SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY

DEPARTMENT OF STATISTICS POSTGRADUATE PROGRAM

Evidence synthesis: from meta-analysis to network meta-analysis with an application in patients with COPD

Adriana Nikolak Thano

A THESIS

Submitted to the Department of Statistics

of the Athens University of Economics and Business

in partial fulfilment of the requirements for

the degree of Master of Science in Statistics

Athens

August 2017

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΜΕΤΑΠΤΥΧΙΑΚΟ

Σύνθεση δεδομένων: από την μέτα-ανάλυση στην μέτα-ανάλυση δικτύων με εφαρμογή σε ασθενείς με ΧΠΑ

Αντριάνα Νικολάκ Θάνο

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα

Αυγούστος 2017

ACKNOWLEDGEMENTS

I would first like to thank my thesis's supervisors, Prof. Ioannis Ntzoufras and Prof. Vasilios Vasdekis of the MSc in Statistics at the Athens University of Economics and Business for their support and guidance towards the direction to work independently to finish this dissertation. This thesis was part of an internship, offered by Mapi Group, working as a research associate and located in Utrecht, the Netherlands, thus I would like to thank Dr. Andreas Karabis for this opportunity and also for the conceptual idea of this thesis' application. I would like to express my very profound gratitude to all my Mapi colleagues for their help and continuous encouragement, and more especially Katerina Papadimitropoulou, Katya Sion and Chrystostomos Kalyvas. Finally, I am thankful to my family, friends and Ioannis for their support, love and patience.

Thank you.

Adriana Thano

ABSTRACT

Adriana Thano

Evidence synthesis: from meta-analysis to network meta-analysis with an application in patients with COPD

August 2017

Evidence synthesis methodologies become essential as more and more analyses are available for a specific research question. This dissertation has been focused on the evidence synthesis methods in helthcare, using randomized control trials (RCT) as a source of evidence. The first method described is the meta-analysis, an overall analysis to pool the treatment effect of two specific treatments being compared directly. The meta-analysis technique has two effect models, the fixed and the random effects, which their differentiation relies on a fundamental assumption over the uncertainty sources; the latter assumes between-study variance in additional to the within-study variance, which is the only source of variability in the fixed effect model. Furthermore, the indirect treatment comparisons (ITC) overcomes the limitation of the meta-analysis, making feasible the comparison of treatments without the requirement of them to be directly compared in an RCT. The ITC uses a common comparator, a treatment which has been compared with the other two treatments of interest, if both indirect and direct evidence are available a pooled estimation can be performed. The ITC and pooled effect methodologies can be considered as mixed treatment comparisons (MTC), however, since they are based on trivial mathematical equations they can not exploit the geometry of the network made by the treatments connected. The last and most important evidence synthesis tool that has been presented is the network meta-analysis, the extension of meta-analysis. A network of multiple treatments, connected directly or indirectly by multiple studies is analyzed simultaneously by fixed or random effects. The dissertation is organized in two parts; the theory of these methods, accompanied with examples in the Bayesian and frequentist prospective for continuous outcomes, and an extensive application in network meta-analysis in patients with COPD, using a publication performed by Mapi [1]. The main scope of this thesis has been to present both in theory and application all the main steps of evidence synthesis and compare the estimations among different approaches and models. As a conclusion, the Bayesian and frequentist approaches deemed to result in approximately same estimations, with the random effects estimations in both cases providing more uncertainty around them.

ΠΕΡΙΛΗΨΗ

Αντριάνα Θάνο

Σύνθεση δεδομένων: από την μέτα-ανάλυση στην μέτα-ανάλυση δικτύων με εφαρμογή σε ασθενείς με ΧΠΑ

Αύγουστος 2017

Οι μεθοδολογίες σύνθεσης δεδομένων αναπτύσονται ολοένα και περισσότερο όσο μεγαλώνει το πλήθος δεδομένων προς καταγραφή και ανάλυση. Ειδικότερα, οι μεθοδολογίες στην σύνθεση δεδομένων στην ιατροφαρμακευτική είναι πρωταρχικής σημασίας· συνδυάζοντας κλινικές δοκιμές ώστε να αποτελέσουν μια ολική εκτίμηση της θεραπείας μέσω της μέτα-ανάλυσης μεθόδου, επίσης επιτρέποντας την σύγκριση διαφόρων φαρμάκων ή θεραπειών που δεν έχουν συγκριθεί άμεσα σε κλινική δοκιμή. Η διπλωματική εργασία αυτή παρουσιάζει τις μεθολογίες της μέτα-ανάλυσης, έμμεσης συγκρισης θεραπειών, απο κοινού (pooled) εκτίμησης και μετα-ανάλυση δικτύων για συνεχή δεδομένα, και παρατίθεται ένα εκτενές παράδειγμα στην μετα-ανάλυση δικτύων σε ασθενείς με χρόνια πνευμονική ανεπάρκεια. Παρουσιάζοντας δύο σχολές στιατιστικής, Μπεύζιανή και κάσσική για τις παραπάνω μεθόδους καθώς και παραδείγματα. Κάθε ένα απο τα οποία έχει αναλυθεί σε μοντέλο σταθερού και τυχαίων επιδράσεων, που διαφέροποιούνται ανάλογως των πηγών αβεβαιότητας που επιτρέπουν. Η μετα-ανάλυση δικτύων, η κορυφή στην πυραμίδα της σύνθεσης δεδομένων, παρουσιάζεται λεπτομερώς για κάθε στατιστική σχολή και μοντέλο επιδράσων θεωριτικά αλλά και έφαρμοσμένα χρησιμοποιώντας μια δημοσιευση της οποίας την ανάλυση είχε υλοποιήσει η Mapi [1]. Ο κύριος σκοπός αυτής της εργασίας ήταν να παρουσιάσει θεωριτικά και εφαρμοσμένα όλα τα βήματα της σύνθεσης δεδομένων και να συγκρίνει τις εκτιμήσεις των δύο διαφορετικών σχολών και μοντέλων παραγόντων. Συμπερασματικά, οι δύο σχολές, Μπευζιανή και κλασσική οδηγούν σε παρόμοια αποτελέσματα, με τις εκτιμήσεις του μοντέλου των τυχαίων παραγόντων να έχουν μεγαλύτερη αβαιβεώτητα.

Contents

List of Tables

List of Figures

0.1 Abbreviations

CI	Confidence Intervals
CrI	Credible Intervals
COPD	Chronic Obstructive Pulmonary Disease
df	Degree of Freedom
DL	DerSimonian and Lierd
FE	Fixed Effect
IRD	Incidence Rate Ratio
IRR	Incidence Rate ratio
MLE	Maximum Likelihood Estimation
NMA	Network Meta-Analysis
RCT	Randmized Control Trial
RE	Random Effects
RD	Relative Difference
RR	Relative Risk
SE	Standard Error
SD	Standard Deviation
SLR	Systematic Literature Research
SMD	Standardize Mean Difference
OR	Odds Ratio
PP	Patient per Protocol

Chapter 1

Introduction

Over the last decades the number of studies conducted in medical and health research has increased dramatically, providing an abundance of evidence on several topics. This large amount of information has enhanced the understanding in these areas albeit creating an issue when published studies provide contradictory evidence on the same topic. Policy makers, healthcare professionals and researchers need to provide reasoning in the direction of evidence. Therefore, the necessity to synthesize all available evidence on a specific research topic, to provide a summary conclusion, is proving to be essential. Evidence synthesis is the approach developed to achieve these summary conclusions. Such methods include meta-analysis, indirect and mixed treatment comparisons and network meta-analysis as main methodologies [2]. These methods aim to synthesize individual study data to generate global summary knowledge about a specific topic, therefore increasing its credibility. Evidence syntheses methods have been applied in a variety of areas, such as: sociology, education, crime and justice, food safety, environmental and healthcare assessment. This dissertation focuses on the healthcare decision making field, specifically in medicine and clinical treatments. Evidence synthesis methods in healthcare that are assessed in this analysis are: meta-analysis indirect and mixed treatment comparison, and network meta-analysis.

The scope of meta-analysis is to combine existing publicly available evidence from studies with a common research question, in such a way to resemble one comprehensive result. In the medical aspect, the interest focuses on comparing treatment effects. All studies conducted measuring the effect size of the treatments of interest should be pooled together in order to provide more credible estimations. Meta-analysis compares the effect of two specific treatments, all studies that are therefore used in the meta-analysis ought to have examined exactly these two treatments. A matter of critical importance is to identify which studies should be considered as appropriate to be combined. To drive this identification of studies, the process of a systematic literature review (SLR) is vetted. The most rational recommendation is to include only randomized control trials (RCT) for appropriate evidence synthesis. In an ideal scenario all combined RCTs should share approximately identical characteristics and have been conducted under the same circumstances to measure

the same treatment effect. This scenario is rather unrealistic and given the presence of heterogeneity across studies there are two different models, one ignoring the heterogeneity and one acknowledging it, namely the fixed effect (FE) and the random effects (RE), respectively. More details on the methods and models of a meta-analysis are discussed in Chapter 2.

Although meta-analysis has been proved useful in many situations, it has its limitations as it can be used solely for pooling of head-to-head comparisons between two specific treatments. As there can be numerous different medications for a disease this restriction advocates the extension of meta-analysis in order to be able to compare more than two treatments, which have not necessarily been directly compared in the published studies. The step further is made by the indirect treatment comparison (ITC), a method in which two treatments which have not been directly compared in a RCT, can be synthesized if they share a common comparator. The pooling of direct and indirect evidence results in a pooled effect, which can be visualized in the constructed network of interlinked studies comparing the treatments both directly and indirectly. When there is a "closed loop" in the network of connected studies, the method is called network meta-analysis. To provide an example of a constructed network note that commonly a new medication is tested only against placebo, when different available treatments are compared against placebo, a network meta-analysis is possible providing the possibility to estimate the difference between the active treatments via the common comparator, placebo. Similarly to meta-analysis, network meta-analysis can be performed under a fixed effect and a random effects approach, which are extensively presented in Chapter 3.

This dissertation provides an introduction to evidence synthesis methods: meta-analysis network meta-analysis in a frequentist and Bayesian framework. Also ITC and pooled effect estimations are analyzed. To provide a clear understanding, an applied example is described. The application is given from International Journal of COPD, authored by Huisman E, Cockle S, Ismaila A, Karabis A and Punekar Y, entitled "Comparative efficacy of combination bronchodilator therapies in COPD: a network meta-analysis" [1], parts of this article are analysed in Chapter 4.

Chapter 2

Meta-Analysis

2.1 Evidence base

The majority of meta-analyses in healthcare utilizes mainly RCTs which are deemed the gold standard in clinical practice and are placed in the top of the pyramid describing the hierarchy of evidence in evidence-based medicine. A RCT is an experimental study examining medical treatments on two or more groups in which the enrolled patients are randomly assigned in order to eliminate confounding parameters, e.g., patients' age, sex or disease years. These studies are typically double-blinded (i.e. neither patients nor researchers are informed which treatment is given in any of the groups). The randomized patient groups are followed up in time and both the primary and the secondary outcomes of interest are recorded. A rigorous RCT is based on a strict study protocol, which is the first step to begin with specifying in detail the procedure to be followed. Meta-analysis is the combination of RCTs therefore prior to performing the analysis, a clear protocol describing the research question, available data (to be synthesized) and methods should be developed.
The credibility of meta-analysis can only be ensured if all evidence relative to study question has been identified. The SLR, the thorough study identification process that researchers conduct, makes this possible. The publication bias cannot be avoided, thus SLRs attempt to deal with any other type of bias, e.g., outcome reporting bias when studies report on the basis of outcome finding [3]. Low study quality can additionally influence the quality of the meta-analysis therefore justification of including or excluding studies should be explained. Heterogeneity is present in all studies given that they have different study and patient characteristics but as a pooling of them is needed these differentiations should be due to randomness and not any systematic factor.

In the case of aggregate data, the minimum information required from each study in order to conduct a meta-analysis is: the treatment effect and its variability, given in any form, e.g., the standard error (SE), the standard deviation (SD) or the confidence intervals (CI). Occasionally, in many trials only the effects and

SEs are reported; in that case we are not able to do further computations and data manipulations, except of computing SD and CI (Cochrane handbook). A more analytic data presentation is a collection of summary statistics for each treatment group in the trial. The most informative type of data are the individual patient data. This form gives flexibility to the analyst since a variety of methodologies can be applied. A meta-analysis can be conducted in any of these cases or their combination as long as the minimum information is provided.

2.1.1 Binary Outcomes

The form of a binary outcome records the presence or the absence of an event, e.g. if a death or a relapse occurred. Several measures are used to report on this type of outcome; the most commonly used ones are the odds ratio (OR), the relative risk (RR) and the risk difference (RD).
The OR can be estimated by the number of patients with a specific disease (n) divided by the number of patients without the disease (N-n) in a group of N patients. Given that for small to moderate sample sizes the sampling distribution is highly skewed, usually the logarithm of OR is used which asymptotically follows a Normal distribution.

$$ln(Odds) = ln(\frac{n}{N-n})$$
$$var(ln(odds)) = 1/n + 1/(N-n)$$

In the need of comparing two groups, as in RCTs, the OR is commonly used as it is able to present all information needed for such comparison. A typical 2x2 table is presented in Table 2.1, this outcome gives a relative measure of chance of event of interest in the forms of the ratio of the odds of an event in two groups.

$$OR = \frac{ad}{bc} \tag{2.1}$$

Table 2.1: A 2x2 table of an RCT for 2 groups

	Event	No event
Treatment	a	b
Placebo	c	d

The interpretation of the OR depends on the nature of examined outcomes, if is a undesirable outcome (death) then a OR less than 1 indicates that the treatment is better than the placebo and an OR greater than 1 indicates that the treatment is less effective than the placebo. If the outcome was positive, e.g., treatment response, the interpretation is reverse. The variance of ln(OR) is :

$$var(ln(OR)) = 1/a + 1/b + 1/c + 1/d \tag{2.2}$$

2.1.2 Continuous Outcomes

For continuous outcomes the use of mean, median or eventually mode is acceptable, but the mean is considered as the most appropriate in the majority of the studies reporting continuous outcomes. The simplest case of reporting continuous outcomes is the absolute difference between means, which is the difference in sample means of effect sizes in each group. In terms of a continuous outcome the treatment effect is given by:

$$T = \mu_t - \mu_c \tag{2.3}$$

μ_t: mean effect in the treatment group, μ_c: mean effect in the control group

The mean effect, μ representing a response, could be a biochemical index, a percentage of improvement or the number of episodes in a disease, or very often is the change from baseline (CFB) value. Usually the parameter of interest is the difference in CFB between the treatment and control group. In cases that all studies estimate the exact same parameter then they can be combined directly in the original scale or in any transformation considered to be appropriate. If one is interested in the progress of patients in a group, then the difference of an effect at a certain time point from the baseline value, (usually considered to be the starting time of the study) provides a safety and efficacy measure in each group.

$$CFB = \mu_{baseline} - \mu_{time}$$

$\mu_{baseline}$: effect in the baseline in the same group,

μ_{time}: effect in a certain time point

Moreover, in order to compare groups, the CFB_t could be the estimation in treatment group and CFB_c respectively in control group, so the T indicates the difference in CFB (DCFB).

$$T = DCFB = CFB_t - CFB_c$$

Given that all patient effects are in the continuous scale it can be assumed that they are observations from a normal distribution with mean effect equal to the average patients' effect and variance the combination of each group variance. The variance of the treatment effect is given by:

$$Var(T) = \sigma^2(1/n_t + 1/n_c)$$

σ^2: the common variance of the groups,

n_t, n_c : the sample size in treatment and control group respectively

It is rather impossible to have the true variance, so the variance is estimated (s^2)from the data; if there is evidence for equal variances only the estimated variance of one group can be used, and in same cases of aggregate data the variance is provided. Alternatively, a pooled estimated variance when the variances in groups differ provides a good estimation.

$$s = \frac{(n_t - 1){s_t}^2 + (n_c - 1){s_c}^2}{n_t + n_c - 2} \tag{2.4}$$

Standardized Mean Difference

It is common that different studies being incorporated into the meta-analysis use different scales to measure the same outcome. In that case a synthesis is possible with transforming the data in order to have the same scaling. The standardized mean difference (SMD) is the approach to handle these data/scales discrepancies. However, it should be kept in mind that the interpretation can become more difficult when combining two types of data. In order to overcome this problem the measure of size effect T is divided with the standard deviation of the trial or its estimation.

$$SMD = \frac{T}{sd} \tag{2.5}$$

The variance in that case is not as simple as in the previous case, the assumption of normality in data is required.

$$Var(SMD) = \frac{n_t + n_c}{n_t n_c} + \frac{SMD^2}{2(n_t + n_c)} \tag{2.6}$$

In addition if the sample sizes in the groups are large and the variances approximately equal then a different formula is preferred.

$$Var(SMD) = \frac{(n_t + n_c)}{n_t n_c} \tag{2.7}$$

2.2 Heterogeneity

In any estimation the sampling error cannot be avoided, estimated effect sizes differ among studies albeit having the same research question and methodology. In meta-analysis, the synthesized studies are considered homogeneous when their effect sizes differ only due to sampling error, i.e., it can be assumed that systematic differences between them do not exist. Patients and study characteristics can have an impact on the estimated treatment difference effect. The placebo effect reflects the effect of individual study and patients characteristics in a RCT; this study effect influences the placebo and treatment group in the same level, all known and unknown factors contributing to this influence are called prognostic factors. The effect size in the intervention group is a consequence of study effect and treatment effect, which are called effect modifiers, e.g. age, sex and disease severity of patients. By allocating the patients in the groups in a random manner, the prognostic factors are considered to be balanced, thus the effect measured is only the treatments effect as Figure 2.1 depicts. As follows naturally, it is desirable for the effect modifiers not to differ significantly

across studies. When all available evidence is identified from the SLR process, statisticians, clinicians and disease experts should examine the feasibility of the analysis.

