the analysis will be presented. This will be followed in Section 8.3 by introducing the first case study (atorvastatin for lowering lipids) where the constancy does not hold. The second case study (lidocaine for reducing propofol-induced pain) will be presented in Section 8.4. Finally, a summary and recommendations will be provided in Section 8.5.

8.2 Methods

8.2.1 Selection of the case studies

The two cases were selected from the Cochrane reviews used in Chapter 5, based on the correlations between the SMD and the year. In the first case study, the correlation was a strong negative correlation (constancy assumption does not hold), and in the second case study, the correlation was a weak positive correlation (constancy assumption does hold). Both case

studies have a similar number of trials, one of them depended on the objective measure (cholesterol reduction) and the other one depended on a subjective measure (pain intensity), and the year difference in both cases was over twenty years.

In each case study, it will be hypothetically assumed that a new treatment (T) has been developed and needs to be tested for non-inferiority compared to the active treatments (atorvastatin and lidocaine in the first and second case studies, respectively) in the year 2020.

Theoretically, if a new treatment (T) is developed, the aim will be to conclude the non-inferiority of this new treatment (T) compared to the active control (C). The use of a placebo in any future trial is not acceptable due to ethical reasons. In both case studies, several placebo-controlled trials that compare the active control (C) to placebo (P) were identified. The aim now is to assess the constancy of the efficacy of the active control compared to the placebo and indirectly to compare the efficacy of the test treatment with that of the placebo.

8.2.2 Assessing the constancy

In each case study, a pairwise meta-regression will be used to assess the constancy assumption by including the year of publication as a covariate in the model. This will assess whether the treatment difference (point estimate) changed over time or not. The fixed effect pairwise meta-regression will be conducted using the R gemtc (Valkenhoef & Kuiper, 2016). Markov Chain Monte Carlo (MCMC) simulation method will be used to calculate the posterior distributions. The number of iterations will be 20,000, there will be one thinning interval, four chains, and the sample size per chain will be 20,000. Half normal priors will be used to cover a range of plausible values, and they seem reasonable for a wide range of diseases and treatments (Schmidli et al. , 2012).

8.2.3 Setting the NI margins

For each case study, two NI margins will be calculated. The unadjusted NI margin, assuming the constancy holds, will be calculated using the lower limit of the 95% CI of the meta-analysis of the placebo and active treatment.

The adjusted NI margin will be calculated from the pairwise meta-regression that adjusts the result for the year of trial conduction. The adjusted non-inferiority margin will be calculated

using the fixed margin approach with an adjusted 95% CI from the placebo-controlled trials in the year of NI trial conducting.

8.2.4 Setting the hypothetical non-inferiority trials

For each case study, two hypothetical NI trials tocompare the active treatment (C) with the test treatment (T) will be formulated. The sample size for the first trial will be calculated based on the unadjusted margin and the second one will be calculated based on the adjusted margin.

A network meta-regression will be used in the analysis phase to indirectly assess if the test treatment was superior to placebo or not (adjusted predictive approach). The results will be compared to the unadjusted non-inferiority margin (assuming the constancy).

To evaluate the efficacy of the use of the 50% of M1 as M2, different percentages of M1 will be compared in the year 2020 both when adjusted for time and while the constancy is assumed. Moreover, the changes of M2 based on the year of conducting (2025, 2030) will be assessed to investigate the validity of using a constant generic 50% of M1 as unadjusted margin.

The methods for each case study will be explained with more details in sections (8.3 and 8.4).

8.3 Atorvastatin for lowering lipids - the constancy assumption does not seem to hold

8.3.1 Background

The review was published in 2015 and aimed to assess the effects of various doses of atorvastatin on body lipids (total serum cholesterol, low-density lipoprotein (LDL), high-density lipoprotein (HDL) and triglycerides) in individuals with and without evidence of cardiovascular disease. It included 296 trials in total (242 are before and after trials, and 54 placebo-controlled trials), with 38,817 patients in total. The main conclusion was that atorvastatin decreases total blood cholesterol and LDL-cholesterol in a linear dose-related manner. In general, the evidence from this review is considered as high-quality evidence and the risk of bias is considered as a moderate risk (Adams et al., 2015). This case study was described previously in Section (5.5.1).

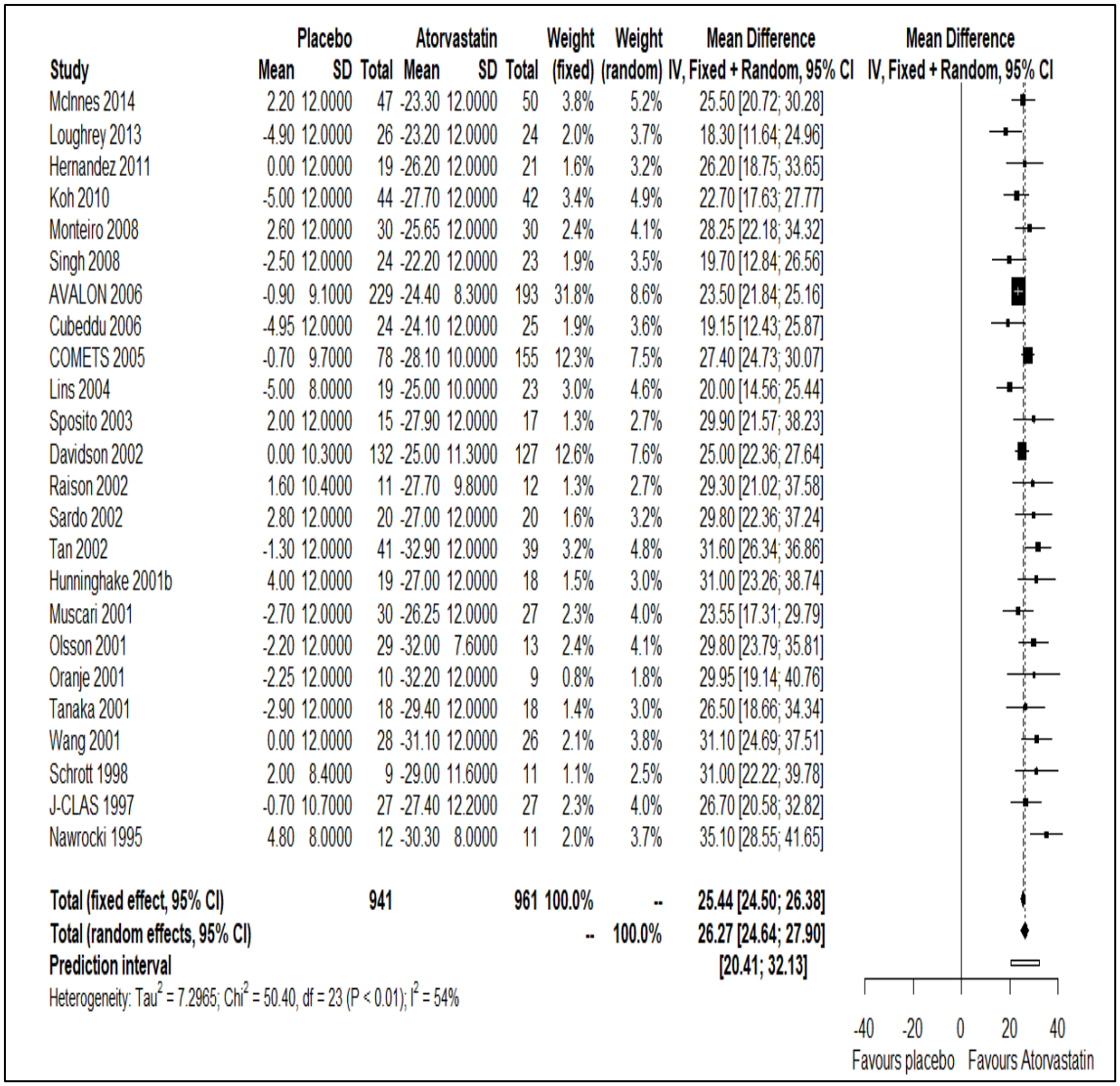

Figure 8-1 Meta-analysis of placebo versus atorvastatin

The meta-analysis used included 24 placebo-controlled trials published from 1995 to 2014. These trials were chosen by the authors of the original review (Adams et al., 2015) and the chosen studies (24 studies) were studies which measured the cholesterol level and used an atorvastatin dose of 10 mg as an active control.

The total sample size was 1902 participants, and the fixed effect model was used to calculate the estimate. The active treatment was atorvastatin 10 mg, and the outcome of the meta-analysis was a reduction in total cholesterol level from the baseline (negative outcome, the more negative the better).

The final estimate was a statistically significant difference between the placebo and the atorvastatin in baseline reduction of total cholesterol. The mean difference for placebo versus atorvastatin was 25.44, 95% CI (24.5: 26.38) (Figure 8.1).

8.3.2 Assessing the constancy and setting the NI margin

The twenty-four placebo-controlled trials that compared the placebo to the atorvastatin were included in the pairwise meta-regression model. The year of publication was the covariate in the model, and the outcome variable was the mean difference between the placebo and the atorvastatin. The fixed effect pairwise meta-regression was conducted using the R gemtc (Valkenhoef & Kuiper, 2016). Markov Chain Monte Carlo (MCMC) simulation method was used to calculate the posterior distributions. The iterations were 20,000, with one thinning interval, four chains, and a sample size per chain of 20,000. Half normal priors were chosen to cover a range of plausible values, and they seem reasonable for a wide range of diseases and treatments (Schmidli et al. , 2012).

Table 8.1 Results from the fixed effect meta-regression analysis

	Mean	Standard Error	95% CrI
Mean Difference	25.88	0.49	(24.92; 26.86)
Year	- 4.84	1.24	(-7.29; -2.41)

Note: the mean difference was in placebo-atorvastatin, Dbar= 61.01, PD=26.01, DIC=87, $I^2 = 23\%$

Table 8.1 shows the results from the fixed effect of meta-regression analysis. The results of meta-regression indicate that the year of publication statistically significantly reduces the difference for reduction of the cholesterol level between the placebo and the atorvastatin. Each year increase in the publication will reduce the mean difference between the placebo and the atorvastatin by 4.84 points (Table 8.1). The results for the random effects meta-regression were the same as for the fixed model, and they are presented in tAppendix (E.1.1)

Assuming the constant variance, the standard error of the point estimate (0.49) from Table 8.1 was used to calculate the confidence intervals for predictions. The prediction per year was calculated using the predict command in R.

95% CI= Mean difference (in a specific year) ± 1.96× Standard error (SE) (8.1)

Hypothetically, if a new treatment (Test) is discovered as an alternative to atorvastatin and needs to be tested for non-inferiority to atorvastatin in 2020, a non-inferiority trial will be designed using the meta-analysis of the historical placebo-controlled trials tocompare the active control with placebo:

The null hypothesis: H0: $\mu_T - \mu_C \leq -\delta$ (8.2)

The alternative hypothesis: Ha: $\mu_T - \mu_C > -\delta$ (8.3)

δ is a percentage (50%) of the lower limit of the mean difference between the placebo and active control (atorvastatin), C is the active control (atorvastatin), T is the test treatment (test), P is the placebo. From the meta-analysis in (Figure 8.1):

μ_P-μ_C 95% CI = 25.44 (24.50; 26.38) and

μ_C-μ_P 95% CI = - 25.44 (- 26.38; -24.50).

Assuming the constancy, to conclude the non-inferiority of the test treatment (T) compared to placebo (P), the lower limit of the 95% CI of the NI trial should be greater than -24.50, or to use the 50% to preserve the effect of the active control, the NI margin (M2) should be greater than -12.25. The -12.25 is the unadjusted margin assuming the constancy holds (Figure 8.2). However, according to the results of the meta-regression (Table 8.1), the year of publication has a negative effect on the efficacy of the atorvastatin compared to the placebo (the constancy does not hold). The treatment difference between the atorvastatin and the placebo is decreased over time. Figure 8.3 represents the bubble plot from the meta-regression analysis.

Using pairwise meta-regression, assuming the constancy does not hold (Figure 8.3), the mean difference between the placebo and the atorvastatin can be extrapolated for 2018, 2019, 2020, 2025, 2030 (the years the trial will possibly be conducted). The prediction by year was calculated from the R predict commands and the 95% CI calculated assuming constant variance with standard error (SE) of 0.49

95 % CI= Mean difference (in specific year) ± 1.96× 0.49.

Table 8.2 Estimates of NI margin using the two different constancy assumptions

Year	μ_C-μ_P (95% CI)	M1	M2
Constancy assumed (estimate from meta-analysis), not adjusted for the year			
2018	- 25.44 (- 26.38; -24.50).	-24.50	-12.25
2019	- 25.44 (- 26.38; -24.50).	-24.50	-12.25
2020	- 25.44 (- 26.38; -24.50).	-24.50	-12.25
2025	- 25.44 (- 26.38; -24.50).	-24.50	-12.25
2030	- 25.44 (- 26.38; -24.50).	-24.50	-12.25
Constancy not assumed (estimate from the meta-regression), adjusted for the year			
2018	-18.79 (-19.72; -17.85)	-17.85	-8.93
2019	-18.29 (-19.35; -17.35)	-17.35	-8.67
2020	-17.79 (-18.73; -16.85)	-16.85	-8.43
2025	-15.30 (-16.22; -14.35)	-14.35	-7.17
2030	-12.80 (-15.30; -11.84)	-11.84	-5.90

μ_C-μ_P is the treatment difference between the atorvastatin and placebo, M1, is the NI margin, the upper limit of the 95 % CI of μ_C-μ_P, M2 is the 50% of M1

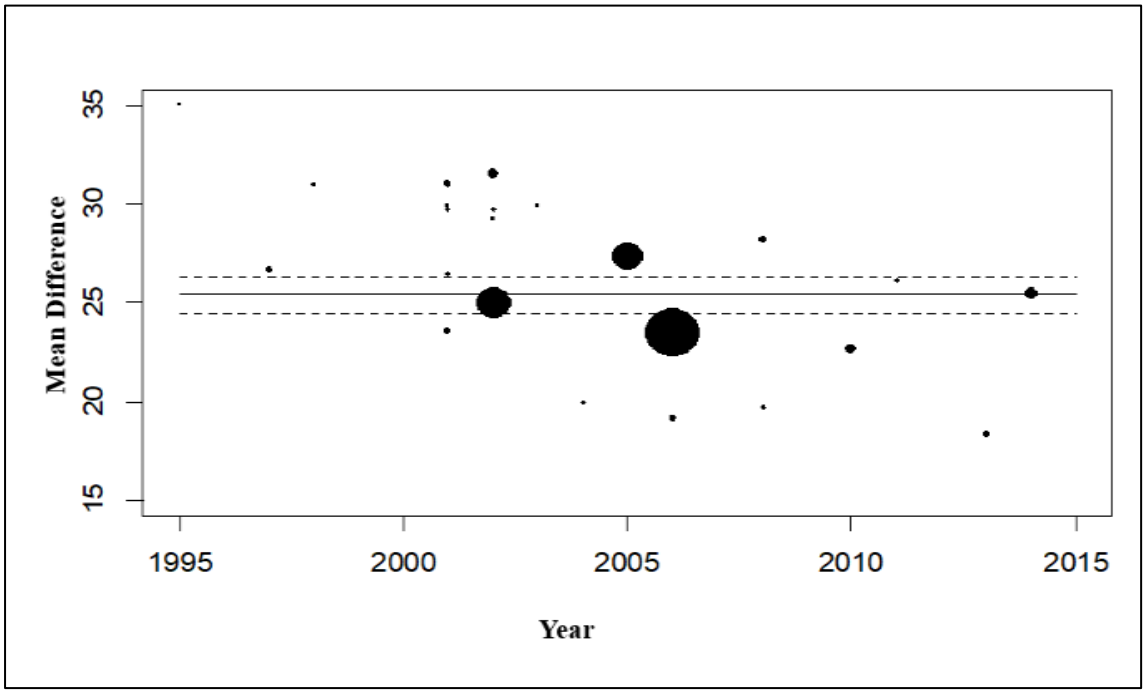

Figure 8-2 Bubble plot for the mean difference between the placebo and atorvastatin when the constancy assumption holds

(Note: the bubble size reflects the sample size, the 95 % CI assuming constant standard error of 0.49)

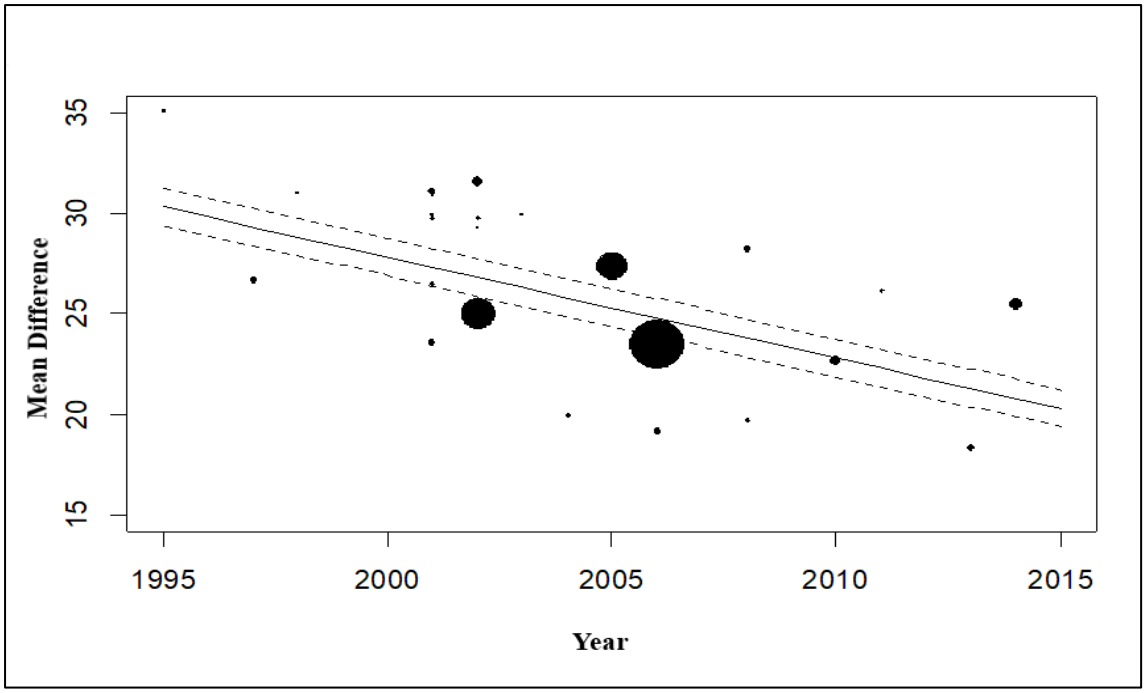

Figure 8-3 Bubble plot for the meta-regression of the mean difference between placebo and atorvastatin

(Note: Bubble size reflects sample size, the 95 % CI assuming constant standard error of 0.49)

Table 8.2 illustrates the difference between the NI margins using the unadjusted 95% CI (assuming the constancy) and the 95% CI from the meta-regression (the constancy does not hold). To conclude the NI of the test treatment (T) compared to atorvastatin, the lower limit of the 95 % CI of the NI trial (μ_T-μ_C) should be greater than M2.

Hypothetically, two NI trials could be designed in 2020, using either the unadjusted margin of -12.25 (assuming the constancy) or the adjusted margin of -8.43 (the constancy does not hold). Using the formula for sample size calculation (Flight & Julious, 2016):

$$n_C = n_T = \frac{(r+1)\sigma^2(Z_{1-\beta}+Z_{1-\alpha})^2}{r((\mu_T-\mu_C)-d)^2} \qquad (8.4)$$

where d is the pre-specified non-inferiority margin, σ^2 is the variance of the mean difference (from Figure 8.1 the standard deviation =12), μ_T is the mean cholesterol reduction in the test treatment, μ_C is the mean cholesterol reduction in the atorvastatin group, r is the allocation ratio between the treatment and control group, and is assumed to be one. α is a type I error, β is a type II error.

Assuming the constancy using the unadjusted NI margin of (d= -12.25), the standard deviation of 12, Type I error (α) of 0.025 and Type II error (β) of 0.1, the sample size assuming zero mean difference between the test treatment and atorvastatin is 21 participants per arm.

$$n_C = n_T = \frac{(2) \times (12)^2 \times (Z_{1-} + Z_{1-0.025})^2}{1 \times \big((0) - 12.25\big)^2}$$

$$n_C = n_T = \frac{21 \times 144}{(12.25)^2}$$

$$n_C = n_T = 20.15$$

$$n_C = n_T = 21$$

When adjusted for time, the adjusted NI margin in 2020 of - 8.43 and standard deviation of 12, Type I error of 0.025 and Type II error of 0.1, the sample size assuming zero mean difference between the test treatment and atorvastatin is 43 participants per arm. In 2025, the sample size will increase to 59 participants per group and to 87 participants in each arm in the year 2030.

8.3.3 Analysis of non-inferiority trial based on the unadjusted margin (NI margin > -12.25)

Suppose the NI trial is conducted in 2020 with a sample size of 21 participants in each arm based on an NI margin of -12.25. The aim is to confirm the non-inferiority of the test treatment compared to atorvastatin with NI margin > -12.25.

Network meta-analysis (NMA) will be used to compare the effect of the test treatment and the atorvastatin, without adjusting for the time (no adjustment in the setting or the analysis), by including the three different treatments (placebo, atorvastatin and the test treatment) in one network. Network meta-regression will assess the treatment difference between the placebo and the test treatment in 2020 (adjusting for the time in the analysis phase).
Figure 8.4 represents the network of the three treatments (placebo, atorvastatin, and the test treatment); the thickness of the lines represents the number of trials that compare the treatments.

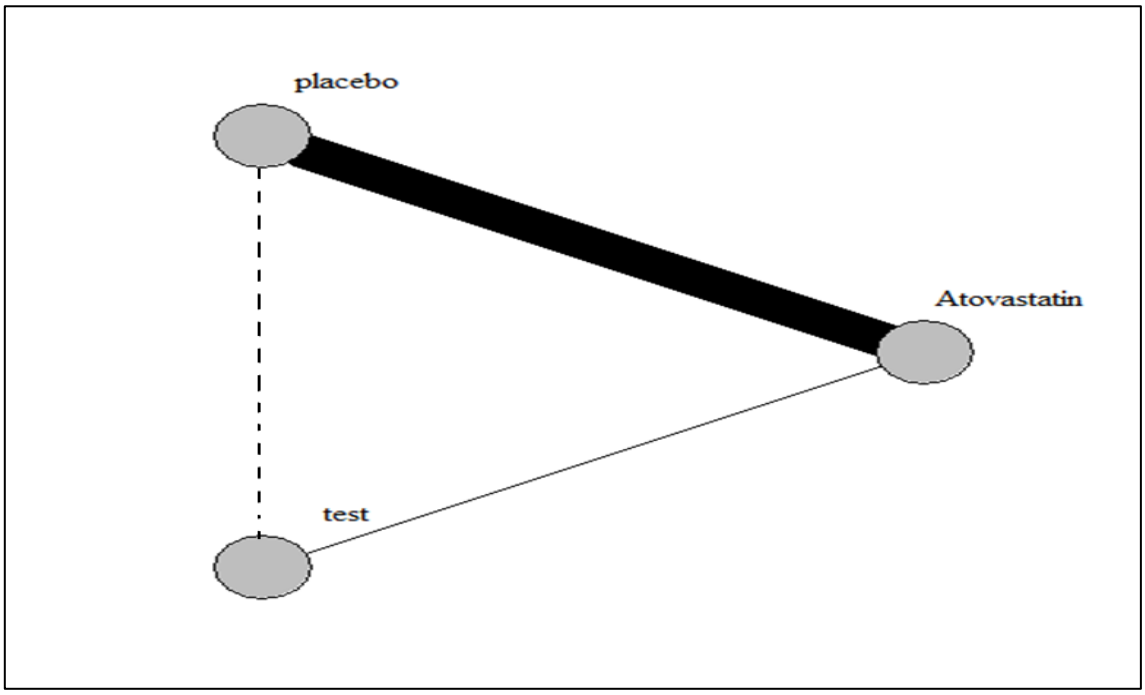

Figure 8-4 Network of atorvastatin, placebo and the test treatment
(Thickness of lines represents the number of trials), circles represent the included treatments, the solid line represents direct comparisons, and the dash line represents the indirect comparison)

The gemtc R package for Bayesian network meta-analysis (Valkenhoef & Kuiper, 2016) was used for conducting both the network meta-analysis and the network meta-regression. Markov Chain Monte Carlo (MCMC) simulation method was used to calculate the posterior distributions. The iterations were 80,000, with one thinning interval, four chainsand a sample size per chain of 40,000, all are the same as those used by Schmidli et al. (2012). Half-normal prior was used to cover a range of plausible values, and they seem reasonable for a wide range of diseases and treatments (Schmidli et al., 2012).

Table 8.3 Comparison of the mean difference between the placebo and test treatment assuming the constancy

μ_t	$\mu_T - \mu_C$ from NI trial (2020) (95% CrI), SE = (3.7)	NMA ($\mu_T - \mu_P$) (95% CrI), SE = (4.22)	NMR ($\mu_T - \mu_P$) in (2020) (95% CrI), SE = (4.22)
18.5	0.0 (-7.2; 7.2)	25.0 (18.0; 33.0)	18.0 (9.6; 26.0)
16.5	-2.0 (-9.2; 5.3)	23.0 (16.0; 31.0)	16.0 (7.6; 24.0)
15.5	-3.0 (-10.0; 4.3)	22.0 (15.0; 30.0)	15.0 (6.6; 23.0)
14.5	-4.0 (-11.0; 3.2)	21.0 (14.0; 29.0)	14.0 (5.6; 22.0)
13.5	-5.0 (-12.3; 2.2)	20.0 (13.0; 28.0)	13.0 (4.6; 21.0)
12.5	-6.0(-13.2; 1.3)	19.0 (12.0; 27.0)	12.0 (3.6; 20.0)
10.5	-8.0 (-15.3; -0.7)	17.0 (10.0; 25.0)	9.8 (1.5; 18.0)
8.5	-10.0 (-17.3; -2.7)	15.0 (8.2; 23.0)	1.8 (-0.5; 16.0)
7.0	-11.4 (-18.5; -4.4)	14.0 (6.8; 21.0)	6.4 (-1.7; 15.0)
1.5	-17.0 (-24.3; -9.7)	8.5 (1.1; 16.0)	0.8 (-7.5; 9.1)
0.0	-18.5 (-25.7; -11.2)	6.9 (-0.4; 14.0)	-0.7 (-8.9; 7.6)

NMA: network meta-analysis, NMR: network meta-regression, the $\mu_t - \mu_c$ refers to the mean difference between the active control and the test treatment, negative sign means the test treatment is less effective than the active control, SE: Standard error, μ_t is the treatment effect in the test group in the NI trial,
light grey = Failure to conclude NI of T versus C, medium grey = C is superior to T, dark grey =T is not superior to placebo.

Table 8.3 illustrates the differences between the placebo and the test treatment, using the NI margin of -12.25 to set the sample size of 21. The results from the NMA assume the constancy in the analysis phase (no changes by time). The results from the NMR represent the predicted results in 2020 (constancy not assumed in the analysis phase).
The first column in the table illustrates μt (the mean of the test treatment), the second column illustrates the results of the NI trial $\mu_T - \mu_C$ with the light shading referring to failure to conclude the non-inferiority of T compared to C and the medium grey shading illustrating that the control is even superior to the test treatment. The last column illustrates the results from the network meta-analysis that indirectly compared the test treatment (T) with the placebo (P) and the dark grey shading illustrates that the test treatment is not superior to the placebo.

From the table (8.3), by assuming the constancy and ignoring the changes of the efficacy of the active control (atorvastatin) and the placebo, using the margin of -12.25 the non-inferiority can be concluded up to four points difference (the efficacy of test treatment is four points less than the atorvastatin). The non-inferiority cannot be concluded if the efficacy of the test treatment is five points less than the active control.

Using NMA (constancy assumed in the analysis phase), the superiority of the test treatment compared to the placebo was concluded, up to a mean effect of the test treatment μ_t =1.5 points (the test treatment was inferior to active control).

Using the NMR (adjusted for time in the analysis phase), the superiority of the test treatment was concluded, with μ_t up to 10.5 points, which is less than the unadjusted one (NMA), but still the test treatment was inferior to the atorvastatin.

8.3.4 Analysis of non-inferiority trial based on the adjusted margin (NI margin < -8.43)

Suppose the NI trial is conducted in 2020 with a sample size of 43 participants in each arm based on an adjusted NI margin of -8.43. The aim is to confirm the non-inferiority of the test treatment compared to atorvastatin with NI margin greater than - 8.43. The three treatments (atorvastatin, test treatment, and placebo) were included in one network (Figure 8.4). NMA will be used to compare the effect of the test treatment compared to placebo without any further adjustment for time (constancy assumed in the analysis phase). Network meta-regression will be used to compare the efficacy of the test treatment compared to the placebo in 2020 (adjusted for time in the analysis phase).

The gemtc R package for Bayesian network meta-analysis (Valkenhoef & Kuiper, 2016) was used for conducting both the network meta-analysis and the network meta-regression. Markov Chain Monte Carlo (MCMC) simulation method was used to calculate the posterior distributions. The iterations were 80,000, with one thinning interval, four chains, and a sample size per chain of 40,000, all are the same as those used by Schmidli et al. (2012). Half-normal prior was used to cover a range of plausible values, and they seem reasonable for a wide range of diseases and treatments (Schmidli et al., 2012).

Table 8.4 illustrates the differences between the placebo and the test treatment, using the NI margin of -8.43 to set the sample size of 43. From the results, by adjusting for time (2020), the non-inferiority of the test treatment compared to atorvastatin was concluded up to three points of difference (the efficacy of test treatment is three points less than that of the atorvastatin). The non-inferiority cannot be concluded if the efficacy of the test treatment is greater than 3.5 points less than the active control.

Using the NMA, (constancy assumed) the superiority of the test treatment compared to placebo was concluded even when the mean effect of the test treatment equalled zero. When adjusted for the time (NMR) the superiority of the test treatment compared to placebo was concluded up to mean effect of test treatment equal to 8.5.

Table 8.4 Comparison of the mean difference between the placebo and test treatment, constancy not assumed

μ_t	$\mu_T - \mu_C$ from NI trial (2020) (95% CrI), SE= 2.58	NMA ($\mu_T - \mu_P$) (95% CrI), SE= 2.67	NMR ($\mu_T - \mu_P$) in (2020) (95% CrI), SE= 3.27
18.5	0.0 (-5.1; 5.1)	25.0 (20.0; 31.0)	18.0 (11.0; 24.0)
16.5	-2.0 (-7.1; 3.1)	23.0 (18.0; 29.0)	16.0 (9.4; 22.0)
15.5	-3.0 (-8.1; 2.1)	22.0 (17.0; 28.0)	15.0 (8.4; 21.0)
14.5	- 4.0 (-9.1; 1.1)	21.0 (16.0; 27.0)	14.0 (7.3; 20.0)
13.5	-5.0 (-10.1; 0.0)	20.0 (15.0; 25.0)	13.0 (6.4; 19.0)
12.5	-6.0 (-11.1; -0.9)	19.0 (14.0; 25.0)	12.0 (5.4; 18.0)
11.5	-7.0 (-12.1; -1.8)	18.0 (13.0; 24.0)	11.0 (4.4; 17.0)
10.5	-8.0 (-13.1; -2.9)	17.0 (12.0; 23.0)	9.8 (3.4; 16.0)
8.5	-10.0 (-15.6; -4.9)	15.0 (10.0; 21.0)	7.8 (1.4; 14.0)
7.0	-11.5 (-16.6; -6.4)	14.0 (8.8; 19.0)	6.3 (-0.1; 13.0)
1.5	-17.0 (-22.0; -11.9)	8.5 (3.3; 14.0)	0.8 (-5.7; 7.2)
0.0	-18.5 (-23.6; -13.5)	6.9 (1.8; 12.0)	-6.9 (-7.1; 5.7)

NMA: network meta-analysis, NMR: network meta-regression, the $\mu_t - \mu_c$ refers to the mean difference between the active control and the test treatment, negative sign means the test treatment is less effective than the active control.
Light grey = Failure to conclude NI of T versus C, medium grey = C is superior to T, dark grey =T is not superior to placebo

8.3.5 The effect of using different percentages of M1 to set M2

As mentioned in Chapter 2, M2 is the actual NI margin using a specific percentage of M1. This percentage is used to protect the constancy assumption (FDA, 2016). As a common generic practice, 50% of M1 is usually used as the M2. However, it is not clear if the use of 50% could be an alternative to the assessment of the constancy and the adjustment for time. Table 8.5 illustrates how the use of a different percentage of M1 could change the results both when assuming the constancy and when adjusting for time.

In 2020, when constancy was assumed (no adjustment for time), using 50% of the M1 instead of M1 as a whole, partially to protect the estimate from the conclusion of non-inferiority of an already inferior test treatment, the difference between the conclusion of non-inferiority and the failure to conclude the superiority to placebo was six points. This difference between the conclusion of non-inferiority and failure to conclude superiority to placebo dropped to three

points using 60% of M1, 0.5 points using 70% of M1. At 80% of M1, the failure to conclude superiority was higher at 12.5 points, while the conclusion of non-inferiority was 11 points (difference of -1.5 points); the conclusion was non-inferiority of an already inferior treatment. At 90% of M1, and if M1 was used, the non-inferiority would be established for an already inferior treatment (Table 8.5).

Table 8.5 Comparison between the unadjusted and adjusted margins when constancy does not hold

	Constancy Assumed	Adjusted for time
M2 = 50% M1		
NI margin	-12.25	-8.43
Sample size $n1 = n2$	21.00	43.00
NI established up to	μ_t=14.50	μ_t=15.50
NI cannot be concluded	μ_t=13.50	μ_t=15.00
Atorvastatin is superior to test treatment	μ_t=10.50	μ_t=13.00
Test treatment not superior to placebo (2020)	μ_t=9.00	μ_t= 7.00
Test treatment not superior to placebo (NMA)	μ_t=0.00	μ_t< 0.00
M2 = 60% M1		
NI margin	-14.70	-10.11
Sample size $n1 = n2$	14.00	30.00
NI established up to	μ_t=13.00	μ_t=14.50
NI cannot be concluded	μ_t=12.50	μ_t=14.00
Atorvastatin is superior to test treatment	μ_t= 9.50	μ_t= 11.50
Test treatment not superior to placebo (2020)	μ_t= 10.00	μ_t= 7.50
Test treatment not superior to placebo (NMA)	μ_t=1.50	μ_t< 0.00
M2 = 70% M1		
NI margin	-17.5	-11.79
Sample size $n1 = n2$	11.00	22.00
NI established up to	μ_t=11.50	μ_t= 14.00
NI cannot be concluded	μ_t=11.00	μ_t= 13.50
Atorvastatin is superior to test treatment	μ_t=7.50	μ_t= 9.50
Test treatment not superior to placebo (2020)	μ_t=11.00	μ_t= 8.50
Test treatment not superior to placebo (NMA)	μ_t=3.00	μ_t<0.00
M2 = 80% M1		
NI margin	-19.60	-13.48
Sample size $n1 = n2$	8.00	17.00
NI established up to	μ_t=11.00	μ_t= 14.00
NI cannot be concluded	μ_t=10.50	μ_t= 13.00
Atorvastatin is superior to test treatment	μ_t=6.50	μ_t=9.50
Test treatment not superior to placebo (2020)	μ_t= 12.50	μ_t= 9.50
Test treatment not superior to placebo (NMA)	μ_t=4.50	μ_t= 0.50
M2 = 90% M1		
NI margin	-22.05	-15.60
Sample size $n1 = n2$	7.00	14.00
NI established up to	μ_t=9.00	μ_t= 12.50
NI cannot be concluded	μ_t= 8.50	μ_t= 11.50
Atorvastatin is superior to test treatment	μ_t= 5.50	μ_t= 9.50
Test treatment not superior to placebo (2020)	μ_t=13.50	μ_t= 9.50
Test treatment not superior to placebo (NMA)	μ_t=5.50	μ_t= 1.50
M2 = M1		
NI margin	-24.50	-16.85
Sample size $n1 = n2$	6.00	11.00
NI established up to	μ_t=8.50	μ_t=12.00

NI cannot be concluded	μ_t=7.50	μ_t= 11.50
Atorvastatin is superior to test treatment	μ_t=4.50	μ_t= 7.50
Test treatment not superior to placebo (2020)	μ_t=6.50	μ_t= 11.50
Test treatment not superior to placebo (NMA)	μ_t=14.50	μ_t= 2.50

When adjusted for time, the adjusted NI margin using 50% of the M1 leads to 8.5 points difference between the conclusion of non-inferiority and failure to conclude the superiority to placebo. This difference is reduced to 7.5 points using 60% of M1, 6 points difference using 70% of M1, 4.5 % using the 80% of M1, 3.5 points using the 90% of M1 and 1.5 points using the whole M1 (Table 8.5). Figures 8.5 and 8.6 illustrate the differences between the use of adjusted and unadjusted margins with different percentages of M1.

Even though using 50% of M1 with the unadjusted margin partially protected from the conclusion of non-inferiority of an already inferior test treatment in 2020, this protection was similar to 70% using the adjusted margin. The use of an unadjusted margin will increase the risk of the conclusion of non-inferiority of an already inferior treatment. The adjusted margin was away from the conclusion of non-inferiority of an inferior treatment even with the use of the whole M1 instead of a percentage of M1. The use of an unadjusted margin led to a false conclusion of non-inferiority of an already inferior test treatment to placebo with the 70% of M1. Using the adjusted margin led to a larger sample size of the planned NI trial.

In 2025, the picture was changed; assuming the constancy and using the NI margin of -12.25 with sample size of 21 led to failure to conclude the non-inferiority with treatment effect of the test treatment ($\mu_t = 13.5$). The superiority of the test treatment compared to placebo cannot be established at $\mu_t = 12$, with the difference between the two being 1.5 points only. The use of 50% of M1 while assuming constancy did not protect against the conclusion of non-inferiority of an already inferior treatment.

By increasing the year difference between the last historical trial and the year of NI trial conducting, the gap between the failure to conclude non-inferiority and failure to conclude the superiority of the test treatment to the placebo was increased. By the year 2030 (15 years difference), the non-inferiority of the test treatment was established up to $\mu_t = 14.5$, while there was failure to establish superiority to placebo at $\mu_t = 15$. In other words, i the non-inferiority of an already inferior treatment was concluded by assuming the constancy without any further adjustment. Moreover, the use of a 50% fraction of M1 did not protect against the false conclusion of non-inferiority.

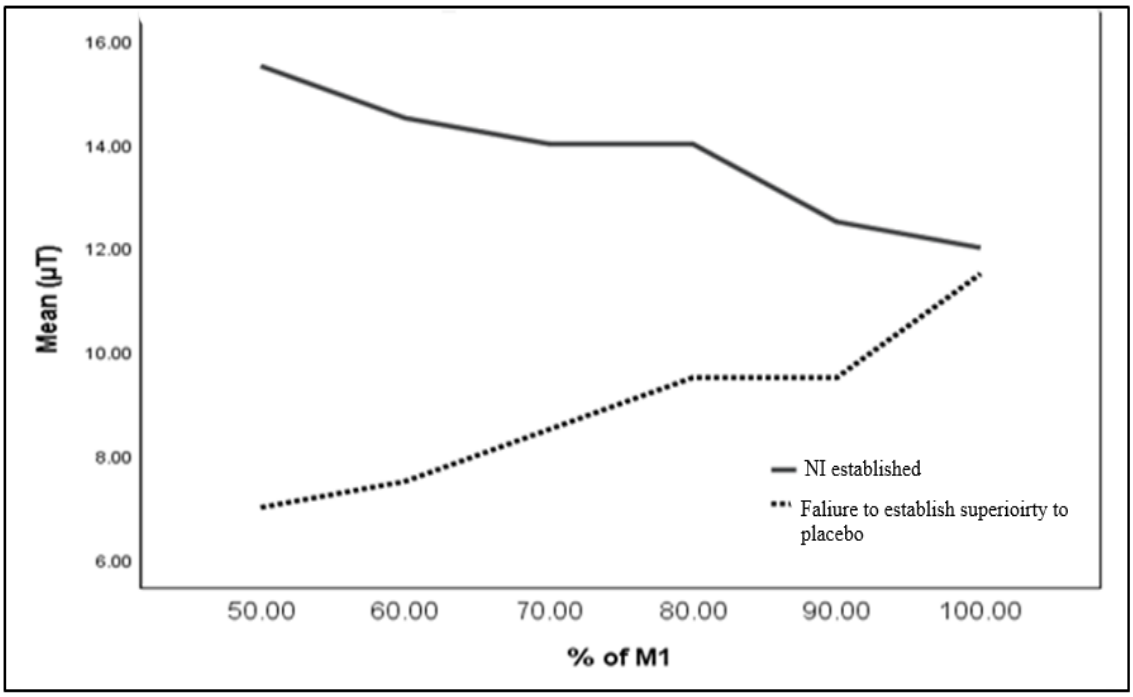

Figure 8-5 Comparison between the different percentages of M1 using the adjusted margin (constancy does not hold)

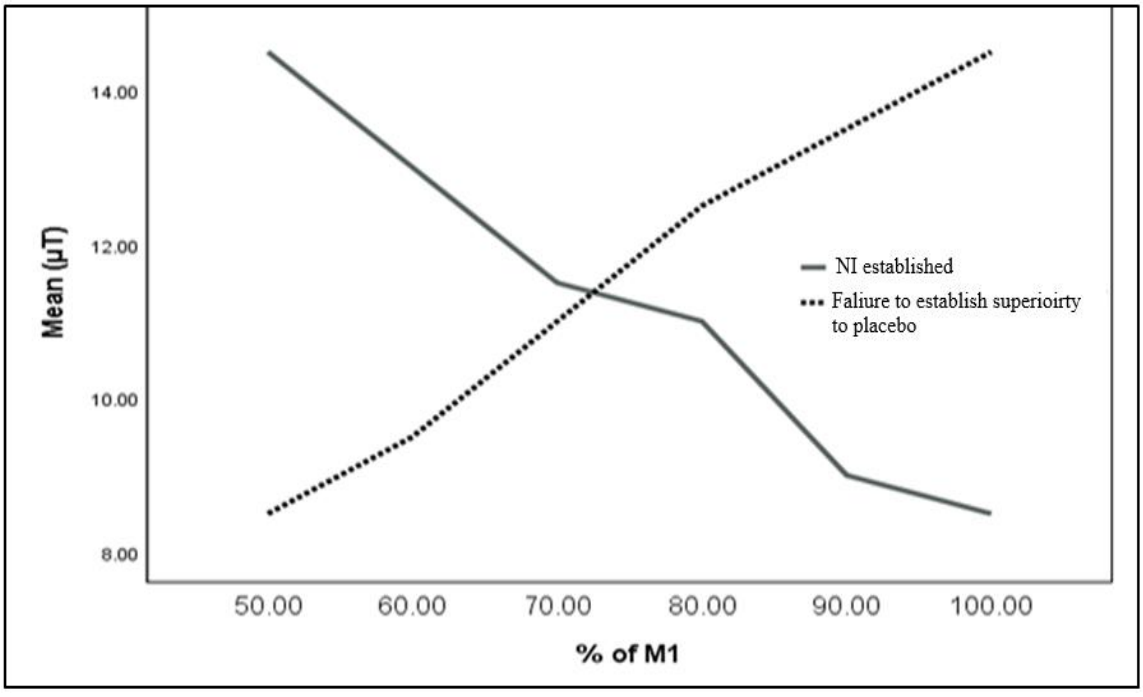

Figure 8-6 Comparison between the different percentages of M1 using the unadjusted margin (constancy assumed)

Table 8.6 Comparison between the unadjusted and adjusted margins for years 2020, 2025, 2030

	Constancy Assumed	Adjusted for time
2020		
NI margin	-12.25	-8.43
Sample size $n1 = n2$	21.00	43.00
NI established up to	μ_t=14.50	μ_t=15.50
NI cannot be concluded	μ_t=13.50	μ_t=15.00
Atorvastatin is superior to test treatment	μ_t=10.50	μ_t=13.00
Test treatment not superior to placebo (NMR)	μ_t=9.00	μ_t= 7.00
Test treatment not superior to placebo (NMA)	μ_t=0.00	μ_t< 0.00
2025		
NI margin	-12.25	-7.17
Sample size $n1 = n2$	21.00	59.00
NI established up to	μ_t=14.50	μ_t=16.00
NI cannot be concluded	μ_t=13.50	μ_t=15.50
Atorvastatin is superior to test treatment	μ_t=10.50	μ_t= 13.50
Test treatment not superior to placebo (NMR)	μ_t=12.00	μ_t= 8.50
Test treatment not superior to placebo (NMA)	μ_t=0.00	μ_t< 0.00
2030		
NI margin	-12.25	-5.90
Sample size $n1 = n2$	21.00	87.00
NI established up to	μ_t=14.50	μ_t= 16.25
NI cannot be concluded	μ_t=13.50	μ_t= 16.00
Atorvastatin is superior to test treatment	μ_t=10.50	μ_t= 13.50
Test treatment not superior to placebo (NMR)	μ_t=15.00	μ_t12.50
Test treatment not superior to placebo (NMA)	μ_t=0.00	μ_t<0.00

When adjusting for time, in 2025, the sample size was increased to 59 participants per arm and the NI margin was -7.17. The non-inferiority of the test treatment compared to the placebo was concluded up to μ_t= 16.00, while the superiority to placebo was concluded up to $\mu_t = 8.5$. In 2030, the sample size increased to 87 participants per arm and the NI margin was -5.90. The non-inferiority of the test treatment compared to the placebo was concluded up to μ_t= 16.25, while the superiority to placebo was concluded up to $\mu_t = 12.5$. (Table 8.6)

In summary, in the case where the constancy assumption does not hold, assuming the constancy and using the unadjusted margin could lead to the conclusion of non-inferiority of an already inferior treatment. The use of a predefined percentage of 50 % M1 instead of the whole M1 cannot protect from the conclusion of non-inferiority of an inferior test treatment, especially when the time difference between the last historical trial and the NI trial is increased. The use of a 50% fraction of M1 cannot be a replacement for the adjustment for the constancy. Using the adjusted margin reduces the chances of the conclusion of the non-inferiority of an inferior treatment regardless of the fraction of the M1 used. In the case of an NI trial, the adjusted margin for time should be used to set M1; M2 should be a matter of clinical judgement and based on the adjusted M1, not used as a tool for protection of the constancy assumption.

8.4 Using lidocaine for reducing propofol-induced pain on the induction of anaesthesia in adults - the constancy assumption seems to hold

8.4.1 Background

This review was updated in 2016. It aimed to investigate the efficacy and adverse effects of lidocaine in reducing high-intensity pain during propofol injection. The review includes 82 multicentre placebo-controlled trials. The quality of evidence is graded as high quality. The main meta-analysis used in this example includes 23 trials. These trials were chosen based on the dose group (low dose group trials). The year difference ranged from 1988 to 2010. The results indicate that the incidence of high-intensity pain in the control group (placebo) was higher than in the lidocaine group (low dose group). The odds of high-intensity pain in the placebo group were 5.16 times higher than in the lidocaine group, 95% CI (4.14; 6.42) (Euasobhon et al., 2016). Figure 8.7 represents the meta-analysis for the comparison between the lidocaine and the placebo.

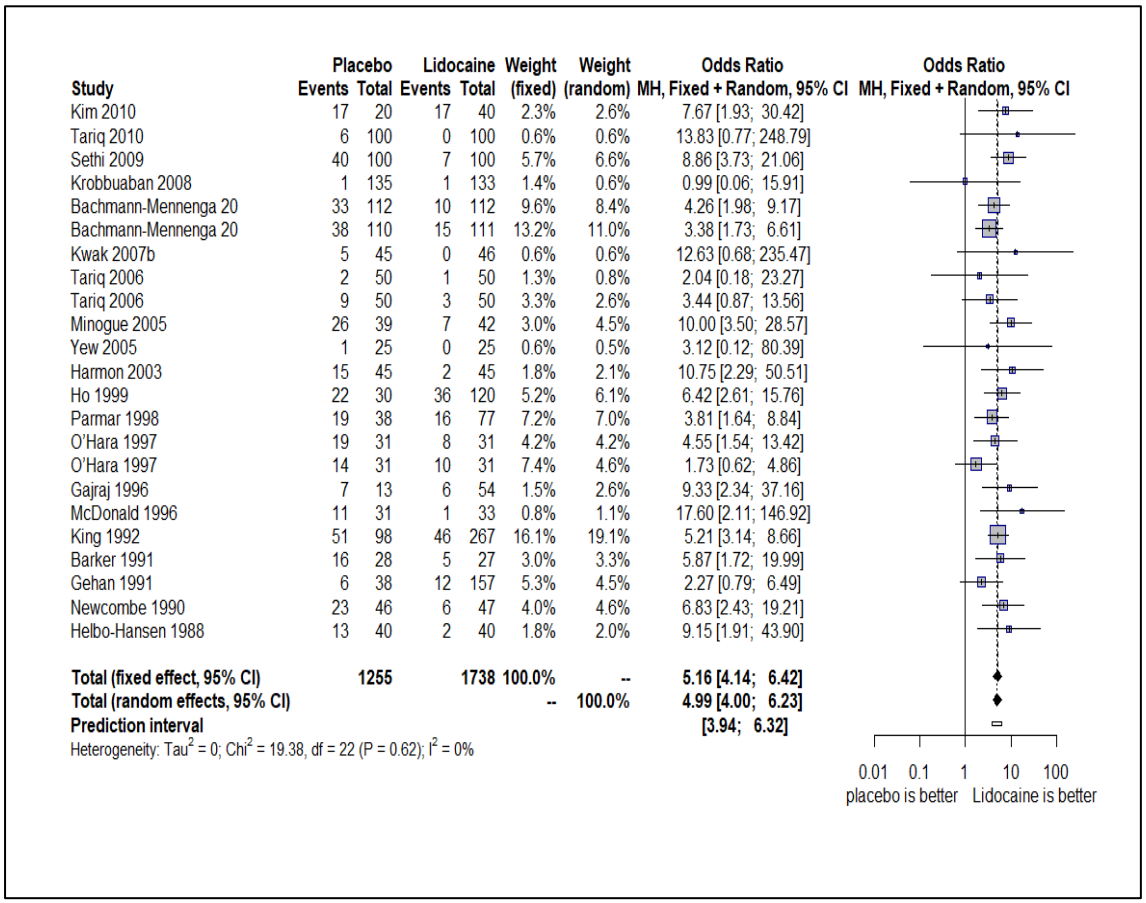

Figure 8-7 Meta-analysis of the pain intensity in placebo versus lidocaine

8.4.2 Assessing the constancy and setting the NI margin

From Chapters 5 and 6, it was concluded that the use of a fixed effect model is recommended in the case of NI trials since it gives less weight to the extreme older trials compared to the random effects model, which will give more weight for smaller studies with extreme results. For that reason, the fixed effect model will be used. The results of the random effects model will be presented in the Appendix (E.3).

Fixed effect pairwise meta-regression was conducted using the R gemtc (Valkenhoef & Kuiper, 2016). Markov Chain Monte Carlo (MCMC) simulation method was used to calculate the posterior distributions. The iterations were 20,000, with one thinning interval, four chains, and a sample size per chain of 20,000. Vague priors used were the same as those used by Schmidli et al. (2012).

Table 8.7 Results of meta-regression of placebo versus lidocaine

	Estimate	Standard Error	95% CrI
Log odds ratio	1.68	0.11	1.46; 1.91)
Year	0.10	0.20	(-0.32; 0.50)

Note: the log odds ratio was between placebo versus Lidocaine, Dbar= 48.86, PD=25.33, DIC=74.20, $I^2 =$ 8 %

The 23 placebo-controlled trials that compared the placebo to the lidocaine were included in the pairwise meta-regression model. The year of publication was the co-variable in the model, and the outcome variable was the log odds ratio between the placebo and the lidocaine.

The results of meta-regression indicate that the year of publication does not affect the point estimate (Table 8.7). Over the 22 years of trial conducting, the effect estimate for the odds of pain in the lidocaine group compared to placebo was constant over time and the effect of the year of publication was not statistically significant (Table 8.7).

Figure 8-8 represents the bubble plot from the fixed effect meta-regression. Based on the meta-regression, the constancy assumption does hold, and the treatment difference between the placebo and the lidocaine is constant over time.

Figure 8.9 would represent the bubble plot if the constancy were assumed and without any further adjustments. Results in both figures are similar.

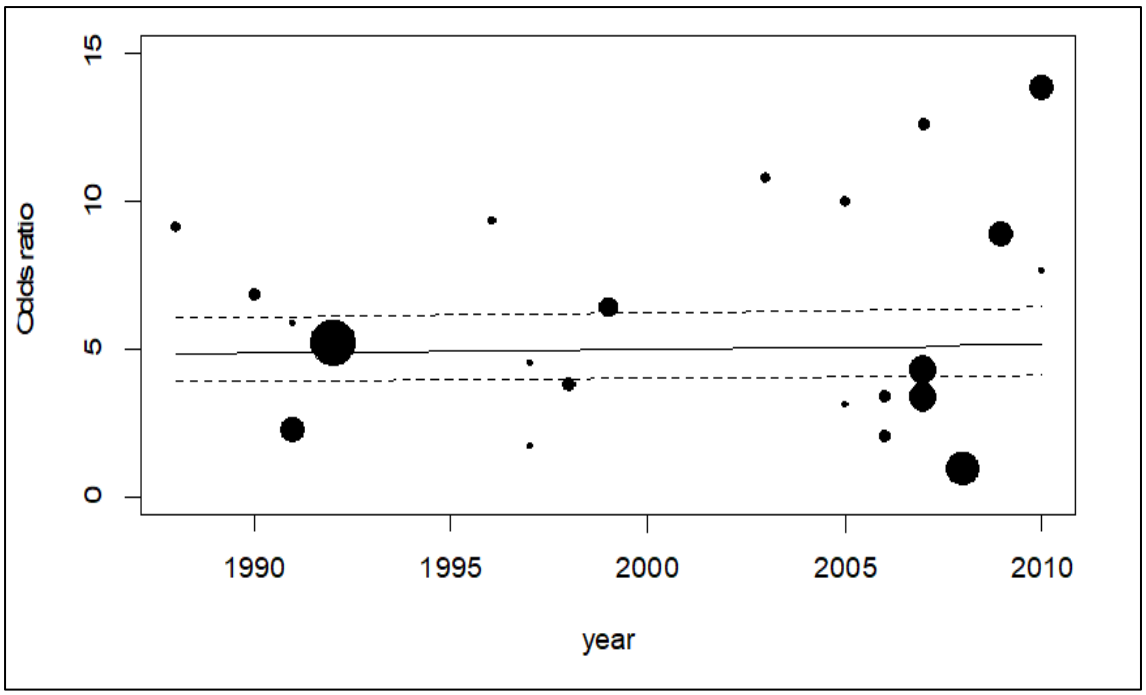

Figure 8-8 Bubble plot for the changes in the odds ratio per year (Constancy not assumed)

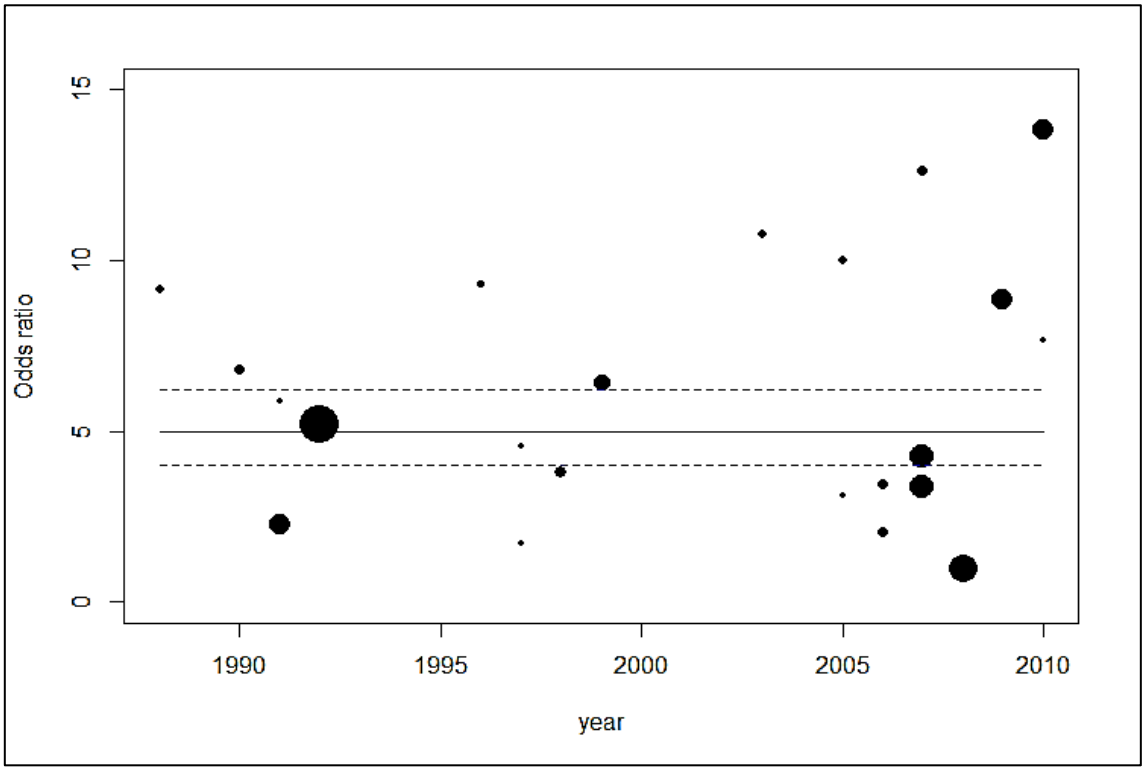

Figure 8-9 Bubble plot for changes in odds ratio (constancy assumed)

Table 8.8 Estimates of NI margin using the adjusted and unadjusted methods

Year	OR (π_P/π_C) 95% CI	M1	M2(1/2 log M1)
Constancy assumed (estimate from meta-analysis), not adjusted for the year			
2018	5.16 (4.14; 6.42)	4.14	2.03
2019	5.16 (4.14; 6.42)	4.14	2.03
2020	5.16 (4.14; 6.42)	4.14	2.03
2025	5.16 (4.14; 6.42)	4.14	2.03
2030	5.16 (4.14; 6.42)	4.14	2.03
Constancy not assumed (estimate from the meta-regression), adjusted for the year			
2018	5.26 (4.22; 6.55)	4.22	2.05
2019	5.27(4.23; 6.57)	4.23	2.06
2020	5.28 (4.24; 6.58)	4.24	2.06
2025	5.40 (4.30; 6.68)	4.30	2.07
2030	5.43 (4.36; 6.77)	4.36	2.09

Suppose in 2020, a new treatment (test) is developed to reduce the pain intensity during propofol injection. To conclude the non-inferiority of the new treatment compared to lidocaine, an NI trial will be designed and conducted in 2020

The null hypothesis: H0: $Odds\ ratio\ \pi_T/\pi_C \geq \delta$ (8.5)

The alternative hypothesis: Ha: $Odds\ ratio\ \pi_T/\pi_C < \delta$ (8.6)

Where δ is the percentage (50%) of the lower limit of the 95% CI between the placebo and active control (lidocaine), C is the active control (lidocaine), T is the test treatment (test), and P is the placebo.

The prediction by year was calculated from the R predict commands and the 95% CI was calculated using the standard error (SE) of 0.11 from Table 8.7

95% CI= Mean difference (in specific year) ± 1.96× 0.11

Table 8.8 illustrates the calculations of the NI margins using the unadjusted 95% CI (assuming the constancy) and the 95% CI from the meta-regression (the constancy does not hold). There was a slight increase in the NI margin even though the constancy assumption over the 22 years

was evident. Increasing the odds ratio means an increase in the pain intensity, which means a decrease in the efficacy of the active control (lidocaine) to reduce the pain compared to placebo.

Hypothetically, an NI trial could be designed in 2020, using either the adjusted margin of 2.06 or the non-adjusted margin; the NI margin 2.03 could be used to calculate the sample size.

Using the formula for sample size calculation (Wang, Chow, & Li, 2002)

$$nT = nC = \left(\frac{\left[Z_{1-\beta}+Z_{1-\frac{\alpha}{2}}\right]^2}{(\log d)^2}\right)\left(\frac{1}{\pi_T(1-\pi_T)}+\frac{1}{\pi_c(1-\pi_c)}\right) \qquad (8.7)$$

1- With the unadjusted NI margin of 2.03, using the formula (8.7), where d is the NI margin (d = 2.03), Type I error of 0.025 and Type II error of 0.1, π_T is the proportion in the treatment group and π_C is the proportion of the control group. The sample size assumes equal relative effects between both treatments (rate of failure of both groups = 0.15) based on the relative effect of the lidocaine (Figure 8.6)

$$\frac{2(1.96+1.282)^2/(log2.03)^2}{(0.15\times0.85)} = 329$$

$$nT = nC = 329$$

2- With the adjusted margin of 2.06, using the formula (8.2) NI margin (odds = 2.06), Type I error of 0.025 and Type II error of 0.1, the sample size assuming equal relative effects between both treatments (rate of failure of both groups = 0.15) based on the relative effect of the lidocaine (Figure 8.6)

$$\frac{2(1.96+1.282)^2/(log2.06)^2}{(0.15\times0.85)} = 316$$

$$nT = nC = 316$$

In the year 2025, the sample size will be reduced to 310 participants per arm and in 2030 the sample size will be 305 participants per arm.

8.4.3 Analysis of non-inferiority trial based on the unadjusted NI margin >2.03

Suppose the NI trial is conducted in 2020 with a sample size of 329 participants in each arm based on NI margin of odds ratio less than 2.03. The aim is to confirm the non-inferiority of the test treatment (Test) compared to lidocaine with NI margin less than 2.03.

The null hypothesis： H0: the upper limit of the 95 CI(π_T/π_C) $\geq$ 2.03

The alternative hypothesis： Ha: the upper limit of the 95 CI $\pi_T/\pi_C < 2.03$

With the constancy assumption holding for this review, the NI margin < 2.03 will be used for both the network meta-analysis (unadjusted analysis) and the network meta-regression (for the year 2020). A network composed of the three treatments (lidocaine, placebo, and test treatment) will be formulated (Figure 8.10). Network meta-analysis (NMA) will be used to compare the efficacy of the test treatment compared to the placebo without any further adjustment. A network meta-regression (NMR) will be used to assess the efficacy of the test treatment compared to placebo in 2020. The gemtc R package for Bayesian network meta-analysis (Valkenhoef & Kuiper, 2016) will be used in the analysis for both the NMA and NMR. Markov Chain Monte Carlo (MCMC) simulation method was used to calculate the posterior distributions. The iterations were 80,000, with one thinning interval, four chains, and a sample size per chain of 40,000, all are the same as those used by Schmidli et al. (2012). Normal half priors used were the same as those used by Schmidli et al. (2012).

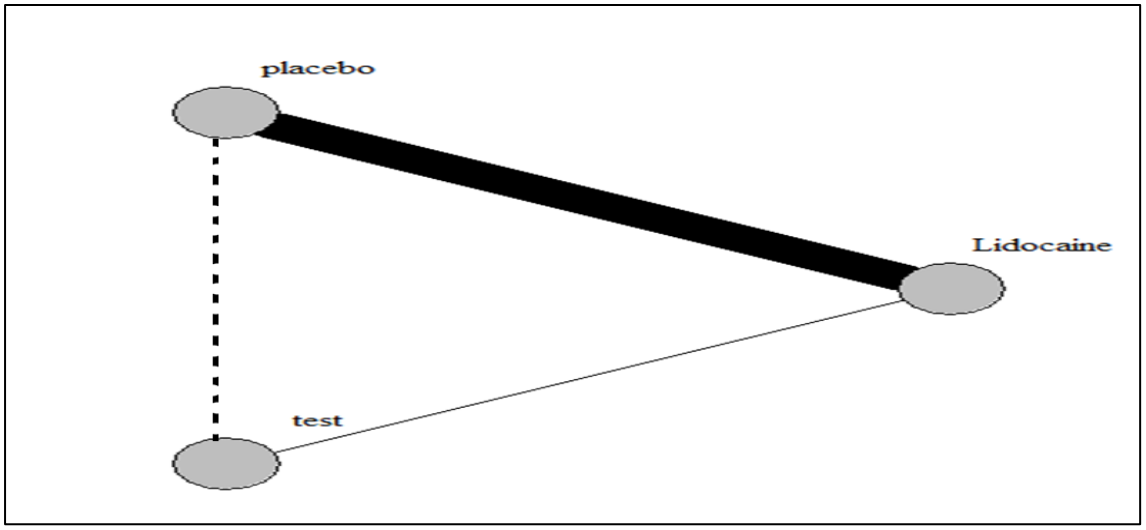

Figure 8-10 Network of lidocaine, placebo and the test treatment

(Thickness of lines represents the number of trials), circles represent the included treatments, the solid line represents direct comparisons, and the dash line represents the indirect comparison)

Table 8.9 Comparison of the odds ratio between the placebo and test treatment assuming the constancy

Risk of failure (π_T)	$OR(\pi_T/\pi_{C)}$ from NI trial (2020) (95% CrI),SE=0.22	NMA, OR (π_P/π_T) (95% CrI),SE= 0.22	NMR, OR (π_P/π_T) in (2020) (95% CrI), SE= 0.41
15.00%	1.0 (0.65; 1.53)	5.31 (3.29; 8.62)	6.28 (2.87; 14.00)
17.00%	1.10 (0.72; 1.67)	4.87 (3.02; 7.87)	5.75 (2.59; 12.9)
19.00%	1.32 (0.88; 2.00)	4.01 (2.53; 6.39)	4.70 (2.13; 10.50)
20.00%	1.40 (0.94; 2.11)	3.80 (2.41; 6.00)	4.40 (1.99; 9.76)
21.00%	1.51 (1.01; 2.27)	3.54 (2.23; 5.54)	4.14 (1.89; 9.45)
23.00%	1.68 (1.13; 2.50)	3.18 (2.01; 4.99)	3.74 (1.70; 8.29)
25.00%	1.89 (1.28; 2.81)	2.83 (1.80; 4.41)	3.27 (1.52; 7.34)
30.00%	2.41 (1.65; 3.55)	2.20 (1.42; 3.41)	2.56 (1.19; 5.50)
35.00%	3.05 (2.11; 4.47)	1.75 (1.12; 2.69)	2.04 (0.94; 4.45)
40.00%	3.75 (2.59; 5.49)	1.42 (0.91; 2.19)	1.67 (0.77; 3.63)

NMA: network meta-analysis, NMR: network meta-regression, the π_t/π_c refers to the odds ratio between the test treatment and the active control, the π_p/π_t refers to the odds ratio between the placebo and the test treatment odds ratio >1 indicates worse outcome (high pain intensity).
Light grey = Failure to conclude NI of T versus C, medium grey = C is superior to T, dark grey =T is not superior to placebo

Table 8.9 illustrates the results of using the unadjusted margin. When assuming the constancy, a non-inferiority margin of 2.03 will be used with a sample size of 329 participants in each arm. The non-inferiority of the test treatment compared to lidocaine was established up to a failure rate (failure to reduce pain) of 19% compared to a failure rate of 15% in the lidocaine group. The non-inferiority could not be established with a failure rate of 20%. Moreover, the inferiority of the test treatment compared to the active control (lidocaine) was evident at a failure rate of 21%. With a failure rate of 35%, the superiority of the test treatment compared to placebo could not be established using the network meta-regression (adjusting for time) or with a failure rate of 40% in the case of network meta-analysis (no adjusting for time).

8.4.4 Analysis of non-inferiority trial based on the adjusted NI margin >2.06

Suppose the NI trial was conducted in 2020 with a sample size of 320 participants in each arm based on NI margin of OR < 2.06. The aim is to confirm the non-inferiority of the test treatment (Test) compared to lidocaine with NI margin of 2.06

$$\text{The null hypothesis：}\ \text{H0: the upper limit of the 95 CI}\ \pi_T/\pi_C \geq 2.06$$

$$\text{The alternative hypothesis：}\ \text{Ha: the upper limit of the 95 CI}\ \pi_T/\pi_C < 2.06$$

The network meta-analysis and network meta-regression were conducted using the gemtc R package for Bayesian network meta-analysis (Valkenhoef & Kuiper, 2016). Markov Chain Monte Carlo (MCMC) simulation method was used to calculate the posterior distributions. The iterations were 80,000, with one thinning interval, four chains, and a sample size per chain of 40,000, all are the same as those used by Schmidli et al. (2012). Vague priors used were the same as those used by Schmidli et al. (2012).

The network meta-analysis will compare the three treatments together with no consideration for the time either in the designing phase (unadjusted NI margin) or the analysis phase (no Covariates included in the model) and will evaluate the efficacy of the test treatment compared to placebo in general (without adjustment). The network meta-regression was used to evaluate the efficacy of the test treatment compared to placebo in 2020.

When adjusting for time, a non-inferiority margin of 2.06 will be used with a sample size of 316 participants in each arm.

Table 8.10 illustrates the results of using the adjusted margin. The non-inferiority of the test treatment compared to lidocaine was established up to failure rate (failure to reduce pain) of 19% compared to a failure rate of 15% in the lidocaine group. The non-inferiority could not be established with a failure rate of 20%. Moreover, the inferiority of the test treatment compared to the active control (lidocaine) was evident at a failure rate of 21%. With a failure rate of 35% the superiority of the test treatment compared to placebo could not be established using the network meta-regression (adjusting for time) or with a failure rate of 40% in the case of network meta-analysis (no adjusting for time).

Table 8.10 Comparison of the odds ratio between the placebo and test treatment, the constancy not assumed (NI margin =2.06)

Risk of failure (π_T)	$OR(\pi_T/\pi_{C)}$ from NI trial (2020) (95% CrI)	NMA, OR (π_P/π_T) (95% CrI)	NMR, OR (π_P/π_T) in (2020) (95% CrI)
15.00%	1.0 (0.64; 1.55)	5.31 (3.27; 8.67)	6.24 (2.80; 14.10)
17.00%	1.15 (0.75; 1.76)	4.63 (2.86; 7.48)	5.42 (2.47; 12.30)
19.00%	1.34 (0.88; 2.04)	3.98 (2.51; 6.38)	4.70 (2.10; 10.60)
20.00%	1.42 (0.94; 2.14)	3.75 (2.49; 6.40)	4.61 (2.06; 10.30)
21.00%	1.50 (1.00; 2.27)	3.54 (2.22; 5.66)	4.13 (1.88; 9.19)
23.00%	1.68 (1.13; 2.52)	3.18 (2.00; 5.01)	3.73 (1.69; 8.44)
25.00%	1.86 (1.25; 2.80)	2.85 (1.80; 4.50)	3.63 (1.51; 7.54)
30.00%	2.41 (1.64; 3.59)	2.21 (1.41; 3.43)	2.60 (1.17; 5.83)
35.00%	3.03 (2.07; 4.49)	1.75 (1.12; 2.71)	2.07 (0.94; 4.43)
40.00%	3.78 (2.60; 5.56)	1.41 (0.90; 2.19)	1.64 (0.75; 3.62)

NMA: network meta-analysis, NMR: network meta-regression, the π_t/π_c refers to the odds ratio between the test treatment and the active control, the π_p/π_t refers to the odds ratio between the placebo and the test treatment odds ratio >1 indicates worse outcome (high pain intensity), SE is the standard error.
Light grey = Failure to conclude NI of T versus C, medium grey = C is superior to T, dark grey =T is not superior to placebo

8.4.5 The effect of different percentages of M1 to set M2

When the constancy assumption holds, both the adjusted and the unadjusted margin yield the same results. The sample size was larger using the unadjusted margin. In the case of constancy assumption hold, using the results from a pairwise meta-analysis of the placebo-controlled trials will lead to unbiased results without the need for any further adjustment. In the case of constancy assumed (NI margin =2.03), the difference in the failure rate between the conclusion of non-inferiority and the failure to conclude superiority was 16 points using the 50% of M1, which was similar using the adjusted margin of 2.06 (50% of M1). The difference was reduced to 15 points using the 60% of M1 in both cases and by the time of using the whole M1 as NI margin the difference was reduced to five points in the unadjusted case and four points with the adjusted margin. Table 8.11 and Figures 8.11 and 8.12 compare the two different margins when the constancy assumption holds (detailed tables are presented in Appendix E).