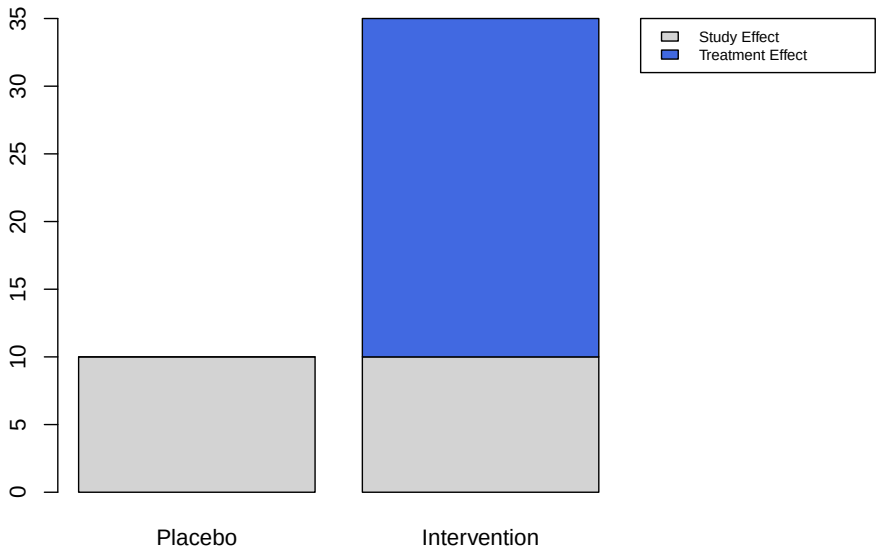


Figure 2.1: The study and treatment effects

Cochrane Q-statistic

The first tool to use in order to examine if a hypothesis is valid is the standard Cochrane Q-statistic test. The null hypothesis is that all k studies share the same true treatment effect, μ.

$$H_o : m_1 = m_2 = \ldots = m_k \qquad \text{k: number of studies}$$
$$H_1 : m_i \neq m_j \qquad i, j = 1, 2 \ldots, k,\ i \neq j$$

The Q statistic is approximately distributed as a χ^2 distribution with k-1 degrees of freedom under the H_0 hypothesis.The power of the test is very low when the number of combined studies is small, implying that the test is not valid in such cases and that heterogeneity may be presented even if the Q statistic is not statistically significant. On the other hand, the null hypothesis can be rejected even if homogeneity exists when the sample sizes in each study is large enough. These reasons suggest to use this tool with caution and in combination with graphical and empirical techniques. A statistically significant result may indicate heterogeneity but a not statistically significant should not be interpreted as a sign of homogeneity. Additional tests should be performed.

I^2 ratio

A widely known measure which uses the Q-statistic value is the I^2. This quantity describes a percentage of variability caused due to heterogeneity and not because of the sampling error in studies. As is mentioned, it is unrealistic to identify identical studies therefore a degree of heterogeneity is always present. The investigation of the degree of heterogeneity is important as it drives the decision on whether the meta-analysis should be conducted. Under these circumstances the critical p-value is rather 0.10 instead of the conventional choice of 0.05. As a conclusion, one may say that it describes the percentage of observed variation in the Q statistic due to heterogeneity rather than sampling error; This quantity does not inherently depend on the number of studies as the Q statistic. In case of the degrees of freedom (df) are greater than Q, then the I^2 ratio is zero.

$$I^2 = \frac{Q-df}{Q}100\%$$

Percentage up to 40% is considered as low evidence of heterogeneity and 75% to 100% indicate highly heterogeneous studies.

Z-score

An other option is a graphical approach, the plot of normalized (z) scores.

$$z_i = \frac{(T_i-\bar{T})}{se(T_i)}$$

As the null hypothesis holds the values of the z-scores should be approximately normally distributed with mean zero and variance equal to one. If a study deviates a lot from this value, it is an indication of differentiation and this study should be examined in more detail for its suitability in the meta-analysis. A typical histogram of z-scores could be helpful to examine the normality.

Forest plot

A widely used graphic to present the meta-analysis results is the forest plot. A forest plot depicts the effect and the CI of each study and also the treatment difference effect estimation. Moreover, it provides information about the heterogeneity of the meta-analysis, by observing whether the CIs overlap across studies. The usual representation lists the studies parallel to the y-axis and each effect size is plotted by its point estimate and the associated CIs are the lateral tips of the point estimate. As it is rational, studies with large CIs are less influential than studies with smaller CIs; the latter are more precise and contribute more in the pooled effect size.

2.3 Frequentist Approach

The field of statistics has three main branches: the Fisherian, the frequentist or classical and the Bayesian. The first one is being gradually abandoned while the rest are gaining ground. The frequentist school is based mainly on Jerzy Neyman and Egon Pearson, and the Bayesian is named after the preach Thomas Bayes. These two schools differ significantly in the way of parameter operation. Under the frequentist approach, the parameters are considered to be fixed values and their estimation is based only on the data given; the critical differentiation comes at this point, as in the Bayesian approach the parameters are random variables and estimations for the parameters are combinations of the observed data and the prior information on these parameters.

Statistical inferences are based on the likelihood of the observed ($L(\theta) or L((y/\theta))$) data (y) given the fixed parameter (θ) being estimated.

$$L(\theta) = f(y/\theta) \sim N(\theta, \sigma^2)$$

The value of θ which maximizes the likelihood is the Maximum Likelihood Estimation (MLE) and thus the estimation of the parameter. In the ordinary case of more available data, the likelihood is the product of the individuals' likelihood assuming that each data is independent. MLE is the only evidence for estimation. In populations distributed by any distribution in the exponential family the MLE calculation is quite simple and predetermined, however in more complex distributions it can be challenging [4].

2.3.1 Fixed Effect Model

The fixed effect model has a fundamental assumption which is the key difference from the random effects model. It assumes that all studies included in the analysis share the same true effect size and they are observations of a common distribution with mean and variance equal to the total treatment difference effect and its variance, respectively. The only source of variability comes from randomness, the within-study error. All study treatment effects are observations from the same normal distribution with mean the true treatment difference effect m and variance σ^2 as the Figure 2.2 illustrates [5].

As every study comes from the same distribution there is only one source of variability, the sampling error. The fixed effect meta-analysis model could be described as follows:

$$T_i = m + e_i \tag{2.8}$$

T: observed treatment effect, m: true treatment effect, e: within-study error

i: 1,2...,k, k: total number of studies

The inverse of variance method is widely used for the estimation of the weights [6]. The sample size could also be used as an naive method, but rather than using the sample size, the inverse variance is preferable as it

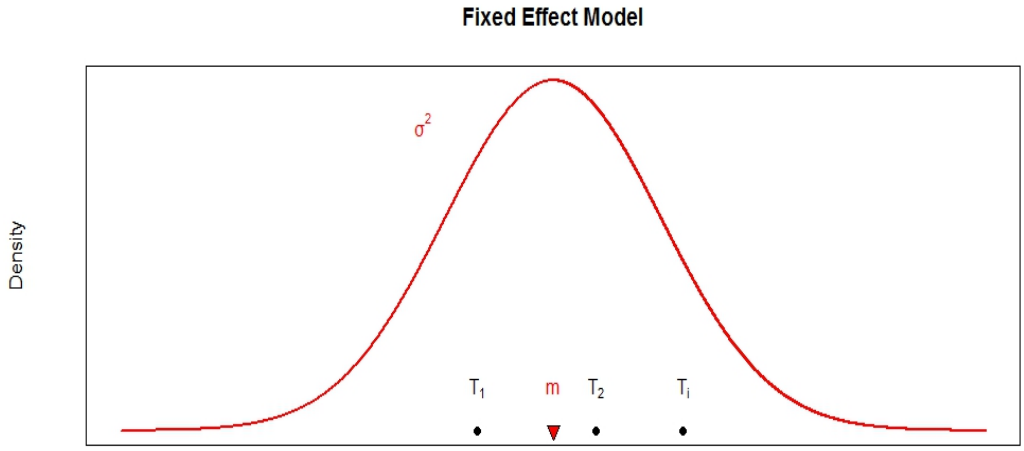


Figure 2.2: Distribution of true effect in fixed effect model. Each effect observation T_i comes from a normal distribution with mean the true treatment effect m and variance σ^2.

is proportional to the sample size and moreover providing information about the dispersion of the population.

$$w_i = 1/v_i \tag{2.9}$$

w: study's weight, v: study's variance

The weighted mean $\bar{T}$, the total treatment effect is the sum of each study effect multiplied by its weight and divided by the sum of the weights, as presented in the equation 2.10.

$$\bar{T} = \frac{\sum_{i=1}^{k} w_i T_i}{\sum_{i=1}^{k} w_i} \tag{2.10}$$

The variance of the total effect is the inverse of the the study weights sum.

$$var(\bar{T}) = \frac{1}{\sum_{i=1}^{k} w_i} \tag{2.11}$$

The CIs are calculated as usual.

$$LowerLimit = \bar{T} - Z_{1-a/2} * SE$$

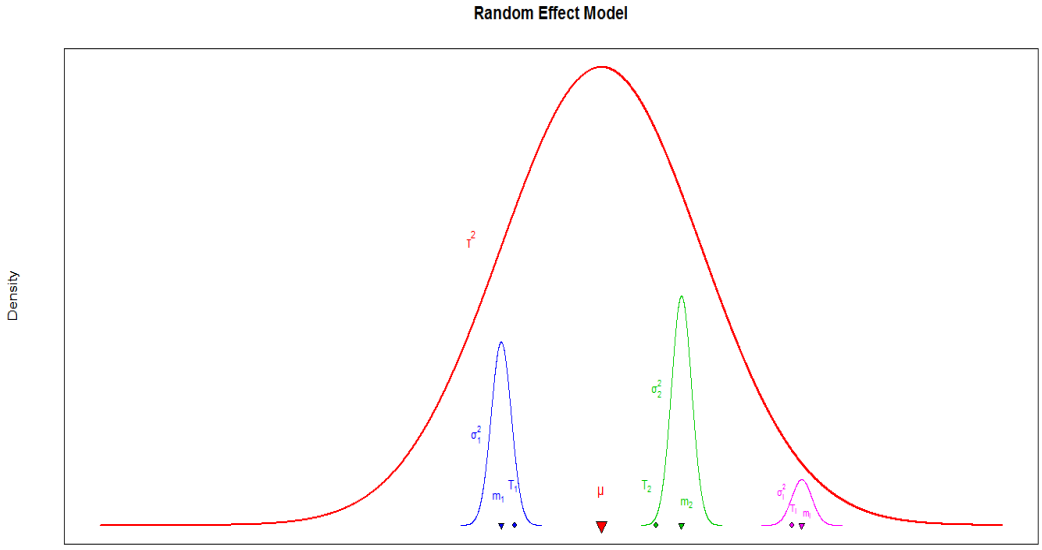
$$UpperLimit = \bar{T} + Z_{1-a/2} * SE$$

Furthermore, the statistical significance of the effect can be examined with Z-test.

$$Z = \frac{\bar{T}}{SE(\bar{T})} \tag{2.12}$$

2.3.2 Random Effects Model

The random effects model is considered to be more realistic compared to the fixed effect, as the assumption of the common true treatment effect in all studies is being abandoned and this is the crucial difference from the fixed effect model. The random effects model is considered in many cases more appropriate as the combined studies share common characteristics, however they could never be identical; many parameters as study and patients characteristics contribute to the differentiation of studies. The random effects model assumes that each study effect is an observation of its own true effect (μ) distribution and all of them are a sample of a greater, overall normal distribution, the total true effect (μ) [5]. This hypothesis provides a second source of variation, the between-study variance (τ^2), determined by the width of the total true effect distribution as Figure 2.3 depicts. Uncertainty for the location of the total effect depends on the magnitude of the between-study variance, the number of studies, and the precision of the individual study estimates [7].



p

Figure 2.3: Distribution of true treatment effect in random effects model. Treatment effect observation T_i is normally distributed over its mean the true individual study effect m_i and variance (within-study) σ^2. They are observations of a overall normal distribution with total true effect μ and variance (between-study) τ^2.

The random effects meta-analysis model could be described as:

$$T_i = m_i + e_i = \mu + \epsilon_i + e_i \tag{2.13}$$

T: observed treatment effect, m: true treatment effect , e: within-study error,

μ: overall true treatment effect, ϵ: between-study error

Equations 2.8 for the fixed and 2.13 for random effects models, respectively represent the general models. The true effect size (m) in equation 2.13, is now distributed about the true effect (μ) for each different study, with a between-studies variance (τ^2) and that actually is the distribution of all true effects. As the within-study variance always exists it is clear that there are two sources of variation, as equation 2.16 present, which are not unrelated, the bigger the within-study variance is the lower the between-study will become. Total variance is desired to be low, so both of them need to have small values.

Between-study variance

The role of between-study variance is of primary importance, this is the reason why many researchers have put an effort to find the best estimator for it. Sixteen estimators of τ^2 are identified [8], the widely used and the basis for others is the DerSimonian and Laird (DL) method [6]. The DL estimator is quite simple to implement, derived from the Cochran's Q-statistic.

$$Q = \sum_{i=1}^{k} w_i (T_i - \bar{T})^2 \tag{2.14}$$

The between-studies variance is the difference of total variance from the expected variance if all studies have the same true effect. The expected variance is the degrees of freedom for the meta-analysis.

$$df = (Numberofstudies) - 1$$

So, the DL estimator of the between-studies variance can be obtained as:

$$\tau^2 = \begin{cases} \frac{Q-df}{U}, & ifQ > df \\ 0, & ifQ \leq df \end{cases} \tag{2.15}$$

The dominator, U, is a scaling factor which ensures that the τ^2 is in the same metric as the within-study variance.

$$U = \sum_{i=1}^{k} w_i - \frac{\sum_{i=1}^{k} {w_i}^2}{\sum_{i=1}^{2} w_i}$$

As it is defined in equation 2.15 when the τ^2 is negative it truncates to zero, and that introduces positive bias into the estimator. Consequently, the DL estimator is positively biased and over-estimates the true value of between-study variance on average. When the number of included RCTs, k, decreases and it is smaller than the df then the random effects model leads to identical results compared to the fixed effect model [9].

There are more estimators for the between-study variance, some of them are based on DL. Among others there are the positive DL method [10], two step estimator with DL [11], non-parametric bootstrap DL

method [10], Hedges and Olkin method [12], maximum likelihood method [7], Rukhin Bayes method [13], fully Bayesian [14], and Hartung and Makambi method [15]. In this analysis only DL will be applied to be in line with the case study described in the publication and presented in Chapter 4.

The variance of a study in the random effects is the sum of two variances:

$$v_i = \tau^2 + \sigma^2 \tag{2.16}$$

As the two different sources of variation are defined, the computational methodology remains the same as equations 2.10, 2.11.

2.4 Bayesian Approach

In the Bayesian approach, any information about a parameter, as history data, previous or pilot survey, or an expert's opinion, can be used as prior information in a probabilistic way either as probability or distribution. Rather than focusing on the likelihood of the data for specific parameters, the Bayesian approach focuses more on drawing inferences about non-specific (non-constant) parameters. Note that any prior must be specified before observing the data. Finally, when the data comes this prior information about the parameters is being updated to a posterior.

As it is obvious this approach is based on the Bayes theorem:

$$P(\theta/\chi) = \frac{P(\chi/\theta)P(\theta)}{P(\chi)} \tag{2.17}$$

θ: parameter χ: data

The likelihood function is denoted by $P(\chi/\theta)$, the prior beliefs are denoted as a probability density function $P(\theta)$. The normalizing constant $P(\chi)$ is quite difficult to be calculated and as it does not contain any parameter can be omitted. The posterior distribution depends only on the prior and the likelihood.

$$P(\theta/\chi) \propto P(\chi/\theta)P(\theta) \tag{2.18}$$

2.4.1 Prior distribution

The choice of prior is based on external evidence, any information regarding the parameters can be used. It is worth mentioning that there is no restriction about it and this is its opponents' main argument. This plethora of choices made the justification of the prior distribution's choice essential. Furthermore, as the amount of provided data increases, the prior belief influences less the posterior and in case that the sample $n- > \infty$ the Bayesian and the frequentist approaches should be quite identical, but if the sample size is small then the prior can affect the result significantly. The selection of the prior distribution can be categorised by

the amount of the existed information about the parameter, and also according to the desired properties.

Conjugate prior

In the case of data derived from any distribution in the exponential family, the choice of conjugate prior is common. All distributions in this family have the right prior resulting to a known posterior distribution from the exponential family as well [16]. As this analysis considers only continuous outcomes only the normal distribution will be described.

Given the likelihood $x/\mu \sim N(\mu, \sigma^2)$, the μ is the parameter of interest, and a normal prior to $\mu \sim N(d, \sigma_0^2)$ leads to a normal posterior.

$$\mu/x \sim N(\frac{\sigma_0^2}{\sigma^2+\sigma_0^2}\mu + \frac{\sigma^2}{\sigma^2+\sigma_0^2}d, \{\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2}\}^{-1}) \tag{2.19}$$

The mean of the posterior is between μ and d; it will be closer to the one with the smallest variance. The ratio $\frac{\sigma_0^2}{\sigma^2+\sigma_0^2}$ could be denoted as k and $\frac{\sigma^2}{\sigma^2+\sigma_0^2}$ as 1-k, and becomes clear that the posterior mean is a weighted average $kx + (1-k)d$ with the weight k determining the exact posterior mean value.

Non-informative/vague prior

When no information is given the use of non-informative (vague) prior is preferred, because of the objectivity that it provides. The posterior distribution takes values only in the range of prior's, thus placing a flat prior distribution over a plausible range is reasonable. It is a common approach to use a quite wider range than what one expects for the likelihood to have. In that way one can avoid being criticised for subjectiveness. Moreover, the Bayesian and the classical approaches should have approximately same results, as the prior is not influencing and the posterior is based more on the likelihood. Characteristic is the example of a normal distribution with extreme variance as N(0, 1000) or even more extreme N(0, 10^6). A specific prior is Jeffreys' prior 2.20 as it is a non-informative one based on Fisher Information and has the property that is invariant in any re-parametrization [17].

$$p(\theta) \sim \sqrt{I(\theta)} \tag{2.20}$$

Highly informative prior

The prior plays a key role in the Bayesian statistics and it is the main differentiation from the frequentist. This differentiation reaches its peak when an expert is confident for his beliefs and uses a highly informative prior. These priors are characterised by small variances, in case that the variance is zero then the posterior has no other chose than being exactly the same number as the prior. In general the more specific the prior distribution is, the more influential will become, and the data will determine less the posterior distribution.

These cases are described by similar posteriors and priors, and the similarity is more intense when there is lack of data. This fact has provoked many concerns whether this is valid or not. It is clear now that a very restrictive prior should be chosen only in specific circumstances and under indisputable evidence. These priors are often reported as highly informative or stubborn priors.

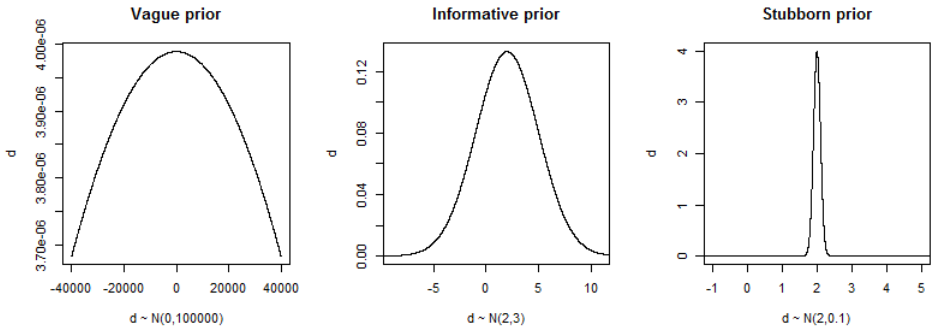


Figure 2.4: Three different normal prior distributions for the parameter d.

Bayes' theorem applies equally to multiparameter models, replacing monoparameter θ with multiparameter θ and the posterior is a joint distribution across all of the θ=$\theta_1, \theta_2 \ldots$.

$$f(\theta_1, \theta_2 \ldots \theta_k | x) \tag{2.21}$$

Inferences for each parameter can be made by averaging over all of the other parameters, the posterior marginal distribution.

$$f(\theta_1|x) = \int_{\theta_2, \theta_3 \ldots \theta_k} f(\theta_1, \theta_2 \ldots \theta_k | x) d\theta_2 \theta_3 \ldots \theta_k \tag{2.22}$$

MCMC simulation

From equation 2.22 one can understand that the algebraically calculations for the posterior distribution could be challenging or even impossible. However, if there was a large amount of observations from this distribution then one could have enough information needed for that specific distribution. Monte Carlo Marcov Chain(MCMC) simulation is the evolution of the Bayesian school, as it provides values from the posterior distribution by drawing values from each parameter repeatedly. Starting with random values for each parameter, the initial values, and then sampling new values based on the previous ones. More specifically if in i iteration there are k parameters $\theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_k^{(i)}$ then the new value is taken from the conditional posterior distribution with fixed the remaining parameters, $f(\theta_1^{(i+1)} | \theta_2^{(i)}, \ldots, \theta_k^{(i)})$ and so on. The algorithm's first steps are far from the target distribution and thus the iterations in the "burn-in" period, the period before sampling from the posterior distribution, are not contribute to the estimation of the posterior mean. It is

important to have convergence to the posterior distribution. In order to ensure that, more than two different initial values, creating more than two different chains, should be used. To implement MCMC simulations used for the Bayesian analysis WinBUGS was used.

Diagnostics

As the sampled values are not taken from the posterior distribution, convergence of the Markov chains need to be assessed. The number of chains needed and samples generated are not certain, a researcher has to explore even more if the model is complicated. The convergence is not given and also if it occurs the time is not known a priori. When the convergence is slow one can let the algorithm run longer or try to accelerate it. A variety of practical procedures have been suggested to check convergence diagnostics [18].

More formally, convergence in a MCMC context is an asymptotic property which implies for a Markov chain that $p_k(\theta)$, the distribution of θ_k, grows to the target distribution $p(\theta|y)$ for $k \to \infty$. In other words it means that, for k large and small ϵ, $d_k \equiv d[p_k(\theta), p(\theta|y)] < \epsilon$, with d(f , g) the distance between two distributions f and g. Theoretical research has focused on establishing conditions under which convergence can be guaranteed. In some simple cases one can provide an expression for k_0 such that $d_k < \epsilon$ for $k > k_0$. For example, Jones and Hobert [19] showed that for the Gaussian case with μ and σ^2 unknown, a k_0 can be specified for the total variation discrepancy measure between two distributions, i.e. $d(f_k, f) = 1/2 \int |f_k(\theta) - f(\theta)| d\theta$. But such theoretical results are hard to establish in most practical cases.

Diagnostics are categorised in checking for stationarity of the chain and verifying the accuracy of the posterior summary measures. After k_0 iterations θ_k is consider to be sampled from the correct posterior distribution, this k is specified in the stationarity step which is equivalent to assessing the burn-in part of the Markov chain. Most convergence diagnostics (graphical and formal) appeal to the stationarity property of a converged chain. In the accuracy step it is verified that the posterior summary measures of interest based on $\theta_k(k = k_0 + 1, ..., n)$ are computed with the desired accuracy.

Trace plot

A simple exploration of the trace plot (given from WinBUGS) gives a first and insightful impression of the characteristics of the Markov chain. Trace plots are produced for each parameter separately and evaluate the chain univariately. In case of stationarity, the trace plot appears as a horizontal strip and the individual moves are hardly discernable. This is the basis of the informal thick pen test [20]. The test involves checking that the trace plot can be covered by a thick pen. Trace plots are the easily understandable, and also clearly show when there are large deviations from stationarity. Moreover, one can infer about the time of convergence as the plot shows how fast the chain explores the posterior.

Autocorrelation plot

MCMC chains are correlated due to Marcov chain that generates the samples, measured by the autocorrelation statistics that show the correlation between sampled values for specified number of iteration (lags). The autocorrelation should drop off with increasing lag, and if it does not do so it indicates slow convergence. High autocorrelations within chains indicate slow mixing and slow convergence. Reparametrizations might help. It might be necessary to increase the thinning interval to achieve a less highly correlated sample. Autocorrelation plots for the example discussed in the section **??** are given in Figures 2.8 and 2.11.

Gelman–Rubin interval diagnostic

Gelman and Rubin's (1992) approach in monitoring convergence is based on detecting when the Markov chains have forgotten their starting points, by comparing several sequences drawn from different starting points and checking that they are indistinguishable [21]. There are many ways to compare parallel sequences, the most obvious approach being to look at overlaid traceplots and see if the two sequences can be distinguished.

A more quantitative approach to answer the question "Are the sequences much farther apart than we could expect, based on their internal variability?" is based on the analysis of variance: Approximate convergence is diagnosed when the variance between the different sequences is no larger than the variance within each individual sequence. Assume we have m parallel simulations each of length n of the variable X. The values are denoted by x_{ij} , i = 1, . . . , m, j = 1 . . . , n. The between-sequence variance B and the within-sequence variance W is computed:

$$W = 1/m \sum_{i=1}^{m} s_i^2 \text{ and } s_i^2 = 1/n \sum_{k=1}^{n} (x_{ik} - \bar{x}_{i.})^2 \text{ and } B = \frac{n}{m-1} \sum_{i=1}^{m} (\bar{x}_{i.} - \bar{x}_{..})^2$$

From the two variance components, two estimates of the variance of X in the target distribution are constructed: First

$$\hat{V} \equiv \hat{var}(X) = \frac{n-1}{n} W + \frac{1}{n} B$$

is an unbiased estimate of the variance when there is stationarity (in which case all $\bar{x}_{i.}$ are unbiased estimates of the true posterior mean). However, when the m chains are not mixing well $\hat{V}$ overestimates $var(X)$ and W often underestimates this variance as long as the chains have not yet explored the whole target distribution. When $n \to \infty$ both $\hat{V}$ and W approach $var(X)$ but from opposite directions. One can now monitor the convergence of the MC by estimating the factor by which the conservative estimate of the distribution of X might be reduced: that is, the ratio between the estimated upper and lower bounds for the standard deviation of X, which is called estimated potential scale reduction or shrink factor:

$$\sqrt{\hat{R}} = \sqrt{\frac{\hat{V}}{W}}$$

As the simulation converges, the shrink factor declines to 1, meaning that the parallel Markov chains are essentially overlapping. If the shrink factor is high, then one should proceed with further simulations.

The Gelman and Rubin diagnostics that can be calculated are the 50% and 97.5% quantiles of the sampling distribution for the shrink factor. These quantiles are estimated from the second half of each chain only. Plots to illustrate these diagnostics are given in the Figures 2.8 and 2.11.

There are more tests to be mentioned as running mean plot, Q–Q plot, and cross-correlation plot for graphical methods and also formal tests as Geweke, Heidelberger–Welch, Raftery–Lewis, and Brooks–Gelman–Rubin diagnostic based on Gelman–Rubin diagnostic [18]. The analyses all diagnostics described is not the scope of this dissertation.

2.4.2 Fixed Effect Model

The idea of meta-analysis remains the same in each approach the only difference is the estimation of parameters. The Bayesian methodology does not trust only the data but uses them to update the prior beliefs about the parameters to posterior [22]

The fixed effect model assumes effects to be random observations of the true effect, so there is only one source of variation as is analysed in section 2.3.1 .

$$Y_i \sim N(d, \sigma_i/n_i) \tag{2.23}$$

i: 1,2...,k, k: total number of studies

The observed effect size, Y_i, of the i^{th} out of k studies, comes from a normal distribution with mean d, the true treatment effect size, which is the parameter of interest and is being estimated by the model. The variance of the distribution is the within-study variance σ_i, and as in the most of cases are unknown they being replaced by the estimated within-study variance, s_i.

Depending on the prior information about the parameter d the distribution could have a variety of choices according to the outcome. So, in the case that the outcome is continuous a normal distribution may be a choice with mean and variance relative to the prior knowledge about it, generally flat priors are preferred when objectivity is required.

$$d \sim N(0, 10^6) \tag{2.24}$$

2.4.3 Random Effects Model

The fixed effect model is considered to be unrealistic for the most of meta-analysis, the assumption of one common true treatment effect can be very restrictive. The patient and study characteristics vary across

studies and a more flexible model is preferable. The alternative model is the random effects, which allows two sources of variation, assuming that each individual study effect is an observation of its own true treatments effect, discussed in detail in section 2.3.2.

The basic assumption of the random effects can be described by the individual effect of the i_th study, Y, as an observation from a normal distribution with true mean $delta_i$, and variance its own within-study variance σ_i. The key point is that all true treatment effects are normally distributed over a total true effect d with between-study variance τ^2, as equation 2.25 present.

$$Y_i \sim N(\delta_i, \sigma_i/n_i) \;\; i = 1, 2, ..., k \tag{2.25}$$

$$delta_i \sim N(d, \tau^2) \;\; i = 1, 2, ..., k \tag{2.26}$$

The estimated parameter are the total true effect size d, and the between-study variance τ^2, thus prior distributions should be specified for these parameters. In section 2.4.1 same decision rules were briefly discussed. In case that one wants to be objective the best choice is a non-informative prior to be used for both parameters, as a vague normal prior to d and a uniform distribution to τ^2. These priors will be used in the application analyses.

$$d \sim N(0, 10^6) \tag{2.27}$$

$$\tau^2 \sim U(0, 10) \tag{2.28}$$

The prior distribution for the between-study variance could have large uncertainty if the number of studies combined is quite small, i.e. <10, under these circumstances alternative prior is recommended. An other widely used prior distribution is the Inverse Gamma (0.001, 0.001).

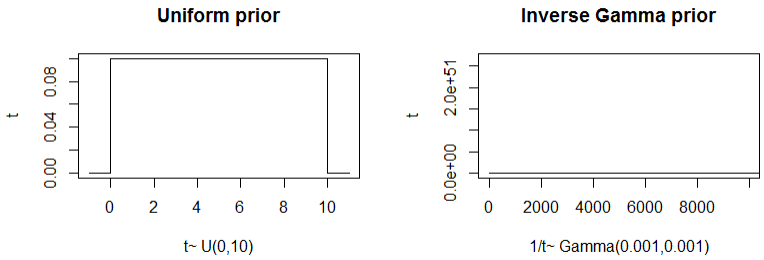


Figure 2.5: Uniform and Inverse Gamma prior distributions for the between-study variance t^2.

Table 2.2: Difference in change from baseline (DCFB) and SE of trough FEV_1 for the 14 fictional studies comparing Tio with the placebo.

Study	DCFB (ml)	SE (ml)
S1	100	15
S2	100	10
S3	140	20
S4	130	18
S5	83	19
S6	110	40
S7	150	14
S8	184	37
S9	102	31
S10	79	17
S11	134	19
S12	127	19
S13	120	15
S14	118	23

2.5 Example

The given meta-analysis example is a fictional example of a basic index in patients with COPD, analyzed under the frequentist and the Bayesian approaches for random and fixed effect models, respectively. The outcome of interest is the forced expiratory volume in 1 second(FEV_1), a marker used to measure lung function, and to monitor chronic obstructive pulmonary disease or other lung diseases over time. The treatments compared are the Tiotropium (Tio) and placebo. The Table **??** illustrates the data for 14 fictional RCTs, given as the difference in change from baseline at the time point of 12 weeks. All of them are statistically significant as obviously non of them cross the axis at zero point, hence Tio is statistically significant efficacious than placebo by means of FEV_1 at 12 weeks in all studies included. Figure 2.6 is a convenient graph to present available evidence comparing Tio to Placebo. The equations below serve to clarify the calculations of the outcome used, the difference in CFB.

$$\mu_{Tio} = \mu_{Tio12} - \mu_{Tio0} \qquad \mu_{Placebo} = \mu_{Placebo12} - \mu_{Placebo0}$$

$$\mu_T = \mu_{Tio} - \mu_{Placebo}$$

All studies provided that Tio had better performance compared with the placebo with minimum value 79 ml and maximum 184 ml. The SE of studies vary from 10 ml to 40 but they do not seem to have big

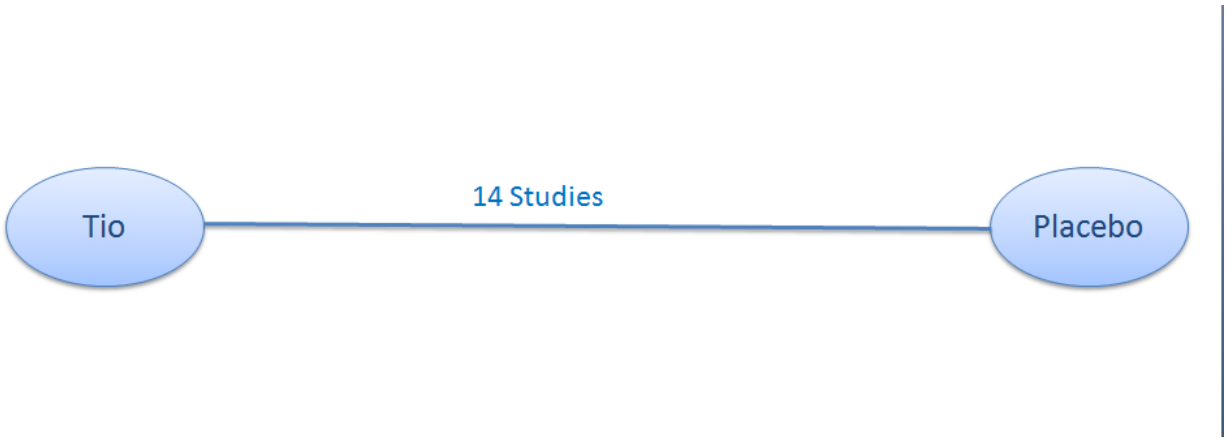


Figure 2.6: The comparison of the effect of Tio with the placebo.

heterogeneity as the CIs overlap. Fourteen studies is deemed to be an adequate sample size. Both fixed and random effects should have approximately the same point estimate size, with random effects providing larger standard error.

The Figure 2.7 illustrates random and fixed effect from both approaches. The Bayesian effect size estimations are quite similar with the random effects having 0.5mL bigger SE. The frequentist random effects estimation provides the biggest SE (6.7mL) and diverse from the other three estimations.
In the Bayesian framework the estimated sizes of total treatment effect are quite similar(RE: 115.8 CrI:103.45 -128.15; FE:115.2 CrI: 105.98 -124.41). The total variance of the random effects is estimated to be 6.3mL (SD: 2.5mL), occurring from the posterior distribution of τ^2 and 5.1mL for the fixed effect which is the same with the frequentist approach, as it is not estimated. A widely used measurement for decision making is the Deviance Information Criterion (DIC), based on likelihood penalizing the amount of parameters being used, is a criterion of model selection. As a model contains a large amount of variables is reasonable to fits these specific data better, in that way becomes more data specific and the predictability is reduced. The DIC values for the two models are approximately equal (RE: 132.7; FE: 135.4) with random effects providing the smallest, hence in a strict follow of the rule it is preferable, but the use of both models is expected in this case. The convergence is of primary importance in the Bayesian perspective, and in order to examine it some diagnostic test were performed, presented in Figure 2.8 and 2.11 for the fixed and random effects respectively. The trace, autocorrelation, and density plots and also the Gelman–Rubin diagram are examined using two chains, and it is clear from that both models had converged and were sampling from the posterior distributions of true effect. The random effects model has a second parameter, the between-variance, for which the same tools were assessed.

In the frequentist framework the estimated sizes of total treatment effect differ more than the Bayesian (RE: 117 CI:103.87-129.93; FE:115.2 CI: 105.98-124.41). The total variance of the random effects is estimated to be 6.7mL (SD: 2.6mL), occurring from the posterior distribution of τ^2 and 5.1mL for the the fixed effect. The Table **??** gives the Q statistic test and the I^2 measure, for both models and their values are quite similar. They both assume heterogeneity, rejecting the hypothesis of Q-statistic test (RE: p.value=0.038; FE: p.value=0.038) and a I^2 around 46%. As it is mentioned the test should not be trusted totally but many features should be taken in consideration, like the previous forest plot which is subjective in general but in that case does not give any evidence to assume heterogeneity as also the I^2.

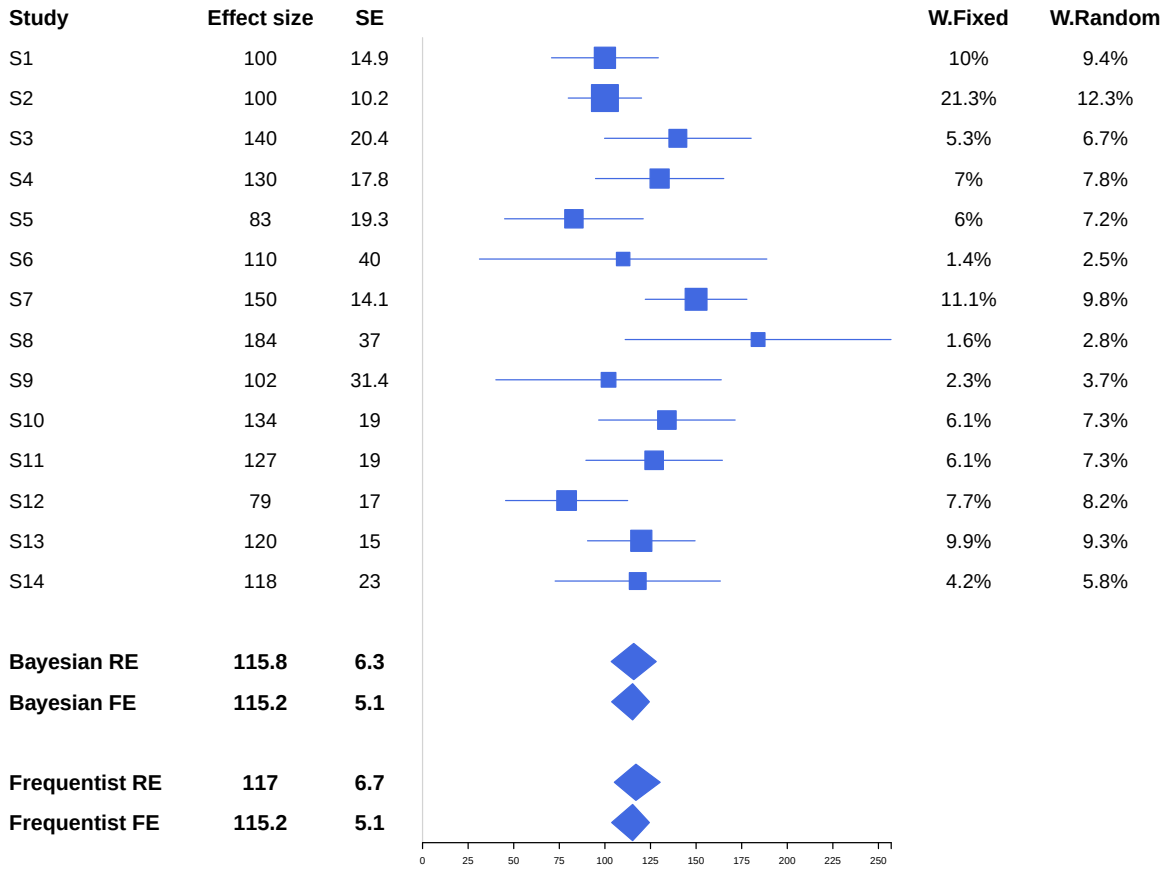


Figure 2.7: Forest plot for random and fixed effect in Bayesian and frequentist approach.