Table 8.11 Comparison between the adjusted and unadjusted margin when the constancy holds

	Constancy Assumed	Adjusted for time
M2 = 50% M1		
NI margin	2.03	2.06
Sample size $n1 = n2$	329.00	316.00
NI established up to	Failure rate = 19.00%	Failure rate =19.00%
NI cannot be concluded	Failure rate = 20.00%	Failure rate =20.00%
Lidocaine is superior to test treatment	Failure rate = 21.00%	Failure rate =21.00%
Test treatment not superior to placebo (2020)	Failure rate = 35.00%	Failure rate =35.00%
Test treatment not superior to placebo (NMA)	Failure rate = 38.00%	Failure rate =38.00%
M2 = 60% M1		
NI margin	2.35	2.38
Sample size $n1 = n2$	227.00	220.00
NI established up to	Failure rate = 20.00%	Failure rate =20.00%
NI cannot be concluded	Failure rate = 21.00%	Failure rate =21.00%
Lidocaine is superior to test treatment	Failure rate = 23.00%	Failure rate =23.00%
Test treatment not superior to placebo (2020)	Failure rate = 35.00%	Failure rate = 35.00%
Test treatment not superior to placebo (NMA)	Failure rate = 37.00%	Failure rate = 36.00%
M2 = 70% M1		
NI margin	2.70	2.75
Sample size $n1 = n2$	167.00	162.00
NI established up to	Failure rate = 21.00%	Failure rate =21.00%
NI cannot be concluded	Failure rate =22.00%	Failure rate = 22.00%
Lidocaine is superior to test treatment	Failure rate =25.00%	Failure rate = 25.00%
Test treatment not superior to placebo (2020)	Failure rate = 33.00%	Failure rate = 33.00%
Test treatment not superior to placebo (NMA)	Failure rate = 38.00%	Failure rate =35.00%
M2 = 80% M1		
NI margin	3.12	3.18
Sample size $n1 = n2$	128.00	124.00
NI established up to	Failure rate =22.00%	Failure rate = 22.00%
NI cannot be concluded	Failure rate =23.00%	Failure rate =23.00%
Lidocaine is superior to test treatment	Failure rate =26.00%	Failure rate =26.00%
Test treatment not superior to placebo (2020)	Failure rate = 32.00%	Failure rate =30.00%
Test treatment not superior to placebo (NMA)	Failure rate =33.00%	Failure rate =33.00%
M2 = 90% M1		
NI margin	3.59	3.67
Sample size $n1 = n2$	101.00	98.00
NI established up to	Failure rate = 23.00%	Failure rate =23.00%
NI cannot be concluded	Failure rate =24.00%	Failure rate =24.00%
Lidocaine is superior to test treatment	Failure rate = 27.00%	Failure rate =26.00%
Test treatment not superior to placebo (2020)	Failure rate = 30.00%	Failure rate =30.00%
Test treatment not superior to placebo (NMA)	Failure rate = 32.00%	Failure rate =30.00%
M2 = M1		
NI margin	4.14	4.24
Sample size $n1 = n2$	82.00	80.00
NI established up to	Failure rate =25.00%	Failure rate =24.00%
NI cannot be concluded	Failure rate =26.00%	Failure rate =25.00%
Lidocaine is superior to test treatment	Failure rate =29.00%	Failure rate =28.00%
Test treatment not superior to placebo (2020)	Failure rate =30.00%	Failure rate =28.00%
Test treatment not superior to placebo (NMA)	Failure rate =30.00%	Failure rate =28.00%

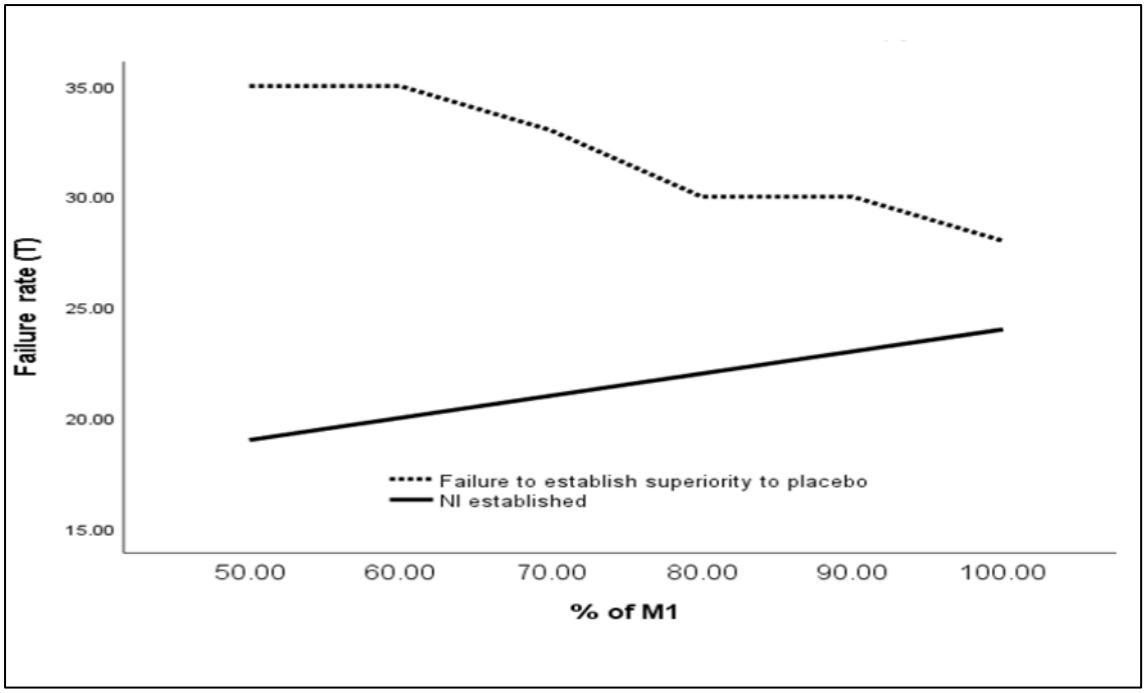

Figure 8-11 Comparison between the different percentages of M1 using the unadjusted margins when the constancy assumption holds

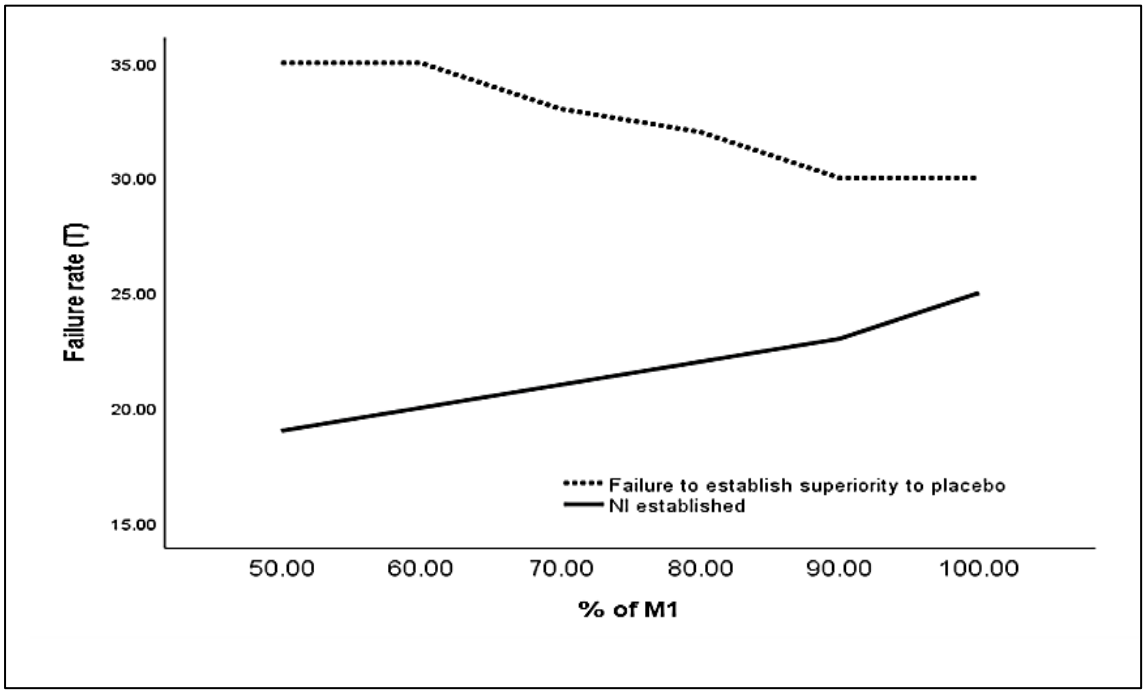

Figure 8-12 Comparison between the different percentages of M1 using the adjusted margins when the constancy assumption holds

Table 8.12 Comparison between the unadjusted and adjusted margins for years 2020, 2025, 2030

	Constancy Assumed	Adjusted for time
2020		
NI margin	2.03	2.06
Sample size $n1 = n2$	329.00	316.00
NI established up to	Failure rate = 19.00%	Failure rate =19.00%
NI cannot be concluded	Failure rate = 20.00%	Failure rate =20.00%
Lidocaine is superior to test treatment	Failure rate = 21.00%	Failure rate =21.00%
Test treatment not superior to placebo (NMR)	Failure rate = 35.00%	Failure rate =35.00%
Test treatment not superior to placebo (NMA)	Failure rate = 38.00%	Failure rate =38.00%
2025		
NI margin	2.03	2.07
Sample size $n1 = n2$	329.00	310.00
NI established up to	Failure rate = 19.00%	Failure rate =19.00%
NI cannot be concluded	Failure rate = 20.00%	Failure rate =20.00%
Lidocaine is superior to test treatment	Failure rate = 21.00%	Failure rate =21.00%
Test treatment not superior to placebo (NMR)	Failure rate = 35.00%	Failure rate = 35.00%
Test treatment not superior to placebo (NMA)	Failure rate = 38.00%	Failure rate = 38.00%
2030		
NI margin	2.03	2.09
Sample size $n1 = n2$	329.00	305.00
NI established up to	Failure rate = 19.00%	Failure rate =19.00%
NI cannot be concluded	Failure rate = 20.00%	Failure rate = 20.00%
Lidocaine is superior to test treatment	Failure rate = 21.00%	Failure rate = 21.00%
Test treatment not superior to placebo (NMR)	Failure rate = 30.00%	Failure rate = 30.00%
Test treatment not superior to placebo (NMA)	Failure rate = 38.00%	Failure rate =38.00%

Table 8.12 illustrates the changes of the treatment effect of the test treatment compared to the lidocaine and the placebo in the years 2020, 2025 and 2030. From the table there were no differences between the treatment effect assuming the constancy and after adjusting for the time. When the constancy is established the use of the proposed method (the adjusted for time method) leads to the same results. Moreover, using the adjusted margins leads to reduction of the sample size needed to establish the non-inferiority.

The use of the unadjusted margin leads to a larger sample size and smaller NI margin which could be considered as conservative (chance of concluding the non-inferiority of an inferior treatment is low) method that could lead to the failure to conclude the NI of an actually non-inferior test treatment.

8.5 Summary and recommendations

In this chapter, pairwise meta-regression was proposed as a method to assess the constancy and set an adjusted NI margin using the year of trial conducting or (publication) as a covariate in the model. The network meta-regression was used in the analysis phase to assess the efficacy of the test treatment compared to placebo in the year of NI trial conducting. Two reviews from the Cochrane reviews discussed in Chapters 5 and 6 were used to validate the proposed method.

In the first example, the constancy did not hold; the treatment difference between the active control (atorvastatin) and the placebo decreased each year. The difference between the unadjusted margin of -12.25 and the adjusted margin of -8.43 in 2020 was 3.82 points, which increased to 5.08 points in the year 2025, and by 2030 the difference increased to 6.35 points.

Using the unadjusted margin of -12.25 led to a smaller sample size of 21 participants compared to 43 participants in 2020 with the adjusted margin. Moreover, the distance between the conclusion of non-inferiority of the test compared to active control and the failure to conclude the superiority of the test treatment compared to the placebo was greater (8.5 points) with the adjusted margin of 50% of M1 (-8.43) compared to six points using the unadjusted margin 50% M1(-12.25). The difference was the same as when using 70% of the M1 of the adjusted margin (Figures 8.5 and 8.6).

The use of a percentage of the M1 in the case of the unadjusted margin could preserve some of the efficacy of the test treatment compared to the active control and partially decreased the chance of conclusion of non-inferiority of an inferior treatment in 2020. However, this was not constant, as by the year 2030 the use of 50% of M1 as a NI margin did not protect against the conclusion of non-inferiority of an already inferior treatment. When the treatment effect of test treatment was 14.5, the non-inferiority of the test treatment compared to placebo was established, while in fact the test treatment was not superior to placebo.

It should be noticed that according to the results from Chapter 5, the relation between the year and the point estimates was not constant. It varies between a positive and negative correlation, with the degree of correlation ranging from strong to weak. That means the use of the generic 50% without any further assessment of the constancy could lead to biased results if the effect of the treatment improved over time.

In the first example, by using the adjusted margin the chosen value of M2 could be from 50 % of M1 to 100% of M1 based only on clinical judgement. Using the unadjusted margin, the clinicians would be forced to choose a NI margin of 50 % or less from M1. In the example above, the use of up to 70% of the adjusted M1 could be considered more appropriate, without any fears of violation of the constancy assumption (since the M1 is already based on the adjusted margin).

The situation was different in the second example, where the constancy was assessed and held over time. The results using the adjusted and unadjusted methods were almost similar. Also, even with the use of M1 as a whole as NI margin, the distance between the conclusion of non-inferiority and failure to conclude superiority to placebo was still stable. The use of the adjusted margin led to smaller sample size and wider margin with the same power to conclude the non-inferiority of the test treatment compared to the unadjusted margin. That means the use of the unadjusted margin leads to more conservative results that could lead to failure to conclude the non-inferiority of an actual non-inferior treatment.

The strength of the proposed method is that it works based on the relation between the treatment estimate and the year of publication. When the relation was strong negative, the adjusted margin was smaller and the sample size was larger using the adjusted margin. While in the second example, where the correlation was weak positive, the adjusted margin was larger and the sample size was smaller than the unadjusted ones. In other words, in comparison with the traditional methods of using the 50% of M1 to protect the assumption of the constancy, the adjusted method using the pairwise meta-regression worked on the base of the magnitude and the direction of the relation between the treatment and the time, not on a fixed percentage of M1.

Designing and conducting NI trials is not straightforward. To reduce the chances of the conclusion of non-inferiority of an already inferior treatment, the constancy should be assessed not assumed. Pairwise meta-regression should be used in the designing phase to assess the constancy of the treatment effect between the placebo and the active control. Based on the results of this assessment, both the NI margin and sample size for the future non-inferiority trial should be determined based on the year of trial conducting. The percentage of M1 that will be used to construct M2 should be based on clinical judgement, not only to secure the constancy assumption, and should be a fraction of the adjusted margin. The chance of

conclusion of non-inferiority of an already inferior treatment was reduced significantly using the adjusted margin compared to the unadjusted margin, especially when the year differences between the NI trial and the historical trials increased.

An important point that should be considered is the fact that the time between the trial design and analysis could vary from one year to up to five years or more. In this case, the year of trial analysis should be used to set the NI margin, not the year of trial design. In both examples used in this chapter, the NI trial was designed in 2018 and the year of the prediction was 2020 as this, not 2018, was the year of the analysis. However, in the cases where the analysis year is delayed beyond the planes, the analysis should be further adjusted for this delay whenever appropriate.

Although the use of pairwise meta-regression to assess the constancy will reduce the chances of type I error (by reducing the chance of conclusion of non-inferiority of an already inferior treatment), there are some situations where the pairwise meta-regression cannot be used. Pairwise meta-regression cannot be used if there were no direct placebo-controlled trials that compared the active treatment to placebo or if there was more than one possible active control to assess. In these cases the alternative will be the network meta-regression.

Network meta-regression can be used in the designing phase to assess the sensitivity of the available active controls, to assess the constancy assumption, to set the NI margin and to calculate the sample size for the future non-inferiority trial. The use of network meta-regression in the designing phase is beyond the scope of this thesis.

Another limitation of the use of pairwise meta-regression is the limited power if the number of included trials is less than ten trials (Thompson & Higgins, 2002). This fact could affect the ability of pairwise meta-regression to assess the constancy. Moreover, the use in this chapter of a hypothetical NI trial based on the information from two Cochrane reviews instead of a real NI trial could be considered as a limitation of this study.

As recomendations, in the designing of an NI trial, the NI margin should be adjusted for time regardless of whether the constancy holds or not; the statistical M1 should be based on the adjusted NI margin, while the fraction of M1 to formulate M2 should be a matter of clinical judgement and based on the adjusted margin M1.

In conclusion, in the case of indirect comparison in general and specifically in any NI trial, the constancy should be assessed not just assumed. Pairwise meta-regression was proposed as a possible solution to adjust for time and is considered as the method of choice for assessing the constancy, setting a non-inferiority margin, and calculating the sample size in the designing phase of the trial. In the analysis phase of the trial, to assess the efficacy of the test treatment compared to placebo, a network meta-regression could be used, adjusted for time. Network meta-regression may provide a solution for the cases where there are no placebo-controlled trials or where there is more than one active control treatment. In conclusion, the proposed method works effectively both in cases when the constancy does not hold, as in the first example, and when the constancy holds (second example).

In the next chapter, final discussion and conclusions will be presented along with recommendations regarding the setting of the NI margin from the indirect comparison.

Chapter 9 Discussion and Conclusion

9.1 Introduction

In medical practice, the superior placebo-controlled randomised trials are the standard to establish the efficacy of a treatment, compared to the placebo group (Fisher, 1999). However, due to changes in medical practice, changes in the patient population and ethical concerns, it has become challenging to apply placebo-controlled trials to test a new treatment. In this situation, NI trials are the alternative to superiority trials. NI trials depend on indirect information from the available historical placebo-controlled trials to establish the superiority of the tested treatment to the putative placebo and from that to conclude the non-inferiority with the active control (D'Agostino et al., 2003).

The three critical assumptions regarding the conducting and analysis of NI trials are: A. assay sensitivity, B. bias minimising (bio-creep and placebo creep) and C. Constancy assumption. These three key assumptions are needed due to the use of indirect comparison between the NI trial and available historical placebo-controlled trials (S. A. Julious, 2011).

Reflecting the challenges and considerations regarding NI trials, the aim of this thesis was to quantify adjusted non-inferiority margins when using retrospective data. The objectives of this thesis are:

- To investigate the methodological and regulatory challenges associated with the planning, conducting and reporting of non-inferiority trials.
- To investigate the changes in the placebo and active treatment effects over time and their impact on the design and analysis of NI trials.
- To quantify and model placebo and active treatment responses over time with recommendations for retrospective comparison back to placebo.
- To propose a method for adjusting for time using indirect comparison in NI trials.

Meeting these objectives will lead at the end to the introduction of the most appropriate method to set and analyse NI trials based on the type of available data that will quantify for the changes in the treatment effect while making an indirect comparison.

Chapters 2 and 3 set the scene for this thesis by describing the methodological and regulatory requirement and challenges associated with the design and conducting of the NI trial. The systematic review in Chapter 4 of the published NI trials in the top medical journals in 2015 provided information on how the NI trials are conducted and reported in medical practice and quantified the importance of historical information in the designing and reporting of NI trials.

Chapter 5 and 6 investigated the changes in the treatment effect of placebo and active treatment over time and how these changes could affect the prediction of any historical placebo-controlled trial.

Chapters 7 and 8 reviewed the possible methods for adjusting for a time in the case of the indirect comparisons and proposed a method to set an adjusted NI margin in two case studies of hypothetical NI trials.

This chapter will discuss the main findings from this thesis in the context of the overall thesis aims and objectives. Strengths and limations of this thesis will be presented as well as recommendations regarding the design and analysis of NI trials.

9.2 Main findings

This section will summarise how this thesis addressed the objectives.

9.2.1 Objective one: Investigate the methodological and regulatory challenges associated with the planning, conducting and reporting of non-inferiority trials

To investigate the methodological challenges associated with the design and analysis of non-inferiority trials, a literature review of the assumptions, challenges, and methods regarding NI trials was conducted in Chapter 2.

The main findings from Chapter 2 were that, in medical fields, RCTs involve not only drug trials, but also include different types of comparisons such as assessing a new treatment, comparing surgical and medical approaches, and comparing different doses of the same treatment.

With regard to NI trials, the three main assumptions that should be considered in designing NI trials are assay sensitivity (A), Bias minimising (Placebo creep and Bio-creep) (B), and

Constancy assumption (C). Violations of any of these assumptions will lead to a biased NI margin and possibly the conclusion of non-inferiority of an inferior treatment. Moreover, choosing the appropriate active control and setting the NI margin are the main challenges in designing NI trials.

Regarding the methods for setting the NI margin, the available methods for setting and analysis of the NI margin are the regulatory methods (fixed margin and the synthesis methods) and predictive methods using the network meta-analysis in the analysis phase. In these methods, to control for the changes in the treatment response of the active control (constancy assumption), the regulatory approaches methods use both the statistically calculated margin (M1) and the smaller clinically significant margin (M2). However, none of the available methods adjusts for the changes in the treatment response or any other possible covariates that could be different between the retrospective data from the placebo-controlled trials and the non-inferiority trials.

Different regulatory guidelines were presented in Chapter 3. All of the guidelines set recommendations on the appropriate designing and conducting of non-inferiority trials but do not impose any enforceable legal responsibilities (FDA, 2016). There was an apparent inconsistency between the guidelines that could negatively affect the quality and reporting of NI trials regarding the definitions and population analysis. Moreover, only the fixed margin approach was approved as a preferred method for setting NI trials and is described as a conservative (chance of concluding the non-inferiority of an inferior treatment is low) approach. The FDA non-inferiority guidelines for industry was the most substantial detailed document in terms of describing the design, setting and analysis of NI trials (FDA, 2016). In addition, all guidelines approve the use of NI design for testing the efficacy but not the safety.

Chapter 4 aimed to investigate the conducting, analysis and reporting of NI trials in clinical practice in regard to the regulatory recommendations. A systematic review of NI trials published in 2015 in four top medical journals was conducted.

The main findings were that 37 NI trials were published in the JAMA, BMJ, Lancet, and NEJM, all of which reported the chosen NI margin. The reporting of NI trials was not compatible with the regulatory guidelines, especially in the blinding, the population included in the analysis and reporting and justification of NI margin used. Sixty per cent of the included trials that reported methods for selection NI margin depended on the historical information

alone or in combination with clinical decisions to set the NI margin. Twenty-four per cent of the included trials did not state the reason behind choosing the NI margin.

The importance of historical information in regard to setting the NI margin from indirect comparison was established both from the literature in Chapters 2 and 3 and from practice in Chapter 4.

By the end of Chapter 4, it was concluded that setting the NI margin depends on the available evidence from the historical placebo-controlled trials. Any possible changes in the efficacy of the active control (assay sensitivity, Section 2.4) or in the effect of placebo (placebo creep, Section 2.8) or in the treatment difference between the placebo group and the active treatment group (constancy assumption, Section 2.5) will lead to a biased NI margin, which could lead to a biased conclusion. The next step was to assess the changes of the treatment effect over time.

9.2.2 Objective two: To investigate the changes in the placebo and active treatment effects over time and their impact on the design and analysis of NI trials

Aiming to investigate the changes in the treatment effect of the placebo group (assess the placebo creep) and active treatment over time, an overview of Cochrane reviews of placebo-controlled trials was conducted in Chapter 5. The correlations between the effect size of a placebo group, active treatment group, and the treatment difference (SMD) and the year of publication were obtained. Besides, the correlations between the sample size and year of publication were obtained. By the end of Chapter 5, it was concluded that the correlations between the treatment difference between placebo group and active control group varied from strong negative to strong positive. The sample size of a trial is positively correlated with the year of publication (R = 0.2), the median placebo effect has a weak positive correlation with the year of publication (R = 0.05), the median active control has a weak negative correlation with year of publication (R = -0.04), and for the treatment difference (SMD) the median correlation was -0.11.

Even though these median correlations are considered weak correlations, the results indicate that, overall, around 58.5% of included reviews had moderate to strong negative correlations regarding the treatment difference between the active control and the placebo. It was possible

to confirm that the effect size of active control and placebo group and the treatment difference between the two are not constant over time. These changes are due to improvement in the placebo effect and decrease in the active treatment effect, which serve to decrease the effect size of the treatment difference between the two.

In the case of NI trials, this conclusion means that the constancy assumption about the treatment difference between the active control and placebo should not be assumed; instead, it should be assessed first, and then the NI margin should be calculated according to the constancy assumption. Also, setting NI margin without adjusting for these changes will lead to a biased estimate from the NI trial that could conclude either the non-inferiority of an inferior treatment or fail to conclude the non-inferiority of an effective treatment. By the end of Chapter 5, two important questions had been raised: first, whether the fixed effect model will be more appropriate for use than a random model; second, what will be the effect of time in the prediction of any future trial? Both questions were answered in Chapter 6.

9.2.3 Objective three: To quantify and model placebo and active treatment responses over time with recommendations for retrospective comparison back to the placebo

To quantify and model the changes in the placebo and active control responses over time, the standardised mean difference for the treatment response between the placebo group and the active treatment from 224 meta-analyses that were reviewed in Chapter 5 were used to build a regression model in Chapter 6.

The predictive power of the model was 55.9%. The three main variables that affected the estimate of any future trial were the point estimate from the meta-analysis of previous trials, the year difference in the meta-analysis, and the year of the predicted trial. Increasing one unit in the estimate from the meta-analysis of the historical trial will lead to an increase in the predicted estimate of the future trial by 0.92. For the year difference in the meta-analysis, increasing the year difference will reduce the predicted estimate of the predicted trial; for each increase in year difference the predicated estimate will be reduced by 0.005. For the year of prediction, for each year increase in the prediction, the predicted estimate of the future trial will be reduced by 0.015. As mentioned earlier, all three variables together explain only 55.9% of the variability of the model. That means the historical data incorporated in a meta-analysis explain only 55.9% of the predicted estimate of any future trial. Using this historical estimate

without further adjustment will lead to biased results, especially when the time differences between the trials were high, which highlights the need for a method to set NI trials that incorporates time (adjust for the time) of trial conducting in the indirect setting of the NI margin. By the end of Chapter 6, the importance was confirmed of including time in the analysis of NI trials. This inclusion of time will not reflect the changes in time only but could also reflect changes in the population characteristics, treatment protocol, and any other changes that can to be measured from studying the demographics or the characteristics of the trials.

Another important conclusion from Chapter 6 related to the type of model used. In NI trials, using a random effect model to estimate the effect of active control compared to placebo will give more weights to the older heterogeneous small studies, which could lead to a biased estimate of the effect of the active control in the current NI trials.

There was no statistically significant difference between the point estimates from the fixed and random models. However, the 95% CI boundaries from the random effects model were statistically significantly wider and different from those of the fixed effect model. In terms of NI trials, the focus was on the 95% CI boundaries, not the point estimate, and for that reason, the use of a fixed effect model in the case of indirect comparison from the meta-analysis of historical trials was found to be most appropriate in the case of NI trials. By the end of Chapter 6, it was concluded that the constancy assumption cannot be assumed; instead, it needs to be assessed.

9.2.4 Objective Four: To propose a method for adjusting for time using indirect comparison in NI trials

To develop a method that could incorporate time in the setting and analysis of NI trials, in Chapter 7 of this thesis, possible ways for adjusting for covariates were reviewed, either by using individual patient data (IPD) or aggregated data (AD). Criteria for the method of adjustment were developed in Chapter 7 based on the conclusions from the previous chapters. These criteria include the ability to adjust for covariates (time), ability to be used in the designing and the analysis phase of the NI trial, to use aggregated data, to assess the sensitivity of the active control, and to offer flexibility and computational ease of use.

Pairwise meta-regression was proposed as a new method to assess the constancy assumption and to set the NI margin in the case where the constancy assumption does not hold. Network meta-regression could be used to assess the constancy assumption and to set the NI margin in the case where the constancy assumption does not hold; additionally, the network meta-regression approach can incorporate different treatments in one network and adjust for any possible co-variables at the same time.

In Chapter 8, pairwise meta-regression was applied in the designing phase of the NI trial and network meta-regression in the analysis phase in two different scenarios of non-inferiority trials. In the first case the constancy assumption did not hold. In this case, using pairwise meta-regression, it was possible to assess the constancy, set the adjusted NI margin, and calculate the sample size of a future NI trial. Network meta-regression was used to assess the efficacy of the test treatment compared to the placebo in the year of NI trial conducting. The obtained results indicate that using the unadjusted margin when the constancy does not hold will lead to biased results and the conclusion of non-inferiority of an inferior treatment. The use of a fixed 50% of M1 instead of the whole M1 was partially protective in the case where constancy was not assumed in the year 2020. However, it was not protective in the year 2030, with a 15-year difference between the historical trials and the NI trial.

In the case of the adjusted margin, the use of the whole M1 was away from the false conclusion of non-inferiority, while with an unadjusted margin of 70% of M1 there was a false conclusion of non-inferiority. In the second case, when the constancy assumption held, the results from the adjusted and unadjusted margin were almost similar. Even with the use of 100% of M1, the results were away from the false conclusion of non-inferiority.

The use of the adjusted margin in both cases was protective from the conclusion of non-inferiority of an inferior treatment. Moreover, using the adjusted margin was protective from the possibility of placebo creep since the adjusted margin will adjust the difference between the placebo group and active treatment group. This could be considered as a strength of this method, which works in different ways regardless of whether the constancy is assumed or not, based on the relation between the time and the treatment effect. In the first example, the NI margin was decreased and the sample size was increased progressively each year based on the fact that the correlation between the time and the treatment effect was a strong negative correlation. The situation was different in the second example, where the NI margin was

slightly increased each year and the sample size was slightly decreased because the correlation between the time and the treatment effect was a weak positive correlation. This method takes into consideration the direction and the magnitude of the changes over time by either increasing or decreasing of the treatment difference. As a conclusion from Chapter 8, M2 should be chosen based on the clinical judgement as a percentage of the adjusted margin with the possibility of use between 100% and 50%, or even less based on the clinical judgement, and should not be obtained to assume (protect) the constancy.

The main findings from this thesis could be summarised into these points: setting the NI margin is the main challenge in a non-inferiority trial; in an NI trial, the constancy should not be assumed but instead should be assessed; using a percentage of the active treatment response (M2) cannot guarantee the constancy.

The changes in the treatment effect over time were mainly due to a decrease in active treatment effect and not due to the improvement of placebo effect only. Pairwise meta-regression is considered as a promising method to assess the constancy, protect from the possibility of placebo creep, set the adjusted margin, and calculate the sample size in the designing phase of the NI trial. Network meta-regression should be used in the analysis phase of the NI trial to assess the efficacy of the test treatment compared to placebo on the year of NI trial conducting. The only limitation of the use of pairwise meta-regression and network meta-regression was the limited power when the number of trials included was less than ten.

9.3 Main thesis strengths

Although the importance of historical trials in relation to conducting NI trials has been reported before, this thesis was among the first to comprehensively incorporate evidence on its importance from the literature, regulations and from medical practice.

To the best of my knowledge, this thesis was the first to conclude that the treatment effect is not constant over time. In addition, this thesis was the first to conclude that these changes in the treatment effect were not due only to the improvement of placebo response; instead, it was a combination of changes in the placebo and the active treatment over time. Even though a reasonable number of publications have investigated changes in the placebo effect over time, to the best of my knowledge, this thesis was the first to compare these changes with the changes

in the active control response over time and the changes in the main treatment effect over time using aggregated data from different therapeutic areas. Moreover, it was the first to incorporate the changes in the treatment effect over time into indirect comparison in NI trials specifically.

To the best of my knowledge, this thesis was the first to use a very comprehensive dataset from different therapeutic areas to assess the relation between the time and the treatment effect. The overview of Cochrane reviews that was used for studying the correlations in Chapter 5 and construction of the weighted regression in Chapter 6 were performed specifically to answer the research question of this thesis: How can time affect the setting of the NI margin?). The data for correlation included 692,753 patients from 2364 placebo-controlled trial aggregated in 236 meta-analyses from 44 different Cochrane groups. The data used to build the regression model were obtained from 681,163 patients from 2310 trials aggregated in 224 meta-analyses. Moreover, a standardised scale (standardised mean difference) was used to measure the estimate from this data. For all these reasons, the results from these analyses are generalisable for different therapeutic areas, and any placebo-controlled trials using any scale of measures.

This thesis was the first to estimate the predictivity of the historical trial to estimate a future trial. It was also the first to conclude that an increase of one unit in the estimate from the meta-analysis of the historical trial will lead to an increase of the predicted estimate of the future trial by 0.92. Moreover, this thesis was the first to measure the negative effect of the year difference between the historical trials and the predicted year of future trial on the predicted estimate of the future trial.

Another important finding from this thesis was that only 55.9% of the predicted estimate of any future trial can be predicted from the estimate from the meta-analysis of historical trials for the same treatment after adjusting for the year of publication of the future trial and the year difference in the historical meta-analysis. These results are considered proof of the changes in the treatment effect over time. Usually, these changes are due to population shifts and changes in the treatment protocols and due to the general improvement in the quality of life. All of these causes cannot be measured by themselves but could be adjusted for in any future trial.

To the best of my knowledge, this thesis was the first to propose the use of pairwise meta-regression to set an adjusted NI margin for time in the case of NI trials. Using the pairwise meta-regression, this thesis was able to assess the constancy, set the NI margin and to calculate the sample size in the designing phase of the NI trial. In the analysis phase, this thesis was the first to compare the placebo and the test treatment indirectly in one network adjusted for the time. The proposed method worked effectively both when the constancy did not hold (the atorvastatin case study) and when the constancy did hold (the lidocaine case study). Finally, this thesis was among the first to use the network meta-regression approach to adjust for the changes in time while setting and analysing NI trials. The adjustment was performed both in the designing phase and in the analysis phase.

One of the most important advantages of the method proposed in this thesis is that the use of the adjusted margin will give clinicians more flexibility to set the fraction of M1 to formulate the M2. By using the adjusted margin, any percentage from M1 will be protected from a false conclusion of NI and the clinician can use any percentage from the adjusted margin based on clinical judgement only and avoid the use of the constant 50% percentage recommended in the FDA regulations.

9.4 Limitations

Due to the scope of the thesis aims and objectives, I did not investigate other situations where there are no direct placebo-controlled trials or where there is more than one active control. Moreover, I did not investigate issues related to the setting of the M2 (clinical margin) as one of the challenges associated with NI trials. Although the historical information is important in setting the NI margin, M2 is the actual margin used in the comparison, and the results of any non-inferiority trial will depend on the chosen M2 that is based on the clinical expert's opinion and usually can be changed by changing the clinical protocol used. However, I was able to demonstrate that M2 is not an alternative for the adjustment for time. Even with the adjusted margin, it is still necessary to set M2 to reflect the clinical opinion.

Biocreep is one of the other important challenges in designing and analysis of NI trials. However, due to the scope of this thesis, biocreep was not discussed in detail. Moreover, I did not investigate whether the proposed adjusted methods could be effective in addressing the possibility of biocreep.

Since only published data were used in this research, these results cannot be generalised to non-published data. On the other hand, according to the evidence synthesis, published data is more widely used than non-published data, which means that the data used related more closely to the real situation.