Table 2.3: Heterogeneity test and measures for frequentist approach, Cochran Q-statistic, I^2, and the p-values for random and fixed effect models.

Frequentist Approach	Q	I^2	p-value
Random Effects	24.09	46%	0.038
Fixed Effect	23.9	45.7%	0.031

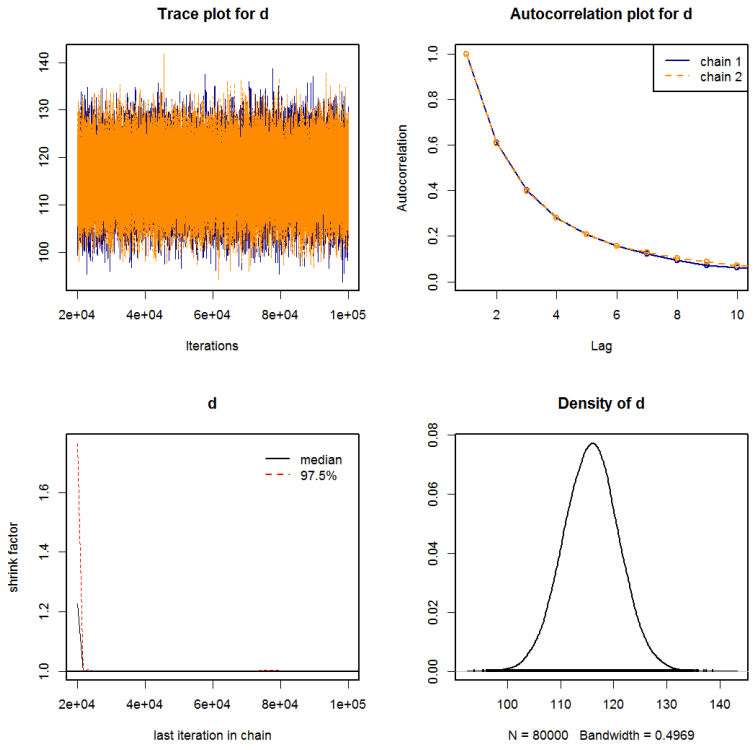


Figure 2.8: Diagnostics test for convergence of true effect (d) for Bayesian fixed effect model.

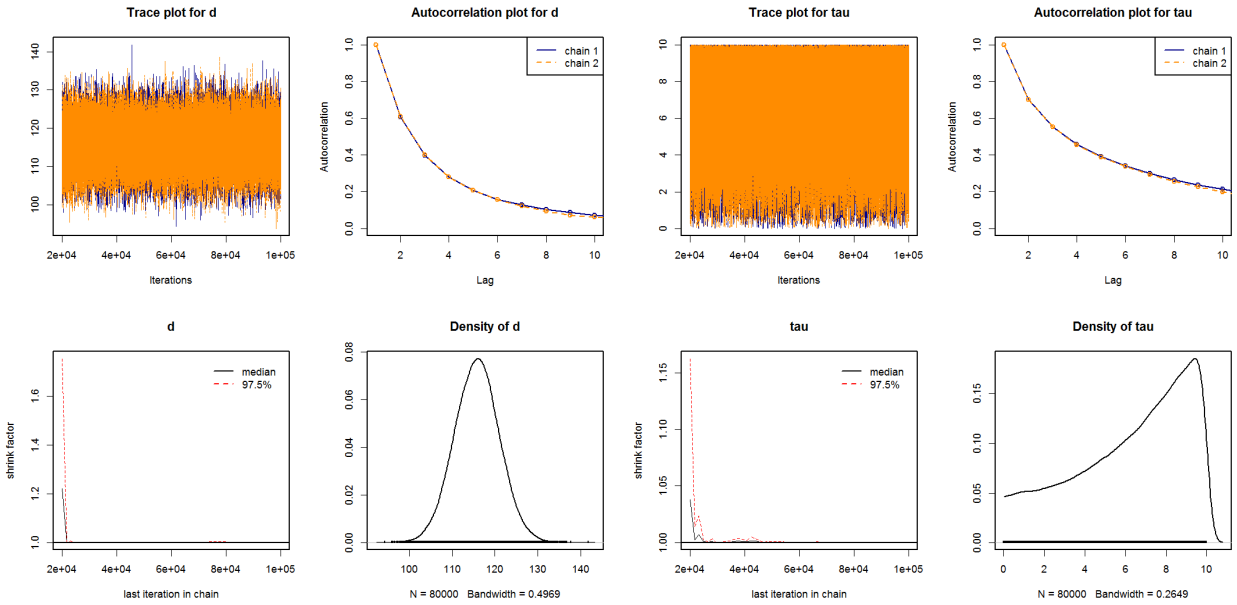


Figure 2.9: True treatment effect (d)

Figure 2.10: Between-variance (τ^2).

Figure 2.11: Diagnostics test for convergence of true effect (d) and between-variance for Bayesian random effects model.

Chapter 3

Indirect and mixed treatment comparison

3.1 Indirect treatment comparison and pooled effect

When the effect of a new treatment is examined versus a placebo in a RCT and results to be efficacious, the need to be tested with the indicative medicine or active treatments arises. Ideally, a well-designed and conducted RCT would simultaneously compare all treatments of interest, however this is rather impossible the decision making is getting complicated [23]. To compare them with the meta-analysis technique, pairwise RCTs for all pairwise combinations of the treatments of interest are required, which is an expensive, time consuming and inefficient way to compare them. Meta-analysis has its limitations comparing up to two treatments; many RCTs can be combined in order to be an "over all" study comparing these specific two interventions, however no information can be retrieved for non directly (head-to-head) compared intervention. As now it is clear, the requirement of head-to-head comparison in meta-analysis is a restrictive requirement. To overcome this fact, the indirect treatment comparison (ITC) is the extension of meta-analysis [24]. The ITC allows the comparison of the effect of two treatments that were not directly examined in a study, by comparing their performance over a common treatment, called as the common comparator, which is compared directly with both of the treatments of interest. Moreover, this comparator is often the placebo or the most commonly used treatment. In meta-analysis, pooled studies should share similarly study and patients characteristics in order to moderate the variability between them, the same applies also in the ITC method.

More specifically even if a direct evidence is not available for the comparison of two particular treatments (A, B) an estimation could be given in case that they have been both compared with a specific treatment(C), the common comparator, as Figure 3.1 illustrates. The effect size calculation for the indirect comparison, is geometrically presented in the Figure 3.2. The indirect treatment effect estimation of A vs B may be

estimated as:

$$\mu_{AB} = \mu_{AC} - \mu_{BC} \tag{3.1}$$

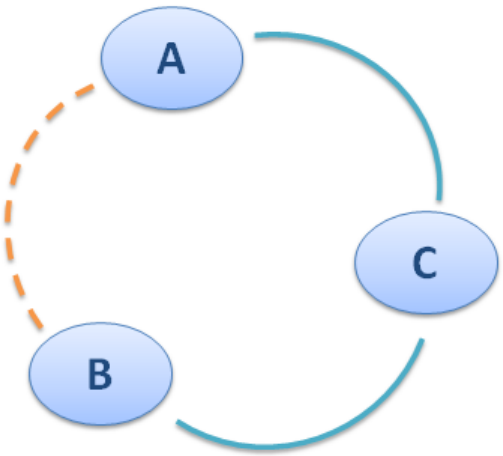


Figure 3.1: ITC diagram comparing treatments A, B, C. Directly comparison of A vs C and B vs C BC, inderctly comparison of A vs B.

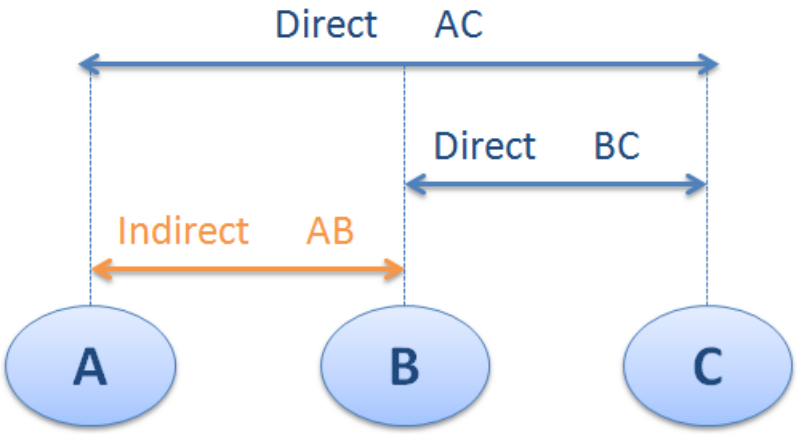


Figure 3.2: Estimation of indirect effect of treatments A and B via treatment C.

By general agreement a direct comparison provides stronger evidence than an indirect, however it cannot be ignored that in cases as only a few direct studies available but plenty of indirect, the volume of indirect

evidence is stronger and more reliable. As the inferences are not based on a RCT the assumption of randomization across RCTs collapses, no pooled SE or SD can be calculated, hence in an ITC the variance is the sum of the directs variances.

$$var(\mu_{AB}) = var(\mu_{AC}) + var(\mu_{BC}) \tag{3.2}$$

$$SD(\mu_{AB}) = \sqrt{SD(\mu_{AC})^2 + SD(\mu_{BC})^2} \tag{3.3}$$

Taking this way of thinking a step further, multiple indirect comparisons can be set in order to have estimations for multiple treatments effect sizes that are not directly compared, as it is presented in Figure 3.2. Adding many steps between the treatments compared increases the uncertainty-variability of the estimation, and the accuracy obviously decreases. The ITC estimation is an equation and has nothing to do with Bayesian and frequentist approach. The estimations that uses can be estimation arriving from any framework or model of meta-analysis methodology.

It is possible both direct and indirect evidence to be available, in that case a pooled effect is estimated. The pooled estimation of a treatment effect size is the combination of direct and indirect evidence with the inverse variance method; each estimation is weighted by the inverse of the variance, as it is defined below:

$$T_{pooled} = \frac{\frac{T_{Direct}}{Var_{T_{Direct}}} + \frac{T_{Indirect}}{Var_{T_{Indirect}}}}{\frac{1}{Var_{T_{Direct}}} + \frac{1}{Var_{T_{Indirect}}}} \tag{3.4}$$

with variance:

$$Var_{T_{pooled}} = \frac{1}{\frac{1}{Var_{T_{Direct}}} + \frac{1}{Var_{T_{Indirect}}}} \tag{3.5}$$

When both types of evidence exist, the result is a weighted average of them. The pooled estimationwill always be in between of the two treatment estimations. Moreover, it will be closer to the one with smaller variance. As it's reasonable the pooled estimation is closer to the (direct or indirect) estimation with the smallest variance. Furthermore, the pooled effect size is estimated with more precision, compared to the ITC. Following this idea, if more than one treatments can be used as the common comparator for an ITC, then all of the different estimations (coming from different paths) can be synthesized together with the direct estimate, if available. The direct estimation can be derived from fixed and random effects models.
The pooled treatment comparison is always available for a connected network, however it cannot be applied in all of them. It is an easy to use method, based on a simple equation 3.4, which doesn't serve for bigger networks, as it is not able to recognize the connection of the treatments, not even in simple cases as in the example below when there are only three treatments connected. Eventually, the restriction of only two treatments effect estimation in a time is the reason of extending this method to the network meta-analysis, in order to be applied in networks of arbitrary size and complexity.

3.1.1 Example

For a better understanding of the ITC and MTC an fictional example is given from the based on the example used in section 2.5. The outcome of interest is again the forced expiratory volume in 1 second(FEV_1) in mL in patients with COPD. A close loop containing three treatments the placebo, the Tio, and the QVA connected by four studies, illustrated in Figure **??**. A meta-analysis comparing Placebo to Tio is already conducted in section 2.5 and the studies' data are given in Table **??**. The Tio compared with the placebo, the one side of the triangle, is connected by 14 studies, the second side compares the effect of QVA to placebo, connected by two studies (S4 and S15), and the third side compares the effect of QVA to Tio, connected also by two studies (S4 and S16), the data for all studies in the close loop are provided in the Table 3.1. The frequentist and Bayesian approaches are permitted, as they are only two studies comparing the treatments in two of the three sides of the loop, random effects model was not deem to be appropriate, nevertheless both models for each approach are presented due to completeness of the example. There are available direct and indirect estimations for all pairwise comparisons, thus for all three of them a pooled estimation could be given, however the pooled estimation of QVA compared with the Tio is chosen to be presented here. As it is obvious, the ITC estimations will be calculated with common comparator the placebo.

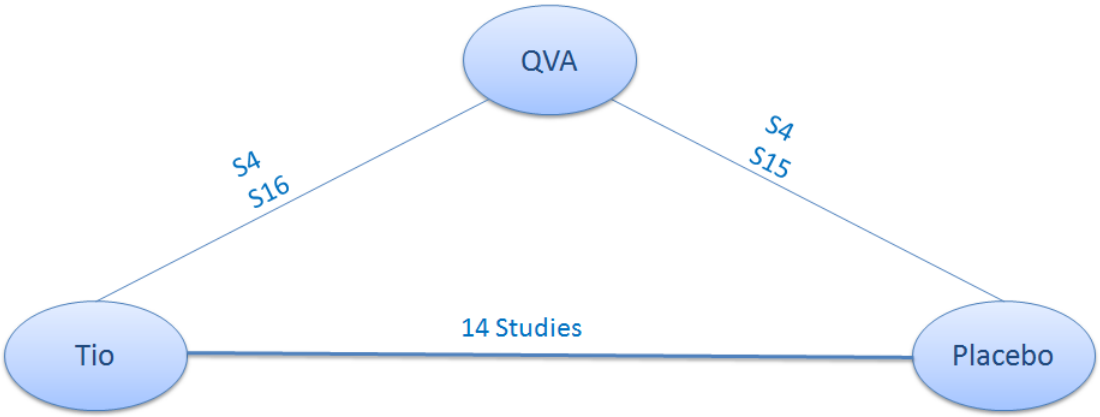


Figure 3.3: Network diagram comparing the 14 studies comparing the effect of Tio to placebo, 2 studies of QVA to placebo, and 2 studies comparing QVA to placebo.

For each side of the loop a meta-analysis is required to pool the evidence. The pooling of Tio vs placebo was examined in the section 2.5, as mentioned before. The next step is to pool the studies S4 and S15 comparing QVA vs placebo, Figure 3.4. Having the pooled direct estimations of Tio vs placebo, and QVA vs placebo, the ITC estimation is available, and it is presented in the Table 3.2 which summarizes the results from both approaches and models. The estimations for the effect of QVA compared with the placebo, and Tio with the placebo are given in the third and forth columns respectively, the last one gives the ITC estimation in each case using the equation **??**. The standard errors have reasonable smaller values in the

Table 3.1: Trough FEV_1 Difference in CFB data for studies in the network.

Study	Treatments	Mean(mL)	SE(mL)
S1	Tio vs Placebo	100	15
S2	Tio vs Placebo	100	10
S3	Tio vs Placebo	140	20
S4	Tio vs Placebo	130	18
S4	QVA vs Placebo	230	18
S4	QVA vs Tio	100	18
S5	Tio vs Placebo	83	19
S6	Tio vs Placebo	110	40
S7	Tio vs Placebo	150	14
S8	Tio vs Placebo	184	37
S9	Tio vs Placebo	102	31
S10	Tio vs Placebo	79	17
S11	Tio vs Placebo	134	19
S12	Tio vs Placebo	127	19
S13	Tio vs Placebo	120	15
S14	Tio vs Placebo	118	23
S15	QVA vs Placebo	163	32
S16	QVA vs Tio	70	14

comparison of Tio with the placebo, 14 studies is consider to provide a more reliable evidence in contrast of only two studies which was the case for QVA compared with the placebo. Fixed effect estimations for the ITC in both frameworks are identical (QVA vs Tio: Bayesian/Frequentist 99mL (SE:16.4mL)) as the estimations in the two meta-analysis (QVA vs placebo: Bayesian/Frequentist 214mL (SE:15.6mL); Tio vs placebo: Bayesian/Frequentist 115mL (SE:5.1mL)). All point estimates are relatively similar, and as it is expected the random effects have bigger uncertainty compared to fixed effect models.

In case that it was only one study examining the QVA with the Tio then the pooled effect would easily had been calculated, however in this case there are two of them, studies S4 and S16, and thus one last meta-analysis should be done, Figure 3.5 depicts the pooled direct evidence from both approaches and models. According to the equations 3.4 and 3.5, the final results are summarized in the Table 3.3. The pooled treatment effect estimations in all cases are more precise and the estimations are closer to the direct estimations as they deem to have smaller SE compare to the indirect ones. It has to be mentioned that the polled methodology can not recognize the connection among the studies, even in simple cases as in this example, a close loop of only three treatments compared. In general, to overcome this disadvantage the network meta-analysis, the extension of the meta-analysis, is presented in the next section.

Table 3.2: Trough FEV_1 Difference in CFB data for studies in the network.

Framework	Model	QVA vs Placebo (SE)	Tio vs Placebo (SE)	QVA vs Tio (SE)
Bayesian	Random effects	213 (16.3)	116 (6.3)	97 (17.5)
	Fixed effect	214 (15.6)	115 (5.1)	99 (16.4)
Frequentist	Random effects	202 (33.1)	117 (6.7)	85 (33.8)
	Fixed effect	214 (15.6)	115 (5.1)	99 (16.4)

Table 3.3: Direct, indirect and pooled treatment effect for QVA compared with the Tio for FEV_1 difference in CFB.

Framework	Model	Direct Estimation (SE)	Indirect (SE)	Pooled effect (SE)
Bayesian	Random effects	82 (11.5)	97 (17.5)	87.71 (6.95)
	Fixed effect	81 (10.9)	99 (16.4)	88.30 (6.55)
Frequentist	Random effects	83 (14.8)	85 (33.8)	83.50 (10.30)
	Fixed effect	81 (10.9)	99 (16.4)	88.33 (6.57)

Figure 3.4: Forest plot for random and fixed effect models in Bayesian and frequentist approach for QVA compared with the Placebo.

Figure 3.5: Forest plot for random and fixed effect models in Bayesian and frequentist approach for QVA compared with the Tio.

3.2 Network meta-analysis

The network meta-analyses (NMA) involves the simultaneous analysis of both direct and indirect comparisons among multiple treatments across multiple studies, usually RCTs [22, 25–28]. A combination of direct and indirect comparisons is the definition of a network of treatments, which is superior to meta-analysis and the final step in hierarchy of evidence synthesis, Figure 3.6 [29].

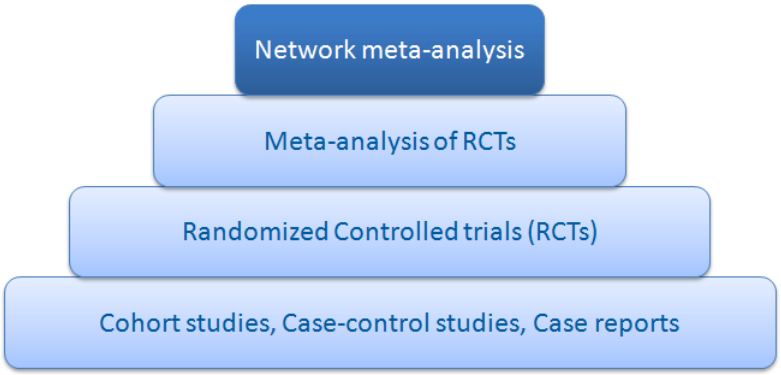


Figure 3.6: Hierarchy of evidence.

As long as all treatments in the network are connected, directly or indirectly all of them can be compared with each other, a very convenient graph illustrating two networks are given in Figure 3.7. It is obvious that as the number of RCTs across the network increases, the evidence increases and the network becomes strengthener. A network connected by one to two trials in all interventions can be very weak and may not be reliable. Prior to the analysis a feasibility assessment should be performed to examine if the evidence provided by the network can be consider as similar enough in order to proceed to the analysis. All RTCs should be chosen with the SLR process as it is mentioned in section 2.1. After the SLR is finalized, the assumptions required for the analysis should be examined, discussed in the next section.

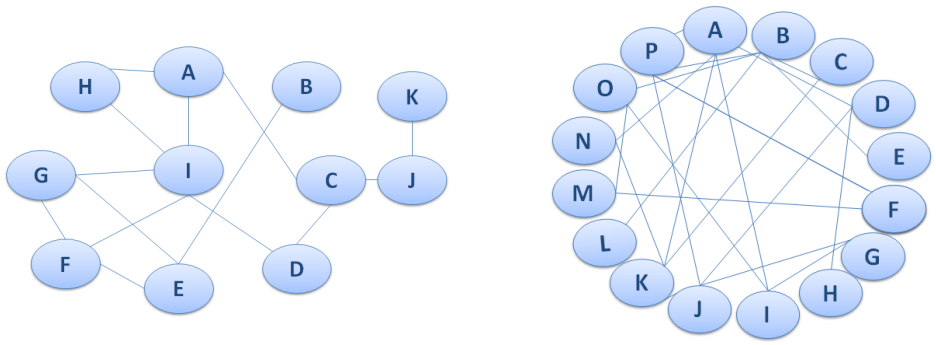


Figure 3.7: Networks containing large number of treatments (letters) and trials (lines).

3.2.1 Assumptions

Any statistical method has to obey specific assumptions which are fundamental and have to hold in order to ensure the validity of the results. The NMA methodology has three basic assumptions, similarity, homogeneity and consistency.The key element in order to meet these assumption is the correct study identification. It is important to mention that these assumptions are correlated, dissimilarities among trials will affect the homogeneity and the consistency of the analysis. The visualization in Figure 3.8 gives a better understanding about the assumptions.

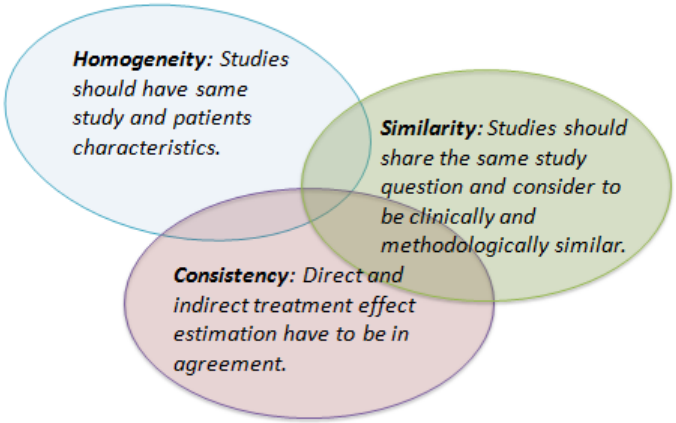


Figure 3.8: Basic assumptions for an NMA.

Similarity

Common criticism of meta-analysis is that the differentiation of combined studies may influence the final results. Similarity assessment aims to ensure that the meta-analysis or the NMA include only studies that are similar enough in order to be pooled together without affecting the effect estimations. It is undeniable that each RCT has unique characteristics and can not be identical with any other. Given that fact a researcher has to find the studies that can be consider as similar and "combinable". The biggest opponents' quarrel is that there are no statistical methods to evaluate the similarity in a network of studies. Hence, one has to be conscious, using all information can proven useful, as historical evidence and mainly experts opinion, an expert clinician can contribute extremely in that phase.

Effect modifiers, discussed in section 2.2 can be more complicated in a network of studies, some of them being compared only indirectly. Taking for granted that the search is taking place in all available databases applying the most relevant key words, PICOS framework is usually adopted to identify these variables. Considering population of the study as a property of primary importance, population treatments applied is the first characteristic examined, followed by the intervention, comparator, outcomes, and settings, as they are given in Table 3.4. PICOS criteria are proved useful in including or excluding studies, and also some researchers use the SPICOS criteria which consider the study design as the first characteristic needs top be checked.

Table 3.4: PICOS criteria

	Description	Possible effect modifiers
P	Population	Demographics, baseline clinical characteristics, disease severity
I	Intervention	Dose, mode of administration, duration
C	Comparator	Active treatment, placebo, concomitant meds
O	Outcomes	Definitions of effect measurements, thresholds, ITT vs. PP
S	Setting	Study design, study duration, location/country, method of outcome assessment

Summarizing the most important and relevant characteristics for the examining disease can also be helpful in understanding the available evidence for making decision, which studies will constitute the network. Tools e.g., summary tables, scatter plots visualizing in a plot the age, sex, and severity by using colors and sizes of the data can be key element in decision making about the similarity of existing evidence.

Homogeneity

A network is consider to be homogeneous when any study in the network could be representative of the NMA results. In other words, when the variability in the studies included in the NMA can be explained by only

the randomness then the network is considered to be homogeneous, in any other case heterogeneity exists among the RCTs. In order to the critical role of homogeneity be understandable it has to be mentioned that if this assumption does not hold then the RCTs combined may not estimating the same treatment effect. The difference in the assumption of similarity and homogeneity may not be distinctive. As discussed in the later section, similarity focuses on study level; it is a qualitative judgment. On the other hand homogeneity refer to outcome level; it is definitely a qualitative judgment. In other words if similarity doesn't hold, the differences in study and patients characteristics plausibly will have an effect on the observed outcomes.The meta-analysis and NMA are based on RCTs, however randomization does not hold across the studies in the network of RCTs. As a result, systematic differences in the distributions of study and patients characteristics across trials can ensue. In general, if there is an imbalance in study and patients characteristics-related effect modifiers across the different types of direct and indirect comparisons in a network meta-analysis, the corresponding indirect comparisons are biased [30, 31].

The assessment of homogeneity is discussed in section 2.2 from the opposite perspective, the heterogeneity, as it is obvious is exactly the same requirement being examined from the other point of view, hence the measurements as the Cochran Q-statistic and the I^2 ratio hold in the NMA as well. From the Bayesian aspect NICE recommends comparing model diagnostics between fixed effect and random effects models, if random effects fits the data better then the analyst can suspect heterogeneity in the network. As the heterogeneity is an indication of discrepancies and existence of effect modifiers across the combined studies the fist task to do is to explore again the studies to examine if the assumption of similarity holds and to identify these effect modifiers. Detection of these effect modifiers may be a challenging task in some cases, however when they are detected a researcher can proceed to a subgroup analysis, random effects model or meta-regression if the amount of combined RCTs is considered to be large enough [32].

Consistency

Considering three trials the Study 1, 2, and 3; which can also be the results of meta-analysis. The are the pairwise comparisons of three different treatments, the treatments A (placebo), B and C, as Figure 3.10 illustrates. They are head-to-head comparisons, thus they create a close loop as it is presented in Figure 3.9.

All the treatments are directly connected to each other thus gives the availability for directed and indirect pairwise comparisons. This fact provides the privilege to examine if the direct and indirect estimations are in line. Ideally these estimation are desired to be identical or as similar as possible, supporting the assumption of consistency in the network. It can be said that when results of indirect and direct comparisons are in agreement, then the network can be consider to be consistent. The Figure 3.11 depicts the estimated effect of ITC estimation, which occurs from the equation 3.2; they are identical. This is an ideal scenario and rather impossible to holds as heterogeneity exist among studies. Once again the need to combine studies

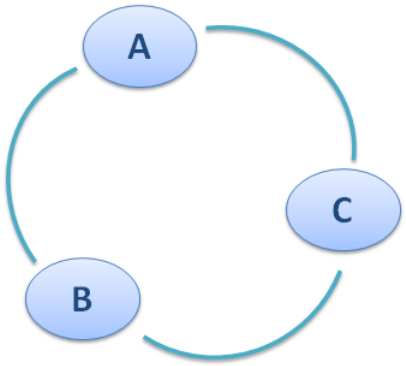


Figure 3.9: Closed loop with three treatments, A, B, and C.

with common study and patient characteristics arises. However, if inconsistency exists, and the estimations are not aligned, is reasonable to consider effect modifiers across the combined studies [33]. The Figure 3.12 presents the different directions of direct and indirect estimations, assuming that the effect size of "Study 3" is different as in Figure 3.10; the ITC overestimate the effect size of the treatment B compared with the A (placebo). In other words combining inconsistent studies may be inappropriate as can lead in invalid estimations. As it is clear now, the ITC methodology can overestimate or underestimate the effect of treatments depending on the studies existing.

3.2.2 Frequentist Approach

For the frequentist network meta-analysis the simple method of ITC can be used, trying to find the most efficient, the shortest, path in order to achieve the minimum variance. The fixed or random effect model depends only on the method estimated in the meta-analysis process, if it was needed . Moreover, if direct evidence is available then the techniques of pooled estimation should be applied. These methods are discussed in detail in section 3.1 and are not of interest in order to conduct and frequentist NMA, as they are not able to understand the geometry of the network and focuses only on the pairwise comparisons. A more precise approach uses the weighted least square regression [34, 35]. The most interesting fact in this methodology is the estimation for multi-arm studies. There are two different approaches to adjust the variances for multi-arm RCTs, the standard approach and the alternative or graph theoretical approach [36]. The adjustment is necessary as a trial with k>2 arms provides k(k-1)/2 pairwise comparisons, which are not independent; they are correlated and this correlation has to be estimated. If this correlation is ignored the estimated effect will be biased as the weights will be incorrect, underestimated. This section will discuss both approaches

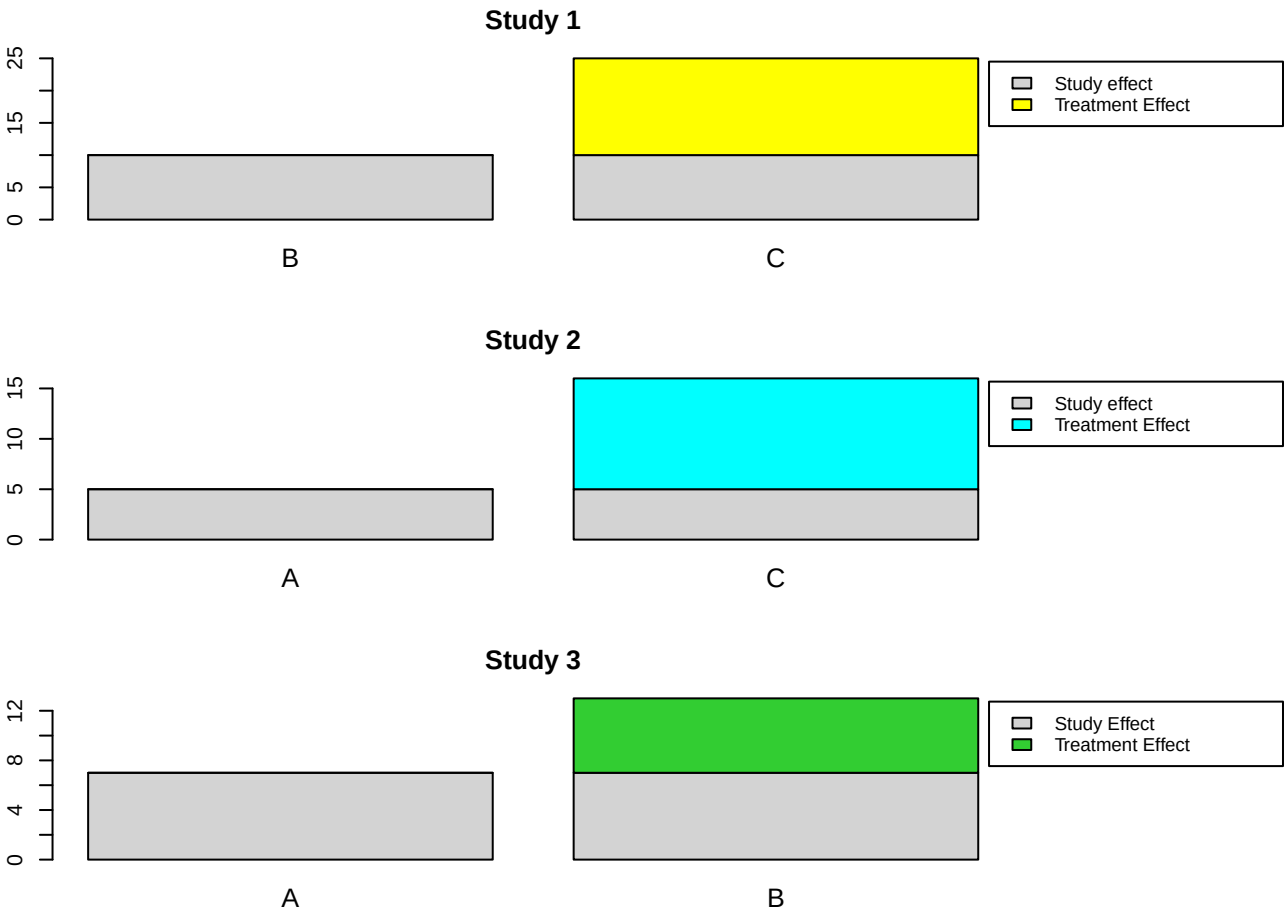


Figure 3.10: Three trilas pairwaise comparing treatments A(placebo), B and C.

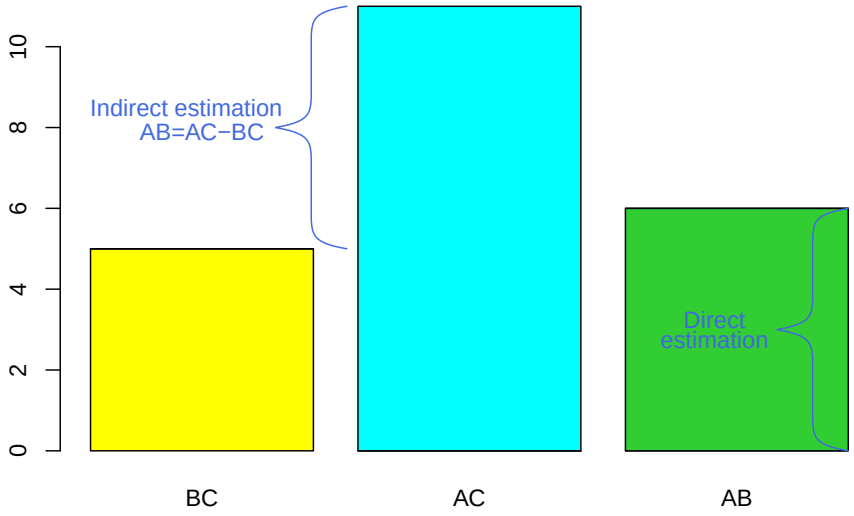


Figure 3.11: Cosistent loop. Direct and indirect estimations are agreement

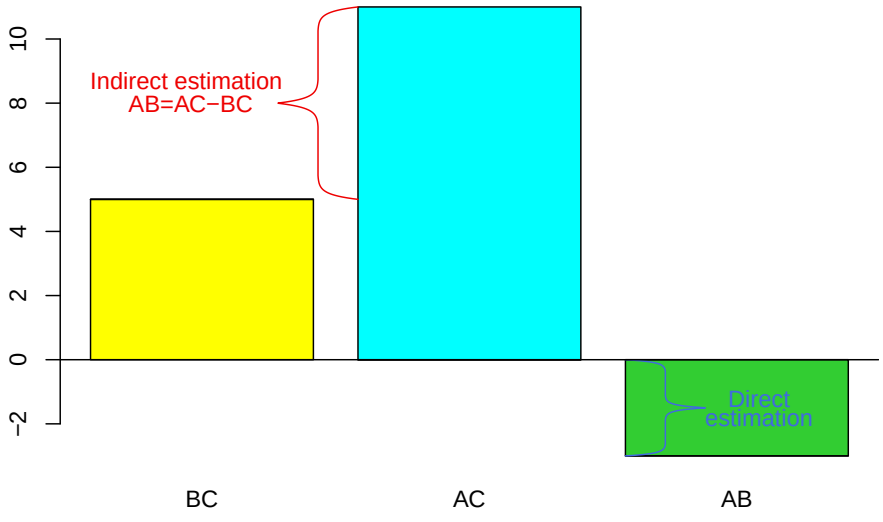


Figure 3.12: Incosistent loop. Direct and indirect estimations are not agreement

focusing more on alternative approach which is performed in the application presented in section 4.6. For both approaches an fictional example is provided.

The ordinary least square assumes equal variances σ^2 for all observations. In contrast, the weighted least square take into consideration the precision of each different observed data point. A study with a large variance contribute less than a study with a smaller variance. The residuals, e_i of a study i are weighted by its precision w_i, thus the weight given in a study is simply the inverse of the variance of its effect. Trials with equal precision will affect equally and in case that all $w_i = 1$ the ordinary least squares estimations are given. The model of weighted least square regression is given below, describes the estimated values $\hat{Y}_i$ as the outcome of the observed variables X_{ij} multiplied by their estimated coefficients β_{ij} with the additional errors residuals, e_i, for each i object and j variable:

$$\hat{Y}_i = \sum_{j=1}^{n} X_{ij}\beta_{ij} + e_i$$

$$e_i = \sqrt{w}_i(Y_i - \hat{Y}_i)$$

The hat matrix B, containing all estimated coefficients β_{ij} is defined by:

$$B = X(X^T C^{-1} X)^{-1} X^T C^{-1} \tag{3.6}$$

As usually X denotes the design matrix and the C represents the covariance-variance matrix of the whole network, which is a block diagonal matrix that consists of an invertible (k-1)x(k-1) blocks (discussed in the next section).