In this research, the year of publication was used as a proxy for the year of trial conducting. This approach could affect the results since trials with negative results tend to be published later than trials with positive results (Rothstein et al., 2006). It was difficult to extract or determine the year of trial conducting, especially for the earlier trials, and for that reason the year reported in the meta-analysis was used as a proxy for the year of the trial conducting.

I was not able to gain access to individual patient data (IPD) and, therefore, in Chapter 7 I could not apply adjusted methods that used IPD, or compare the results from methods that used aggregated data (AD) to those from methods that used IPD, for example, the one and two stage adjusted fixed margin and synthesis methods proposed by Xu et al. (2014).

Furthermore, to test the proposed method I used hypothetical NI trials and not real trials. This could be considered as a limitation to the proposed method. However, the historical data was real data from Cochrane reviews and my aim concerned the design of an NI trial not the analysis and for that reason using the hypothetical NI trial was considered more appropriate. In addition, the method was tested for different years and with different percentages of M1, which could be considered as a strength of this thesis. Finally, usually in clinical trials, the time lag between designing a study and the analysis will vary. Using the year of design as the predicted year for setting the NI margin and sample size calculation could lead to biased results. Estimating the year of analysis and setting the NI margin based on that could be the answer to the problem. However, the changes in the analysis year could lead to changes in the NI margin and the conclusion of NI trial, especially when the constancy assumption does not hold.

9.5 Future work

Based on the results of this thesis, the future planned work will be to extend the work to include the use of network meta-regression in setting and analysis of non-inferiority trials where there is no direct placebo-controlled trial to compare the active control with placebo. The investigation could be extended to other situations where pairwise meta-regression and network

meta-regression cannot be used (if the number of trials is less than ten). Additionally, valid comparisons could be conducted between the IPD approaches for adjusted indirect comparison and the AD approaches and more investigations are needed in regard to the use of fixed or random effects model in setting the NI margin. And finally, future work could investigate whether the proposed approach can address the possibility of biocreep.

9.6 Recommendations

Based on the results of this thesis, it is recommended that, when possible, a placebo arm should be included in the trial design to ensure the sensitivity of the active control, adjust for the constancy, and reduce the risk of biases. The choice of a non-inferiority design should be fully justified.

Additionally, there should be adjustment for time using the pairwise meta-regression regardless of the constancy assumption. In the case where the constancy does not hold, using the adjusted margin will protect from the conclusion of NI of an inferior test treatment. In the case of the constancy assumption, the use of the adjusted margin will reduce the sample size.

It is further recommended to investigate the possible presence of any other effect modifiers (other than time) for any indirect comparison and adjusting for these possible effects, and modifiers should be the standard for any indirect comparison. Moreover, the available regulatory guidelines should include methods for adjusting for indirect comparison when setting the NI margin.

Finally, to ensure that the clinical NI margin M2 is appropriate. M2 should be a fraction of the adjusted NI margin M1 regardless of whether the constancy assumption holds or not. This fraction should be based on the clinical opinion not used as a method to ensure the constancy.

9.7 Overall conclusions

Designing and conducting a non-inferiority trial is associated with methodological, statistical, and regulatory challenges. The main challenge is the need to borrow information from historical trials to conclude the relationship between the putative placebo and the test treatment. To avoid any violation of the methodological assumption regarding NI trials, firm regulatory guidelines to control the conducting and reporting of NI trials are needed. Despite the

methodological and regulatory challenges, NI trials present an excellent alternative to superiority trials when conducting the latter is not possible.

This thesis aimed to investigate the constancy assumption and its effect on setting the NI margin. One of the main conclusions was that the changes that occur in the treatment effect over time are due not only to improvement in the placebo response, but mainly due to a decrease in the efficacy of the active control, both of which will lead to a decrease of the treatment difference between the active control and the placebo. This thesis was also able to confirm that there is a deficiency in reporting the setting and choosing of the NI margin in the published NI trials, which leads to misinterpretation of the results of NI trials.

As a solution for the adjusting for a time in NI trials, this thesis recommends the use of pairwise meta-regression and network meta-regression approaches to assess the constancy assumption and to set and analyse NI trials when the constancy assumption does not hold. Including in the synthesis of this network not only the placebo-controlled trials but also all relevant trials that compare all possible active controls either to each other or to placebo, with all possible treatments, will provide a valid comparison between all active controls and the test treatment and the placebo both directly and indirectly. Moreover, this method will rank the treatments' efficacy based on their relative effectiveness. The network meta-regression approach could also be used in the designing phase to choose the most appropriate active control for a determined sample size of the proposed NI trial and to set the NI margin for the fixed margin approach or synthesis approach. In the analysis phase, network meta-regression will provide a consistent comparison between all available treatments (assay sensitivity), will adjust the estimate for the time (constancy adjustment), and will compare both the direct and indirect evidence (bias minimise) and investigate the presence of bio-creep in the NI trial. With the adjusted NI margin, the chosen M2 was more flexible since even the use of 100% of M1 was protected from the false conclusion of non-inferiority of an inferior treatment. Using the unadjusted margin when the constancy assumption does not hold will restrict the clinical choice to using either 50% or less to avoid the false conclusion of non-inferiority of an inferior treatment.

References

Adams, S., Sekhon, S., & Wright, J. (2015). Lipid-lowering efficacy of rosuvastatin (Review). *Cochrane Database of Systematic Reviews*, (11). https://doi.org/10.1002/14651858.CD008226.pub3.www.cochranelibrary.com

Agid, O., Siu, C. O., Potkin, S. G., Kapur, S., Watsky, E., Vanderburg, D., … Remington, G. (2013). Meta-regression analysis of placebo response in antipsychotic trials, 1970-2010. *American Journal of Psychiatry*, *170*(11), 1335–1344. https://doi.org/10.1176/appi.ajp.2013.12030315

Ardehali, A., Esmailian, F., Deng, M., Soltesz, E., Hsich, E., Naka, Y., … Kobashigawa, J. (2015). Ex-vivo perfusion of donor hearts for human heart transplantation (PROCEED II): a prospective, open-label, multicentre, randomised non-inferiority trial. *Lancet (London, England)*, *385*(9987), 2577–2584. https://doi.org/10.1016/S0140-6736(15)60261-6

Bachelez, H., van de Kerkhof, P. C. M., Strohal, R., Kubanov, A., Valenzuela, F., Lee, J.-H., … Wolk, R. (2015). Tofacitinib versus etanercept or placebo in moderate-to-severe chronic plaque psoriasis: a phase 3 randomised non-inferiority trial. *Lancet (London, England)*, *386*(9993), 552–561. https://doi.org/10.1016/S0140-6736(14)62113-9

Barone, M. A., Widmer, M., Arrowsmith, S., Ruminjo, J., Seuc, A., Landry, E., … Gülmezoglu, A. M. (2015). Breakdown of simple female genital fistula repair after 7 day versus 14 day postoperative bladder catheterisation: a randomised, controlled, open-label, non-inferiority trial. *Lancet (London, England)*, *386*(9988), 56–62. https://doi.org/10.1016/S0140-6736(14)62337-0

Beecher. (1955). THE POWERFUL PLACEBO. *Journal of the American Medical Association*, *159*(17), 1602. https://doi.org/10.1001/jama.1955.02960340022006

Behringer, K., Goergen, H., Hitz, F., Zijlstra, J. M., Greil, R., Markova, J., … Engert, A. (2015). Omission of dacarbazine or bleomycin, or both, from the ABVD regimen in treatment of early-stage favourable Hodgkin's lymphoma (GHSG HD13): an open-label, randomised, non-inferiority trial. *Lancet (London, England)*, *385*(9976), 1418–1427. https://doi.org/10.1016/S0140-6736(14)61469-0

Bensdorp, A. J., Tjon-Kon-Fat, R. I., Bossuyt, P. M. M., Koks, C. A. M., Oosterhuis, G. J. E., Hoek, A., … van Wely, M. (2015). Prevention of multiple pregnancies in couples with unexplained or mild male subfertility: randomised controlled trial of in vitro fertilisation with single embryo transfer or in vitro fertilisation in modified natural cycle compared with intrauterine inse. *BMJ (Clinical Research Ed.)*, *350*, g7771. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4288434&tool=pmcentrez&rendertype=abstract

Bernard, L., Dinh, A., Ghout, I., Simo, D., Zeller, V., Issartel, B., … Mulleman, D. (2015). Antibiotic treatment for 6 weeks versus 12 weeks in patients with pyogenic vertebral osteomyelitis: an open-label, non-inferiority, randomised, controlled trial. *Lancet (London, England)*, *385*(9971), 875–882. https://doi.org/10.1016/S0140-6736(14)61233-2

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*(2), 135–160. https://doi.org/10.1191/096228099673819272

Blonde, L., Jendle, J., Gross, J., Woo, V., Jiang, H., Fahrbach, J. L., & Milicevic, Z. (2015). Once-weekly dulaglutide versus bedtime insulin glargine, both in combination with prandial insulin lispro, in patients with type 2 diabetes (AWARD-4): a randomised, open-label, phase 3, non-inferiority study. *Lancet (London, England)*, *385*(9982), 2057–2066. https://doi.org/10.1016/S0140-6736(15)60936-9

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. John Wiley & Sons, Ltd.

Borzak, S., & Ridker, P. M. (1995). Discordance between meta-analyses and large-scale randomized, controlled trials. Examples from the management of acute myocardial infarction. *Annals of Internal Medicine*, *123*(11), 873–877. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7486471

Bown, S. R. (2003). *Scurvy : how a surgeon, a mariner and a gentleman solved the greatest medical mystery of the age of sail*. Summersdale.

Büller, H. R., Bethune, C., Bhanot, S., Gailani, D., Monia, B. P., Raskob, G. E., … Weitz, J. I. (2015). Factor XI antisense oligonucleotide for prevention of venous thrombosis. *The New England Journal of Medicine*, *372*(3), 232–240. https://doi.org/10.1056/NEJMoa1405760

Burns, N. (Nancy A., & Grove, S. K. (2007). *Understanding nursing research : building an evidence-based practice*. Saunders Elsevier.

Caro, J. J., & Ishak, K. J. (2010). No Head-to-Head Trial? Simulate the Missing Arms. *PharmacoEconomics*, *28*(10), 957–967. https://doi.org/10.2165/11537420-000000000-00000

Chan, A.-W., Tetzlaff, J. M., Altman, D. G., Laupacis, A., Gøtzsche, P. C., Krleža-Jerić, K., … Moher, D. (2013). SPIRIT 2013 Statement: Defining Standard Protocol Items for Clinical Trials. *Annals of Internal Medicine*, *158*(3), 200. https://doi.org/10.7326/0003-4819-158-3-201302050-00583

CHMP. (2005). *GUIDELINE ON THE CHOICE OF THE NON-INFERIORITY MARGIN*, *European Medicines Agency Evaluation of Medicines for Human Use CHMP*. (October 2005), 0–3. https://doi.org/10.1002/sim.2584

Cooper, N. A. M., Clark, T. J., Middleton, L., Diwakar, L., Smith, P., Denny, E., … Daniels, J. (2015). Outpatient versus inpatient uterine polyp treatment for abnormal uterine bleeding: randomised controlled non-inferiority study. *BMJ (Clinical Research Ed.)*, *350*, h1398. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4370502&tool=pmcentrez&rendertype=abstract

Cooper, N. J., Sutton, A. J., Morris, D., Ades, A. E., & Welton, N. J. (2009). Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Statistics in Medicine*, *28*(14), 1861–1881. https://doi.org/10.1002/sim.3594

Cox, T. M., Drelichman, G., Cravo, R., Balwani, M., Burrow, T. A., Martins, A. M., … Puga, A. C. (2015). Eliglustat compared with imiglucerase in patients with Gaucher's disease type 1 stabilised on enzyme replacement therapy: a phase 3, randomised, open-label, non-inferiority trial. *Lancet (London, England)*, *385*(9985), 2355–2362. https://doi.org/10.1016/S0140-6736(14)61841-9

CPMP. (2000). *Points To Consider on Switching Between Superiority and Non-Inferiority Discussion in the Efficacy Working Party (Ewp) Transmission To the Cpmp Release for Consultation Re-Submission To the Efficacy Working Party*. (July), 0–10.

D'Agostino, R. B., Massaro, J. M., & Sullivan, L. M. (2003). Non-inferiority trials: Design concepts and issues - The encounters of academic consultants in statistics. *Statistics in Medicine*, *22*(2), 169–186. https://doi.org/10.1002/sim.1425

DerSimonian, R., & Kacker, R. (2007). Random-effects model for meta-analysis of clinical trials: an update. *Contemporary Clinical Trials*, *28*(2), 105–114. https://doi.org/10.1016/j.cct.2006.04.004

Detollenaere, R. J., den Boon, J., Stekelenburg, J., IntHout, J., Vierhout, M. E., Kluivers, K. B., & van Eijndhoven, H. W. F. (2015). Sacrospinous hysteropexy versus vaginal hysterectomy with suspension of the uterosacral ligaments in women with uterine prolapse stage 2 or higher: multicentre randomised non-inferiority trial. *BMJ (Clinical Research Ed.)*, *351*, h3717. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4512203&tool=pmcentrez&rendertype=abstract

Dold, M., & Kasper, S. (2015). Increasing placebo response in antipsychotic trials: a clinical perspective. *Evidence Based Mental Health*, *18*(3), 77–79. https://doi.org/10.1136/eb-2015-102098

Donegan, S., Welton, N. J., Tudur Smith, C., D'Alessandro, U., & Dias, S. (2017). Network meta-analysis including treatment by covariate interactions: Consistency can vary across covariate values. *Research Synthesis Methods*, *8*(4), 485–495. https://doi.org/10.1002/jrsm.1257

Donegan, S., Williamson, P., D'Alessandro, U., & Tudur Smith, C. (2013). Assessing key assumptions of network meta-analysis: a review of methods. *Research Synthesis Methods*, *4*(4), 291–323. https://doi.org/10.1002/jrsm.1085

Douketis, J. D., Spyropoulos, A. C., Kaatz, S., Becker, R. C., Caprini, J. A., Dunn, A. S., … Ortel, T. L. (2015). Perioperative Bridging Anticoagulation in Patients with Atrial Fibrillation. *New England Journal of Medicine*, *373*(9), 150622051516008. https://doi.org/10.1056/NEJMoa1501035

Dranitsaris, G., Jelincic, V., & Choe, Y. (2011). Meta regression analysis to indirectly compare dalteparin to enoxaparin for the prevention of venous thromboembolic events following total hip replacement. *Thrombosis Journal*, *9*(1), 3. https://doi.org/10.1186/1477-9560-9-3

Eckert, L., & Falissard, B. (2006). *Using meta-regression in performing indirect-comparisons: comparing escitalopram with venlafaxine XR*. *22*(11). https://doi.org/10.1185/030079906X148625

Eckert, L., & Lançon, C. (2006). Duloxetine compared with fluoxetine and venlafaxine: use of meta-regression analysis for indirect comparisons. *BMC Psychiatry*, *6*(1), 30. https://doi.org/10.1186/1471-244X-6-30

Eikelboom, J. W., Anand, S. S., Malmberg, K., Weitz, J. I., Ginsberg, J. S., & Yusuf, S. (2000). Unfractionated heparin and low-molecular-weight heparin in acute coronary syndrome without ST elevation: a meta-analysis. *The Lancet*, *355*(9219), 1936–1942. https://doi.org/10.1016/S0140-6736(00)02324-2

Ellis, S. G., Kereiakes, D. J., Metzger, D. C., Caputo, R. P., Rizik, D. G., Teirstein, P. S., … Stone, G. W. (2015). Everolimus-Eluting Bioresorbable Scaffolds for Coronary Artery Disease. *The New England Journal of Medicine*, *373*(20), 1905–1915. https://doi.org/10.1056/NEJMoa1509038

Enthoven, W., Roelofs, P., Deyo, R., Van Tulder, M., & Koes, B. (2016). Non-steroidal anti-inflammatory drugs for chronic low back pain (Review). *Cochrane Database of Systematic Reviews 2016*, (2), 69. https://doi.org/10.1002/14651858.CD012087.www.cochranelibrary.com

Euasobhon, P., Dej-arkom, S., Siriussawakul, A., Muangman, S., Sriraj, W., Pattanittum, P., & Lumbiganon, P. (2016). Lidocaine for reducing propofol-induced pain on induction of anaesthesia in adults. *Cochrane Database of Systematic Reviews*, *2016*(2). https://doi.org/10.1002/14651858.CD007874.pub2

Everitt, B. S. (2002). The Cambridge Dictionary of Statistics. In *Journal of Chemical Information and Modeling* (2nd ed., Vol. 53). https://doi.org/10.1017/CBO9781107415324.004

Everson-Stewart, S., & Emerson, S. S. (2010). Bio-creep in non-inferiority clinical trials. *Statistics in Medicine*, *29*(27), 2769–2780. https://doi.org/10.1002/sim.4053

FDA. (2010). Guidance for Industry Non-Inferiority Clinical Trials (DRAFT GUIDANCE). *Draft Guidance for Industry*, (February), 50.

FDA. (2016). *Non-Inferiority Clinical Trials to Establish Effectiveness Guidance for Industry Non-Inferiority Clinical Trials to Establish Effectiveness Guidance for Industry*. (November).

Fisher, L. D. (1999). *ADVANCES IN CLINICAL TRIALS*.

Fleming, T. R. (2008). *Current issues in non-inferiority trials ‡*. (January 2007), 317–332. https://doi.org/10.1002/sim

Fleshman, J., Branda, M., Sargent, D. J., Boller, A. M., George, V., Abbas, M., … Nelson, H. (2015). Effect of Laparoscopic-Assisted Resection vs Open Resection of Stage II or III Rectal Cancer on Pathologic Outcomes: The ACOSOG Z6051 Randomized Clinical Trial. *JAMA*, *314*(13), 1346–1355. https://doi.org/10.1001/jama.2015.10529

Flight, L., & Julious, S. A. (2016). Practical guide to sample size calculations: non-inferiority and equivalence trials. *Pharmaceutical Statistics*, *15*(1), 80–89. https://doi.org/10.1002/pst.1716

Fraguas, D., Díaz-Caneja, C. M., Pina-Camacho, L., Umbricht, D., & Arango, C. (2018). Predictors of Placebo Response in Pharmacological Clinical Trials of Negative Symptoms in Schizophrenia: A Meta-regression Analysis. *Schizophrenia Bulletin*. https://doi.org/10.1093/schbul/sbx192

Furukawa, T. A., Cipriani, A., Atkinson, L. Z., Leucht, S., Ogawa, Y., Takeshima, N., … Salanti, G. (2016). Placebo response rates in antidepressant trials: a systematic review of published and unpublished double-blind randomised controlled studies. *The Lancet Psychiatry*, *3*(11), 1059–1066. https://doi.org/10.1016/S2215-0366(16)30307-8

Furukawa, T. A., Cipriani, A., Leucht, S., Atkinson, L. Z., Ogawa, Y., Takeshima, N., … Salanti, G. (2018). Is placebo response in antidepressant trials rising or not? A reanalysis of datasets to conclude this long-lasting controversy. *Evidence-Based Mental Health*, *21*(1), 1–3. https://doi.org/10.1136/eb-2017-102827

Gallahan, W. C., Case, D., & Bloomfeld, R. S. (2010). An analysis of the placebo effect in Crohn's disease over time. *Alimentary Pharmacology & Therapeutics*, *31*(1), 102–107. https://doi.org/10.1111/j.1365-2036.2009.04125.x

GAO. (2010). *New Drug Approval: FDA's Consideration of Evidence from Certain Clinical Trials*. Retrieved from https://www.gao.gov/products/GAO-10-798

Geisler, W. M., Uniyal, A., Lee, J. Y., Lensing, S. Y., Johnson, S., Perry, R. C. W., … Kerndt, P. R. (2015). Azithromycin versus Doxycycline for Urogenital Chlamydia trachomatis Infection. *The New England Journal of Medicine*, *373*(26), 2512–2521. https://doi.org/10.1056/NEJMoa1502599

GLASS, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, *5*(10), 3–8. https://doi.org/10.3102/0013189X005010003

Goldstein, J. N., Refaai, M. A., Milling, T. J., Lewis, B., Goldberg-Alberts, R., Hug, B. A., & Sarode, R. (2015). Four-factor prothrombin complex concentrate versus plasma for rapid vitamin K antagonist reversal in patients needing urgent surgical or invasive interventions: a phase 3b, open-label, non-inferiority, randomised trial. *Lancet (London, England)*, *385*(9982), 2077–2087. https://doi.org/10.1016/S0140-6736(14)61685-8

Gross, J. G., Glassman, A. R., Jampol, L. M., Inusah, S., Aiello, L. P., Antoszyk, A. N., … Beck, R. W. (2015). Panretinal Photocoagulation vs Intravitreous Ranibizumab for Proliferative Diabetic Retinopathy: A Randomized Clinical Trial. *JAMA*, *314*(20), 2137–2146. https://doi.org/10.1001/jama.2015.15217

Guido Schwarzer. (2015). *netmeta package | R Documentation*. Retrieved from https://www.rdocumentation.org/packages/netmeta/versions/0.9-8

Healey, J. S., Hohnloser, S. H., Glikson, M., Neuzner, J., Mabo, P., Vinolas, X., … Connolly, S. J. (2015). Cardioverter defibrillator implantation without induction of ventricular fibrillation: a single-blind, non-inferiority, randomised controlled trial (SIMPLE). *Lancet (London, England)*, *385*(9970), 785–791. https://doi.org/10.1016/S0140-6736(14)61903-6

Higgins, J. P., & Green, S. (2008). Cochrane Handbook for Systematic Reviews of. In *The Cochrane Collaboration*. https://doi.org/10.1002/9780470712184

Hm, W., Jl, T., Xp, L., Jtf, L., Pc, L., Woo, J., & Li, Y. (2009). Acupuncture for stroke rehabilitation (Review). *Cochrane Review*, (1), 1–26. https://doi.org/10.1002/14651858.CD004131.pub3.www.cochranelibrary.com

Hoaglin, D. C., Hawkins, N., Jansen, J. P., Scott, D. A., Itzler, R., Cappelleri, J. C., … Barrett, A. (2011). Conducting Indirect-Treatment-Comparison and Network-Meta-Analysis Studies: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 2. *Value in Health*, *14*(4), 429–437. https://doi.org/10.1016/J.JVAL.2011.01.011

Hróbjartsson, A., & Gøtzsche, P. C. (2001). Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *N Engl J Med*, *344*(21), 1594–1602. https://doi.org/10.1056/NEJM200105243442106

IBM Corp. (2016). *IBM SPSS Statstics for windows*. Armonk, NY: IBM Corp.

ICH. (1998). *ICH E9*.

ICH. (2017). E9 (R1) Estimands and Sensitivity Analysis in Clinical Trials. *Guidance*, *9*(June).

ICH, I. C. on H. (2001). CHOICE OF CONTROL GROUP AND RELATED ISSUES IN CLINICAL TRIALS E10. *Federal Register*, (July), 66(93), 24390-24391.

IntHout, J., Ioannidis, J. P. A., Borm, G. F., & Goeman, J. J. (2015). Small studies are more heterogeneous than large ones: a meta-meta-analysis. *Journal of Clinical Epidemiology*, *68*(8), 860–869. https://doi.org/10.1016/J.JCLINEPI.2015.03.017

Ioannidis, J. P. A., & Lau, J. (2001). Evolution of treatment effects over time: Empirical insight from recursive cumulative metaanalyses. *Proceedings of the National Academy of Sciences*, *98*(3), 831–836. https://doi.org/10.1073/pnas.021529998

Ishak, K. J., Proskorovsky, I., & Benedict, A. (2015). Simulation and Matching-Based Approaches for Indirect Comparison of Treatments. *PharmacoEconomics*, *33*(6), 537–549. https://doi.org/10.1007/s40273-015-0271-1

Jansen, J. P., Fleurence, R., Devine, B., Itzler, R., Barrett, A., Hawkins, N., … Cappelleri, J. C. (2011). Interpreting Indirect Treatment Comparisons and Network Meta-Analysis for Health-Care Decision Making: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 1. *Value in Health*, *14*(4), 417–428. https://doi.org/10.1016/j.jval.2011.04.002

Joura, E. A., Giuliano, A. R., Iversen, O.-E., Bouchard, C., Mao, C., Mehlsen, J., … Luxembourg, A. (2015). A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women. *The New England Journal of Medicine*, *372*(8), 711–723. https://doi.org/10.1056/NEJMoa1405044

Julious, S. (2011). The ABC of non-inferiority margin setting from indirect comparisons. *Pharmaceutical Statistics*, *10*(5), 448–453. https://doi.org/10.1002/pst.517

Julious, S. A., & Wang, S. J. (2008). How Biased are Indirect Comparisons, Particularly When Comparisons are Made over Time in Controlled Trials? *Therapeutic Innovation & Regulatory Science*, *42*(6), 625–633. https://doi.org/10.1177/009286150804200610

Kamper, S. J., & Williams, C. M. (2013). The placebo effect: Powerful, powerless or redundant? *British Journal of Sports Medicine*, *47*(1), 6–9. https://doi.org/10.1136/bjsports-2012-091472

Kaul, U., Bangalore, S., Seth, A., Arambam, P., Abhaychand, R. K., Patel, T. M., … Bahuleyan, C. G. (2015). Paclitaxel-Eluting versus Everolimus-Eluting Coronary Stents in Diabetes. *The New England Journal of Medicine*, *373*(18), 1709–1719. https://doi.org/10.1056/NEJMoa1510188

Kehoe, S., Hook, J., Nankivell, M., Jayson, G. C., Kitchener, H., Lopes, T., … Swart, A.-M. (2015). Primary chemotherapy versus primary surgery for newly diagnosed advanced ovarian cancer (CHORUS): an open-label, randomised, controlled, non-inferiority trial. *Lancet (London, England)*, *386*(9990), 249–257. https://doi.org/10.1016/S0140-6736(14)62223-6

Kent, F., Ambler, G. K., Bosanquet, D. C., Twine, C. P., Bell, R., Bicknell, C. D., … Zayed, H. (2018). The Safety of Device Registries for Endovascular Abdominal Aortic Aneurysm Repair: Systematic Review and Meta-regression. *European Journal of Vascular and Endovascular Surgery*, *55*(2), 177–183. https://doi.org/10.1016/j.ejvs.2017.11.013

Kerr, C. e., Milne, I., & Kaptchuk, T. j. (2008). William Cullen and a missing mind-body link in the early history of placebos. *Jrsm*, *101*(2), 89–92. https://doi.org/10.1258/jrsm.2007.071005

Khan, A., Fahl Mar, K., Faucett, J., Khan Schilling, S., & Brown, W. A. (2017). Has the rising placebo response impacted antidepressant clinical trial outcome? Data from the US Food and Drug Administration 1987-2013. *World Psychiatry*, *16*(2), 181–192. https://doi.org/10.1002/wps.20421

Khan, A., Fahl Mar, K., Schilling, J., & Brown, W. A. (2018a). Does the rising placebo response impact antihypertensive clinical trial outcomes? An analysis of data from the Food and Drug Administration 1990-2016. *PLOS ONE*, *13*(2), e0193043. https://doi.org/10.1371/journal.pone.0193043

Khan, A., Fahl Mar, K., Schilling, J., & Brown, W. A. (2018b). Magnitude and pattern of placebo response in clinical trials of antiepileptic medications: Data from the Food and Drug Administration 1996–2016. *Contemporary Clinical Trials*, *64*(October 2017), 95–100. https://doi.org/10.1016/j.cct.2017.10.017

Khan, A., Fahl Mar, K., Schilling, J., & Brown, W. A. (2018c). Magnitude and Pattern of Placebo Response in Clinical Trials of Oral Antihyperglycemic Agents: Data From the Food and Drug Administration 1999-2015. *Diabetes Care*, dc171316. https://doi.org/10.2337/dc17-1316

Kienle, G. S., & Kiene, H. (1997). the Powerful Placebo Effect- Fact or Fiction, Gunver, 1997.Pdf. *Journal of Clinical Epidemiology*, *50*(12), 1311–1318.

Kirsch, I. (2013). The placebo effect revisited: Lessons learned to date. *Complementary Therapies in Medicine*, *21*(2), 102–104. https://doi.org/10.1016/j.ctim.2012.12.003

Koshi, E., & Short, C. (2007). Placebo Theory and Its Implications for Research and Clinical Practice: A Review of the Recent Literature. *Pain Practice*, *7*(1), 4–20.

Kutner, M. H., Nachtsheim, C., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th editio). McGraw-Hill/Irwin.

Le Henanff, A., Giraudeau, B., Baron, G., & Ravaud, P. (2006). Quality of Reporting of Noninferiority and Equivalence Randomized Trials. *JAMA*, *295*(10), 1147. https://doi.org/10.1001/jama.295.10.1147

Le Page, E., Veillard, D., Laplaud, D. A., Hamonic, S., Wardi, R., Lebrun, C., … Edan, G. (2015). Oral versus intravenous high-dose methylprednisolone for treatment of relapses in patients with multiple sclerosis (COPOUSEP): a randomised, controlled, double-blind, non-inferiority trial. *Lancet (London, England)*, *386*(9997), 974–981. https://doi.org/10.1016/S0140-6736(15)61137-0

Lee, K., Bacchetti, P., & Sim, I. (2008). Publication of Clinical Trials Supporting Successful New Drug Applications: A Literature Analysis. *PLoS Medicine*, *5*(9), e191. https://doi.org/10.1371/journal.pmed.0050191

LeLorier, J., Grégoire, G., Benhaddad, A., Lapierre, J., & Derderian, F. (1997). Discrepancies between Meta-Analyses and Subsequent Large Randomized, Controlled Trials. *New England Journal of Medicine*, *337*(8), 536–542. https://doi.org/10.1056/NEJM199708213370806

Leucht, S., Arbter, D., Engel, R. R., Kissling, W., & Davis, J. M. (2009). How effective are second-generation antipsychotic drugs? A meta-analysis of placebo-controlled trials. *Molecular Psychiatry*, *14*(4), 429–447. https://doi.org/10.1038/sj.mp.4002136

Lewis, J. a. (2001). Switching between superiority and non-inferiority: an introductory note. *British Journal of Clinical Pharmacology*, *52*(3), 221. https://doi.org/10.1046/j.0306-5251.2001.01397.x

Liang, W., Wu, X., Fang, W., Zhao, Y., Yang, Y., Hu, Z., … Zhang, L. (2014). Network Meta-Analysis of Erlotinib, Gefitinib, Afatinib and Icotinib in Patients with Advanced Non-Small-Cell Lung Cancer Harboring EGFR Mutations. *PLoS ONE*, *9*(2), e85245. https://doi.org/10.1371/journal.pone.0085245

Lin, J., Gamalo-Siebers, M., & Tiwari, R. (2016). Non-inferiority and networks: inferring efficacy from a web of data. *Pharmaceutical Statistics*, *15*(1), 54–67. https://doi.org/10.1002/pst.1729

Linde, K., Allais, G., Brinkhaus, B., Fei, Y., Mehring, M., Vertosick, E. A., … White, A. R. (2016). Acupuncture for the prevention of episodic migraine. *Cochrane Database of Systematic Reviews*, *2016*(6). https://doi.org/10.1002/14651858.CD001218.pub3

Lu, G., & Ades, A. E. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine*, *23*(20), 3105–3124. https://doi.org/10.1002/sim.1875

Lu, G., & Ades, A. E. (2006). Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association*, *101*(474), 447–459. https://doi.org/10.1198/016214505000001302

Lumley, T. (2002). Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine*, *21*(16), 2313–2324. https://doi.org/10.1002/sim.1201

Machin, D., Campbell, M. J., & Walters, S. J. (2008). *Medical Statistics: A Textbook for the Health Sciences* (4th ed.). https://doi.org/10.1198/tas.2008.s274

Mauri, L., & D'Agostino, R. B. (2017). Challenges in the Design and Interpretation of Noninferiority Trials. *New England Journal of Medicine*, *377*(14). https://doi.org/10.1056/NEJMra1510063

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2006). *Introduction to linear regression analysis*. Wiley-Interscience.

Murphy, S. A., Gibson, C. M., Morrow, D. A., Van de Werf, F., Menown, I. B., Goodman, S. G., … Braunwald, E. (2007). Efficacy and safety of the low-molecular weight heparin enoxaparin compared with unfractionated heparin across the acute coronary syndrome spectrum: a meta-analysis. *European Heart Journal*, *28*(17), 2077–2086. https://doi.org/10.1093/eurheartj/ehm224

Nie, L., & Soon, G. (2010). A covariate-adjustment regression model approach to noninferiority margin definition. *Statistics in Medicine*, *29*(10), n/a-n/a. https://doi.org/10.1002/sim.3871

Nielsen, A. (2017). Acupuncture for the Prevention of Tension-Type Headache (2016). *Explore: The Journal of Science and Healing*, *13*(3), 228–231. https://doi.org/10.1016/j.explore.2017.03.007

OASIS investigators. (2006). Comparison of Fondaparinux and Enoxaparin in Acute Coronary Syndromes. *New England Journal of Medicine*, *354*(14), 1464–1476. https://doi.org/10.1056/NEJMoa055443

Odem-Davis, K., & Fleming, T. R. (2013). Adjusting for Unknown Bias in Noninferiority Clinical Trials. *Statistics in Biopharmaceutical Research*, *5*(3), 248–258. https://doi.org/10.1080/19466315.2013.795910

Odem-Davis, K., & Fleming, T. R. (2015). A Simulation Study Evaluating Bio-Creep Risk in Serial Noninferiority Clinical Trials for Preservation of Effect. *Statistics in Biopharmaceutical Research*, *7*(1), 12–24. https://doi.org/10.1080/19466315.2014.1002627

Oler, A., Whooley, M. A., Oler, J., & Grady, D. (1996). Adding heparin to aspirin reduces the incidence of myocardial infarction and death in patients with unstable angina. A meta-analysis. *Jama*, *276*(10), 811–815. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8769591

Oppegaard, K. S., Qvigstad, E., Fiala, C., Heikinheimo, O., Benson, L., & Gemzell-Danielsson, K. (2015). Clinical follow-up compared with self-assessment of outcome after medical abortion: a multicentre, non-inferiority, randomised, controlled trial. *Lancet (London, England)*, *385*(9969), 698–704. https://doi.org/10.1016/S0140-6736(14)61054-0

Ouzzani, M., Hammady, H., & Fedorowics, Zbys, Elmagarmid, A. (2016). *Rayyan. Systematic review*. https://doi.org/10.1186/s13643-016-0384-4

Park, S.-J., Ahn, J.-M., Kim, Y.-H., Park, D.-W., Yun, S.-C., Lee, J.-Y., … Ong, T. K. (2015). Trial of everolimus-eluting stents or bypass surgery for coronary disease. *The New England Journal of Medicine*, *372*(13), 1204–1212. https://doi.org/10.1056/NEJMoa1415447

Paul, M., Bishara, J., Yahav, D., Goldberg, E., Neuberger, A., Ghanem-Zoubi, N., … Leibovici, L. (2015). Trimethoprim-sulfamethoxazole versus vancomycin for severe infections caused by meticillin resistant Staphylococcus aureus: randomised controlled trial. *BMJ (Clinical Research Ed.)*, *350*, h2219. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4431679&tool=pmcentrez&rendertype=abstract

Paul, M., & Leibovici, L. (2014). Systematic review or meta-analysis? Their place in the evidence hierarchy. *Clinical Microbiology and Infection*, *20*(2), 97–100. https://doi.org/10.1111/1469-0691.12489

Pfeffer, M. A., Claggett, B., Diaz, R., Dickstein, K., Gerstein, H. C., Køber, L. V, … Tardif, J.-C. (2015). Lixisenatide in Patients with Type 2 Diabetes and Acute Coronary Syndrome. *The New England Journal of Medicine*, *373*(23), 2247–2257. https://doi.org/10.1056/NEJMoa1509225

Phillippo, D. M., Ades, A. E., Dias, S., Palmer, S., Abrams, K. R., & Welton, N. J. (2018). Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal. *Medical Decision Making : An International Journal of the Society for Medical Decision Making*, *38*(2), 200–211. https://doi.org/10.1177/0272989X17725740

Piaggio, G., Elbourne, D. R., & Altman, D. G. (2006). Reporting of noninferiority and equivalence randomized trials. *JAMA: The Journal of …*, *295*(10), 1147–1151. Retrieved from http://jama.ama-assn.org/content/295/10/1152.short

Piaggio, G., Elbourne, D. R., Pocock, S. J., Evans, S. J. W., Altman, D. G., & CONSORT Group, for the. (2012). Reporting of Noninferiority and Equivalence Randomized Trials. *JAMA*, *308*(24), 2594. https://doi.org/10.1001/jama.2012.87802

Pinto, A. (2013). Considerations for planning and designing meta-analysis in oral medicine. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, *116*(2), 194–202. https://doi.org/10.1016/J.OOOO.2013.02.024

Postma, D. F., van Werkhoven, C. H., van Elden, L. J. R., Thijsen, S. F. T., Hoepelman, A. I. M., Kluytmans, J. A. J. W., … Bonten, M. J. M. (2015). Antibiotic treatment strategies for community-acquired pneumonia in adults. *The New England Journal of Medicine*, *372*(14), 1312–1323. https://doi.org/10.1056/NEJMoa1406330

R Development Core Team. (2008). *R: A language and environment for statistical computing*. https://doi.org/ISBN 3-900051-07-0

Radford, J., Illidge, T., Counsell, N., Hancock, B., Pettengell, R., Johnson, P., … Barrington, S. (2015). Results of a trial of PET-directed therapy for early-stage Hodgkin's lymphoma. *The New England Journal of Medicine*, *372*(17), 1598–1607. https://doi.org/10.1056/NEJMoa1408648

Rahman, N. M., Pepperell, J., Rehal, S., Saba, T., Tang, A., Ali, N., … Miller, R. F. (2015). Effect of Opioids vs NSAIDs and Larger vs Smaller Chest Tube Size on Pain Control and Pleurodesis Efficacy Among Patients With Malignant Pleural Effusion: The TIME1 Randomized Clinical Trial. *JAMA*, *314*(24), 2641–2653. https://doi.org/10.1001/jama.2015.16840

Raungaard, B., Jensen, L. O., Tilsted, H.-H., Christiansen, E. H., Maeng, M., Terkelsen, C. J., … Lassen, J. F. (2015). Zotarolimus-eluting durable-polymer-coated stent versus a biolimus-eluting biodegradable-polymer-coated stent in unselected patients undergoing percutaneous coronary intervention (SORT OUT VI): a randomised non-inferiority trial. *Lancet (London, England)*, *385*(9977), 1527–1535. https://doi.org/10.1016/S0140-6736(14)61794-3

Rehal, S., Morris, T. P., Fielding, K., Carpenter, J. R., & Phillips, P. P. J. (2016). Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals. *BMJ Open*, *6*(10), e012594. https://doi.org/10.1136/bmjopen-2016-012594

Rheims, S., Perucca, E., Cucherat, M., & Ryvlin, P. (2011). Factors determining response to antiepileptic drugs in randomized controlled trials. A systematic review and meta-analysis. *Epilepsia*, *52*(2), no-no. https://doi.org/10.1111/j.1528-1167.2010.02915.x

Rosenfield, K., Jaff, M. R., White, C. J., Rocha-Singh, K., Mena-Hurtado, C., Metzger, D. C., … Scheinert, D. (2015). Trial of a Paclitaxel-Coated Balloon for Femoropopliteal Artery Disease. *The New England Journal of Medicine*, *373*(2), 145–153. https://doi.org/10.1056/NEJMoa1406235

Rothmann, M. D., Wiens, B. L., Chan, I. S. F., Crc, H., & Group, F. (2012). *Design and Analysis of Non-Inferiority Trials*.