For a better understanding of the methods a fictional example of a 3-arm and a 2-arm study was implemented. A 3-arm study, S_1 and two 2-arm studies, S_2. S_1 performing a close loop, comparing the placebo, the Tiotropium (Tio), and the combination of Tiotropium and Olodaterol (Tio/Olo). The Table 3.5 presents the difference in CFB values and their SEs in mL of trough FEV_1 considering a time point of 12 weeks. According to these comparisons consider the matrix H, an edge-vertex matrix, the number of rows correspond the number of pairwise comparisons, m, and the number of columns indicates the number of treatments, n. The number of comparisons of a study with k arms is k(k-1)/2. For the fictional example given in Table 3.5, the corresponding matrix H is:

$$H = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \\ 1 & -1 & 0 \end{pmatrix} \tag{3.7}$$

Each row of the matrix H sums to zero, the three first rows corresponds to S_1 study, as it is 3-arm study (m=3), and for study S_2 there is only one row, the last one, as it is a 2-arm study and thus only one comparison is possible. For the S_1 study the need to adjust the SE arises. To adjust the uncertainty for k-arm, k>2, studies there are two methods as mentioned before, fist the standard method is discussed.

Table 3.5: Fictional example: a 3-arm study (S_1) and a 2-arm study (S_2) comparing three treatments, Tio/Olo, Tio, and Placebo.

Study	Treatment1	Tretament 2	Difference in CFB (SE)
S_1	Tio/Olo	Tio	28 (19.1)
	Tio/Olo	Placebo	162 (19.1)
	Tio	Placebo	135 (19.7)
S_2	Tio/Olo	Tio	76 (14.9)

Fixed effect

Standard approach

In an attempt to adjust the variances for a k-arm, k>2 study in the standard approach, its variance-covariance matrix C_k has to be calculated. The matrix Σ of the equation 3.8 is the diagonal matrix of arm-based, as it is illustrated below.

$$C_k = H_k \Sigma H_k^T \tag{3.8}$$

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix}$$

The C_3 matrix for the S_1 study is:

$$C_3 = \begin{pmatrix} \sigma_1 + \sigma_2 & \sigma_1 & \sigma_2 \\ \sigma_1 & \sigma_1 + \sigma_3 & \sigma_3 \\ -\sigma_2 & \sigma_3 & \sigma_2 + \sigma_3 \end{pmatrix}$$

The Table 3.5 provides estimations and variances of the contrasts for the two fictional studies S_1 and S_2. In more analytic form the table provides the contrasts as $\sigma_1+\sigma_2 = 18.8$, $\sigma_1+\sigma_3 = 19.1$, and $\sigma_2+\sigma_3 = 19.7$. Using the B_3, we can observe that the H matrix, by omitting the last row, it is obvious that the C_3 corresponding S_1 study is:

$$C_3 = \begin{pmatrix} 18.8 & 9.1 & 9.7 \\ 9.1 & 19.1 & 10 \\ -9.7 & 10 & 19.7 \end{pmatrix} \tag{3.9}$$

The C_3 is a 3x3 matrix containing the contrasts variances in the diagonal, as well as their covariances, e.g. $Cov(Tio/Olo, Tio) = \sigma_1 = 9.1$. In other words, it is adjusted accordingly to take into consideration the correlation existing among the effect estimations in k-arm studies. As for the H, the hat matrix, the inverse of

C_k is needed, however the C_k matrix has a rank k-1 and thus not invertible. The standard approach suggests dimensionality reduction by omitting all except the first k-1 rows and columns, leading to a (k-1)x(k-1) $\tilde{C_k}$ which is invertible, completely justifying the reason why the method is also named dimension reduction. It is obvious that there is no need for this process is k=2, for 2-arm studies.

The C and X matrices for the overall network are:

$$C = \begin{pmatrix} 18.8 & 9.1 & 0 \\ 9.1 & 19.1 & 0 \\ 0 & 0 & 14.9 \end{pmatrix} \quad X = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \end{pmatrix} \tag{3.10}$$

Finally, using the equation 3.6 the hat matrix B can be estimated, and the effect estimations Bx are defined by multiplying the B with the corresponding estimations of the Table 3.5:

$$B = \begin{pmatrix} 0.44 & 0.00 & 0.56 \\ -0.27 & 1 & 0.27 \\ 0.44 & 0 & 0.56 \end{pmatrix} \quad Bx = \begin{pmatrix} 54.3 \\ 174.9 \\ 54.3 \end{pmatrix} \tag{3.11}$$

Graph theoretical approach

The graph-theoretical networks that have been applied to electrical networks also work well in network meta-analysis [36]. The idea of frequentist network meta-analysis is described by these networks. For a meta-analytic network, the vertices (nodes) correspond to treatments and the edges to existing evidence between treatments. As already mentioned in section 2.3 each direct study estimation is weighted according to its inverse variance, the precision. The variance of a comparison corresponds to the resistance (R) of the edge in the network, whereas the precision corresponds to the conductance. In the simple case for a parallel connection of edges between two nodes in an electrical network, the effective resistance is given by the inverse of the sum of resistances $\frac{1}{R} = \sum_k \frac{1}{R_k}$. This is the case of pairwise meta-analysis, where the total precision (inverse variance) estimate of the total effect ($\bar{T}$) is given by the sum of the precisions from all m studies combined $\frac{1}{V(\bar{T})} = \sum_{k=1}^{m} \frac{1}{V_{Tk}}$.

The standard approach reduces the dimensions to obtain invertible variance-covariance matrices, the alternative approach reduces the weights of all comparisons in advance and then uses the B matrix without reduction. As the B matrix is not invertible, the Moore-Pensrose pseudoinverse is considered. The B matrix is given by:

$$B = H(H^T W H)^+ H^T W \tag{3.12}$$

The W matrix is a diagonal matrix containing all observed inverse variances weights of all comparisons. The corresponding variance-covariance matrix C of the standard approach is the Laplacian matrix L, defined as:

$$L = H^T W H \tag{3.13}$$

For a network with n treatments L is an nxn matrix, not invertible as has rank n-1. The weights of the treatment comparisons are the negative non-diagonal values. For each k-arm study with k>2, the L matrix should be calculated separately for each k-arm study. The increased variances, reduced weights, can be obtained from L^+ matrix using the V matrix kxk, which contains the observed variances of all comparisons:

$$L^+ = -\frac{1}{2k^2} H^T H V H^T H \tag{3.14}$$

The reduced weights are the negative non-diagonal entries of L matrix, which is trivially the $L = (L^+)^+$.

The example discussed below is the same as in the standard approach. First the reduced weights of S_1 study should be calculated. The corresponding matrix V is:

$$V = \begin{pmatrix} 0 & \sigma_1 + \sigma_2 & \sigma_1 + \sigma_3 \\ \sigma_2 + \sigma_1 & 0 & \sigma_2 + \sigma_3 \\ \sigma_3 + \sigma_1 & \sigma_3 + \sigma_2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 18.8 & 19.1 \\ 18.8 & 0 & 19.7 \\ 19.1 & 19.7 & 0 \end{pmatrix} \tag{3.15}$$

The non-adjusted weights of the study are the inverse of theirs variances, and then (0.0532, 0.0524, and 0.0508). The adjusted, reduced weights derived from 3.14 are 0.0362, 0.0351 and 0.0329, the negative no-diagonal entries find below:

$$L = \begin{pmatrix} 0.0713 & -0.0362 & -0.0351 \\ -0.0362 & 0.0691 & -0.0329 \\ -0.0351 & -0.0329 & 0.0680 \end{pmatrix} \tag{3.16}$$

As mentioned, the W matrix is a nxn matrix, n indicates the number of studies, thus for the fictional data in Table 3.5 the adjusted weights for the 3-arm study S_1, are the weights obtained from the matrix L, for the 2-arm study S_2 there is no need for adjustment, its weight is simply the inverse of its variance.

$$W = \begin{pmatrix} 0.0362 & 0 & 0 & 0 \\ 0 & 0.0362 & 0 & 0 \\ 0 & 0 & 0.0329 & 0 \\ 0 & 0 & 0 & 0.0680 \end{pmatrix} \tag{3.17}$$

Finally, the results estimated from the alternative approach are identical with the standard approach, using the equation 3.12, and calculating the effect estimations by multiplying the matrix B with all the observed treatment comparison effects given in Table 3.5, can be found that :

$$B = \begin{pmatrix} 0.30 & 0.14 & -0.14 & 0.56 \\ 0.15 & 0.58 & 0.42 & 0.27 \\ -0.16 & 0.44 & 0.56 & -0.29 \\ 0.30 & 0.14 & -0.14 & 0.56 \end{pmatrix} \quad Bx = \begin{pmatrix} 54.3 \\ 174.9 \\ 120.9 \\ 54.3 \end{pmatrix} \tag{3.18}$$

Random Effects

The two approaches are equivalent, since both of the each study's contribution are the same, thus the total residual heterogeneity Q-statistic is the same in the standard and the alternative approach, a fact that can be proven based on the following equation:

$$C_k = H\Sigma H^T = HL^+H^T$$

Based on the fictional example, the matrix C_3 for the study S_1 (matrix 3.9), is identical if instead of Σ the L^+ is used.

$$L^+ = \begin{pmatrix} 6.23 & -3.07 & -3.16 \\ -3.07 & 6.43 & 3.36 \\ -3.16 & -3.36 & 6.52 \end{pmatrix}$$

$$C_3 = H\Sigma H^T = HL^+H^T = \begin{pmatrix} 18.8 & 9.1 & 9.7 \\ 9.1 & 19.1 & 10 \\ -9.7 & 10 & 19.7 \end{pmatrix} \tag{3.19}$$

As the matrix C_k is not invertible, dimension reduction works with the (k-1)x(k-1) reduced matrix $\tilde{C_k}$, which can be described by matrix notation with the use of two additional matrices, P_1 and P_2. P_1 is a $(k-1)xm_k$ containing the first k-1 unit vectors at the first k-1 columns and zero columns for the following $m_k - (k-1)$ columns, and P_2 is a (k-1)xk matrix having a zero column followed by k-1 unit vector columns, an example for k=3 is given below.

$$P_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, P_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Using P_1 and P_2 matrices the reduced variance-covariance matrix of the contrasted can be written as $\tilde{C_k} = P_1H\Sigma H^TP_1^T = P_1HL^+H^TP_1^T$. At this point, using the lemma below, it can be said that both approaches lead to identical variance-covariance matrix of the contrasts.

Lemma: The inverse of covariance matrix of a complete graph is identical to its matrix, reduced by the first row and column.

$$\tilde{C}_k^{-1} = P_2LP_2^T \tag{3.20}$$

Since the two approaches lead clearly to the same results, referring to the estimation of effect and its variance, the need to estimate the residual heterogeneity measured buy the Q-statistic arises. The contribution of each study can be defined in general as the weighted squared deviation of the observed from the estimated effect. The difference of the estimated effect (Bx) from the observed effect(x) is defined as x_d. Thus, using the aforementioned lemma, the Q-statistic for a study i is equal in both approaches and can be defined as:

$$Q_i = x_{id}^T C_i^{-1} x_{id} = x_{id}^T P_2 L_i P_i^T x_{id} = x_{id}^T P_2 H_i^T W_i H_i P_2^T x_{id} \tag{3.21}$$

Finally, the random effects model can be presented, the graph theoretical perspective has been decided to be discussed, as it has been used in the application of this thesis. The random effects model simply differs from the fixed effect in the additional between-study variance, the common variance parameter τ^2. In the frequentist approach this parameter is calculated by:

$$\tau^2 = max(\frac{Q - df}{tr((I - B)UW)}, 0) \tag{3.22}$$

where df are the degrees of freedom defined as:

$$df = \sum_k (k-1)h_k - (n-1) \tag{3.23}$$

The number of studies with k arms is denoted by n_k, I is a mxm (m: number of comparisons), the identity matrix U is a block diagonal matrix derived from the mxm matrix $HH^{T/2}$. The random effects model adds the τ^2 parameter to the observed sampling variance of each single comparison in the network before any necessary weigh reduction. The v vector containing all observed variances of all m comparisons can be defined in a general form as:

$$v = (\sigma_1 + \tau^2, \sigma_2 + \tau^2, \ldots, \sigma_m + \tau^2) \tag{3.24}$$

3.2.3 Bayesian Approach

The Bayesian approach of network meta-analysis is deemed to be more flexible compared to the frequentist framework, as it uses models to estimate the treatment effects. The correlations induced by multi-arm studies, the estimation of predictive intervals and the ranking probabilities are straightforward in this approach, thus Bayesian framework could be preferable. Topics discussed in section 2.4 applies also for network meta-analysis methodology. The main difference with the meta-analysis can be noticed to the number of parameters needed. The simple meta-analysis requires only one prior, the true treatment effect. Network meta-analysis is a collection of many treatments connected by RCTs, some of the treatment effects are chosen to be the *basic parameters* and the rest of them as the *functional parameters*. The latter are linear combination of the basics, and thus only the basics parameters have priors and are considered as random variables. The choice of priors is according to the knowledge and evidence available as discussed in section 2.4.1.

In generalized linear models, a likelihood is defined in terms of parameters (γ), while a link function (g) maps the parameters of interest in $\Re$. There are different likelihoods and link functions for different outcomes, the Table 3.6 illustrates the most frequently used. In a more general form the network meta-analysis is based on equation 3.26, in which μ_i is the baseline treatment effect for the i specific trial. For the i trial, the b arm is the control treatment arm (b=1, the first arm) and thus the $\delta_{i,bk} I_{k \neq 1}$ is the treatment effect of the k-arm to

the control-arm of the study i. Priors should be given to all trial baseline effects and basic parameters. For example in a 2-arm study I, priors should be given to the baseline effect μ_I and to the treatment effect $\delta_{I,12}$.

Table 3.6: The link functions, the likelihoods, the outcomes, and the measurements for the most frequent used generalized linear models. The parameter is denoted by γ and the link function by θ.

Link	Link function $\theta = g(\gamma)$	Likelihood	Outcome	Measure
Identity	γ	Normal	Continuous	MD, SMD, CFB, difference in CFB
Logit	$ln(\frac{\gamma}{1-\gamma})$	Binomial	Binary	OR, RR
Probit	$\Phi^{-1}(\gamma)$	Multinomial		
cloglog	$ln(-ln(1-\gamma))$			
Log	$ln(\gamma)$	Poisson	Rate	IRR, IRD

$$g(\gamma) = \theta_{ik} = \mu_i + \delta_{i,bk} I_{k \neq 1} \tag{3.25}$$

$$\delta_{i,12} \sim N(d_{12}, \sigma^2) \tag{3.26}$$

Fixed Effect

The basic idea for the Bayesian network meta-analysis is described by the connected network illustrated in Figure 3.13. In this example all studies are deemed as 2-arms in order to be well understood. The treatment A is the reference treatment and as it is connected directly with treatments B, C and D, creates three basic parameters, AB, AC, and AD. These parameters are named basics as they are related to the reference treatment A, all the remaining contrasts BC, BD, and CD can be defined in terms of basic parameters and therefore are the functional parameters. To all the basic parameters priors need to be given, as well as to the study effects [37].

Figure 3.13: Connected network comparing four interventions(A,B,C,D) with three basic (AB, AC, AD) and three functianl parameters (BD, BC, DC) with 6 studies (lines).

More specifically the functional parameters can be derived from the basics as:

$$d_{BC} = d_{AC} - d_{AB}$$
$$d_{BD} = d_{AD} - d_{AB}$$
$$d_{CD} = d_{AD} - d_{AC}$$

The sample size should be large enough to allow referring the Central Limit Theorem, assuming normality over the mean. The fixed effect model could be described as the extension of the simple pairwise meta-analysis. The Y_{ik}, indicates the effect of k treatment in the i study and from a normal distribution as described below:

$$Y_{ik} \sim N(\theta_{ik}, \sigma_{ik}^2) \tag{3.27}$$

$$\theta_{ik} = \mu_i + d_{i,bk} I_{k \neq 1} \tag{3.28}$$

$$\mu_i, d_{AB}, d_{AC}, d_{AD} \sim N(0, 10^6) \tag{3.29}$$

The observed effect Y_{ik} of the k intervention in the study i, comes from a normal distribution with true mean θ_{ik} and variance the standard error σ_{ik}^2 of it. The normal distribution requires identity link function, and thus the model for the linear predictor is the baseline effect of the study μ_i and the difference of the first, "the baseline intervention", b, to the k arm, as the equation 3.28 shows. Finally, vague priors are given to all baseline effects and basic parameters due to objectiveness. The fixed effect doesn't assume variability among the RCTs, the true treatment effects are consider to be fixed and not derived from an individual study distribution, e.g. $d_{i,12} = d_{12}$ and not $d_{i,12} \sim N(d_{12}, \sigma_i)$ for any i study.

The model described before is convenient with time point data, when the effects for each RCT are reported at the time point of interest with any measure of uncertainty. In case of reported measurements at the baseline and at a pre-specified time point, the CFB is preferred, as presented in section 2.1.2. The best case scenario is when CFB values are reported with an uncertainty measure, then the researcher can follow the model as previously. When only baseline and at time point effects reported then calculations have to be done and the calculated variance should take into consideration the correlation among arms in the same study. An other way to present aggregated data, probably the most common, is the difference between arms with any measure according to the data examined as, OR, RR, MD or IRR, and its variance or even in cases only the p-value. It can be assumed that the mead difference is approximately normally distributed, therefore the link function remains the same, however the model of the equation 3.28 changes as there is no baseline effect and the model is obtained as $\theta_{ik} = d_{i,bk} I_{k \neq 1}$. A very crucial issue to which it should be given great concern is when more than study has more than two arms, in that case the reported mean differences are reported based on a reference treatment and all of them are correlated and that require adjustment to the likelihood. The last of cases to report continuous outcomes is the SMD, most preferred when outcomes reported in a variety of scales, mainly in psychological and neurological fields.

Random Effects

The differentiation of fixed and random effects is the sources of variation allowed. The random effects model assumes two sources of variance, the within-study variance, in any study included and also the between-study variance, allowing the treatment difference effects not being equal but coming from a distribution with mean the study effect and variance the between-study variance. Additional to the basic parameters and baseline effects, a prior should be given to the between-study variance, most common a vague prior is used. The likelihood, model and priors are given in way to be consisted with the fixed effect network meta-analysis, continuous outcome is again assumed.

$$Y_{ik} \sim N(\theta_{ik}, \sigma_{ik}^2/n_{ik}) \tag{3.30}$$

$$\theta_{ik} = \mu_i + \delta_{i,bk} I_{k\neq 1} \tag{3.31}$$

$$\delta_{i,bk} \sim N(d_{ik}, \tau^2) \tag{3.32}$$

$$\begin{gathered} \mu_i, d_{AB}, d_{AC}, d_{AD} \sim N(0, 10^6) \\ \tau^2 \sim U(0,2) \end{gathered} \tag{3.33}$$

It is obvious that the random effects model is more complicated, it is preferred in occasions of large enough number of RCTs included in the network. The random effects estimations gives similar point estimates as fixed effect model, however the variability is greater. The additional variability derives from the fact that the differences of effects are assumed to come from a normal distribution with mean the treatment differences and variance the between-study variance, equation 3.32, which is estimated and thus has to be given prior. As in fixed effect model, priors are given to baseline effect and basic parameters and also prior is given to the between-study variance, equation 3.33.

The way to handle different ways of reported aggregated data is of course the same. Once again, for multi-arm studies the correlation among effects should be taken into consideration.

Variance adjustment for multi multi-arm studies

Models discussed holds for 2-arms studies, they do not account for the multi-arm studies correlation. In general, the consistency equations in a s-arm study are described as:

$$\begin{aligned} d_{2,3} &= d_{1,3} - d_{1,2} \\ d_{2,4} &= d_{1,4} - d_{1,2} \\ d_{2,5} &= d_{1,5} - d_{1,2} \\ &\vdots \\ d_{(s-1),s} &= d_{1,s} - d_{1,(s-1)} \end{aligned} \tag{3.34}$$

The basic parameters are the $d_{1,2}, d_{1,3}, \ldots, d_{1,s}$, and all the remaining contrasts are functional to them. For instance lets consider the effect of the treatment 3 compared with the treatment 2 of the i study:

$$\delta_{i,(2,3)} \sim N(d_{2,3}, \sigma^2_{2,3})$$

The functional parameter $d_2, 3$ derives from the basic parameters $d_{1,3}, d_{1,2}$, as $d_{2,3} = d_{1,3} - d_{1,2}$ and thus

$$\sigma^2_{2,3} = \sigma^2_{1,2} + \sigma^2_{1,3} - 2\rho_{2,3}\sigma^2_{1,2}\sigma^2_{1,3} \tag{3.35}$$

The equation 3.35, in which the ρ represents the correlation between the relative effect of arm's 3 treatment, clarifies the need of adjustment for multi-arm studies [38]. For simplicity reasons it is common to assume equal variances, $\sigma^2_{1,2} = \sigma^2_{1,3} = \sigma^2_{2,3} = \sigma^2$, and thus the correlation between any two treatment contrasts is 0.5.

In a general form the random effects, assuming equal between-study variance for a k-arm study i are coming from a (k-1) multivariate normal distribution as below:

$$\delta_i = \begin{pmatrix} \delta_{i,(1,2)} \\ \delta_{i,(1,3)} \\ \vdots \\ \delta_{i,(1,k)} \end{pmatrix} \sim N_{i,s-1} \left[\begin{pmatrix} d_{i,(1,2)} \\ d_{i,(1,3)} \\ \vdots \\ d_{i,(1,k)} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 & \ldots & \sigma^2/2 \\ \vdots & \ddots & & \vdots \\ \sigma^2/2 & \sigma^2/2 & \ldots & \sigma^2 \end{pmatrix} \right]$$

the vector δ_i is the random effects of treatments comparisons for the specific study i and it is a multivariate normal. The conditional univariate distribution for the random effects of k-arm study is:

$$\delta_{i,1,k}| \begin{pmatrix} \delta_{i,(1,2)} \\ \delta_{i,(1,3)} \\ \vdots \\ \delta_{i,(1,k-1)} \end{pmatrix} \sim N((d_{1,k} - d_{1,1}) + \frac{1}{k-1} \sum_{j=1}^{k-1} [\delta_{i,(1,j)} - (d_{1,j-d_{1,1}})], \frac{k}{2(k-1)\sigma^2}) \tag{3.36}$$

The equation 3.36 should be used to estimate the random effects for each multi-arm study, in that way the between-arm correlation among the parameters will be taken into account. As the fixed effect models doesn't make any assumption for the differences it is no need for adjustment in that case.

Treatment differences in multi-arm studies

An other important issue should be examined, in the majority of multi-arm studies the effects are reported as mean differences from the baseline-reference treatment, usually placebo. Since differences are taken relative to the same baseline arm, an adjustment to the likelihood is needed as the correlation is inherent in the data. A RTC with k-arms produce k-1 correlated treatment differences. The variances (V) of treatments compared

with the baseline is required. A 3-arm study will report 2 treatment differences (AB, AC) compared with the baseline (treatment A) and their covariance is obtained from the equation 3.37

$$Var(Y_{AB} - Y_{AC}) = Var(Y_{AB}) + Var(Y_{AC}) - 2Cov(Y_{AB}, Y_{AC}) \tag{3.37}$$

It is known from the original data that $Var(Y_{AB}) = Var(Y_A) + Var(Y_B)$ and $Var(Y_{AC}) = Var(Y_A) + Var(Y_C)$, and thus the covariance of the reported treatment differences is simple the variance of the common arm.

$$Cov(Y_{AB}, Y_{AC}) = var(Y_A)$$

An example from the overall network of the application in COPD patients is used to describe the adjustment of a 3-arm study. The study used as example is the S1 (same as in the frequantist), a 3-arm study with placebo treatment acts as the baseline-reference treatment for the frequentist approach.

The data for the study are presented in the Table **??** as the differences in CFB of FEV_1 in mL at 12 weeks are given as the differences in CFB. The SE of Tio vs Placebo is 19.7mL, and Tio/Olo vs placebo is 19.1mL. Then the SE of Tio/Olo to Tio, the covariance of the two differences reported is the SE^2 of placebo, 196mL, thus the covariance of the 2.

$$SE_{(Tio/Olo\ vs\ Placebo - Tio\ vs\ Placebo)} = SE_{(Tio/Olo\ vs\ Placebo} + SE_{Tio\ vs\ Placebo} - 2\rho_{Tio/Olo\ vs\ Placebo, Tio\ vs\ Placebo} \tag{3.38}$$

Chapter 4

Application in COPD

4.1 Disease characteristics

Chronic obstructive pulmonary disease (COPD) characterized by the development of airway obstruction, manifesting as a decline in lung function, breathlessness and exacerbations [39]. The disease is usually progressive and associated with an enhanced chronic inflammatory response in the airways and lungs. It is the fourth leading cause of death worldwide and a major cause of chronic morbidity and mortality throughout the world. The COPD burden is expected to increase over the coming decades because of continued exposure to COPD risk factors and aging of the population.

The goals of effective COPD management are to prevent disease progression, relieve symptoms, prevent and treat complications and exacerbations and to improve health status. Current guidelines (ref) recommend the use of long-acting bronchodilators, as they are more effective at producing maintained symptom relief than short-acting bronchodilators. The choice between long-acting β_2-agonists (LABAs), long-acting anticholinergic (LAMA), and fixed combination of LABA plus ICS (inhaled corticosteroids) in one inhaler depends on the individual patient response in terms of symptom relief and side effects. Also, combining bronchodilators of different pharmacology classes may be considered as a better alternative to manage COPD compared to increasing the dose of a single bronchodilator.

4.2 Objectives

This analysis used the publication of Huisman. E, Comparative efficacy of combination bronchodilator therapies in COPD: a network meta-analysis, 2015 [1], in which a Bayesian network meta-analysis was performed by MAPI in request of GlaxoSmithKline (GSK). By this time GSK was developing a closed combination of an inhaled long-acting muscarinic antagonist (LAMA) and long-acting $beta_2$ agonist (LABA) for COPD, umeclidinium/vilanterol (UMEC/VI), delivered using the Ellipta® dry-powder inhaler. The clinical efficacy

and safety of UMEC/VI 62.5/25 is being evaluated in a comprehensive Phase III program. The anticipated target population for UMEC/VI is mild to moderate COPD patients who have a history of no or very few exacerbations, including patients who are symptomatic on monotherapy.

A clear value proposition, supported by robust comparative clinical evidence will be required to address the needs of payers in different markets. GSK asked Mapi to conduct a robust comparative assessment of UMEC/VI with all of its potential comparators. Therefore a comprehensive systematic review (SLR) was performed to identify randomized controlled trials (RCTs) in COPD patients eligible for maintenance therapy. A feasibility assessment was performed to evaluate the appropriateness of evidence synthesis for UMEC/VI 62.5/25 with its comparators (based on the results of the systematic literature review). The trial evidence should be used to perform a Bayesian network meta-analysis.

The objective of the analysis performed for this dissertation has been to compare a network meta-analysis in frameworks analysed, Bayesian and frequentist, random and fixed models were conducted in any of them for the outcome of interest the trough FEV_1 at 12 and 24 weeks. The paper presented three more outcomes, the rescue medication use in puffs per day, the St George's Respiratory Questionnaire total score, and the transitional dyspnea index focal score, which were not of interest as they are all continuous outcomes, as FEV_1, and the analysis does not differfor the methodological prospective. Moreover, the aspect of this analysis has been to assess the relative efficacy of umeclidinium 62.5 mg in combination with vilanterol 25 mg (UMEC/VI 62.5/25) versus indacaterol 150mg + tiotropium 18mg (IND 150 + TIO 18), tiotropium 18mg + salmeterol 50mg (TIO 18 + SAL 50), tiotropium 18mg + formoterol 10mg or 12mg (TIO 18 + FOR 10), and indacaterol/glycopyrronium 110/50mg (QVA149).

4.3 Data

The search for relevant RCTs was performed on $14^{th} - 18^{th}$ April in 2014, and resulted in screening of 4720 registries and 3006 abstracts. Both abstract and full-text screening has been performed independently based on the PICOS criteria by two researchers, and the discrepancies in their decisions were evaluated by a third researcher independently, in order to avoid any bias in selection of the publications. The process of the systematic literature review is presented in the flow-chart below and resulted in 26 trials with 77 citations. Finally, for the outcome of interest FEV_1, 22 RCTs were identified to report it in total, 8 of them at 12 weeks, 4 at 24 weeks and 10 for both time points. Only one study was 3-arm study, the SHINE [40], all the remaining 21 were 2-arm study.

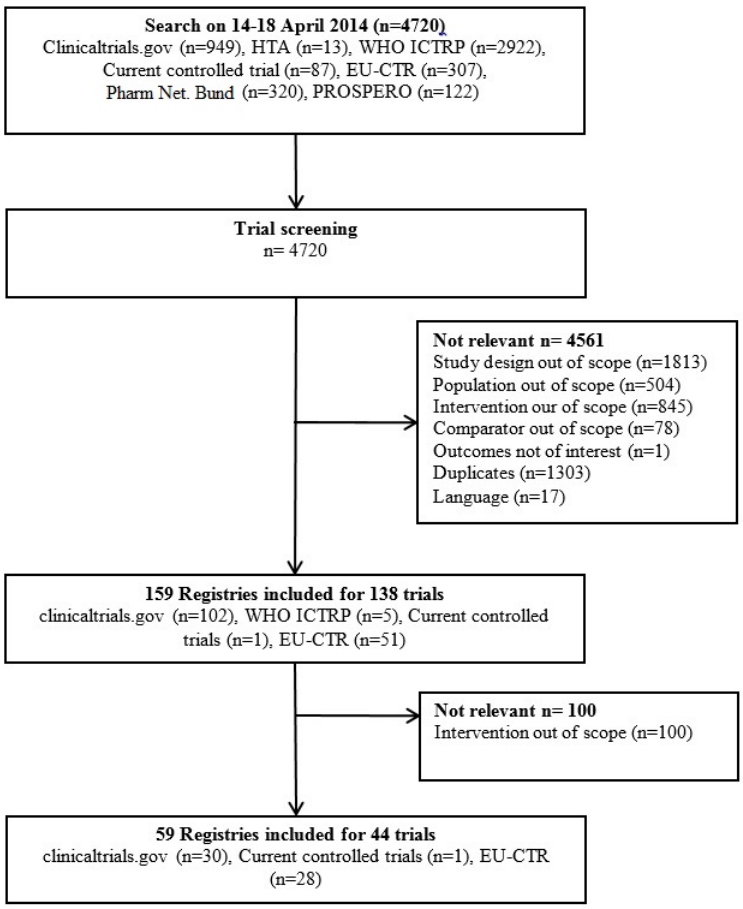


Figure 4.1: Flow chart of registries.

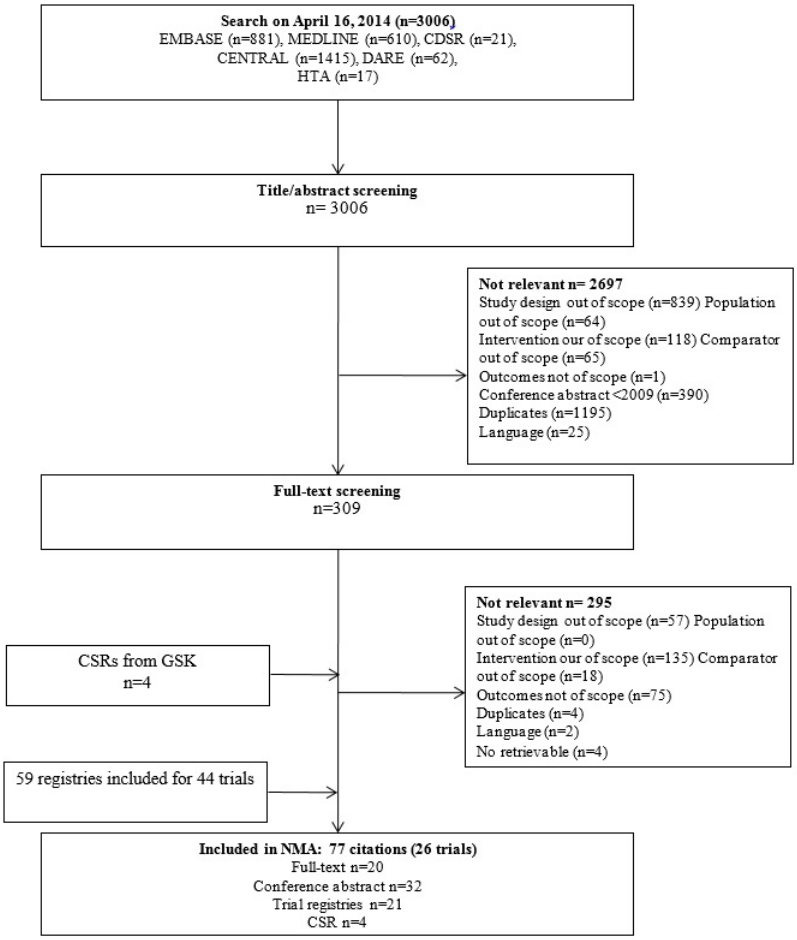


Figure 4.2: Flow chart of studies.

An overview of study characteristics is presented in Table 4.1. All studies were randomized, double blind, multicenter trials. Two studies, SPARK [41] and SHINE [40] included tiotropium 18mcg as an open label arm, the other arms of these trials were double-blind, DB2113374 [42], DB2113360 [43] and ZEP117115 [44] studies were double-blind, double-dummy trials that had trough FEV_1 as primary efficacy endpoint. DB2113373 [45] and SHINE [40] also had trough FEV_1 as primary objective. Tashkin 2009 [46] focused on the area under the curve for FEV_1 measured 0-4 hours after dosing as primary efficacy variable. The primary outcome of INTRUST 1 [47] and INTRUST 2 [47] was the standardized area under the curve of FEV_1 from 5 minutes to 8 hours post dose. The primary objective of SPARK [41] and the Aaron 2007 [48] were the exacerbation rate, and the ENLIGHTEN [49] study had the number of treatment emergent adverse events as primary endpoint. The SPARK [41] (64 weeks), Aaron 2007 [48] (52 weeks) and ENLIGHTEN [49] (52 weeks), Chan 2007 [50] (48 weeks), UPLIFT [51] (4 years), GLOW 2 [52] (52 weeks) studies were considerable longer than the other trials, that had a duration of 26 weeks (SHINE [40], Donohue 2010 [53]) or 24 weeks (DB2113360 [43],DB2113374 [42], DB2113373 [45], ZEP117115 [44], Niewoehner 2005 [54], Brusasco 2003 [55], Donohue 2002 [56]), 13 weeks (Cassaburi 2000 [57]), or 12 weeks (INTRUST 1 [47], INTRUST 2 [47],

Tashkin 2009 [46], Verkindre 2006 [58], Covelli 2005 [59], Moita 2008 [60]). All trials included patients with COPD older than 40 years and with a smoking history of more than 10 pack-years. The SPARK [41] study included more severe COPD patients than the other studies. SPARK included patients with stage III or IV COPD (severe or very severe COPD according to GOLD guidelines) with a FEV_1/FVC ratio of less than 70% and a FEV_1 of less than 50% of predicted normal values, and at least one exacerbation in the previous year requiring treatment with systemic corticosteroids, antibiotics or both. In contrast, the other studies included moderate or severe (stage II or III) COPD patients with a FEV_1/FVC ratio of less than 70% and a FEV_1 of less than 70% of predicted normal values, and no inclusion criterion related to the number of exacerbations in the previous year. The Table 4.2 presents the treatments,and the duration of each study included in the network as also the inclusion criteria and the background treatment . As a result of these different inclusion criteria, differences were observed in the SPARK [41] study compared with the other trials in COPD severity at baseline. For the same reason, the proportion of patients using ICS at baseline in the SPARK [41] study was higher (75%) than in other studies (52% on average). In addition, the FEV_1 percentage predicted was lower at baseline in the SPARK study (37.2% predicted) versus other studies (FEV_1 ranging from 42.1% predicted to 59.4% predicted across treatment arms). The other patient characteristics were homogeneous across trials. The percentage of males in each study arm ranged from 65% to 77.3% and the mean age was 61.9 to 65.0 years across treatment arms within the trials. Long acting bronchodilators had to be discontinued before entering all trials. The key patient characteristics are presented in Table 4.2. A summary statistics for the main patient characteristics for the studies included in the NMA reporting the trough FEV_1 values is presented in Table 4.3. A visualization of the mean trough FEV_1, age, treatment group and pack years of smoking at baseline is given in Figure 4.3, duo to not reported values 19 points is missing out of 45 in total.

Table 4.1: Key study characteristics at baseline for studies included in NMA, reporting FEV_1 value.