Rothmann, M., Li, N., Chen, G., Chi, G. Y. H. H., Temple, R., & Tsou, H.-H. H. (2003). Design and analysis of non-inferiority mortality trials in oncology. *Statistics in Medicine*, *22*(2), 239–264. https://doi.org/10.1002/sim.1400

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2006). Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. https://doi.org/10.1002/0470870168

Salminen, P., Paajanen, H., Rautio, T., Nordström, P., Aarnio, M., Rantanen, T., … Grönroos, J. M. (2015). Antibiotic Therapy vs Appendectomy for Treatment of Uncomplicated Acute Appendicitis. *JAMA*, *313*(23), 2340. https://doi.org/10.1001/jama.2015.6154

Sankoh, A. J. (2008). A note on the conservativeness of the confidence interval approach for the selection of non-inferiority margin in the two-arm active-control trial. *Statistics in Medicine*, *27*(19), 3732–3742. https://doi.org/10.1002/sim.3256

Sax, P. E., Wohl, D., Yin, M. T., Post, F., DeJesus, E., Saag, M., … McCallister, S. (2015). Tenofovir alafenamide versus tenofovir disoproxil fumarate, coformulated with elvitegravir, cobicistat, and emtricitabine, for initial treatment of HIV-1 infection: two randomised, double-blind, phase 3, non-inferiority trials. *Lancet (London, England)*, *385*(9987), 2606–2615. https://doi.org/10.1016/S0140-6736(15)60616-X

Schiller, P., Burchardi, N., Niestroj, M., & Kieser, M. (2012). Quality of reporting of clinical non-inferiority and equivalence randomised trials - update and extension. *Trials*, *13*(1), 1. https://doi.org/10.1186/1745-6215-13-214

Schmidli, H., Wandel, S., & Neuenschwander, B. (2012). The network meta-analytic-predictive approach to non-inferiority trials. *Statistical Methods in Medical Research*, *22*(2), 219–240. https://doi.org/10.1177/0962280211432512

Schwarzer, G. (2007). meta: An R package for meta-analysis. *R News*, *7*(3), 40–45. Retrieved from https://cran.r-project.org/doc/Rnews/Rnews_2007-3.pdf

Scott, D. L., Ibrahim, F., Farewell, V., O'Keeffe, A. G., Walker, D., Kelly, C., … Kingsley, G. H. (2015). Tumour necrosis factor inhibitors versus combination intensive therapy with conventional disease modifying anti-rheumatic drugs in established rheumatoid arthritis: TACIT non-inferiority randomised controlled trial. *BMJ (Clinical Research Ed.), 350*, h1046. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4358851&tool=pmcentrez&rendertype=abstract

Se, H., Gr, C., Kg, W., Jj, B., Sn, M., Hetrick, S. E., … Merry, S. N. (2016). *Cognitive behavioural therapy (CBT), third-wave CBT and interpersonal therapy (IPT) based interventions for preventing depression in children and adolescents (Review) Cognitive behavioural therapy (CBT), third-wave CBT and interpersonal therapy (* . (8), 10–13. https://doi.org/10.1002/14651858.CD003380.pub4.Copyright

Shapiro A. K., M. L. A. (1978). *The placebo effect in medical and psychological therapies. In Handbook of psychotherapy and behavior*.

Signorovitch, J. E., Sikirica, V., Erder, M. H., Xie, J., Lu, M., Hodgkins, P. S., … Wu, E. Q. (2012). Matching-Adjusted Indirect Comparisons: A New Tool for Timely Comparative Effectiveness Research. *Value in Health*, *15*(6), 940–947. https://doi.org/10.1016/j.jval.2012.05.004

Signorovitch, J. E., Wu, E. Q., Yu, A. P., Gerrits, C. M., Kantor, E., Bao, Y., … Mulani, P. M. (2010). Comparative Effectiveness Without Head-to-Head Trials. *PharmacoEconomics*, *28*(10), 935–945. https://doi.org/10.2165/11538370-000000000-00000

Snapinn, S. M. (2000). Noninferiority trials. *Current Controlled Trials in Cardiovascular Medicine*, *1*(1), 19–21. https://doi.org/10.1186/cvm-1-1-019

Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What Are We Weighting For? *Journal of Human Resources*, *50*(2), 301–316. https://doi.org/10.3368/jhr.50.2.301

Stéphan, F., Barrucand, B., Petit, P., Rézaiguia-Delclaux, S., Médard, A., Delannoy, B., … Bérard, L. (2015). High-Flow Nasal Oxygen vs Noninvasive Positive Airway Pressure in Hypoxemic Patients After Cardiothoracic Surgery: A Randomized Clinical Trial. *JAMA*, *313*(23), 2331–2339. https://doi.org/10.1001/jama.2015.5213

Suda, K. J., Hurley, A. M., McKibbin, T., & Motl Moroney, S. E. (2011). Publication of noninferiority clinical trials: Changes over a 20-year interval. *Pharmacotherapy*, *31*(9), 833–839. https://doi.org/10.1592/phco.31.9.833

Talbot, M. (2000). The placebo prescription. *New Yourk Times*.

Thomas, W. J. (2001). Informed Consent, the Placebo Effect, and the Revenge of Thomas Percival. *Journal of Legal Medicine*, *22*. Retrieved from https://heinonline.org/HOL/Page?handle=hein.journals/jlm22&id=343&div=23&collection=journals

Thompson, S. G., & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, *21*(11), 1559–1573. https://doi.org/10.1002/sim.1187

Thompson, S. G., & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine*, *18*(20), 2693–2708. https://doi.org/10.1002/(SICI)1097-0258(19991030)18:20<2693::AID-SIM235>3.0.CO;2-V

Tonin, F. S., Rotta, I., Mendes, A. M., & Pontarolo, R. (2017). Network meta-analysis: a technique to gather evidence from direct and indirect comparisons. *Pharmacy Practice*, *15*(1), 943. https://doi.org/10.18549/PharmPract.2017.01.943

TransMedics, I. (2008). *Proceed-II-Rev-1.6.pdf*.

Urban, P., Meredith, I. T., Abizaid, A., Pocock, S. J., Carrié, D., Naber, C., … Morice, M.-C. (2015). Polymer-free Drug-Coated Coronary Stents in Patients at High Bleeding Risk. *The New England Journal of Medicine*, *373*(21), 2038–2047. https://doi.org/10.1056/NEJMoa1503943

Valkenhoef, V., & Kuiper, J. (2016). *gemtc: Network Meta-Analysis Using Bayesian Methods*.

Wagenlehner, F. M., Umeh, O., Steenbergen, J., Yuan, G., & Darouiche, R. O. (2015). Ceftolozane-tazobactam compared with levofloxacin in the treatment of complicated urinary-tract infections, including pyelonephritis: a randomised, double-blind, phase 3 trial (ASPECT-cUTI). *Lancet (London, England)*, *385*(9981), 1949–1956. https://doi.org/10.1016/S0140-6736(14)62220-0

Walsh, B. T., Seidman, S. N., Sysko, R., & Gould, M. (2002). Placebo Response in Studies of Major Depression. *JAMA*, *287*(14), 1840. https://doi.org/10.1001/jama.287.14.1840

Wang, H., Chow, S. C., & Li, G. (2002). On sample size calculation based on odds ratio in clinical trials. *Journal of Biopharmaceutical Statistics*, *12*(4), 471–483. https://doi.org/10.1081/BIP-120016231

Wangge, G., Klungel, O. H., Roes, K. C. B., de Boer, A., Hoes, A. W., & Knol, M. J. (2010). Room for improvement in conducting and reporting non-inferiority randomized controlled trials on drugs: A systematic review. *PLoS ONE*, *5*(10), e13550. https://doi.org/10.1371/journal.pone.0013550

We, S. R., Koog, Y. H., Park, M. S., & Min, B. Il. (2012). Placebo effect was influenced by publication year in three-armed acupuncture trials. *Complementary Therapies in Medicine*, *20*(1–2), 83–92. https://doi.org/10.1016/j.ctim.2011.10.003

White, I. R., Barrett, J. K., Jackson, D., & Higgins, J. P. T. (2012). Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Research Synthesis Methods*, *3*(2), 111–125. https://doi.org/10.1002/jrsm.1045

Whitehead, A. (2002). *Meta-Analysis Of Controlled Clinical Trials*. https://doi.org/10.1002/0470854200

Witte, S., Schmidli, H., O'Hagan, A., & Racine, A. (2011). Designing a non-inferiority study in kidney transplantation: A case study. *Pharmaceutical Statistics*, *10*(5), 427–432. https://doi.org/10.1002/pst.511

Xu, S., Barker, K., Menon, S., & D'Agostino, R. B. (2014). Covariate Effect on Constancy Assumption in Noninferiority Clinical Trials. *Journal of Biopharmaceutical Statistics*, *24*(6), 1173–1189. https://doi.org/10.1080/10543406.2014.941993

Zeitouni, M., Kerneis, M., Nafee, T., Collet, J.-P., Silvain, J., & Montalescot, G. (2018). Anticoagulation in Acute Coronary Syndrome-State of the Art. *Progress in Cardiovascular Diseases*, *60*(4–5), 508–513. https://doi.org/10.1016/J.PCAD.2018.01.004

Appendices

Appendix A Chapter 2

This appendix includes the alternative models used in the analysis of the OASIS trial and the R codes for Chapter 2

A. 1 Random effects network meta-analysis

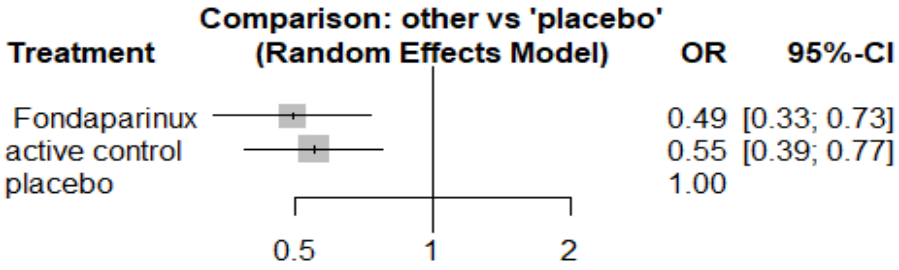

Figure A. 1 Random effect network meta-analysis of OASIS trial

A. 2 Codes for network meta-analysis

```
study<-c("Theroux 88", "RISC Group 90", "Cohen 90", "Cohen 94", "Holdright 94", "GurfinkelUFH 1995", "GurfinkelLMWH 1995", "FRISCI11997", "OASIS52006")
arm1<-c(rep("placebo",8),rep(" Fondaparinux",1))
arm2<-c(rep("active control",9))
#event1<- number of events in arm1(active control)
#event2<- number of events in arm2
event1<-c(4.5, 7.5, 1.5, 9.5, 40.5, 7.5,7.5, 36.5,619.5 )
event2<-c(2.5, 3.5, 0.5, 4.5, 42.5, 4.5, 0.5,13.5, 682.5)
# n1<- total number of patients in arm 1
#n2<- total number of patients in arm 2
n2<-c(123, 211,38, 106, 155, 71, 69,757,10022 )
n1<-c(122, 190, 33, 110, 132, 74, 74,757, 10058)
data <- data.frame(study, arm1, arm2, event1, event2, n1, n2)
library(netmeta)
data
net1<- pairwise(list(arm1, arm2), list(event1, event2), n =list(n1, n2), studlab=study,data=data, sm="OR")
net1
nma1 <- netmeta(TE,seTE, treat1, treat2, studlab,sm="OR", data=net1)
nmar <- netmeta(TE,seTE, treat1, treat2, studlab,sm="OR", comb.random = TRUE, data=net1)
nmar
summary (nmar)
forest.netmeta(nmar, reference.group=" Fondaparinux")
forest.netmeta(nmar, reference.group="active control")
forest.netmeta(nmar, reference.group="placebo")
summary(nmar)
netgraph(nmar, points=TRUE, cex.points=9, cex=1.5)
netrank(nmar, small.values="good")
```

Appendix B Chapter 4

This appendix includes the characteristics of the included trials in the systematic review, the extraction form discussed in Chapter 4 and a poster presented to the STC conference.

B. 1 Characteristics of the included reviews

Study Name	Fund	Blinding	CI	NI margin	Analysis	Conclusion
			BMJ			
OPT, (Cooper et al., 2015)	Public	Open Label	2 sided 95%	Clinical judgement	primary ITT, sensitivity PP	NI established
(Bensdorp et al., 2015)	Public	Open Label	2 sided 95%	Clinical judgement	PP	NI established
(Mical Paul et al., 2015)	Public	Open Label	2 sided 95%	Not stated	primary ITT, sensitivity PP	NI did not conclude
TACIT, (Scott et al., 2015)	Public	Open Label	2 sided 95%	Clinical judgement	ITT	NI established
(Detollenaere et al., 2015)	Public	Open Label	2 sided 95%	Previous studies	primary ITT, sensitivity PP	NI established
			JAMA			
BiPOP, (Stéphan et al., 2015)	Public	Open Label	2 sided 95%	Both clinical and previous data	ITT	NI established
APPAC, (Salminen et al., 2015)	Public	Open Label	2 sided 95%	Both clinical and previous data	primary ITT, sensitivity PP	NI did not conclude
ACOSOGZ 6051, (Fleshman et al., 2015)	Public	Open Label	1 sided 95%	Both clinical and previous data	primary ITT, sensitivity PP	NI did not conclude
(Gross et al., 2015)	Public	Open Label	2 sided 95%	Both clinical and previous data	primary ITT, sensitivity PP	NI established
(Rahman et al., 2015)	Public	Open Label	1 sided 95%	Both clinical and previous data	primary ITT, sensitivity PP	NI established
			Lancet			
COPOUSEP, (Le Page et al., 2015)	Both	Double-blinded	2 sided 90%	Not stated	Primary PP, sensitivity ITT	NI established
(Goldstein et al., 2015)	Private	Open Label	2 sided 95%	Clinical judgement	primary ITT, sensitivity PP	superiority established
SIMPLE, (Healey et al., 2015)	Both	Single-blinded	2 sided 95%	Previous studies	primary ITT, sensitivity PP	NI established
SORT OUT VI, (Raungaard et al., 2015)	Private	Open Label	2 sided 95%	Previous studies	ITT	NI established
GHSG HD13, (Behringer et al., 2015)	Both	Open Label	2 sided 95%	Previous studies	primary ITT, sensitivity PP	NI did not conclude
(Bernard et al., 2015)	Public	Open Label	2 sided 95%	Regulatory guidelines	primary ITT, sensitivity PP	NI established

(Oppegaard et al., 2015)	Public	Open Label	2 sided 95%	Clinical judgement	primary ITT, sensitivity PP	NI established
(Bachelez et al., 2015)	Private	Double blinded	2 sided 95%	Not stated	ITT	NI established
AWARD 4, (Blonde et al., 2015)	Private	Open Label	2 sided 95%	Not stated	ITT	NI established
CHORUS, (Kehoe et al., 2015)	Public	Open Label	1 sided 90%	Both clinical and previous data	primary ITT, sensitivity PP	NI established
ASPECT-cUTI, (Wagenlehner et al., 2015)	Private	Double-blinded	2 sided 95%	Clinical judgement	primary ITT, sensitivity PP	superiority established
(Barone et al., 2015)	Public	Open Label	2 sided 95%	Both clinical and previous data	PP	NI established
(Sax et al., 2015)	Private	Double blinded	2 sided 95%	Not stated	PP	NI established
PROCEED II, (Ardehali et al., 2015)	Private	Open Label	1 sided 95%	Not stated	primary ITT, sensitivity PP	NI established
(Cox et al., 2015)	Private	Open Label	1 sided 95%	Previous studies	PP	NI established
New England Journal of Medicine						
(Geisler et al., 2015)	Public	Open Label	1 sided 90%	Regulatory guidelines	PP	NI did not conclude
BEST, (Park et al., 2015)	Public	Open Label	2 sided 95%	Not stated	ITT	NI did not conclude
(Joura et al., 2015)	Private	Double blinded	2 sided 95%	Not stated	Primary PP, sensitivity ITT	NI established
FXI-ASO, (Büller et al., 2015)	Private	Open Label	2 sided 90%	Both clinical and previous data	Primary PP, sensitivity ITT	superiority established
ELIXA, (Pfeffer et al., 2015)	Private	Double blinded	2 sided 95%	Regulatory guidelines	ITT	NI established
Tuxedo, (Kaul et al., 2015)	Private	single blinded	2 sided 95%	Previous studies	primary ITT, sensitivity PP	NI did not conclude
12EU01, (Urban et al., 2015)	Private	Double blinded	1 sided 97.5%	Previous studies	primary ITT, sensitivity PP	superiority established
BRIDGE, (Douketis et al., 2015)	Public	Double-blinded	2 sided 95%	Previous studies	Primary PP, sensitivity ITT	NI established
LEVANT 2, (Rosenfield et al., 2015)	Private	single blinded	2 sided 95%	Previous studies	ITT	NI established
ABSORB III, (Ellis et al., 2015)	Private	single blinded	2 sided 95%	Regulatory guidelines	Primary PP, sensitivity ITT	NI established
RAPID, (Radford et al., 2015)	Public	Open Label	2 sided 95%	Both clinical and previous data	primary ITT, sensitivity PP	NI did not conclude
CAP-START, (Postma et al., 2015)	Public	Open Label	2 sided 90%	Not stated	ITT	NI established

B. 2 The extraction form used for the systematic review conducted in Chapter 4:

Study Id:-

Title	
Authors	
Publication date	
Sponsor	
Type of study	
Clinicaltrial.gov	
Ethics	
Aim of the study	

1- Trial characteristics

Phase of Trial	
Sample size (justification)	
Blinding	
Duration of study	
Single centre or multicentre	
Placebo Active control	
Test drug	
Arms	
Inclusion criteria	
Exclusion criteria	
Intervention	
Primary end point	
Stat analysis(intent to treat or per protocol)	
Conclusion	

Power of the study	
Confidence interval	
Interim analysis	

2- NI margin:-

NI Margin	
Method of NI margin	
M1	
M2	
Reporting of NI margin	
Assay sensitivity	
Constant assumption	
Placebo creep	
Indirect comparison	

Additional information:-

B. 3 Poster presented to STC conference

The University Of Sheffield.

An investigation of the methods used to design, analyse and quantify non-inferiority margins in four medical journals in a 12-month time period

Enass M. Duro, Steven Julious, Kate Ren
emduro1@Sheffield.ac.uk

Design, Trials & Statistics

Introduction

Non-inferiority trials first appeared in the late 90s. Since then, the number of non-inferiority trials conducted each year has increased.

From January 2015 to December 2015 more than 350 NI trials were published (Figure 1).

Several regulatory guidelines were established to advise on the conducting and reporting of active control and NI trials. However, none of these guidelines establishes any enforceable responsibilities. Instead, they gave only advice and guidelines.

The main challenges associated with NI trials are:-

- Choosing an appropriate active comparator
- The subjectivity in determining the NI margin
- Using indirect comparison to compare the efficacy of the test treatment with the historical placebo

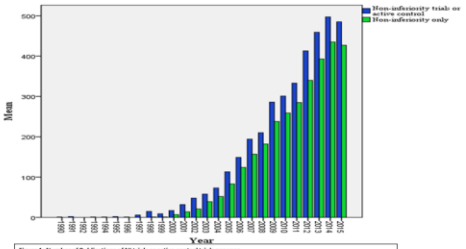

Figure 1: Number of Publications of NI trials or active control trials per year
Note: the search done Pub Med on April/ 2016 with search terms:- Search non-inferior* OR noninferior* OR("active-controlled")* Filters: Randomized Controlled Trial; Humans

Objectives and Methods

- The main aim of this review is to investigate the design, analysis, interpretation and reporting of non-inferiority trials in four top medical journals. A secondary objective is to compare the trials according to the source of funding.
- A search for Non-inferiority trials in the Pub Med database published between 1/1/2015 and 31/12/2015 was performed.
- The inclusion criteria: Non-inferiority trials that were randomised clinical trials, done on adult humans, published in English and full text available (Figure 2).
- Data extraction and analysis: a standardised data extraction form was created.
- Data were extracted and revised by the main investigator.
- Data were analysed using SPSS 22 (SPSS; www.spss.com).
- Clinical trial registries were used to fill out any missing information

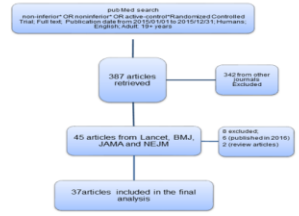

Figure 2: Flow chart for the data extraction process

Results

Out of 37 articles included in the analysis, 15 were published in **The Lancet**, 12 in **the New England Journal of Medicine**, 5 in **BMJ** and 5 in **JAMA**.

- 24 (64.9 %) of the studies were open label studies Study (Table1).
- 29 (78.4%) trials had two arm parallel design. 11 (29.7%) of the trials conduct an interim analysis.
- All the trials reported their NI margin. The methods for determining NI margin were not clear in five (13.55%) trials.
- 14 (37.8%) trials used intent to treat analysis (ITT) as the primary analysis and per protocol analysis (PP) as the sensitivity analysis. 13 (35.1%) trials used both, 8 (21.6%) trials used only intent to treat analysis. Only one trial used per protocol as primary analysis.
- All trials gave a justification for choosing active control as it a standard treatment or best standard treatment. The main differences between publicly and private funded trials presented in table2 .

Table 1: Characteristics of the included NI trials

	Number of Trials	%
Type of drug		
Anti-infective	6	16.2%
Cardiovascular	12	32.4%
Oncology	4	10.8%
Others	15	40.6%
funding		
Public	18	48.6%
Private	15	40.5%
Both	4	10.8%
Blinding		
Open label	24	64.9%
Single blinded	5	13.5%
Double blinded	8	21.6%
End point		
Efficacy	17	45.9%
Safety	1	2.8%
Both	19	51.3%
Conclusion		
NI established	24	64.9%
NI not established	5	13.5%
Superiority stablished	8	21.6%

Table 2: Comparison between publicly and Private funded NI trials

	Public trials	Private trials	both
Methods to determine NI margin			
Not stated	1(5.9%)	3(20%)	1(25%)
Based on previous trials and clinical judgment	9 (50%)	3(20%)	0
Calculated by the investigator from previous trials	6(33.3%)	4(26.7%)	3(75%)
Based on clinical judgment only	1(5.6%)	1(6.7%)	0
Based on guidelines	1(5.6%)	4(26.7%)	0
Blinding			
Open label	17 (94.4%)	6 (40%)	1 (25%)
Single blinding	0	3 (20%)	2 (50%)
Double blinding	1(5.6%)	6 (40%)	1 (25%)
Type of study			
Anti-infective	4(22.2%)	2(13.33%)	0
Cardiovascular and haemostasis	1 (5.6%)	9 (60%)	2 (50%)
Oncology	3 (16.7%)	0	1 (25%)
Gynaecology	4 (22.2%)	0	0
Others	6 (33.33%)	4 (26.6%)	1 (25%)
Type of primary analysis			
Intent to Treat (ITT)	3 (16.7%)	5 (33.3%)	0
Per protocol (PP)	1 (5.6%)	0	0
ITT(primary), PP (sensitivity)	8 (44.4%)	4 (26.7%)	2 (50%)
Both	5 (27.8%)	6 (40%)	2 (50%)
Not clear	1 (5.6%)		
Conclusion			
NI established	13(72.2%)	9(60%)	2(50%)
NI not established	5 (27.8%)	1 (6.7%)	2 (50%)
Superiority established	0	5 (33.3%)	0

Conclusion

- Non-inferiority trials were established as a result of ethical considerations regarding including placebo arm in a clinical trial where there is available standard treatment.
- Most of published NI trials in the four journals did not follow the regulatory guidelines regarding Blinding, type of primary analysis and using two NI margins. Pharmaceutical companies tend to use more liberal methods to prove non-inferiority than the publically funded.
- There is a need to improve the conduction, interpretation and inference of published. NI trials

•References

- D'Agostino RB-S, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. Stat Med 2003 Jan 30; 22 (2):169-86.
- ICH Expert Working Group. (2001) Choice of control group and related issues in clinical trials. E-10 .Available at: ftp://www.ich.org/Readmh/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E10/Step4/E10_Guideline.pdf
- Committee for Medicinal Products for Human Use (CHMP). (2005) Guideline on the choice of the non-inferiority margin. EMEA/CPMP/EWP/2158/99. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf
- Centre for Drug Evaluation and Research (CDER) Centre for Biologics Evaluation and Research (CBER). (2010) DRAFT GUIDANCE: Guidance for industry non-inferiority clinical trials Available at http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM029140.pdf

Appendix C Chapter 5

This appendix includes the results for Spearman correlations and SPSS codes for Chapter 5

C. 1 Spearman Correlations

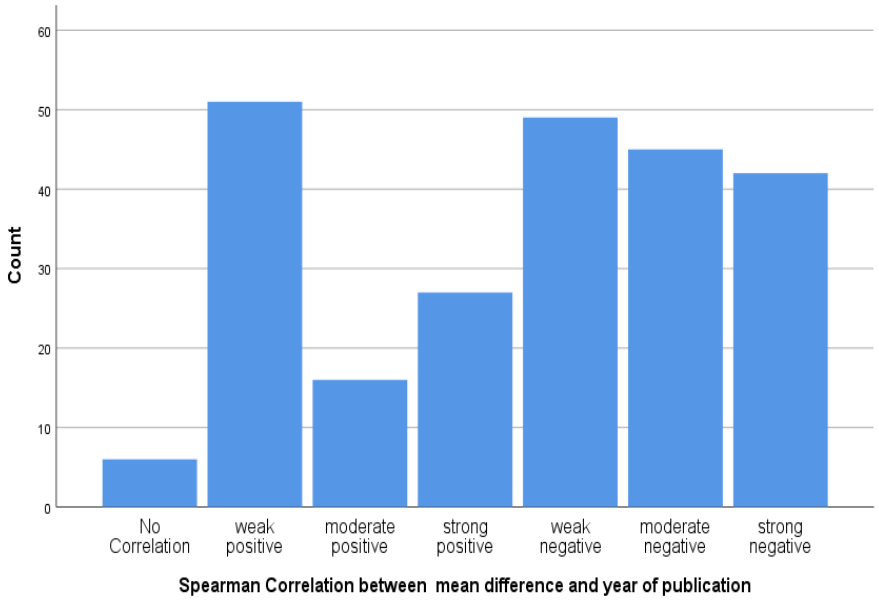

Figure C. 1 Spearman Correlation between the Standardised Mean Difference and the Year of Publication

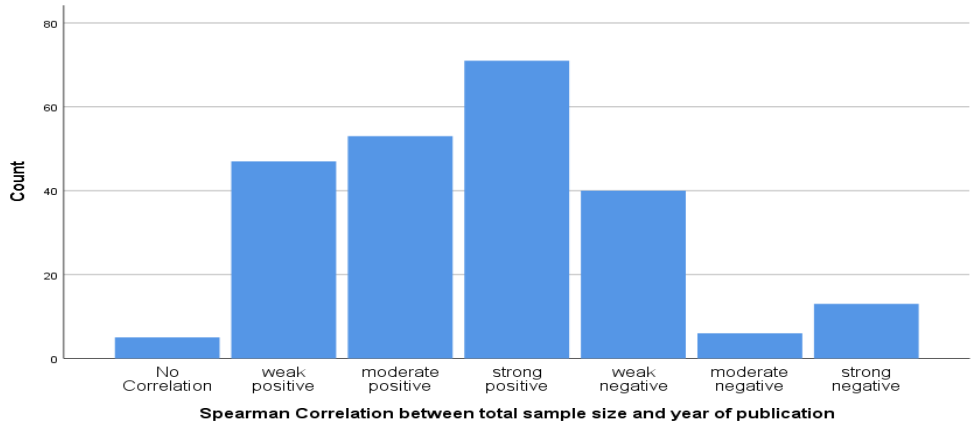

Figure C. 2 Spearman Correlation between the Sample Size and the Year of Publication

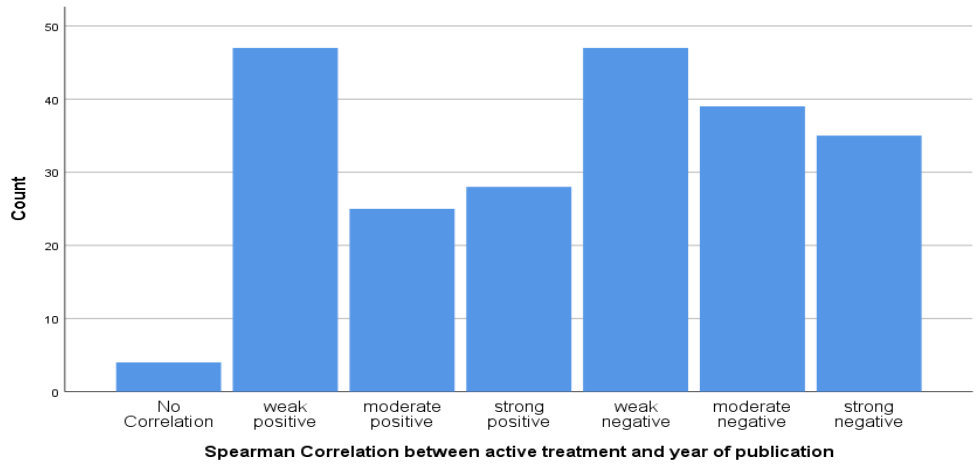

Figure C. 3 Spearman Correlation between the Active Treatment and the Year of Publication

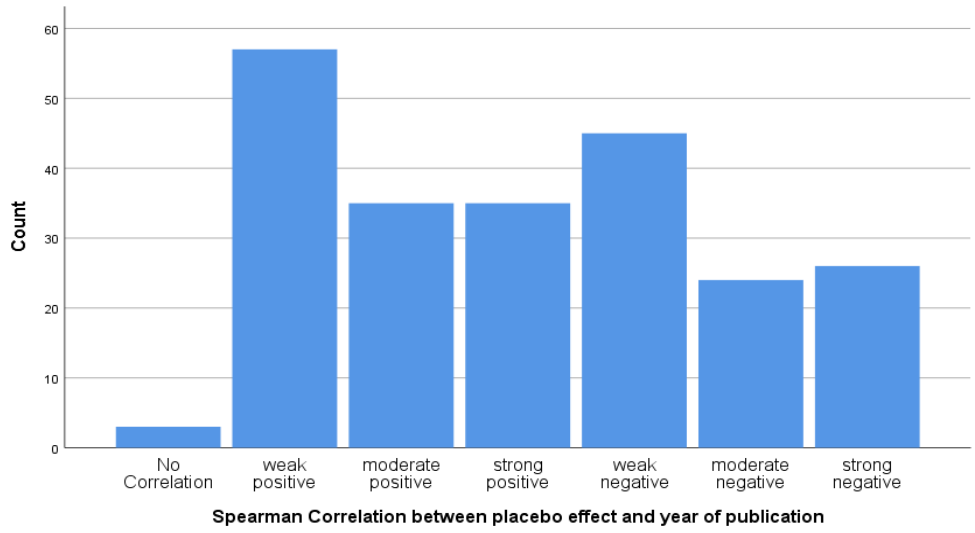

Figure C. 4 Spearman Correlation between the Placebo and the Year of Publication

C. 2 SPSS Codes for Correlations

C. 2. 1 Pearson correlation

```
Encoding: UTF-8.
sort cases by CD.
split file by CD.
*2. Show only variable labels in output table.
set tvars labels.
CORRELATIONS
  /VARIABLES=year d
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE.
* OMS.
DATASET DECLARE  d1.
OMS
  /SELECT TABLES
  /IF COMMANDS=['Correlations'] SUBTYPES=['Correlations']
  /DESTINATION FORMAT=SAV NUMBERED="CD"
   OUTFILE='d1' VIEWER=YES
  /TAG='CD'.
CORRELATIONS
  /VARIABLES=year d
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE.
omsend tag = ['CD'].
* OMS.
```

C. 2. 2 Partial Correlation

```
*2. Show only variable labels in output table.
set tvars labels.
*3. Create correlation table.
PARTIAL CORR
  /VARIABLES=year P  BY N
  /SIGNIFICANCE=TWOTAIL
  /MISSING=LISTWISE.
* OMS.
DATASET DECLARE  ParP.
OMS
  /SELECT TABLES
  /IF COMMANDS=['Partial Corr'] SUBTYPES=['Correlations']
  /DESTINATION FORMAT=SAV NUMBERED=TableNumber_
   OUTFILE='ParP' VIEWER=YES
  /TAG='CD'.
PARTIAL CORR
  /VARIABLES=year P  BY N
  /SIGNIFICANCE=TWOTAIL
  /MISSING=LISTWISE.
omsend tag = ['CD'].
* Encoding: UTF-8.
*1. Split file by study major (pyschology and so on).
```

C. 2. 3 Spearman Correlation

```
OMS
  /SELECT TABLES
  /IF COMMANDS=['Non Par Corr'] SUBTYPES=['Correlations']
  /DESTINATION FORMAT=SAV NUMBERED="CD"
   OUTFILE='d1' VIEWER=YES
  /TAG='CD'.
NONPAR CORR
  /VARIABLES=year d
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE.
omsend tag = ['CD'].
*N
* OMS.
```

C. 2. 4 Bubble plots and meta-regression codes

```
library(meta)
library (metafor)
library(foreign)
CD=file.choose()
CD= read.spss(CD, to.data.frame = TRUE)
View(CD)
meta2<- metacont(N1, M1, SD1, N2, M2, SD2,data=CD, sm="MD", studlab=paste(ID))
M3<-forest(meta2,layout="RevMan5", comb.random=TRUE,
        label.right="Favours experimental ", col.label.right="black",
        label.left="Favours placebo", col.label.left="black", lab.e= "expiremental", lab.c= "Placebo",
        prediction=FALSE, digits.sd = 2)
a<-mean(CD$M1)
b<-mean(CD$M2)
mu1 <- update(meta2, byvar = year)
m2<-metareg(mu1)
names (meta2)
bubble(m2, lwd = 2, col.line = "black", xlim = c(1980, 2015),ylim = c(0, 25), regline=TRUE, xlab= "Year", ylab= "Mean
difference", pch=21, col="black"
      ,bg = "skyblue")
```

Appendix D Chapter 6

This appendix includes graphs for the differences between the different datasets used and the results of the different alternative regression models used in Chapter 6 in addition to the R Codes

D. 1 Differences between the difference datasets extracted

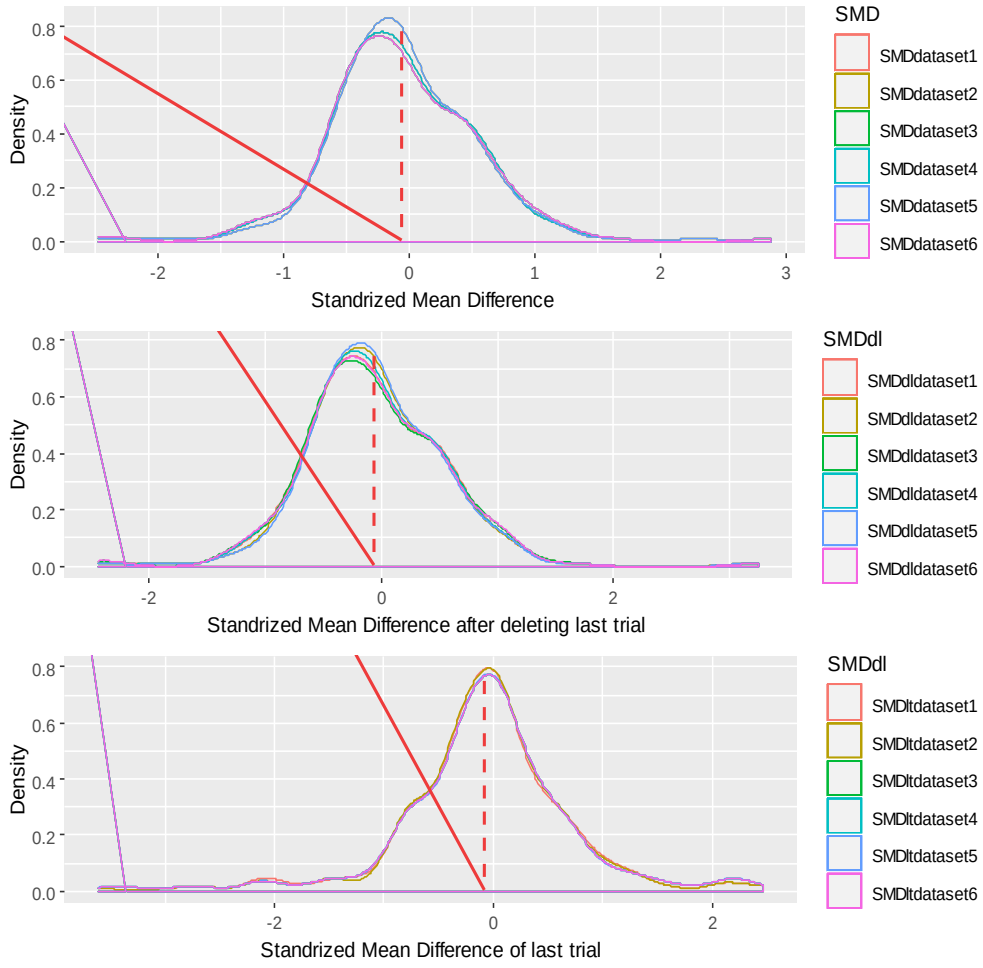

Figure D. 1 Differences between the six datasets

D. 2 Main regression model:

Checking the regression model adequacy (The main model using the training dataset):

Outliers: An analysis of the residuals was carried out. The residuals were normally distributed (min= - 2.97, Max= 1.6) Mean (SD) = -0.027 (0.52); the histogram for the residuals indicate the data contained approximately normally distributed errors. In addition, the Q-Q plot of residuals confirms the conclusion of randomly normal distributed residuals. The maximum Cook's distance was 0.34 which indicates no possible influential cases.

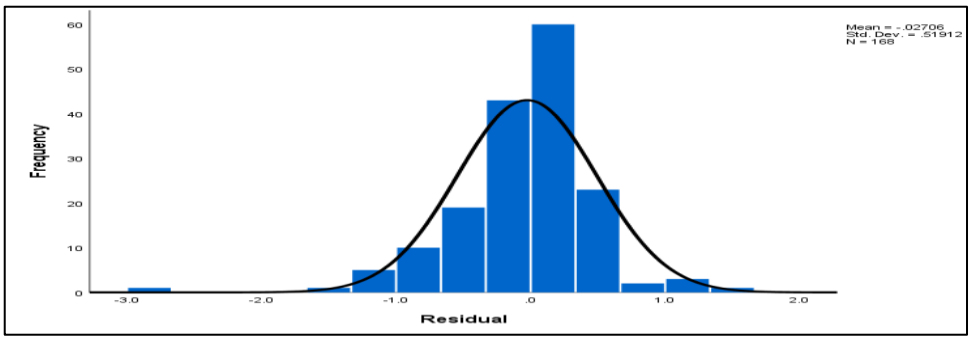

Figure D. 2 Histogram of the Residuals

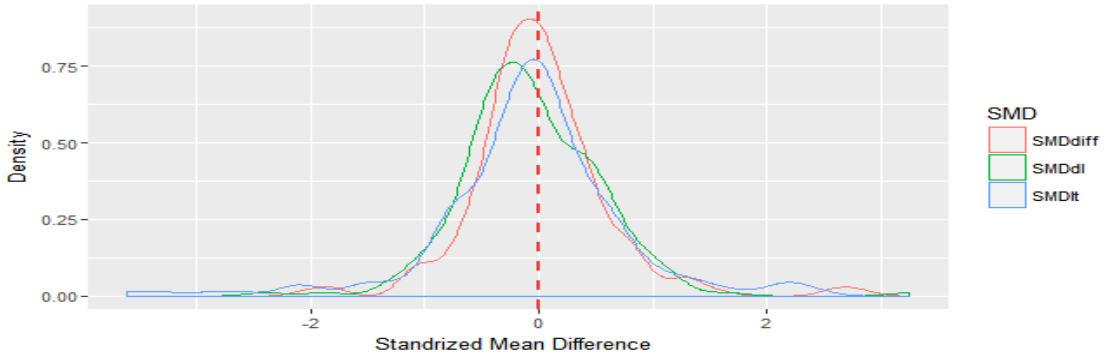

Figure D. 3 The frequency distribution for the differences between SMD of previous trials and the last trial

Collinearity: Tests to see if the data met the assumption of collinearity and both the variance inflation factor (VIF) and tolerance were below the concerned levels (Kutner et al., 2005). They

indicated that multicollinearity was not a concern (SMDdl; Tolerance = 0.748, *VIF* = 1.02; predicted year (Ylt) Tolerance = 0.97, *VIF* = 1.03, Year difference (Ymeta); Tolerance= 0.99, VIF = 1.009). The assumption of independent errors was met with the Durbin-Watson value = 2.2

Homoscedasticity, linearity and random normally distributed errors:

Both the response variable (SMDlt) and the predictor variable (SMDdl) were normally distributed (Figure 6.10). The assumption of linearity between these two variables was met (Figure 6.11).

Regarding the assumptions of homoscedasticity, the scatter plot of predicted values against the residuals showed that the data met the assumptions of homogeneity of variance and linearity. There was a random pattern in the plot for the predicted and the predicted values (Figure 6.9).

It was observed that all the assumptions of multiple regression model were met. In addition, the model was adequately presented.

Model validation

Leave one out cross validation (LOOCV)

Using this method, in each sample one case is the leave out and then the developed model is tested on the leave out case. There were 221 samples in total. The results of the resampling R^2 for the test dataset were 0.55, RMSE= 0.49, MAE= 0.35.

K-fold cross-validation

In the K-fold cross validation the data are divided into k fold and each time one fold works as the test set and the K-1 fold works as the training set. 10 folds was considered an acceptable number of folds and the results from this method were:

R^2 for the test dataset was 0.57, RMSE= 0.47, MAE= 0.35

Repeated K-fold cross-validation

In this method the k fold cross validation is repeated and the average results will be taken. For the model three repeats were used and k=10. The results were:

R^2 for the test dataset was 0.56 RMSE= 0.48, MAE= 0.35

As noticed from the different methods used for validation, the original model I developed was valid and accurate.

D. 3 Unweighted regression model using training dataset

The results of the unweighted regression model are presented in the table below

Table D. 1: Summary of the regression model to predict SMD of last trial using SMD from previous meta-analysis

Model	B	Std. Error	Beta	t	Significance	95.0% CI of β
(Constant)	26.97	13. 89		1.94	0.054	(-0.462; 54.41)
SMDdl	0.98	0.072	0.729	13.68	< 0.001	(0.839; 1.121)
Ymeta	-0.003	0.003	-0.04	-0.75	0.45	(-0.009 ;-0.001)
Ylt	-0.013	0.007	-0.105	-1.94	0.054	(- 0.027; - 0.000)

Weighted Least Squares Regression - Weighted by sample size of the historical meta-analyses. Dependent Variable: ***SMDlt;*** *Standardised mean difference of predicted trial,* ***Ymeta****: year difference between first and last trials in the meta-analysis,* ***SMDdl****= Standardised mean difference of historical trials,* ***Ylt****= year of publication of the predicted trial*

The results of the final regression model include 168 reviews. The model indicated that SMD from the meta-analysis of all trials deleted last (SMDdl). Year difference and the year of predicted trial (Ylt) explain 52.8% of the variance in the model (R^2= 0.537, F (89.784), P <0.0001). SMDdl statistically significantly predicted the SMDlt (β = 0.98, P < 0.0001), for each unit increase in SMDdl the SMDlt increased by 0.917 units. The year difference in the meta-analysis was not statistically significantly associated with the predicted estimate or the predicted year increase.

D. 4 Regression model using the whole dataset:

Table D. 2: Summary of the regression model to predict SMD of last trial using SMD from previous meta-analysis

	B	Std. Error	Beta			Lower Bound	Upper Bound
(Constant)	35.925	9.4		3.8	.0.00	17.309	54.54
SMDdl	0.917	.057	0.732	16.055	.0000	0.804	1.03
Ydiff	-0.007	.002	-0.162	-3.566	.000	-0.010	-0.003
Ylt	-0.018	0.005	-0.173	-3.794	0.00	-0.027	- 0.009

Weighted Least Squares Regression - Weighted by Ndl. Dependent Variable: Standardised mean difference last trial SMDlt, Ydiff= year difference between last trial and the oldest one, SMDdl= Standardised mean difference after deleting last trial, Ylt= year of publication of the predicted trial

The results of final regression model include 221 reviews, two reviews were potential outliers and excluded from the final model to improve the R^2 (Kutner et al., 2005). One case was considered as as influential case using Cook's difference = 0.415 and was excluded from the model (Kutner et al., 2005).

The model indicated the SMD from the meta-analysis of all trials deleted last (SMDdl). Year difference and the year of predicted trial (Ylt) explain 55.4% of the variance in the model (R^2= 0.554, F (89.784), P <0.0001). SMDdl statistically significantly predicted the SMDlt (β = 0.917, P < 0.0001), for each unit increase in SMDdl the SMDlt increased by 0.917 units. For the year difference between the oldest trial and the predicted trial, for every one year increase in the difference the SMDlt decreased by -0.007 (β = - 0.007, P = 0.001). The year of the predicted trial (Ylt) statistically significantly predicted the SMDlt (β= - 0.018, P = 0.001).

The final regression model was

$$Y(SMDlt) = 35.925 + 0.917(SMDdl) - 0.007(Ydiff) - 0.018(Ylt)$$

D. 4. 1 Checking the regression model adequacy

Outliers: An analysis of the residuals was carried out and three possible outliers were detected. Removing these from the model improved the model's predictability. With these outliers the residual were (Min = -3.06, Max = 2.21). Regarding the influence, the maximum Cook's distance was 0.48. According to Kutner et al., if the removing of a potential influential case from the model changes the model inference, this case is considered an influential case and should be omitted from the model (Kutner et al., 2005). After removing the three outliers the residuals were reduced to (Min= - 1.59, Max= 2.2). The R^2 value improved from 50% to 55.4%. Figure 7.9 describes the changes in the model after removing the potential influential and outlier cases.

Collinearity: Tests to see if the data met the assumption of collinearity were conducted (Table 7.1); both the variance inflation factor (VIF) and tolerance were below the concerned levels (Kutner et al., 2005). They indicated that multicollinearity was not a concern (SMDdl, Tolerance = 0.99, *VIF* = 1.01; last year (Ylt) Tolerance = 0.98, *VIF* = 1.014, Year difference Tolerance= 0.99, VIF = 1.009). The assumption of independent errors was met with the Durbin-Watson value = 1.86.

Homoscedasticity, linearity and random normally distributed errors:

Both the response variable (SMDlt) and the predictor variable (SMDdl) were normally distributed. The assumption of linearity between these two variables was met.

The histogram of residuals indicated that the data contained approximately normally distributed errors. In addition, both the normal P-P plot and Q plot of residuals confirmed the conclusion of randomly normal distributed residuals.

Regarding the assumptions of homoscedasticity, the scatter plot of predicted values against the residuals showed that the data met the assumptions of homogeneity of variance and linearity. There was a random pattern in the plot for the predicted and the adjusted predicted values.

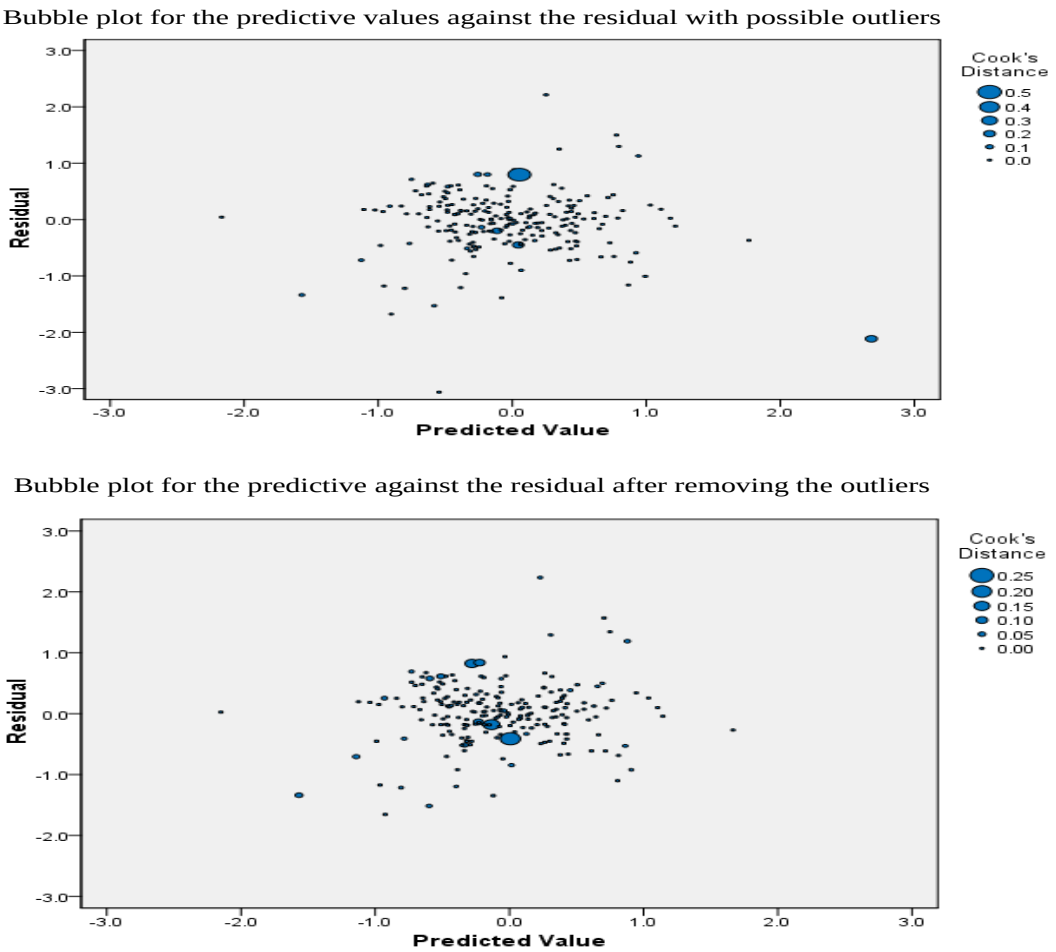

Figure D. 4. Bubble plots before and after removing the potential outliers and influential cases, the bubble size reflects Cook's D

D. 5 Regression analysis using the dataset seven (removing all studies with more than one last trial)

Table D. 3. Summary of the regression analysis

Model	B	Std. Error	Beta	t	Significance	95.0% CI of β
(Constant)	39.43	12.068		3.268	0.001	(-15.678; 63.256)
SMDdl	0.83	0.074	0.655	11.25	< 0.001	(0.69; 0.984)
Ymeta	-0.009	0.006	-1.86	-3.256	0.001	(-0.014 ;-0.004)
Ylt	-0.02	0.003	-0.197	-3.383	0.001	(- 0.031; - 0.008)

A sensitivity analysis was done after removing all trials that contained more than one last trial. The remainder totalled 177 meta-analyses. The results were similar to the model from the whole dataset and there was no difference between the two datasets.

Measuring the agreement between the predicted value and the observed SMDlt

The Bland Altman plot (Bland & Altman, 1999) was used to measure the agreement between the predicted and the observed values in the training dataset and between the observed value in the test dataset and the prediction from the regression model built from the training dataset. The Bland Altman from the training dataset is presented in Figure 6.10. There is a good agreement between the two estimates, with the average mean difference (bias) = -0.03 and the limits of agreement LOA = (-1.04; 0.99). The critical difference was 1.2. As mentioned earlier in Section 6.5.2, the interpretation of Bland Altman depends on the visual inspection of the graph and pre-specified level of agreement. In this case, the most important factor is that the mean difference is close to 0 and the 95 % CI contains zero. On that basis, I can confirm that there is a good agreement between the observed and the predictive values in the model for the training dataset.

The Bland Altman plot for the agreement between the predicted and the observed values ispresented in Figure 6.11. The average mean difference (bias) = 0.06 and the limits of agreement LOA = (-1.14; 1.27). The critical difference was 1.2. which confirms that there is a good agreement between the observed and the predictive values in the model. However, it was noticed that the agreement was higher in the training dataset compared to the test dataset. That is because the model was built using the training dataset, and it would be expected to have a better agreement than the test dataset.

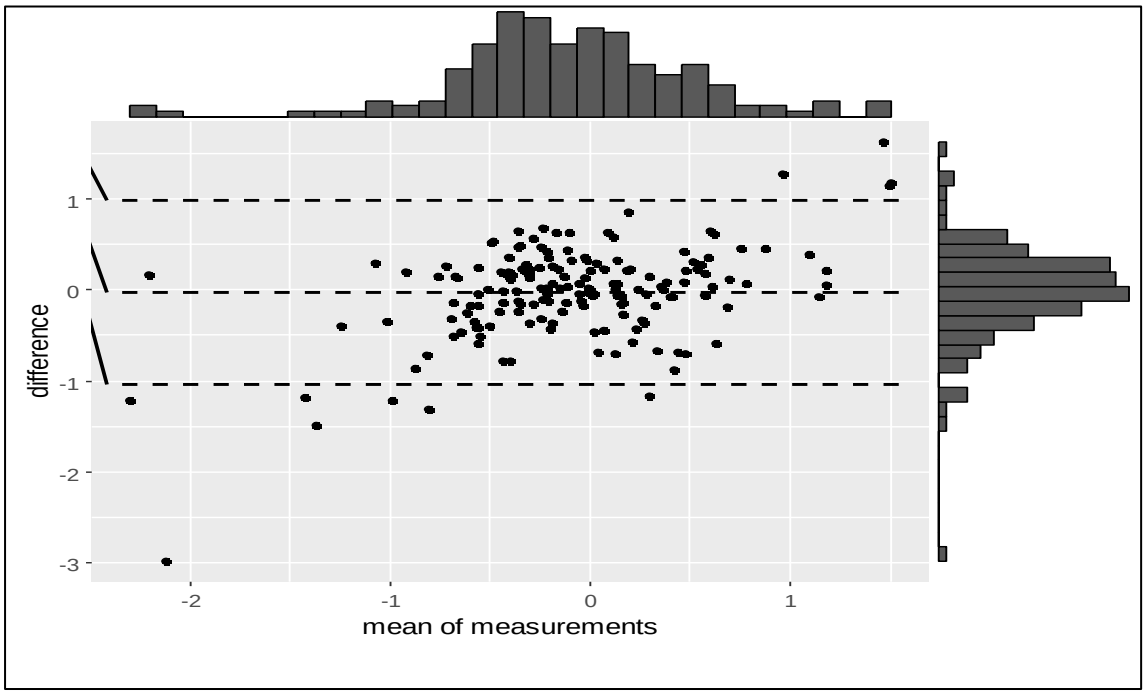

Figure D5. Bland Altman plot for the agreement between the observed and the predicted SMD in the training dataset

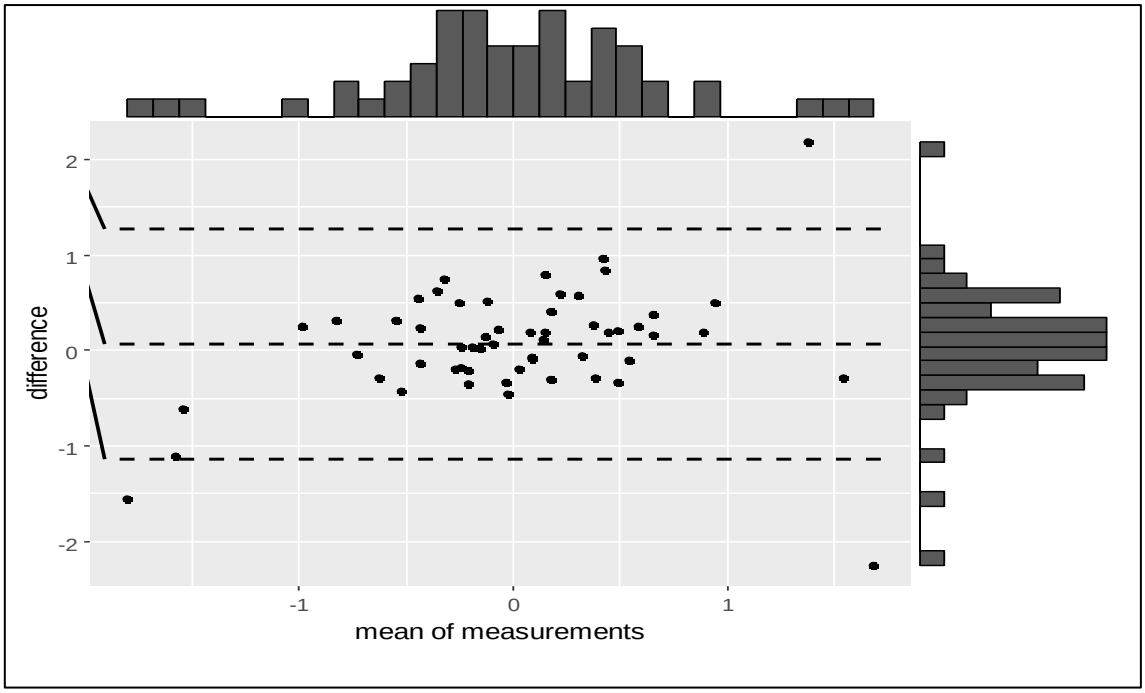

Figure D6. Bland Altman plot for the agreement between the observed and the predicted SMD in the training dataset

D. 6 R codes in Chapter 6

```
# regression, split and boots and Bland Altman
Library (tidyverse)
Library (caret)
Library (modelr)
Library (broom)
Library (boot)
Library (ggplot2)
Library (simpleboot)
Library (boot)
Library (foreign)
Options (digits=1)
all=file.choose()
all= read.spss(all, to.data.frame = TRUE)
names (all)
View (all)
M1<- lm(SMDlt~SMDdl+ y3+ Lastyear, weights = Ndl, data = all)
AIC(M1)
BIC(M1)
Weights (M1)
```

```
Layout (matrix(c(1,2,3,4),2,2)) # optional 4 graphs/page
Plot (lm(SMDlt~SMDdl+ y3+
        Lastyear, weights = Ndl, data = all))
Print (M1)
P1<-predict(M1)
all<- data.frame(all, P1)
  R2 = rsquare(M1, data = all)
  RMSE = rmse(M1, data = all)
  MAE = mae(M1, data = all)
predictions <- M1 %>% predict(all)
data.frame(
  R2 = R2(predictions, all$SMDlt),
  RMSE = RMSE(predictions, all$SMDlt),
  MAE = MAE(predictions, all$SMDlt))
Library (broom)
Glance (M1)
Layout (matrix(c(1,2,3,4),2,2)) # optional 4 graphs/page
Plot (M1)
# Split the data into training and test set
Data (all)
## 75% of the sample size
smp_size <- floor(0.75 * nrow(all))
## set the seed to make your partition reproducible
set.seed(123)
train_ind <- sample(seq_len(nrow(all)), size = smp_size)
train <- all[train_ind, ]
test <- all[-train_ind, ]
View (train)
View (test)
str (train)
str (test)
# Build the model
t1<- lm(SMDlt~SMDdl+ y3+ Lastyear, weights = Ndl, data = train)
print (t1)
P2<-predict(t1)
train<- data.frame(train, P2)
glance(t1)
layout(matrix(c(1,2,3,4),2,2)) # optional 4 graphs/page
plot(t1)
qqPlot(t1, data=train, layout=c(1, 3))
# Make predictions and compute the R2, RMSE and MAE
predictions <- t1 %>% predict(test)
test1<-data.frame(test, predictions)
View(test1)
par(mfrow=c(1,1))
plot(test1$predictions, test1$SMDlt, type="p",
    col="darkblue")
cor(test1$predictions, test1$SMDlt, method = c("pearson", "kendall", "spearman"))
RMSE (test1$predictions, test1$SMDlt)/mean(test.data$SMDlt)
Plot (test1$predictions, test1$SMDlt)
Cor (test1$predictions,test1$SMDlt )
R2 = R2(test1$predictions, test1$SMDlt)
RMSE = RMSE(test1$predictions, test1$SMDlt)
MAE = MAE(test1$predictions, test1$SMDlt)
#bootstrap
data(train)
attach(train)
set.seed(10)
lmodel <- lm(SMDlt~SMDdl+ y3+ Lastyear, weights = Ndl)
lboot <- lm.boot(lmodel, R = 2000)
summary(lboot)
```

```
print (lboot)
# LOOCV
# Define training control
train.control <- trainControl(method = "LOOCV")
# Train the model
model1 <- train(SMDlt~SMDdl+ y3+ Lastyear, weights = Ndl, data = all, method = "lm",
  trControl = train.control)
# Summarize the results
print(model1)
#K-fold cross-validation
# Define training control
set.seed(10)
train.control <- trainControl(method = "cv", number = 10)
# Train the model
model12 <- train(SMDlt~SMDdl+ y3+ Lastyear, weights = Ndl, data = all, method = "lm",
            trControl = train.control)
# Summarize the results
print(model12)
#Repeated K-fold cross-validation
set.seed(10)
train.control <- trainControl(method = "repeatedcv",
                      number = 10, repeats = 3)
# Train the model
model23 <- train(SMDlt~SMDdl+ y3+ Lastyear, weights = Ndl, data = all, method = "lm",
   trControl = train.control)
# Summarize the results
Print (model23)
# bootstrapping with 1000 replications
library(simpleboot)
library (boot)
data (all)
attach (all)
set.seed (123)
lmodel <- lm(SMDlt~SMDdl+ y3+ Lastyear, weights = Ndl)
lboot <- lm.boot(lmodel, R = 1000)
summary(lboot)
w <- all$Ndl
lbootw <- lm.boot(lmodel, R = 1000, weights = w)
summary(lbootw)
lboot2 <- lm.boot(lmodel, R = 1000, rows = FALSE)
summary(lboot2)
#Bland Altman
library(BlandAltmanLeh)
Pa<-bland.altman.plot(group1= all$SMDlt, group2= all$P1,data= all,   xlab="Means", ylab="Differences",
conf.int=.95)
library(BlandAltmanLeh)
Pa<-bland.altman.plot(group1= train$SMDlt, group2= train$P2,data= train,   xlab="Means",
ylab="Differences", conf.int=.95)
Pa<-bland.altman.plot(group1= test1$SMDlt, group2= test1$predictions,data= test1,   xlab="Means",
ylab="Differences", conf.int=.95)
```

```
#last trial fixed and random models
library(meta)
library(foreign)
LT=file.choose()
```

```
LT= read.spss(LT, to.data.frame = TRUE)
View (LT)
meta1<- metagen (d, SE, sm="SMD", data=LT, byvar = (CD))
names(meta1)
View(meta1)
#create datafram for the fixed effect model
meta1$bylevs #CD
meta1$k.all.w #k
meta1$pval.fixed.w
meta1$TE.fixed.w #SMD
meta1$lower.fixed.w #lower bound of 95% CI
meta1$upper.fixed.w  #upper bound of 95% CI
result <- data.frame(meta1$bylevs,meta1$k.all.w,
meta1$TE.fixed.w,meta1$lower.fixed.w,meta1$upper.fixed.w,meta1$pval.fixed.w)
View(result)
library("dplyr")
names(result)
LTfixed<-rename(result, CD=meta1.bylevs, k=meta1.k.all.w, Pvalue= meta1.pval.fixed.w, SMD=meta1.TE.fixed.w,
lower95=meta1.lower.fixed.w,upper95= meta1.upper.fixed.w)
View(LTfixed)
write.table(LTfixed, file="LTfixed.csv", sep=",")
#create datafram for the Random effect model
names(meta1)
meta1$bylevs #CD
meta1$k.all.w #k
meta1$pval.random.w# p value
meta1$TE.random.w #SMD
meta1$lower.random.w #lower bound of 95% CI
meta1$upper.random.w  #upper bound of 95% CI
resultr <- data.frame(meta1$bylevs,meta1$k.all.w,
meta1$TE.random.w,meta1$lower.random.w,meta1$upper.random.w,meta1$pval.random.w)
View(resultr)
library("dplyr")
names(resultr)
LTrandom<-rename(resultr, CD=meta1.bylevs, k=meta1.k.all.w, Pvalue= meta1.pval.random.w,
SMD=meta1.TE.random.w, lower95=meta1.lower.random.w,upper95= meta1.upper.random.w)
View(LTrandom)
write.table(LTrandom, file="LTrandom.csv", sep=",")
```

```
# conduct meta-analysis after deleting last trial both fixed and random

Library (meta)
Library (foreign)
ttdl=file.choose()
ttdl= read.spss(ttdl, to.data.frame = TRUE)
View (ttdl)
meta2<- metagen (d, SEd, sm="SMD", data=ttdl, byvar = (CD))
names(meta2)
View(meta2)
#create datafram for the fixed effect model
meta2$bylevs #CD
meta2$k.all.w #k
meta2$pval.fixed.w
meta2$TE.fixed.w #SMD
meta2$lower.fixed.w #lower bound of 95% CI
```

```
meta2$upper.fixed.w  #upper bound of 95% CI
result2 <- data.frame(meta2$bylevs,meta2$k.all.w,
meta2$TE.fixed.w,meta2$lower.fixed.w,meta2$upper.fixed.w,meta2$pval.fixed.w)
View(result2)
library("dplyr")
names(result2)
DLfixed<-rename(result2, CD=meta2.bylevs, k=meta2.k.all.w, Pvalue= meta2.pval.fixed.w, SMD=meta2.TE.fixed.w,
lower95=meta2.lower.fixed.w,upper95= meta2.upper.fixed.w)
View(DLfixed)
write.table(DLfixed, file="DLfixed.csv", sep=",")
#create datafram for the Random effect model
names(meta2)
meta2$bylevs #CD
meta2$k.all.w #k
meta2$pval.random.w# p value
meta2$TE.random.w #SMD
meta2$lower.random.w #lower bound of 95% CI
meta2$upper.random.w  #upper bound of 95% CI
result2r <- data.frame(meta2$bylevs,meta2$k.all.w,
meta2$TE.random.w,meta2$lower.random.w,meta2$upper.random.w,meta2$pval.random.w)
View(result2r)
library("dplyr")
names(result2r)
DLrandom<-rename(result2r, CD=meta2.bylevs, k=meta2.k.all.w, Pvalue= meta2.pval.random.w,
SMD=meta2.TE.random.w, lower95=meta2.lower.random.w,upper95= meta2.upper.random.w)
View(DLrandom)
write.table(DLrandom, file="DLrandom.csv", sep=",")
```

```
library(meta)
library(foreign)
TT=file.choose()
dataset= read.spss(TT, to.data.frame = TRUE)
View (dataset)
TT<-dataset
View (TT)
metaTT <- metamean(N, d, V, data=TT, byvar = (CD))
mu1<-metareg(metaTT, year+N, intercept = TRUE)
library(meta)
library(foreign)
TTDL=file.choose()
dataset= read.spss(TTDL, to.data.frame = TRUE)
View (dataset)
TTDL<-dataset
View (TTDL)
metaTTDL <- metamean(N, d, V, data=TTDL, byvar = (CD))
```

```
# Codes for the Bland Altman plots
library(ggplot2)
library(grid)
library(gridExtra)
library(BlandAltmanLeh)
```

```
library(blandr)
ba.stats <- bland.altman.stats(fr2$lower95E3f,fr2$lower95E3r)
print(ba.stats)
ba.stats <- bland.altman.stats(fr2$upper95E3f,fr2$upper95E3r)
print(ba.stats)
ba.stats <- bland.altman.stats(fr2$SMDE3f,fr2$SMDE3r)
print(ba.stats)
ba.stats <- bland.altman.stats(fr2$distanceF,fr2$distanceR)
print(ba.stats)
P6<- ggplot(fr2, aes(x = mean, y = difference,size=N,  fill= I2)) +
 geom_point(shape = 21)+ theme_bw()+ ylim(-2,2)+ xlab("mean for the main estimate")+   ylab("Difference of the main
estimate")
       p7<- p6+geom_hline(yintercept=-0.01024521, color = "darkblue")
       p8<- p7+ geom_hline(yintercept=-0.2289256, linetype="dashed", color = "darkblue")
       p9<- p8+ geom_hline(yintercept=0.2084352, linetype="dashed", color = "darkblue")


 p16 <- ggplot(fr2, aes(x = meandistance, y = diffdistance,size=N,  fill= I2)) +
         geom_point(shape = 21)+
         theme_bw()+ ylim(-2,2)+xlab("mean width 95% CI")+   ylab("Difference width 95% CI")
       P17<- p16 +geom_hline(yintercept=-0.49649510, color = "darkblue")
       p18<- P17+ geom_hline(yintercept=-0.09179477, linetype="dashed", color = "darkblue")
       p19<- p18+ geom_hline(yintercept=0.3129056, linetype="dashed", color = "darkblue")
 p26 <- ggplot(fr2, aes(x = meanup, y = diffup,size=N,  fill= I2)) +
         geom_point(shape = 21)+
         theme_bw()+ ylim(-2,2)+ xlab("mean upper 95% CI")+   ylab("Difference upper 95% CI")
       P27<- p26 +geom_hline(yintercept=-0.07669147, color = "darkblue")
       p28<- P27+ geom_hline(yintercept=-0.41099310, linetype="dashed", color = "darkblue")
       p29<- p28+ geom_hline(yintercept=0.25761016, linetype="dashed", color = "darkblue")

p36 <- ggplot(fr2, aes(x = meanlower, y = difflower,size=N,  fill= I2)) +
         geom_point(shape = 21)+
         theme_bw()+ ylim(-2,2)+ xlab("mean lower 95% CI")+   ylab("Difference lower 95% CI")
       P37<- p36 +geom_hline(yintercept=0.0971819, color = "darkblue")
       p38<- P37+ geom_hline(yintercept=-0.2434670, linetype="dashed", color = "darkblue")
       p39<- p38+ geom_hline(yintercept=0.4378308, linetype="dashed", color = "darkblue")

       grid.arrange(p9, p19,p29, p39, ncol = 2)
   # number of trials
 p6 <- ggplot(fr2, aes(x = mean, y = difference,size=Trials,  fill= I2)) +
         geom_point(shape = 21)+
         theme_bw()+ ylim(-2,2)+xlab("mean of the main estimate")+   ylab("Difference of the main estimate")
       P7<- p6 +geom_hline(yintercept=-0.01024521, color = "darkblue")
       p8<- P7+ geom_hline(yintercept=-0.2289256, linetype="dashed", color = "darkblue")
       p9<- p8+ geom_hline(yintercept=0.2084352, linetype="dashed", color = "darkblue")
 p16 <- ggplot(fr2, aes(x = meandistance, y = diffdistance,size=Trials,  fill= I2)) +
         geom_point(shape = 21)+
         theme_bw()+ ylim(-2,2)+xlab("mean width 95% CI")+   ylab("Difference width 95% CI")
       P17<- p16 +geom_hline(yintercept=-0.49649510, color = "darkblue")
       p18<- P17+ geom_hline(yintercept=-0.09179477, linetype="dashed", color = "darkblue")
       p19<- p18+ geom_hline(yintercept=0.3129056, linetype="dashed", color = "darkblue")
p26 <- ggplot(fr2, aes(x = meanup, y = diffup,size=Trials,  fill= I2)) +
         geom_point(shape = 21)+
         theme_bw()+ ylim(-2,2)+ xlab("mean upper 95% CI")+   ylab("Difference upper 95% CI")
       P27<- p26 +geom_hline(yintercept=-0.07669147, color = "darkblue")
       p28<- P27+ geom_hline(yintercept=-0.41099310, linetype="dashed", color = "darkblue")
       p29<- p28+ geom_hline(yintercept=0.25761016, linetype="dashed", color = "darkblue")

 p36 <- ggplot(fr2, aes(x = meanlower, y = difflower,size=Trials,  fill= I2)) +
         geom_point(shape = 21)+
         theme_bw()+ ylim(-2,2)+ xlab("mean lower 95% CI")+   ylab("Difference lower 95% CI")
```

```
P37<- p36 +geom_hline(yintercept=0.0971819, color = "darkblue")
p38<- P37+ geom_hline(yintercept=-0.2434670, linetype="dashed", color = "darkblue")
p39<- p38+ geom_hline(yintercept=0.4378308, linetype="dashed", color = "darkblue")

grid.arrange(p9, p19,p29, p39, ncol = 2)
```

```
library(meta)
library (metafor)
library(foreign)
tte3=file.choose()
tte3= read.spss(tte3, to.data.frame = TRUE)
tt<-data.frame(tte3$CD, tte3$Name, tte3$year, tte3$N, tte3$weight, tte3$d, tte3$SEd, tte3$Vd)
View (tt)
library("dplyr")
names(tt)
tt<-rename(tt, CD=tte3.CD, Name=tte3.Name, year= tte3.year, N=tte3.N, weight=tte3.weight,d= tte3.d, SEd=tte3.SEd,
Vd=tte3.Vd)
tt2<-View(tt)
group_by(tt,tt$CD)
metacd <- unique(tt$CD)
res.rma<-rma.mv(d, Vd, mods= ~ year, data=tt, subset= (CD))
for (i in tt) {
  res.rma <- rma.mv(d, Vd, mods =  ~ year  , data=tt, subset = (CD == "[i]"))
}
get.n.Name(data)
daply(tt, "CD", Name)
```

```
#total trials fixed and random models
library(meta)
library(foreign)
tte3=file.choose()
tte3= read.spss(tte3, to.data.frame = TRUE)
View (tte3)
meta1<- metagen (d, SEd, sm="SMD", data=tte3, byvar = (CD))
names(meta1)
View(meta1)
#create datafram for the fixed effect model
meta1$bylevs #CD
meta1$k.all.w #k
meta1$pval.fixed.w
meta1$TE.fixed.w #SMD
meta1$lower.fixed.w #lower bound of 95% CI
meta1$upper.fixed.w  #upper bound of 95% CI
result <- data.frame(meta1$bylevs,meta1$k.all.w,
meta1$TE.fixed.w,meta1$lower.fixed.w,meta1$upper.fixed.w,meta1$pval.fixed.w)
View(result)
library("dplyr")
names(result)
E3fixed<-rename(result, CD=meta1.bylevs, k=meta1.k.all.w, Pvalue= meta1.pval.fixed.w, SMD=meta1.TE.fixed.w,
lower95=meta1.lower.fixed.w,upper95= meta1.upper.fixed.w)
View(E3fixed)
write.table(E3fixed, file="E3fixed.csv", sep=",")
#create datafram for the Random effect model
names(meta1)
meta1$bylevs #CD
meta1$k.all.w #k
meta1$pval.random.w# p value
```

```
meta1$TE.random.w #SMD
meta1$lower.random.w #lower bound of 95% CI
meta1$upper.random.w  #upper bound of 95% CI
resultr <- data.frame(meta1$bylevs,meta1$k.all.w,
meta1$TE.random.w,meta1$lower.random.w,meta1$upper.random.w,meta1$pval.random.w)
View(resultr)
library("dplyr")
names(resultr)
E3random<-rename(resultr, CD=meta1.bylevs, k=meta1.k.all.w, Pvalue= meta1.pval.random.w,
SMD=meta1.TE.random.w, lower95=meta1.lower.random.w,upper95= meta1.upper.random.w)
View(E3random)
write.table(E3random, file="E3random.csv", sep=",")
```

Appendix E Chapter 8

E. 1 Atorvastatin for lowering lipids

This appendix includes the results for the random effects meta-regression, the detailed tables for the different percentage of M1 for the both examples and the R codes used in Chapter 8 and a special case of NI trial (the OASIS trial)

E. 1. 1 Meta-regression using Random effects model

```
Iterations = 5001:25000
Thinning interval = 1
Number of chains = 4
Sample size per chain = 20000
1. Empirical mean and standard deviation for each variable,
  plus standard error of the mean:
                    Mean     SD  Naive SE Time-series SE
d.Atorvastatin.placebo -25.9078 0.5367 0.0018975      0.023378
sd.d                 0.3073 0.2245 0.0007937      0.007489
B                   4.8246 1.2582 0.0044483      0.011986
2. Quantiles for each variable:
                     2.5%     25%     50%     75%    97.5%
d.Atorvastatin.placebo -26.94422 -26.2698 -25.8878 -25.5548 -24.8677
sd.d                 0.01521   0.1262   0.2634   0.4468   0.8303
B                   2.36472   3.9758   4.8267   5.6753   7.2756
-- Model fit (residual deviance):
   Dbar      pD     DIC
59.57558 26.62229 86.19788
48 data points, ratio 1.241, I^2 = 21%
```

E. 1. 2 Use of different percentages of M1 to set M2

Table E.1 Constancy assumed, 60% NI margin=14.7, n1=n2=14

μ_t	$\mu_T - \mu_C$ from NI trial (2020) (95% CrI), se=4.9	NMA ($\mu_T - \mu_P$) (95% CrI), se= 4.5	NMR ($\mu_T - \mu_P$) in (2020) (95% CrI), se= 4.95
18.5	0.0 (-8.9; 8.9)	25.0 (17.0; 34.0)	18.0 (8.1; 28.0)
16.5	-2.0 (-10.9; 6.9)	23.0 (15.0; 32.0)	16.0 (6.1; 26.0)
14.5	-4.0 (-12.9; 4.9)	21.0 (13.0; 30.0)	14.0 (4.1; 24.0)
13.5	-5.0 (-14.1; 3.8)	20.0 (12.0; 29.0)	13.0 (3; 22.0)
13.0	-5.5 (-14.4; 3.4)	20.0 (11.0; 29.0)	12.0 (2.6; 22.0)
12.5	-6.0(-14.8; 2.89)	19.0 (11.0; 28.0)	12.0 (2.2; 22.0)
10.5	-8.0 (-16.7; 0.9)	17.0 (8.5; 26.0)	9.9 (0.26; 20.0)
10.0	-8.5 (-17.0; 0.28)	17.0 (8.0; 26.0)	9.3 (-0.4; 19.0)
9.5	-9.0 (-18.0; -.0.1)	16.0 (7.5; 25)	8.8 (-0.8; 19)
1.5	-17.0 (-26.0; -8.1)	8.4 (-0.5; 17.0)	0.9 (-9.0; 11)

Table E.2 Constancy assumed, 70% NI margin= 17.5, n1=n2= 11

μ_t	$\mu_T - \mu_C$ from NI trial (2020) (95% CrI), se=5.13	NMA ($\mu_T - \mu_P$) (95% CrI), se= 5.16	NMR ($\mu_T - \mu_P$) in (2020) (95% CrI), se= 5.15
18.5	0.0 (-10.1; 10.1)	25.0 (15.0; 36.0)	18.0 (7.0; 29.0)
16.5	-2.0 (-12.0; 8.1)	23.0 (13.0; 34.0)	16.0 (5.1; 27.0)
14.5	-4.0 (-14.1; 6.8)	21.0 (11.0; 32.0)	14.0 (3.1; 25.0)
12.5	-6.0 (-16.0; 4.1)	19.0 (9.3; 30.0)	12.0 (1.0; 23.0)
11.5	-7.0 (-17.1; 3.0)	18.0 (8.3; 28.0)	11.0 (0.0; 22.0)
11.0	-7.5 (-17.5; 2.5)	18.0 (8.0; 28.0)	10.0 (-0.4; 21.0)
9.5	-9.0 (-19.1; 1.1)	16.0 (6.3; 26.0)	8.8 (-2.0; 20.0)
7.5	-10.9 (-21.0; -0.9)	14.0 (4.4; 25.0)	6.9 (-3.9; 18.0)
3.5	-14.9 (-25.0; -5.0)	10.0 (0.4; 21.0)	2.9 (-7.9; 14.0)
3.0	-15.6 (-25.6; -5.8)	9.9 (-0.2; 20.0)	2.3 (-8.6; 13.0)

Table E.3 Constancy assumed, 80% NI margin= 19.6, n1=n2= 8

μ_t	$\mu_T - \mu_C$ from NI trial (2020) (95% CrI), se=6.0	NMA ($\mu_T - \mu_P$) (95% CrI), se= 6.01	NMR ($\mu_T - \mu_P$) in (2020) (95% CrI), se= 6.03
18.5	0.0 (-11.9; 11.6)	26.0 (14.0; 37.0)	18.0 (5.3; 30.0)
16.5	-2.0 (-13.7; 9.6)	23.0 (12.0; 35.0)	16.0 (3.4; 28.0)
14.5	-4.0 (-15.7; 7.7)	21.0 (9.7; 33.0)	14.0 (1.5; 26.0)
12.5	-6.0 (-17.6; 5.7)	19.0 (7.6; 31.0)	12.0 (-0.5; 24.0)
11.5	-7.0 (-18.7; 4.8)	18.0 (6.8; 30.0)	11.0 (-1.6; 23.0)
10.5	-8.0 (-19.8; 3.6)	17.0 (5.5; 29.0)	9.8 (-2.6; 22.0)
9.5	-9.0 (-20.7; 2.8)	16.0 (4.6; 28.0)	8.9 (-3.5; 21.0)
7.5	-11.0 (-22.8; 0.8)	14.0 (2.7; 26.0)	6.8 (-5.6; 19.0)
6.5	-12.0 (-23.7; -0.16)	13.0 (1.6; 25.0)	5.8 (-6.5; 18.0)
4.5	-14.0 (-25.8; -2.2)	11.0 (-0.4; 23.0)	3.9 (-8.7; 16.0)

Table E.4 Constancy assumed, 90% NI margin= 22.05, n1=n2= 7

μ_t	$\mu_T - \mu_C$ from NI trial (2020) (95% CrI), se=6.36	NMA ($\mu_T - \mu_P$) (95% CrI), se= 6.3	NMR ($\mu_T - \mu_P$) in (2020) (95% CrI), se= 6.6
18.5	0.0 (-12.5; 12.4)	25.0 (13.0; 38.0)	18.0 (4.7; 31.0)

14.5	-4.0 (-16.6; 8.5)	21.0 (8.8; 34.0)	14.0 (0.7; 27.0)
13.5	-5.0 (-17.7; 7.7)	20.0 (7.7; 33.0)	13.0 (-0.5; 26.0)
12.5	-6.0 (-18.6; 6.6)	19.0 (6.7; 32.0)	12.0 (-1.4; 25.0)
9.5	-9.0 (-21.6; 3.5)	16.0 (3.8; 29.0)	8.8 (-4.3; 22.0)
9.0	-9.5 (-22.0; 3.1)	16.0 (3.3; 28.0)	8.3 (-4.8; 22.0)
8.5	-10.0 (-22.5; 2.6)	15.0 (2.9; 28.0)	7.8 (-5.3; 21.0)
7.5	-11 (-23.5; 1.68)	14.0 (1.9; 27.0)	6.8 (-6.3; 20.0)
5.5	-13.0 (-25.6; -0.4)	13.0 (-0.3; 25.0)	4.8 (-8.4;18.0)

Table E.5 Constancy assumed, M2= M1= 24.5, n1=n2= 6

μ_t	$\mu_T - \mu_C$ from NI trial (2020) (95% CrI), se=6.9	NMA ($\mu_T - \mu_P$) (95% CrI), se= 6.9	NMR ($\mu_T - \mu_P$) in (2020) (95% CrI), se= 7.1
18.5	0.0 (-13.7; 13.5)	26.0 (12.0; 39.0)	18.0 (3.6; 32.0)
14.5	-4.0 (-17.6; 9.5)	21.0 (7.8; 35.0)	14.0 (-0.3; 28.0)
12.5	-6.0 (-19.5; 7.5)	19.0 (5.8; 33.0)	12.0 (-2.2; 26.0)
9.5	-9.0 (-22.5; 4.6)	16.0 (2.8; 30.0)	8.9 (-5.2; 23.0)
8.5	-10.0 (-23.6; 3.4)	15.0 (1.9; 29.0)	7.7 (-6.4; 22.0)
7.5	-11 (-24.5; 2.5)	14.0 (0.6; 28.0)	6.8 (-7.3; 21.0)
6.5	-11.9 (-25.7; 1.6)	14.0 (-0.2; 27.0)	5.9 (-8.4;20.0)
4.5	14.0 (-27.6; -0.4)	11.0 (-2.1; 25.0)	3.8 (-10.0; 18.0)

Table E.6 Constancy not assumed, 60% NI margin=10.11, n1=n2= 30

μ_t	$\mu_T - \mu_C$ from NI trial (2020) (95% CrI), se=3.08	NMA ($\mu_T - \mu_P$) (95% CrI), se= 3.14	NMR ($\mu_T - \mu_P$) in (2020) (95% CrI), se= 3.68
18.5	0.0 (-6.0; 6.1)	25.0 (19.0; 32.0)	18.0 (11; 25.0)
16.5	-2.0 (-8.1; 4.1)	23.0 (17.0; 30.0)	16.0 (8.5; 23.0)
15.5	-3.0 (-9.1; 3.1)	22.0 (16.0; 29.0)	15.0 (7.5; 22.0)
14.5	-4.0 (-10.1; 2.1)	21.0 (15.0; 28.0)	14.0 (6.6; 21.0)
12.5	-6.0 (-12.1; 0.1)	19.0 (13.0; 26.0)	12.0 (4.6; 19.0)
11.5	-7.0(-13.1; -1.0)	18.0 (12.0; 25.0)	11.0 (3.6; 18.0)
9.5	-9.0 (-15.0; -2.8)	16.0 (10.0; 23.0)	8.9 (1.6; 16.0)
7.5	-11.0 (-17.1; -4.9)	14.0 (8.3; 21.0)	6.8 (-0.4; 14.0)
3.5	-15.0 (-21.0; -.8.9)	10.0 (4.3; 17)	2.8 (-4.4; 10.0)
0.5	-18.0 (-24.1; -11.9)	7.4 (1.2; 14.0)	-0.2 (-7.5; 7.0)

Table E.7 Constancy not assumed, 70% NI margin= 11.79, n1=n2= 22

μ_t	$\mu_T - \mu_C$ from NI trial (2020) (95% CrI), se=3.6	NMA ($\mu_T - \mu_P$) (95% CrI), se= 3.6	NMR ($\mu_T - \mu_P$) in (2020) (95% CrI), se= 4.13
18.5	0.0 (-7.0; 7.1)	25.0 (18.0; 32.0)	18.0 (9.7; 26.0)
16.5	-2.0 (-9.1; 5.1)	23.0 (16.0; 30.0)	16.0 (7.7; 24.0)
14.5	-4.0 (-11.1; 3.1)	21.0 (14.0; 29.0)	14.0 (5.7; 22.0)
14.0	-4.5 (-11.5; 2.5)	21.0 (14.0; 28.0)	13.0 (5.3; 21.0)

13.5	-5.0 (-12.2; 2.1)	20.0 (13.0; 28.0)	13.0 (4.7; 21.0)
11.5	-7.0(-14.1; 0.0)	18.0 (11.0; 26.0)	11.0 (2.7; 19.0)
9.5	-9.0 (-16.2; -1.9)	16.0 (9.3; 24.0)	8.8 (0.6; 17.0)
8.5	-10.0 (-17.1; -2.9)	15.0 (8.3; 23.0)	7.8 (-0.2; 16.0)
3.5	-15.0 (-22.0; -.7.8)	10.0 (3.3; 18)	2.8 (-5.3; 11.0)
0.5	-18.0 (-25.1; -10.9)	7.5 (0.2; 15.0)	-0.2 (-8.3; 8.0)

Table E.8 Constancy not assumed, 80% NI margin= 13.48, n1=n2= 17

μ_t	$\mu_T - \mu_C$ from NI trial (2020) (95% CrI), se=3.6	NMA ($\mu_T - \mu_P$) (95% CrI), se= 3.6	NMR ($\mu_T - \mu_P$) in (2020) (95% CrI), se= 4.13
18.5	0.0 (-8.1; 8.1)	25.0 (17.0; 34.0)	18.0 (8.8; 27.0)
16.5	-2.0 (-10.1; 6.1)	23.0 (15.0; 32.0)	16.0 (6.8; 25.0)
14.5	-4.0 (-12.1; 4.1)	21.0 (13.0; 30.0)	14.0 (4.9; 23.0)
13.0	-5.5 (-13.5; 2.6)	20.0 (12.0; 28.0)	12.0 (3.4; 21.0)
11.5	-7.0(-15.1; 1.0)	18.0 (10.0; 26.0)	11.0 (1.9; 20.0)
9.5	-9.0 (-17.1; -0.9)	16.0 (8.3; 25.0)	8.8 (-0.2; 18.0)
7.5	-11.0 (-19.2; -3.0)	14.0 (6.4; 23.0)	6.8 (-2.3; 16.0)
3.5	-15.0 (-23.0; -.6.8)	10.0 (2.3; 19)	2.9 (-6.1; 12.0)
0.5	-18.0 (-26.1; -10.0)	7.5 (-0.6; 16.0)	-0.2 (-9.2; 8.7)

Table E.9Constancy not assumed, 90% NI margin= 15.6, n1=n2= 14

μ_t	$\mu_T - \mu_C$ from NI trial (2020) (95% CrI), se=3.6	NMA ($\mu_T - \mu_P$) (95% CrI), se= 3.6	NMR ($\mu_T - \mu_P$) in (2020) (95% CrI), se= 4.13
18.5	0.0 (-8.9; 8.9)	25.0 (16.0; 34.0)	18.0 (8.8; 28.0)
16.5	-2.0 (-10.9; 6.7)	23.0 (14.0; 32.0)	16.0 (6.0; 25.0)
14.5	-4.0 (-12.9; 4.9)	22.0 (13.0; 30.0)	14.0 (4.1; 24.0)
12.5	-6.0 (-14.8; 2.9)	19.0 (11.0; 28.0)	12.0 (2.2; 22.0)
11.5	-7.0 (-15.9; 1.9)	18.0 (9.5; 27.0)	11.0 (1.0; 21.0)
10.5	-8.0 (-16.8; 1.1)	17.0 (8.4; 26.0)	9.9 (0.2; 20.0)
9.5	-9.0 (-17.9; 0.0)	16.0 (7.5; 25.0)	8.8 (-0.8; 19.0)
8.5	-10.0 (-18.9; -1.0)	15.0 (6.5; 24.0)	7.9 (-2.0; 18.0)
1.5	-17.0 (-25.9; -8.2)	8.4 (-0.5; 17.0)	0.8 (-8.9; 11.0)

Table E.10 Constancy not assumed, NI margin=M1= 16.85 , n1=n2= 11

μ_t	$\mu_T - \mu_C$ from NI trial (2020) (95% CrI), se=3.6	NMA ($\mu_T - \mu_P$) (95% CrI), se= 3.6	NMR ($\mu_T - \mu_P$) in (2020) (95% CrI), se= 4.13
18.5	0.0 (-10.1; 10.1)	25.0 (15.0; 36.0)	18.0 (7.0; 29.0)
16.5	-2.0 (-12.0; 8.0)	23.0 (13.0; 33.0)	16.0 (5.0; 27.0)
14.5	-4.0 (-14.0; 6.0)	21.0 (11.0; 31.0)	14.0 (3.1; 25.0)
12.5	-6.0 (-16.0; 4.0)	20.0 (10.0; 30.0)	12.0 (1.1; 23.0)
12.0	-6.5 (-16.6; 3.5)	19.0 (8.9; 29.0)	11.0 (0.5; 22.0)

11.5	-7.0(-17.3; 3.0)	18.0 (8.4; 28.0)	11.0 (0.0; 22.0)
10.5	-8.0 (-18.0; 2.0)	17.0 (7.4; 27.0)	9.8 (-0.9; 21.0)
7.5	-11.0 (-20.9; -0.9)	15.0 (4.3; 25.0)	6.8 (-3.9; 18.0)
2.5	-16.0 (-26.0; -5.9)	9.4 (-0.7; 19)	1.8 (-8.9; 13.0)

E. 1. 3 Use of 50% of M1 to set M2 in the years of 2025, 2030

Table E.11 Comparison of the mean difference between the placebo and test treatment assuming the constancy(2025), NI=-12.25, N= 21

μ_t	$\mu_T - \mu_C$ from NI trial (95% CrI), SE = (3.7)	NMA ($\mu_T - \mu_P$) (95% CrI), SE = (3.74)	NMR ($\mu_T - \mu_P$) in (95% CrI), SE = (4.53)
18.5	0.0 (-7.2; 7.2)	25.0 (18.0; 33.0)	15.0 (6.4; 24.0)
16.5	-2.0 (-9.2; 5.3)	23.0 (16.0; 31.0)	13.0 (4.5; 22.0)
15.5	-3.0 (-10.0; 4.3)	22.0 (15.0; 30.0)	12.0 (3.4; 21.0)
14.5	-4.0 (-11.0; 3.2)	21.0 (14.0; 29.0)	11.0 (2.5; 20.0)
13.5	-5.0 (-12.3; 2.2)	20.0 (13.0; 28.0)	10.0 (1.6; 19.0)
12.5	-6.0 (-13.2; 1.3)	19.0 (12.0; 27.0)	9.4 (0.5; 18.0)
10.5	-8.0 (-15.3; -0.7)	17.0 (10.0; 25.0)	7.5 (-1.5; 16.0)
8.5	-10.0 (-17.3; -2.7)	15.0 (8.2; 23.0)	5.3 (-3.6; 14.0)
7.0	-11.4 (-18.5; -4.4)	14.0 (6.8; 21.0)	3.9 (-5.1; 13.0)
1.5	-17.0 (-24.3; -9.7)	8.5 (1.1; 16.0)	-1.7 (-11.0; 7.2)
0.0	-18.5 (-25.7; -11.2)	6.9 (-0.4; 14.0)	-3.2 (-12.0; 5.8)

NMA: network meta-analysis, NMR: network meta-regression, the $\mu_t - \mu_c$ refers to the mean difference between the active control and the test treatment, negative sign means the test treatment is less effective than the active control, SE: Standard error, μ_t is the treatment effect in the test group in the NI trial

Table E.12 Comparison of the mean difference between the placebo and test treatment assuming the constancy(2030), NI=-12.25, N=21

μ_t	$\mu_T - \mu_C$ from NI trial (2025) (95% CrI), SE = (3.7)	NMA ($\mu_T - \mu_P$) (95% CrI), SE = (3.74)	NMR ($\mu_T - \mu_P$) in (2025) (95% CrI), SE = (4.53)
18.5	0.0 (-7.2; 7.2)	25.0 (18.0; 33.0)	12.9 (3.2; 22.5)
16.5	-2.0 (-9.2; 5.3)	23.0 (16.0; 31.0)	10.9 (1.32; 20.5)
15.5	-3.0 (-10.0; 4.3)	22.0 (15.0; 30.0)	9.9 (0.27; 19.7)
14.5	-4.0 (-11.0; 3.2)	21.0 (14.0; 29.0)	8.9 (-0.7; 18.6)
13.5	-5.0 (-12.3; 2.2)	20.0 (13.0; 28.0)	7.8 (-1.9; 17.5)
12.5	-6.0 (-13.2; 1.3)	19.0 (12.0; 27.0)	6.9 (-2.8; 16.5)
10.5	-8.0 (-15.3; -0.7)	17.0 (10.0; 25.0)	4.8 (-4.7; 14.5)
8.5	-10.0 (-17.3; -2.7)	15.0 (8.2; 23.0)	2.8(-6.8; 12.5)
7.0	-11.4 (-18.5; -4.4)	14.0 (6.8; 21.0)	1.3(-8.3; 11.0)
1.5	-17.0 (-24.3; -9.7)	8.5 (1.1; 16.0)	-4.1(-13.8; 5.5)
0.0	-18.5 (-25.7; -11.2)	6.9 (-0.4; 14.0)	-5.6(-15.2; 4.0)

NMA: network meta-analysis, NMR: network meta-regression, the $\mu_t - \mu_c$ refers to the mean difference between the active control and the test treatment, negative sign means the test treatment is less effective than the active control, SE: Standard error, μ_t is the treatment effect in the test group in the NI trial

Table E.13 Comparison of the mean difference between the placebo and test treatment, constancy not assumed (2025) NI=-7.17, N= 59

μ_t	$\mu_T - \mu_C$ from NI trial (2025) (95% CrI), SE= 2.20	NMA ($\mu_T - \mu_P$) (95% CrI), SE= 2.26	NMR ($\mu_T - \mu_P$) in (2025) (95% CrI), SE= 3.46
18.5	0.0 (-4.3; 4.3)	25.0 (21.0; 30.0)	15.4 (8.6; 22.2)
16.5	-2.0 (-6.3; 2.3)	23.0 (19.0; 28.0)	13.3 (6.6; 20.0)
16.0	-2.5 (-6.8; 1.8)	22.9 (18.5; 27.4)	12.9 (6.2; 19.5)
15.5	-3.0 (-7.3; 1.3)	22.0 (18.0; 27.0)	12.4 (5.6; 19.0)
13.5	-5.0 (-9.3; -0.7)	20.0 (16.0; 25.0)	10.3 (3.6; 17.1)
12.5	-6.0 (-10.3; -1.7)	19.0 (15.0; 24.0)	9.3 (2.5; 16 .0)
11.5	-7.0 (-11.3; -2.7)	18.0 (14.0; 23.0)	8.4 (1.6; 15.2)
10.5	-8.0 (-12.3; -3.7)	17.0 (13.0; 22.0)	7.4 (0.6; 14.1)
8.5	-10.0 (-14.3; -5.7)	15.0 (11.0; 20.0)	5.4 (-1.3; 12.2)
7.0	-11.5 (-15.8; -7.1)	14.0 (9.5; 18.0)	3.8 (-2.8; 10.6)
1.5	-17.0 (-21.3; -12.6)	8.5 (4.1; 13.0)	-1.6 (-8.2; 4.9)
0.0	-18.5 (-22.8; -14.1)	6.9 (2.5; 11.0)	-3.2 (-9.8; 3.6)

NMA: network meta-analysis, NMR: network meta-regression, the $\mu_t - \mu_c$ refers to the mean difference between the active control and the test treatment, negative sign means the test treatment is less effective than the active control

Table E.14 Comparison of the mean difference between the placebo and test treatment, constancy not assumed (2030) NI=-5.9, N= 87

μ_t	$\mu_T - \mu_C$ from NI trial (2030) (95% CrI), SE= 1.8	NMA ($\mu_T - \mu_P$) (95% CrI), SE= 1.8	NMR ($\mu_T - \mu_P$) in (2030) (95% CrI), SE= 3.7
18.5	0.0 (-3.5; 3.5)	25.0 (22.0; 29.0)	13.0 (5.6; 20.2)
16.5	-2.0 (-5.6; 1.5)	23.0 (20.0; 27.0)	10.8 (3.6; 18.1)
16.0	-2.5 (-6.0; 1.1)	22.9 (19.2; 26.6)	10.3 (3.17; 1.7)
15.5	-3.0 (-6.5; 0.6)	22.0 (19.0; 26.0)	9.8 (2.5; 17.2)
13.5	-5.0 (-8.5; -1.4)	20.0 (17.0; 24.0)	7.6 (0.6; 15.1)
12.5	-6.0 (-9.5; -2.4)	19.0 (16.0; 23.0)	6.8 (-0.6; 14.0)
11.5	-7.0 (-10.6; -3.5)	18.0 (15.0; 22.0)	5.9 (-1.4; 13.2)
10.5	-8.0 (-11.6; -4.5)	17.0 (14.0; 21.0)	4.9 (-2.41; 12.2)
8.5	-10.0 (-13.6; -6.4)	15.0 (12.0; 19.0)	2.9 (-4.3; 10.3)
7.0	-11.5 (-15.0; -7.9)	14.0 (10.0; 18.0)	1.4 (-5.8; 8.6)
1.5	-17.0 (-20.6; -13.5)	8.5 (4.8; 12.0)	-4.2 (-11.6; 3.1)
0.0	-18.5 (-22.0; -15.0)	6.9 (3.2; 11.0)	-5.7 (-13.1; 1.7)

NMA: network meta-analysis, NMR: network meta-regression, the $\mu_t - \mu_c$ refers to the mean difference between the active control and the test treatment, negative sign means the test treatment is less effective than the active control

E. 2 Lidocaine for pain reduction

E. 2. 1 Random effects pairwise meta-regression

```
Results on the Log Odds Ratio scale

Iterations = 5001:25000
Thinning interval = 1
Number of chains = 4
Sample size per chain = 20000
1. Empirical mean and standard deviation for each variable,
  plus standard error of the mean:
                Mean     SD  Naive SE Time-series SE
d.Lidocaine.placebo 1.7053 0.1328 0.0004696      0.003390
sd.d             0.1858 0.1389 0.0004912      0.005236
B               0.1187 0.2575 0.0009103      0.002957
2. Quantiles for each variable:
                  2.5%     25%    50%    75%  97.5%
d.Lidocaine.placebo  1.455225  1.61537 1.7016 1.7906 1.9774
sd.d             0.004903  0.07111 0.1612 0.2728 0.5043
B               -0.383180 -0.05159 0.1175 0.2885 0.6317
-- Model fit (residual deviance):

   Dbar      pD     DIC
46.92977 27.93494 74.86470
```

E. 2. 2 Use of different percentages of M1 to set M2

Table E.15 60% of M1, NI margin= 2.35 , n1=n2=227 (Assuming constancy)

Risk of failure (π_T)	$OR(\pi_T/\pi_{C)}$ from NI trial (2020) se= 0.26	NMA, OR (π_P/π_T) se= 0.28	NMR, OR (π_P/π_T) in (2020) se= 0.43
15.00%	1.00 (0.60; 1.70)	5.30 (3.00; 9.30)	6.30 (2.70; 15.00)
17.00%	1.10 (0.69; 1.90)	4.70 (2.70; 8.10)	5.50 (2.40; 13.00)
19.00%	1.30 (0.80; 2.20)	4.00 (2.40; 6.90)	4.70 (2.00; 11.00)
20.00%	1.40 (0.86; 2.29)	3.80 (2.20; 6.50)	4.40 (1.90; 10.00)
21.00%	1.47 (0.90; 2.40)	3.61 (2.10; 6.13)	4.23 (1.85; 9.92)
22.00%	1.56 (0.97; 2.54)	3.43 (2.02; 5.78)	4.00 (1.77; 9.29)
23.00%	1.68 (1.05; 2.70)	3.18 (1.87; 5.36)	3.70 (1.63; 8.55)
25.00%	1.85 (1.16; 2.98)	2.88 (1.69; 4.85)	3.37 (1.47; 7.75)
30.00%	2.41 (1.53; 3.85)	2.21 (1.31; 3.67)	2.60 (1.15; 5.98)
35.00%	3.01 (1.92; 4.79)	1.77 (1.06; 2.90)	2.08 (0.91; 4.70)
37.00%	3.24 (2.08; 5.11)	1.64 (0.98; 2.70)	1.930.86; 4.32)

Table E.16. 70% of M1, NI margin= 2.70 , n1=n2=167 (Assuming constancy)

Risk of failure (π_T)	$OR(\pi_T/\pi_{C)}$ from NI trial (2020) se= 0.31	NMA, OR (π_P/π_T) se= 0.32	NMR, OR (π_P/π_T) in (2020) se= 0.46
15.00%	1.00 (0.54; 1.82)	5.33 (2.83; 10.00)	6.22 (2.55; 15.50)
17.00%	1.14 (0.64; 2.05)	4.68 (2.52; 8.72)	5.45 (2.21; 13.5)
19.00%	1.29 (0.73; 2.28)	4.12 (2.23; 7.57)	4.87 (2.01; 12.00)
21.00%	1.50 (0.86; 2.62)	3.56 (1.94; 6.52)	4.18 (1.73; 10.10)
22.00%	1.55 (0.89; 2.72)	3.45 (1.88; 6.23)	4.02 (1.66; 9.79)
23.00%	1.66 (0.96; 2.93)	3.20 (1.74; 5.83)	3.76 (1.58; 9.20)
25.00%	1.83 (1.06; 3.18)	2.91 (1.61; 5.23)	3.40 (1.45; 8.15)
30.00%	2.39 (1.42; 4.14)	2.22 (1.23; 3.94)	2.63 (1.09; 6.30)
33.00%	2.76 (1.64; 4.71)	1.92 (1.07; 3.38)	2.25 (0.96; 5.43)
38.00%	2.98 (1.76; 5.10)	1.78 (0.99; 3.13)	2.11 (0.89; 4.99)

Table E.18. 80% of M1, NI margin= 3.12 , n1=n2=128 (Assuming constancy)

Risk of failure (π_T)	$OR(\pi_T/\pi_{C)}$ from NI trial (2020) se= 0.34	NMA, OR (π_P/π_T) se= 0.36	NMR, OR (π_P/π_T) in (2020) se= 0.49
15.00%	1.00 (0.51; 1.98)	5.32 (2.59; 10.9)	6.27 (2.38; 16.50)
17.00%	1.12 (0.57; 2.20)	4.74 (2.33; 9.54)	5.53 (2.14; 14.60)
19.00%	1.32 (0.69; 2.55)	4.04 (2.00; 8.05)	4.76 (1.86; 12.30)
21.00%	1.45 (0.76; 2.79)	3.65 (1.82; 7.16)	4.29 (1.69; 11.00)
22.00%	1.59 (0.85; 3.05)	3.35 (1.71; 6.47)	3.88 (1.55; 9.84)
23.00%	1.66 (0.88; 3.18)	3.22 (1.63; 6.24)	3.75 (1.45; 9.60)
25.00%	1.81 (0.98; 3.45)	2.93 (1.49; 5.67)	3.42 (1.37; 8.61)
26.00%	1.97 (1.08; 3.71)	2.70 (1.36; 5.18)	3.17 (1.26; 7.88)
30.00%	2.38 (1.31; 4.43)	2.24 (1.16; 4.27)	2.62 (1.06; 6.65)
33.00%	2.76 (1.53; 5.16)	1.93 (1.00; 3.66)	2.23 (0.89; 5.57)

Table E.19. 90% of M1, NI margin= 3.59 , n1=n2=101 (Assuming constancy)

Risk of failure (π_T)	$OR(\pi_T/\pi_{C)}$ from NI trial (2020) se= 0.39	NMA, OR (π_P/π_T) se= 0.40	NMR, OR (π_P/π_T) in (2020) se= 0.52
15.00%	1.00 (0.46; 2.18)	5.29 (2.38; 11.8)	6.23 (2.22; 17.5)
17.00%	1.16 (0.55; 2.46)	4.59 (2.10; 9.95)	5.40 (1.94; 14.9)
19.00%	1.32 (0.64; 2.78)	4.03 (1.87; 8.60)	4.70 (1.72; 13.10)
21.00%	1.48 (0.73; 3.08)	3.58 (1.66; 7.57)	4.22 (1.56; 11.30)
23.00%	1.68 (0.82; 3.48)	3.20 (1.50; 6.70)	3.72 (1.38; 9.93)
24.00%	1.77 (0.88; 3.64)	3.03 (1.41; 6.29)	3.51 (1.33; 9.16)
25.00%	1.86 (0.93; 3.80)	2.87 (1.34; 5.90)	3.38 (1.25; 8.93)

27.00%	2.06 (1.04; 4.18)	2.58 (1.24; 5.27)	3.02 (1.15; 7.98)
30.00%	2.40 (1.22; 4.87)	2.23 (1.07; 4.50)	2.58 (0.98; 6.83)
31.00%	2.49 (1.27; 5.05)	2.13 (1.01; 4.32)	2.50 (0.94; 6.62)
32.00%	2.60 (1.33; 5.23)	2.03 (0.97; 4.09)	2.41 (0.92; 6.30)

Table E.20. 100% of M1, NI margin= 4.14 , n1=n2=82 (Assuming constancy)

Risk of failure (π_T)	$OR(\pi_T/\pi_{C)}$ from NI trial (2020) se= 0.43	NMA, OR (π_P/π_T) se= 0.45	NMR, OR (π_P/π_T) in (2020) se= 0.55
15.00%	1.00 (0.43; 2.33)	5.35 (2.21; 12.9)	6.22 (2.14; 18.70)
17.00%	1.09 (0.47; 2.57)	4.88 (2.02; 11.5)	5.69 (1.87; 16.9)
19.00%	1.30 (0.57; 2.99)	4.12 (1.74; 9.55)	4.79 (1.63; 13.8)
21.00%	1.51 (0.69; 3.40)	3.54 (1.52; 8.08)	4.11 (1.47; 11.8)
23.00%	1.61 (0.74; 3.58)	3.29 (1.42; 7.47)	3.84 (1.36; 11.0)
25.00%	1.85 (0.85; 4.13)	2.88 (1.25; 6.40)	3.35 (1.19; 9.44)
26.00%	1.97 (0.92; 4.34)	2.69 (1.17; 5.94)	3.17 (1.13; 8.84)
27.00%	2.08 (0.98; 4.59)	2.55 (1.11; 5.68)	3.00 (1.09; 8.39)
29.00%	2.23 (1.05; 4.89)	2.41 (1.07; 5.24)	2.80 (1.01; 7.65)
30.00%	2.36 (1.12; 5.17)	2.26 (0.99; 4.89)	2.60 (0.94; 7.30)

Table E.21. 60% of M1, NI margin= 2.38 , n1=n2=220 (adjusted for time)

Risk of failure (π_T)	$OR(\pi_T/\pi_{C)}$ from NI trial (2020) se= 0.27	NMA, OR (π_P/π_T) se= 0.29	NMR, OR (π_P/π_T) in (2020) se= 0.43
15.00%	1.00 (0.58; 1.69)	5.32 (2.99; 9.46)	6.28 (2.66; 14.80)
17.00%	1.19 (0.72; 2.01)	4.51 (2.57; 7.85)	5.21 (2.26; 12.20)
19.00%	1.34 (0.82; 2.23)	3.98 (2.27; 6.88)	4.60 (1.99; 10.80)
20.00%	1.42 (0.86; 2.35)	3.73 (2.15; 6.47)	4.38 (1.91; 10.50)
21.00%	1.55 (0.94; 2.54)	3.47 (2.01; 5.90)	4.05 (1.77; 9.43)
23.00%	1.72 (1.06; 2.83)	3.11 (1.80; 5.31)	3.60 (1.56; 8.43)
25.00%	1.90 (1.18; 3.07)	2.80 (1.64; 4.74)	3.29 (1.43; 7.57)
30.00%	2.44 (1.54; 3.94)	2.18 (1.28; 3.65)	2.55 (1.12; 5.92)
35.00%	3.07 (1.95; 4.91)	1.73 (1.03; 2.89)	2.02 (0.88; 4.62)
36.00%	3.26 (2.07; 5.20)	1.63 (0.97; 2.70)	1.92 (0.85; 4.33)

Table E.22 70% of M1, NI margin= 2.75 , n1=n2=162 (adjusted for time)

Risk of failure (π_T)	$OR(\pi_T/\pi_{C)}$ from NI trial (2020) se= 0.31	NMA, OR (π_P/π_T) se= 0.32	NMR, OR (π_P/π_T) in (2020) se= 0.46

15.00%	1.00 (0.54; 1.84)	5.33 (2.82; 10.20)	6.24 (2.50; 15.30)
17.00%	1.15 (0.63; 2.08)	4.63 (2.47; 8.72)	5.40 (2.23; 13.50)
19.00%	1.30 (0.72; 2.34)	4.11 (2.18; 7.68)	4.77 (1.96; 11.8)
21.00%	1.52 (0.86; 2.71)	3.50 (1.88; 6.40)	4.09 (1.70; 10.1)
23.00%	1.69 (0.97; 2.99)	3.15 (1.70; 5.76)	3.70 (1.52; 8.77)
25.00%	1.88 (1.08; 3.30)	2.86 (1.55; 5.19)	3.30 (1.38; 8.02)
30.00%	2.40 (1.40; 4.18)	2.21 (1.22; 3.95)	2.62 (1.10; 6.26)
33.00%	2.76 (1.62; 4.76)	1.93 (1.07; 3.41)	2.26 (0.95; 5.39)
35.00%	3.01 (1.78; 5.24)	1.77 (0.98; 3.13)	2.040.85; 4.82)

Table E.23. 80 % of M1, NI margin= 3.18 , n1=n2=124 (adjusted for time)

Risk of failure (π_T)	$OR(\pi_T/\pi_{C)}$ from NI trial (2020) se= 0.35	NMA, OR (π_P/π_T) se= 0.37	NMR, OR (π_P/π_T) in (2020) se= 0.51
15.00%	1.00 (0.49; 2.01)	5.33 (2.57; 11.10)	6.21 (2.34; 16.7)
17.00%	1.19 (0.61; 2.37)	4.43 (2.16; 9.12)	5.22 (2.00; 13.80)
19.00%	1.33 (0.68; 2.62)	4.00 (1.97; 8.08)	4.68 (1.79; 12.20)
21.00%	1.55 (0.82; 3.00)	3.45 (1.72; 6.82)	4.00 (1.57; 10.30)
22.00%	1.63 (0.85; 3.13)	3.30 (1.64; 6.58)	3.85 (1.49; 9.97)
23.00%	1.69 (0.9; 3.27)	3.11 (1.56; 6.09)	3.68 (1.42; 9.43)
25.00%	1.86 (0.98; 3.57)	2.88 (1.45; 5.60)	3.35 (1.32 (8.61)
30.00%	2.47 (1.33; 4.66)	2.16 (1.10; 4.13)	2.49 (0.98; 6.30)
33.00%	2.77 (1.50; 5.25)	1.94 (0.98; 3.68)	2.26 (0.88; 5.69)

Table E.24. 90% of M1, NI margin= 3.67 , n1=n2=98 (adjusted for time)

Risk of failure (π_T)	$OR(\pi_T/\pi_{C)}$ from NI trial (2020) se= 0.40	NMA, OR (π_P/π_T) se= 0.41	NMR, OR (π_P/π_T) in (2020) se= 0.53
15.00%	1.00 (0.45; 2.20)	5.32 (2.35; 12.10)	6.32 (2.23; 18.0)
17.00%	1.16 (0.53; 2.55)	4.60 (2.05; 10.10)	5.35 (1.92; 15.40)
19.00%	1.34 (0.63; 2.85)	3.97 (1.76; 8.59)	4.70 (1.71; 13.00)
21.00%	1.52 (0.73; 3.22)	3.52 (1.61; 7.57)	4.14 (1.50; 11.30)
23.00%	1.72 (0.83; 3.61)	3.10 (1.43; 6.59)	3.61 (1.33; 9.73)
24.00%	1.81 (0.88; 3.81)	2.94 (1.37; 6.19)	3.45 (1.25; 9.25)
25.00%	1.92 (0.94; 4.06)	2.80 (1.29; 5.85)	3.20 (1.18; 8.64)
26.00%	2.03 (1.01; 4.28)	2.64 (1.22; 5.56)	3.08 (1.12; 8.30)
30.00%	2.48 (1.25; 5.10)	2.15 (1.00; 4.41)	2.49 (0.92; 6.60)

Table E.25. 100% of M1, NI margin= 4.24 , n1=n2= 80 (adjusted for time)

Risk of failure (π_T)	$OR(\pi_T/\pi_{C)}$ from NI trial (2020) se= 0.44	NMA, OR (π_P/π_T) se= 0.46	NMR, OR (π_P/π_T) in (2020) se= 0.57
15.00%	1.00 (0.42; 2.43)	5.31 (2.13; 13.30)	6.22 (2.03; 19.30)
17.00%	1.21 (0.51; 2.88)	4.38 (1.78; 10.60)	5.13 (1.71; 15.60)
19.00%	1.42 (0.62; 3.30)	3.74 (1.55; 8.86)	4.41 (1.50; 13.00)
21.00%	1.54 (0.68; 3.59)	3.47 (1.45; 8.04)	4.04 (1.38; 11.70)
23.00%	1.79 (0.80; 4.11)	2.98 (1.27; 6.84)	3.49 (1.21; 10.10)

25.00%	1.91 (0.87; 4.35)	2.79 (1.19; 6.38)	3.27 (1.13; 9.27)
26.00%	2.05 (0.93; 4.66)	2.61 (1.12; 5.92)	3.03 (1.04; 8.62)
27.00%	2.17 (0.99; 4.91)	2.45 (1.06; 5.47)	2.87 (1.01; 8.20)
28.00%	2.31 (1.06; 5.23)	2.27 (0.97; 5.11)	2.69 (0.93; 7.52)

E. 2. 3 Use of 50% of M1 in the years 2025, 2030

Table E.26. Comparison of the odds ratio between the placebo and test treatment assuming the constancy (2025), NI =2.03

Risk of failure (π_T)	$OR(\pi_T/\pi_{C)}$ from NI trial (2025) (95% CrI),SE=0.22	NMA, OR (π_P/π_T) (95% CrI),SE= 0.22	NMR, OR (π_P/π_T) in (2025) (95% CrI), SE= 0.41
15.00%	1.0 (0.65; 1.53)	5.31 (3.29; 8.62)	6.47 (2.60; 16.70)
17.00%	1.10 (0.72; 1.67)	4.87 (3.02; 7.87)	5.60 (2.23; 14.70)
19.00%	1.32 (0.88; 2.00)	4.01 (2.53; 6.39)	4.88 (1.96; 12.10)
20.00%	1.40 (0.94; 2.11)	3.80 (2.41; 6.00)	4.60 (1.86; 11.70)
21.00%	1.51 (1.01; 2.27)	3.54 (2.23; 5.54)	4.30 (1.75; 11.10)
23.00%	1.68 (1.13; 2.50)	3.18 (2.01; 4.99)	3.81 (1.55; 9.97)
25.00%	1.89 (1.28; 2.81)	2.83 (1.80; 4.41)	3.42 (1.36; 8.68)
30.00%	2.41 (1.65; 3.55)	2.20 (1.42; 3.41)	2.69 (1.10; 6.74)
35.00%	3.05 (2.11; 4.47)	1.75 (1.12; 2.69)	2.10 (0.86; 5.40)
40.00%	3.75 (2.59; 5.49)	1.42 (0.91; 2.19)	1.75 (0.72; 4.34)

NMA: network meta-analysis, NMR: network meta-regression, the π_t/π_c refers to the odds ratio between the test treatment and the active control, the π_p/π_t refers to the odds ratio between the placebo and the test treatment odds ratio >1 indicates worse outcome (high pain intensity

Table E.27. Comparison of the odds ratio between the placebo and test treatment assuming the constancy (2030), NI =2.03

Risk of failure (π_T)	$OR(\pi_T/\pi_{C)}$ from NI trial (2030) (95% CrI),SE=0.22	NMA, OR (π_P/π_T) (95% CrI),SE= 0.22	NMR, OR (π_P/π_T) in (2030) (95% CrI), SE= 0.41
15.00%	1.0 (0.65; 1.53)	5.31 (3.29; 8.62)	6.63 (2.35; 19.10)
17.00%	1.10 (0.72; 1.67)	4.87 (3.02; 7.87)	5.92 (2.07; 17.70)
19.00%	1.32 (0.88; 2.00)	4.01 (2.53; 6.39)	5.08 (1.78; 14.50)
20.00%	1.40 (0.94; 2.11)	3.80 (2.41; 6.00)	4.82 (1.75; 14.20)
21.00%	1.51 (1.01; 2.27)	3.54 (2.23; 5.54)	4.36 (1.60; 12.70)
23.00%	1.68 (1.13; 2.50)	3.18 (2.01; 4.99)	4.03 (1.41; 11.70)
25.00%	1.89 (1.28; 2.81)	2.83 (1.80; 4.41)	3.53 (1.22; 9.97)
30.00%	2.41 (1.65; 3.55)	2.20 (1.42; 3.41)	2.77 (0.97; 7.94)
35.00%	3.05 (2.11; 4.47)	1.75 (1.12; 2.69)	2.20 (0.75; 6.49)
40.00%	3.62 (2.44; 5.44)	1.42 (0.91; 2.19)	1.79 (0.62; 5.22)

NMA: network meta-analysis, NMR: network meta-regression, the π_t/π_c refers to the odds ratio between the test treatment and the active control, the π_p/π_t refers to the odds ratio between the placebo and the test treatment odds ratio >1 indicates worse outcome (high pain intensity)

Table E.28. Comparison of the odds ratio between the placebo and test treatment assuming the constancy (2025), NI =2.07, N= 310

Risk of failure (π_T)	$OR(\pi_T/\pi_{C)}$ from NI trial (2025) (95% CrI),SE=0.23	NMA, OR (π_P/π_T) (95% CrI),SE= 0.25	NMR, OR (π_P/π_T) in (2025) (95% CrI), SE= 0.47
15.00%	1.00 (0.64; 1.55)	5.33 (3.27; 8.72)	6.52 (2.63; 16.30)
17.00%	1.15 (0.75; 1.77)	4.62 (2.84; 7.44)	5.55 (2.18; 14.20)
19.00%	1.32 (0.87; 2.03)	4.04 (2.50; 6.54)	4.84 (1.92; 12.40)
20.00%	1.40 (0.93; 2.14)	3.81 (2.37; 6.09)	4.66 (1.88; 12.00)
21.00%	1.51 (1.0; 2.30)	3.52 (2.18; 5.59)	4.35 (1.69; 11.00)
23.00%	1.69 (1.13; 2.55)	3.13 (1.96; 4.95)	3.82 (1.52; 9.67)
25.00%	1.89 (1.27; 2.85)	2.82 (1.77; 4.46)	3.37 (1.35; 8.48)
30.00%	2.41 (1.63; 3.60)	2.22 (1.40; 3.47)	2.66 (1.07; 6.83)
35.00%	3.04 (2.09; 4.54)	1.75 (1.11; 2.72)	2.14 (0.84; 5.32)
40.00%	3.74 (2.56; 5.56)	1.42 (0.90; 2.21)	1.71 (0.69; 4.36)

NMA: network meta-analysis, NMR: network meta-regression, the π_t/π_c refers to the odds ratio between the test treatment and the active control, the π_p/π_t refers to the odds ratio between the placebo and the test treatment odds ratio >1 indicates worse outcome (high pain intensity)

Table E.29. Comparison of the Odds ratio between the placebo and test treatment assuming the constancy (2030), NI =2.09, N= 305

Risk of failure (π_T)	$OR(\pi_T/\pi_{C)}$ from NI trial (2025) (95% CrI),SE=0.23	NMA, OR (π_P/π_T) (95% CrI),SE= 0.25	NMR, OR (π_P/π_T) in (2025) (95% CrI), SE= 0.54
15.00%	1.00 (0.64; 1.56)	5.34 (3.25; 8.80)	6.56 (2.31; 19.30)
17.00%	1.16 (0.75; 1.80)	4.60 (2.81; 7.48)	5.65 (2.31; 16.80)
19.00%	1.32 (0.87; 2.02)	4.03 (2.50; 6.50)	5.15 (1.75; 15.20)
20.00%	1.41 (0.93; 2.14)	3.78 (2.35; 6.08)	4.71 (1.66; 13.40)
21.00%	1.53 (1.00; 2.33)	3.48 (2.18; 5.56)	4.44 (1.53; 12.80)
23.00%	1.72 (1.14; 2.60)	3.11 (1.93; 4.36)	3.88 (1.38; 11.80)
25.00%	1.94 (1.30; 2.95)	2.75 (1.71; 4.36)	3.46 (1.19; 8.12)
30.00%	2.44 (1.65; 3.65)	2.19 (1.37; 3.44)	2.75 (0.96; 8.12)
35.00%	3.06 (2.08; 4.57)	1.75 (1.11; 2.74)	2.12 (0.75; 5.95)
40.00%	3.77 (2.58; 5.59)	1.42 (0.91 2.21)	1.77 (0.65; 5.23)

NMA: network meta-analysis, NMR: network meta-regression, the π_t/π_c refers to the odds ratio between the test treatment and the active control, the π_p/π_t refers to the odds ratio between the placebo and the test treatment odds ratio >1 indicates worse outcome (high pain intensity)

E. 3 R Codes for the network meta-regression and network meta-analysis

```
#Pairwise meta-regression Atorvastatin
library (foreign)
library (gemtc)
library (rjags)
library (coda)
library(jagsUI)
library(igraph)
# Create a new network by specifying all information.
treatments <- read.table(textConnection(  '
  id  description
  Atorvastatin   "active-control"
  placebo   "placebo"    '
), header=TRUE)
data <- read.table(textConnection(
  ' study      treatment         mean     std.dev  sampleSize
  McInnes     Atorvastatin       23.3     12       50
  McInnes     placebo           -2.2     12       47
  Loughrey    Atorvastatin       23.2     12       24
  Loughrey    placebo            4.9      12       26
  Hernandez   Atorvastatin       26.2     12       21
  Hernandez   placebo            0        12       19
  Koh         Atorvastatin       27.7     12       42
  Koh         placebo            5        12       44
  Monteiro    Atorvastatin       25.65    12       30
  Monteiro    placebo           -2.6      12       30
  Singh       Atorvastatin       22.2     12       23
  Singh       placebo            2.5      12       24
  AVALON      Atorvastatin       24.4     8.3      193
  AVALON      placebo            0.9      9.1      229
  Cubeddu     Atorvastatin       24.1     12       25
  Cubeddu     placebo            4.95     12       24
  COMETS      Atorvastatin       28.1     10       155
  COMETS      placebo            0.7      9.7      78
  Lins        Atorvastatin       25       10       23
  Lins        placebo            5        8        19
  Sposito     Atorvastatin       27.9     12       17
  Sposito     placebo           -2        12       15
  Davidson    Atorvastatin       25       11.3     127
  Davidson    placebo            0        10.3     132
  Raison      Atorvastatin       27.7     9.8      12
  Raison      placebo           -1.6      10.4     11
  Sardo       Atorvastatin       27       12       20
  Sardo       placebo           -2.8      12       20
  Tan         Atorvastatin       32.9     12       39
  Tan         placebo            1.3      12       41
  Hunninghake Atorvastatin       27       12       18
  Hunninghake placebo           -4        12       19
  Muscari     Atorvastatin       26.25    12       27
  Muscari     placebo            2.7      12       30
  Olsson      Atorvastatin       32       7.6      12
  Olsson      placebo            2.2      12       29
  Oranje      Atorvastatin       32.2     12       9
  Oranje      placebo            2.25     12       10
  Tanaka      Atorvastatin       29.4     12       18
  Tanaka      placebo            2.9      12       18
```

```
  Wang       Atorvastatin       31.1    12       26
  Wang        placebo          0       12       28
  Schrott     Atorvastatin      29       11.6     11
  Schrott    placebo          -2       8.4      9
  J-CLAS      Atorvastatin      27.4      12.2     27
  J-CLAS     placebo          0.7       10.7     27
  Nawrocki    Atorvastatin      30.3     8        11
  Nawrocki    placebo          -4.8     8        12
  ' ), header=TRUE)
year <- read.table(textConnection(  '
  study         year
  McInnes       2014
  Loughrey       2013
  Hernandez      2011
  Koh           2010
  Monteiro       2008
  Singh         2008
  AVALON         2006
  Cubeddu        2006
  COMETS         2005
  Lins          2004
  Sposito        2003
  Davidson       2002
  Raison         2002
  Sardo         2002
  Tan           2002
  Hunninghake    2001
  Muscari        2001
  Olsson         2001
  Oranje         2001
  Tanaka         2001
  Wang           2001
  Schrott        1998
  J-CLAS         1997
  Nawrocki       1995    '), header=TRUE)
network <- mtc.network(data, description="network", treatments=treatments, studies = year)
# sd ~ half-Normal(mean=0, sd=0.25)
# network meta-regression
model <- mtc.model(network)
regressor <- list(coefficient= 'shared',
                  variable='year',
                  control= 'Atorvastatin')
model <- mtc.model(network,
                   type="regression",
                   regressor=regressor,
                   hy.prior=hy.prior,linearModel="fixed")
result <- mtc.run(model)
summary (result)
plot (result)
forest(relative.effect(result, 'placebo', covariate = 2020))
forest(relative.effect(result, 'Atorvastatin', covariate = 2020), digit=3)
summary(relative.effect(result, 'placebo', covariate = 2020))
```

```
# Network meta-regression and network meta-analysis for the mean differences (Atorvastatin)
library (foreign)
library (gemtc)
library (rjags)
library (coda)
library(jagsUI)
library(igraph)
```

```
# Create a new network by specifying all information.
treatments <- read.table(textConnection(  '
  id  description
 Atorvastatin   "active-control"  placebo   "placebo"
              test   "test treatment"  '
), header=TRUE)
data <- read.table(textConnection(  '
              study      treatment         mean     std.dev  sampleSize
              McInnes     Atorvastatin       23.3     12       50
              McInnes     placebo          -2.2     12       47
              Loughrey    Atorvastatin       23.2     12       24
              Loughrey    placebo           4.9     12       26
              Hernandez   Atorvastatin       26.2     12       21
              Hernandez   placebo           0       12       19
              Koh         Atorvastatin       27.7     12       42
              Koh         placebo           5       12       44
              Monteiro    Atorvastatin       25.65    12       30
              Monteiro    placebo          -2.6     12       30
              Singh       Atorvastatin       22.2     12       23
              Singh       placebo           2.5      12       24
              AVALON      Atorvastatin       24.4     8.3      193
              AVALON      placebo           0.9      9.1      229
              Cubeddu     Atorvastatin       24.1     12       25
              Cubeddu     placebo           4.95     12       24
              COMETS      Atorvastatin       28.1     10       155
              COMETS      placebo           0.7      9.7      78
              Lins        Atorvastatin       25       10       23
              Lins        placebo           5        8        19
              Sposito     Atorvastatin       27.9     12       17
              Sposito     placebo          -2       12       15
              Davidson    Atorvastatin       25       11.3     127
              Davidson     placebo          0        10.3     132
              Raison      Atorvastatin       27.7     9.8      12
              Raison      placebo          -1.6     10.4     11
              Sardo       Atorvastatin       27       12       20
              Sardo        placebo          -2.8     12       20
              Tan         Atorvastatin       32.9     12       39
              Tan         placebo           1.3      12       41
              Hunninghake Atorvastatin       27       12       18
              Hunninghake placebo           -4       12       19
              Muscari     Atorvastatin       26.25    12       27
              Muscari     placebo           2.7      12       30
              Olsson      Atorvastatin       32       7.6      12
              Olsson      placebo           2.2      12       29
              Oranje      Atorvastatin       32.2     12       9
              Oranje      placebo           2.25     12       10
              Tanaka      Atorvastatin       29.4     12       18
              Tanaka      placebo           2.9      12       18
              Wang        Atorvastatin       31.1     12       26
              Wang         placebo          0        12       28
              Schrott     Atorvastatin       29       11.6     11
              Schrott     placebo          -2       8.4      9
              J-CLAS      Atorvastatin       27.4     12.2     27
              J-CLAS      placebo           0.7      10.7     27
              Nawrocki    Atorvastatin       30.3     8        11
              Nawrocki    placebo          -4.8     8        12
              NI2020      Atorvastatin       18.5     12       22
              NI2020       test             7        12       22'
            ), header=TRUE)
            year <- read.table(textConnection(  '
             study          year
```

```
 McInnes        2014
Loughrey        2013
 Hernandez      2011
 Koh           2010
 Monteiro       2008
 Singh         2008
 AVALON         2006
 Cubeddu        2006
 COMETS         2005
 Lins          2004
 Sposito        2003
 Davidson       2002
 Raison         2002
 Sardo          2002
 Tan           2002
 Hunninghake    2001
 Muscari        2001
 Olsson         2001
 Oranje         2001
 Tanaka         2001
 Wang           2001
 Schrott        1998
 J-CLAS         1997
 Nawrocki       1995
 NI2020         2020'
), header=TRUE)
network <- mtc.network(data, description="network", treatments=treatments, studies = year)
# sd ~ half-Normal(mean=0, sd=0.25)
# network meta-regression
model <- mtc.model(network)
regressor <- list(coefficient= 'shared',
              variable='year',
              control= 'Atorvastatin')
model <- mtc.model(network,
               type="regression",
               regressor=regressor,
               hy.prior=hy.prior,linearModel="fixed")
result <- mtc.run(model)
forest(relative.effect(result, 'placebo', covariate = 2020))
forest(relative.effect(result, 'Atorvastatin', covariate = 2020), digit=3)
## network meta-analysis
model1<-mtc.model(network, hy.prior=hy.prior,linearModel="fixed")
result1<- mtc.run(model1)
forest(relative.effect(result1, 'placebo'))
summary(relative.effect(result, 'test', covariate = 2020))
summary(relative.effect(result1, 'placebo'))
```

```
#pairwise meta-regression for the lidocaine example
              # Network meta-regression for binary data
              library (foreign)
              library (gemtc)
              library (rjags)
              library (coda)
              library(jagsUI)
              library(igraph)
              # Create a new network by specifying all information.
              treatments <- read.table(textConnection('
              id  description
              Lidocaine   "active-control"
              placebo   "placebo"'
```

```
), header=TRUE)
data <- read.table(textConnection(  '
study              treatment      responders  sampleSize
Kim                Lidocaine       17         40
Kim                placebo         17         20
Tariq              Lidocaine       0          100
Tariq              placebo         6          100
Sethi              Lidocaine       7          100
Sethi              placebo         40         100
Krobbuaban         Lidocaine       1          133
Krobbuaban         placebo         1          135
Bachmann-Mennenga  Lidocaine       10         112
Bachmann-Mennenga  placebo         33         112
Bachmann-Mennenga1 Lidocaine       15         111
Bachmann-Mennenga1 placebo         38         110
Kwak               Lidocaine       0          46
Kwak               placebo         5          45
Tariq1             Lidocaine       1          50
Tariq1             placebo         2          50
Tariq2             Lidocaine       3          50
Tariq2             placebo         9          50
Minogue            Lidocaine       7          42
Minogue            placebo         26         39
Yew                Lidocaine       0          25
Yew                placebo         1          25
Harmon             Lidocaine       2          45
Harmon             placebo         15         45
Ho                 Lidocaine       36         120
Ho                 placebo         22         30
Parmar             Lidocaine       16         77
Parmar             placebo         19         38
OHara              Lidocaine       8          31
OHara              placebo         19         31
OHara1             Lidocaine       10         31
OHara1             placebo         14         31
Gajraj             Lidocaine       6          54
Gajraj             placebo         7          13
McDonald           Lidocaine       1          33
McDonald           placebo         11         31
King               Lidocaine       46         267
King               placebo         51         98
Barker             Lidocaine       5          27
Barker             placebo         16         28
Gehan              Lidocaine       12         157
Gehan              placebo         6          38
Newcombe           Lidocaine       6          47
Newcombe           placebo         23         46
Helbo-Hansen       Lidocaine       2          40
Helbo-Hansen       placebo         13         40'
), header=TRUE)
year <- read.table(textConnection(  '
 study             year
 Kim               2010
Tariq              2010
Sethi              2009
Krobbuaban         2008
Bachmann-Mennenga  2007
Bachmann-Mennenga1 2007
Kwak               2007
Tariq1             2006
Tariq2             2006
```

```
Minogue            2005
Yew               2005
Harmon             2003
Ho                1999
Parmar             1998
OHara              1997
OHara1             1997
Gajraj            1996
McDonald           1996
King              1992
Barker             1991
Gehan              1991
Newcombe           1990
Helbo-Hansen       1988    '
), header=TRUE)
network <- mtc.network(data, description="network", treatments=treatments, studies = year)
model <- mtc.model(network)
# network meta-regression
regressor <- list(coefficient= 'shared',
             variable='year',
             control= 'Lidocaine')
# sd ~ half-Normal(mean=0, sd=0.32)
hy.prior <- mtc.hy.prior(type="std.dev", distr="dhnorm", 0, 9.77)
model <- mtc.model(network,
              type="regression",
              regressor=regressor,
              hy.prior=hy.prior,
              linearModel="fixed")
result <- mtc.run(model)
summary (result)
forest(relative.effect(result, 'test', covariate=2020),digits=3)
forest(relative.effect(result, 'Lidocaine', covariate=2020), digits=3)
summary (relative.effect(result, 'test', covariate=2020),digits=3)
# network meta-analysis
model1<-mtc.model(network, hy.prior=hy.prior,linearModel="fixed")
result1<- mtc.run(model1)
summary (result1)
forest(relative.effect(result1, 'test'), digits=3)
```

```
# Network meta-regression and network meta-analysis for binary data
library (foreign)
library (gemtc)
library (rjags)
library (coda)
library(jagsUI)
library(igraph)
# Create a new network by specifying all information.
treatments <- read.table(textConnection('
id  descriptionLidocaine   "active-control"
test     "test tratment"
placebo   "placebo"'
), header=TRUE)
data <- read.table(textConnection(  '
study              treatment      responders  sampleSize
Kim               Lidocaine       17          40
Kim               placebo         17          20
Tariq             Lidocaine       0           100
Tariq             placebo         6           100
Sethi             Lidocaine       7           100
Sethi             placebo         40          100
```

```
Krobbuaban          Lidocaine       1       133
Krobbuaban          placebo         1       135
Bachmann-Mennenga   Lidocaine       10      112
Bachmann-Mennenga   placebo         33      112
Bachmann-Mennenga1  Lidocaine       15      111
Bachmann-Mennenga1  placebo         38      110
Kwak                Lidocaine       0       46
Kwak                placebo         5       45
Tariq1              Lidocaine       1       50
Tariq1              placebo         2       50
Tariq2              Lidocaine       3       50
Tariq2              placebo         9       50
Minogue             Lidocaine       7       42
Minogue             placebo         26      39
Yew                 Lidocaine       0       25
Yew                 placebo         1       25
Harmon              Lidocaine       2       45
Harmon              placebo         15      45
Ho                  Lidocaine       36      120
Ho                  placebo         22      30
Parmar              Lidocaine       16      77
Parmar              placebo         19      38
OHara               Lidocaine       8       31
OHara               placebo         19      31
OHara1              Lidocaine       10      31
OHara1              placebo         14      31
Gajraj              Lidocaine       6       54
Gajraj              placebo         7       13
McDonald            Lidocaine       1       33
McDonald            placebo         11      31
King                Lidocaine       46      267
King                placebo         51      98
Barker              Lidocaine       5       27
Barker              placebo         16      28
Gehan               Lidocaine       12      157
Gehan               placebo         6       38
Newcombe            Lidocaine       6       47
Newcombe            placebo         23      46
Helbo-Hansen        Lidocaine       2       40
Helbo-Hansen        placebo         13      40
NI2020              Lidocaine       53      316
NI2020              test            53      316 '
), header=TRUE)
year <- read.table(textConnection( '
  study             year
  Kim               2010
Tariq               2010
Sethi               2009
Krobbuaban          2008
Bachmann-Mennenga   2007
Bachmann-Mennenga1  2007
Kwak                2007
Tariq1              2006
Tariq2              2006
Minogue             2005
Yew                 2005
Harmon              2003
Ho                  1999
Parmar              1998
OHara               1997
OHara1              1997
```

```
Gajraj              1996
McDonald            1996
King               1992
Barker              1991
Gehan               1991
Newcombe            1990
Helbo-Hansen        1988
 NI2020             2020'
), header=TRUE)
network <- mtc.network(data, description="network", treatments=treatments, studies = year)
model <- mtc.model(network)
# network meta-regression
regressor <- list(coefficient= 'shared',
                  variable='year',
                  control= 'Lidocaine')
# sd ~ half-Normal(mean=0, sd=0.32)
hy.prior <- mtc.hy.prior(type="std.dev", distr="dhnorm", 0, 9.77)
model <- mtc.model(network,
                   type="regression",
                   regressor=regressor,
                   hy.prior=hy.prior,
                   linearModel="fixed")
result <- mtc.run(model)
summary (result)
forest(relative.effect(result, 'test', covariate=2020),digits=3)
forest(relative.effect(result, 'Lidocaine', covariate=2020), digits=3)
summary (relative.effect(result, 'test', covariate=2020),digits=3)
# network meta-analysis
model1<-mtc.model(network, hy.prior=hy.prior,linearModel="fixed")
result1<- mtc.run(model1)
forest(relative.effect(result1, 'test'), digits=3)
```