Study	Treatments	Trial Duration	Inclusion criteria	Background treatment
DB2113360 [43]	Tiotropium; 18mg; OD Vilanterol 25mg + Umeclidinium 62.5mg Vilanterol 25 mg + Umeclidinium 125mg	24 weeks	Outpatient; $\geq$ 40 years old; diagnosed with COPD, post-salbutamol $FEV_1 \leq 7\%$ and post-salbutamol FEV_1/FVC ratio < 0.7. Smoking history $\geq$10 pack-years	Allowed: ICS at a dose of up to 1000 mcg/day of FP or equivalent, salbutamol/albuterol as rescue Not allowed: LABAs, short acting β_2-agonists, short acting anticholinergics and SABA/ICS combination products

Continued on next page

Table 4.1 – *Continued from previous page*

Study	Treatments	Trial Duration	Inclusion criteria	Background treatment
DB2113374 [42]	Tiotropium; 18mg; OD Vilanterol 25mg + Umeclidinium 62.5mg Vilanterol 25 mg + Umeclidinium 125mg	24 weeks	Outpatient;≥40 years old; diagnosed with COPD; post-salbutamol FEV_1/FVC ratio of <0.70 and a post-salbutamol FEV_1 of ≤70%; Smoking history ≥10 pack-years	Allowed: ICS at a dose of up to 1000 mcg/day of FP or equivalent, salbutamol/albuterol as rescue Not allowed: LABAs, oral short acting and long acting β_2-agonists, inhaled short acting β_2-agonists, inhaled short acting anticholinergics and SABA/ICS combination products
DB2113373 [45]	Placebo Vilanterol 25mg + Umeclidinium 62.5mg	24 weeks	Outpatient; ≥40 years old; diagnosed with COPD; post-salbutamol FEV_1/FVC ratio of <0.70 and a post-salbutamol FEV_1 of ≤70%; Smoking history ≥ pack-years	Allowed: ICS at a dose of up to1000 mcg/day of FP or equivalent, salbutamol/albuterol as rescue Not allowed: LABAs, LABA/ICS combination products, short acting β_2-agonists, short acting anticholinergics and SABA/ICS combination products

Continued on next page

Table 4.1 – *Continued from previous page*

Study	Treatments	Trial Duration	Inclusion criteria	Background treatment
ZEP117115 [44]	Tiotropium; 18mg; OD Vilanterol 25mg + Umeclidinium 62.5mg	24 weeks	Outpatient; $\geq$ 40 years old; diagnosed with COPD; post-salbutamol FEV_1/FVC ratio of <0.70 and a post-salbutamol FEV_1 of $\leq$70%; Smoking history $\geq$ pack-years	Allowed: ICS at a dose of up to1000 mcg/day of FP or equivalent, salbutamol/albuterol as rescue Not allowed: LABAs, LABA/ICS combination products, oral short acting and long acting β_2-agonists, inhaled short acting β_2-agonists, inhaled short acting anticholinergics and SABA/ICS combination products
INTRUST 1 [47]	Tiotropium; 18mg; OD Indacaterol 150mg + Tiotropium; 18mg; OD	12 weeks	$\geq$ 40 years old; Post-bronchodilator FEV_1 $\leq$65% and $\geq$ 30%. Post-bronchodilator FEV_1/FVC < 70%. Smoking history $\geq$ 10 pack years	Allowed: ICS monotherapy, salbutamol/albuterol as rescue Not allowed: LABAs, short acting β_2-agonists (except those prescribed in the study), theophylline, anticholinergics

Continued on next page

Table 4.1 – *Continued from previous page*

Study	Treatments	Trial Duration	Inclusion criteria	Background treatment
INTRUST 2 [47]	Tiotropium; 18mg; OD Indacaterol 150mg + Tiotropium; 18mg; OD	12 weeks	$\geq$ 40 years old; Post-bronchodilator $FEV_1 \leq 65\%$ and $\geq 30\%$. Post-bronchodilator FEV_1/FVC $< 70\%$. Smoking history $\geq$ 10 pack years	Allowed: ICS monotherapy, salbutamol/albuterol as rescue. Not allowed: LABAs, short acting β_2-agonists (except those prescribed in the study), theophylline, anticholinergics
Aaron 2007 [48]	Tiotropium 18mg OD + salmeterol 25mg 2puffs BID Tiotropium 18mg OD + placebo 2 puffs BID Salmeterol/Fluticasone; 25/250mg/puff; two puffs BID + Tiotropium; 18mg; OD	52 weeks	$\geq$ 35 years old; diagnosis of moderate or severe COPD; $\geq$ 1 exacerbation of COPD requiring systemic steroids or antibiotics in previous 12 months; smoking history $\geq$ 10 pack years; Post-bronchodilator $FEV_1 \leq 65\%$; FEV_1/FVC $< 70\%$.	Allowed: albuterol for relief of symptoms Not allowed: ICS, LABA, anticholinergics

Continued on next page

Table 4.1 – *Continued from previous page*

Study	Treatments	Trial Duration	Inclusion criteria	Background treatment
ENLIGHTEN [49]	Placebo QVA149 (110 mg indacaterol/50 mg glycopyrronium); OD	52 weeks	$\geq$ 40 years old; diagnosis of moderate or severe COPD (stage II or III according to GOLD 2008 criteria); post-bronchodilator $FEV_1 < 80\%$ and $\geq 30\%$. Post-bronchodilator FEV_1/FVC < 0.70. Smoking history $\geq$ 10 pack years	Allowed: Albuterol as rescue medication, ICS monotherapy Not allowed: long-acting bronchodilators (LABA, LAMA, theophylline), short-acting muscarinic antagonists

Continued on next page

Table 4.1 – *Continued from previous page*

Study	Treatments	Trial Duration	Inclusion criteria	Background treatment
SPARK [41]	Tiotropium; 18mg; OD QVA149 (110 mg indacaterol/50 mg glycopyrronium); OD	64 weeks	$\geq$ 40 years old; diagnosis of severe or very severe COPD (stage III or IV according to GOLD 2008 criteria); post-bronchodilator $FEV_1 < 50\%$; FEV_1/FVC < 0.70; $\geq$1 exacerbation in the previous 12 months requiring systemic corticosteroids or antibiotics; smoking history $\geq$ 10 pack years	Allowed: salbutamol, stable dose of ICS Not allowed: long-acting bronchodilators

Continued on next page

Table 4.1 – *Continued from previous page*

Study	Treatments	Trial Duration	Inclusion criteria	Background treatment
SHINE [40]	Placebo Tiotropium; 18mg; OD QVA149 (110 mg indacaterol/50 mg glycopyrronium); OD	26 weeks	$\geq$ 40 years old; diagnosis of moderate or severe COPD (stage II or III according to GOLD 2008 criteria); post-bronchodilator FEV_1 < 80% and $\geq$ 30%. Post-bronchodilator FEV_1/FVC < 0.70. Smoking history $\geq$ 10 pack years	Allowed: Salbutamol/albuterol as rescue medication, inhaled or intranasal corticosteroids in constant doses Not allowed: LABA, LAMA, LABA/ICS
Tashkin 2009 [46]	Tiotropium 18mg OD + Formoterol 12mg BID Tiotropium bromide 18mg OD + Placebo BID	12 weeks	aged $\geq$40 years; post-bronchodilator FEV_1 <70% and >30% predicted normal or >0.75 L, whichever was less, at run-in; FEV_1/FVC<0.70	Continued use of prior stable ICS regimens and systemic corticosteroids for the treatment of exacerbations was permitted throughout the study. All patients were provided with albuterol inhalers for use as rescue medication

Continued on next page

Table 4.1 – *Continued from previous page*

Study	Treatments	Trial Duration	Inclusion criteria	Background treatment
Chan 2007 [50]	Tiotropium 18mg; OD Placebo	48 weeks	$\geq$ 40 years old; $\geq$10 pack-years; $FEV_1 \leq 65\%$; FEV_1/FVC $\leq$ 70%; included if $\geq$ 1 exacerbation previous yr but not in 6 wks prior (later amended to incl 1 exacerbation in past 2 yrs)	Allowed: Stable dose oral corticosteroids, ICS, theophylline preparations, mucolytic preparations (not containing bronchodilators), LABAs
UPLIFT [51]	Tiotropium 18mg; OD Placebo	4 years	$\geq$ 40 years old; >10 pack-years; $FEV_1 \leq 70\%$;FEV_1/FVC $\leq$ 70%; excluded if exacerbation 4 wks prior	Allowed: All respiratory medications, except other inhaled anticholinergic drugs
Niewoehner 2005 [54]	Tiotropium 18mg; OD Placebo	6 months	$\geq$40 years old; $\geq$10 pack-years; $FEV_1 \leq 60\%$; FEV_1/FVC $\leq$ 70%;excluded if not recovered from exacerbation $\geq$ 30 days prior	Allowed: All other respiratory medications (including ICS and LABAs) Not Allowed: Open-label anticholinergic bronchodilator
Brusasco 2003 [55]	Tiotropium 18mg; OD Placebo	24 weeks	>40 years old; >10 pack-years; $FEV_1 \leq 65\%$; FEV_1/FVC $\leq$ 70%	-

Continued on next page

Table 4.1 – *Continued from previous page*

Study	Treatments	Trial Duration	Inclusion criteria	Background treatment
Donohue 2002 [56]	Tiotropium 18mg; OD Placebo	24 weeks	≥40 years old; >10 pack-years; FEV_1 ≤ 60%;FEV_1/FVC ≤ 70%	Allowed: Usual ICS and oral steroids; Not Allowed: Inhaled anticholinergic LABAs
Donohue 2010 [53]	Tiotropium 18mg; OD Placebo	26 weeks	Patients aged 40 years or older with a smoking history of 20 pack-years or more and a diagnosis of moderate-to-severe COPD (GOLD criteria) were enrolled. Post-bronchodilator (within 30 min of inhaling albuterol 360 mg) forced expiratory volume in 1 second (FEV_1) <80% and ≥30% predicted and FEV_1/forced vital capacity (FVC) <70%.	Patients could continue inhaled corticosteroid (ICS) monotherapy if stable for 1 month before screening; dose and regimen were to remain stable throughout the study. Before the start of the run-in period, treatment with anticholinergic bronchodilators or with β_2-agonists was discontinued with appropriate washout, and patients receiving fixed-combination β_2-agonist/ICS were switched to ICS monotherapy at an equivalent dose. All patients were supplied with albuterol for use as needed.

Continued on next page

Table 4.1 – *Continued from previous page*

Study	Treatments	Trial Duration	Inclusion criteria	Background treatment
GLOW 2 ([52]	Glycopyrronium 50mg OD Tiotropium 18mg; OD Placebo	52 weeks	Males and females ≥40 years, with a smoking history of ≥10 pack-yrs, a diagnosis of moderate-to-severe stable COPD, post-bronchodilator FEV_1 ≥30% and <80% of the predicted normal, and post-bronchodilator FEV_1/FVC<0.70 were enrolled.	Allowed: inhaled or intranasal corticosteroids and H1-antagonists; Salbutamol/albuterol as rescue medication. Not allowed: LAMAs (min 7 days before run-in); LABAs or LABA/ICS combinations (min 48h before run-in).
Verkindre 2006 [58]	Tiotropium 18mg; OD Placebo	12 weeks	FEV_1 ≤ 50%;FEV_1/SVC≤ 70%;residual volume ≥ 125%; excluded if unstable doses oral corticosteroid 6 wks prior	Allowed: Stable doses oral corticosteroids, ICS, theophylline preparations, mucolytic agents; Not Allowed: Use of SABAs, oral ß2-agonists, or LABAs
Casaburi 2000 [57]	Tiotropium 18mg; OD Placebo	13 weeks	FEV_1 ≤ 65%;FEV_1/FVC ≤ 70%; ≥ 40 years of age; diagnosis of COPD defined by ATS; smoking history of >10 pack-years	Allowed: stable doses of theophylline, ICS, oral prednisone Not Allowed: Other inhaled or oral bronchodilators

Continued on next page

Table 4.1 – *Continued from previous page*

Study	Treatments	Trial Duration	Inclusion criteria	Background treatment
Covelli 2005 [59]	Tiotropium 18mg; OD Placebo	12 weeks	$FEV_1 \leq 60\%$;FEV_1/FVC $\leq 70\%$;excluded if exacerbation in prior 6 wks	Allowed: ICS, LABAs and theophyllines Not Allowed: Cromones, leukotriene antagonists, and inhaled anticholinergics
Moita 2008 [60]	Tiotropium 18mg; OD Placebo	12 weeks	$FEV_1 \leq 70\%$;FEV_1/FVC $\leq 70\%$;excluded if ≥ 3 exacerbations previous year or exacerbation in 6wks prior	Allowed: LABAs, theophylline, mucolytics, ICS, stable doses oral corticosteroids. Temporary increases in theophylline or oral steroids for exacerbations; Not Allowed: theophylline 24 h preparation

Table 4.2: Key patient characteristics at baseline for studies included in NMA, reporting FEV_1 value.

Study	Treatments	ITT	Male (%)	Age (sd)	Current smokers (%)	Severe or very severe (%)	ICS use (%)	COPD Duration mean (sd)	Pack years (sd)	FEV_1 % predicted (sd)
DB2113360 [43]	Tiotropium; 18mg; OD	208	67	62.6 (9.39)	48	53	45	NR	41.9 (24.44)	47.8 (13.36)
	Vilanterol 25mg + Umeclidinium 62.5mg	212	70	63 (8.67)	46	50	44	NR	44.8 (27.65)	48.0 (12.94)
	Vilanterol 25 mg + Umeclidinium 125mg	214	71	62.9 (8.87)	58	53	48	NR	43.5 (24.98)	47.2 (12.79)
DB21133747 [42]	Tiotropium; 18mg; OD	215	71	65.2 (8.3)	47	52	53	NR	54.0 (31.59)	47.4 (13.10)
	Vilanterol 25mg + Umeclidinium 62.5mg	217	65	65 (8.62)	42	51	47	NR	47.8 (26.13)	47.7 (13.55)
	Vilanterol 25 mg + Umeclidinium 125mg	215	69	63.8 (8.51)	45	59	53	NR	46.9 (24.90)	47.1 (12.88)
DB2113373 [45]	Placebo	280	70	62.2 (9.04)	54	58	49	NR	47.2 (27.21)	46.7 (12.71)

Continued on next page

Table 4.2 – *Continued from previous page*

Study	Treatments	ITT	Male (%)	Age (sd)	Current smokers (%)	Severe or very severe (%)	ICS use (%)	COPD Duration mean (sd)	Pack years (sd)	FEV_1 % predicted
	Vilanterol 25mg + Umeclidinium 62.5mg	413	74	63.1 (8.71)	49	51	51	NR	46.5 (25.80)	47.8 (13.19)
ZEP117115 [44]	Tiotropium; 18mg; OD	451	67	62.7 (8.50)	54	58	53	NR	44.4 (25.03)	46.5 (12.76)
	Vilanterol 25mg + Umeclidinium 62.5mg	454	68	61.9 (8.41)	59	60	54	NR	44.1 (24.44)	46.2 (13.02)
INTRUST 1 [47]	Tiotropium; 18mg; OD	564	67	63.4 (9.22)	36	53	52	6.6 (6.45)	47.2 (26.58)	48.9 (11.46)
	Indacaterol; 150mg; OD + Tiotropium; 18mg; OD	570	70	64.0 (9.07)	40	53	52	7.1 (6.12)	47.2 (25.86)	48.3 (9.70)
INTRUST 2 [47]	Tiotropium; 18mg; OD	570	68	62.8 (8.98)	43	54	51	7.1 (6.26)	46.3 (24.64)	48.6 (9.76)
	Indacaterol; 150mg; OD + Tiotropium; 18mg; OD	572	63	63.1 (8.83)	38	54	57	7.3 (6.48)	46.2 (25.52)	48.6 (9.74)
Aaron 2007 [48]	Tiotropium; 18mg; OD	156	53.8	68.1 (8.9)	27	NR	25	11.3 (8.8)	51.8 (28.0)	42.1 (13.5)

Continued on next page

Table 4.2 – *Continued from previous page*

Study	Treatments	ITT	Male (%)	Age (sd)	Current smokers (%)	Severe or very severe (%)	ICS use (%)	COPD Duration mean (sd)	Pack years (sd)	FEV_1 % predicted
	Tiotropium 18mg OD + salmeterol 25mg 2puffs BID	148	57.4	67.6 (8.2)	24.3	NR	34.9	NR	48.7 (27.1)	41.2 (13.0)
	Salmeterol/ Fluticasone; 25/250mg/puff; two puffs BID + Tiotropium; 18mg; OD	145	57.9	67.5 (8.9)	32	NR	27	10.3 (8.1)	50.3 (23.1)	42.2 (12.2)
ENLIGHTEN [49]	Placebo	113	76.1	62.9 (8.14)	45	19	39	5.46 (5.1)	38.1 (15.93)	59.43 (12.5)
	QVA149 (110 mg indacaterol/50 mg glycopyrronium); OD	226	77.3	62.5 (8.81)	45	31	46	5.82 (5.74)	36.3 (16.01)	56.39 (13.27)
SPARK [41]	Tiotropium 18mg; OD	742	75	63.6 (7.8)	37	100c	76	7.2 (5.5)	47 (28)	37.4 (8.1)

Continued on next page

Table 4.2 – *Continued from previous page*

Study	Treatments	ITT	Male (%)	Age (sd)	Current smokers (%)	Severe or very severe (%)	ICS use (%)	COPD Duration mean (sd)	Pack years (sd)	FEV_1 % predicted
	QVA149 (110 mg indacaterol/50 mg glycopyrronium); OD	741	76	63.1 (8.1)	38	100c	75	7.2 (5.8)	45 (23)	37.0 (8.1)
SHINE [40]	Placebo	234f	72.8	64.4 (8.6)	40	32	58	6.4 (5.7)	NR	55.2 (12.7)
	Tiotropium 18mg; OD	483	75.0	63.5 (8.7)	39	38	59	6.1 (5.5)	NR	55.1 (13.5)
	QVA149 (110 mg indacaterol/50 mg glycopyrronium); OD	475f	76.4	64.0 (8.9)	40	34	56	6.0 (5.5)	NR	55.7 (13.2)
Tashkin 2009 [46]	Tiotropium 18mg OD + Formoterol 12mg BID	124	65	63.8 (8.7)	49	NR	27	NR	NR	NR

Continued on next page

Table 4.2 – *Continued from previous page*

Study	Treatments	ITT	Male (%)	Age (sd)	Current smokers (%)	Severe or very severe (%)	ICS use (%)	COPD Duration mean (sd)	Pack years (sd)	FEV_1 % predicted
	Tiotropium bromide 18mg OD + Placebo BID	131	68	63.9 (8.5)	46	NR	27	NR	NR	NR
Chan 2007 [50]	Tiotropium; 18mg; OD	608	59	67.0 (8.7)	32	NR	66	9.9 (8.1)	50.2 (22.6)	0.39 (0.13)
	Placebo	305	61	67.0 (9.1)	30	NR	71	9.9 (7.9)	51.0 (26.3)	0.39 (0.14)
UPLIFT [51]	Tiotropium; 18mg; OD	2987	75	65.0 (8.4)	29	52	62	9.9 (7.6)	49.0 (28.0)	0.40 (0.12)
	Placebo	3006	74	65.0 (8.5)	30	53	62	9.7 (7.4)	48.4 (27.9)	0.39 (0.12)
Niewoehner 2005 [54]	Tiotropium; 18mg; OD	914	98	67.6 (8.7)	29	NR	61	12.2 (10.4)	67.4 (35.4)	0.36 (0.13)
	Placebo	915	99	68.1 (8.5)	30	NR	58	11.9 (10.5)	69.4 (36.6)	0.36 (0.13)
Brusasco 2003 [55]	Tiotropium; 18mg; OD	402	77	63.8 (8.0)	NR	NR	NR	9.0 (7.3)	44.1 (22.9)	0.39 (0.12)
	Placebo	400	76	64.6 (8.6)	NR	NR	NR	9.8 (7.4)	42.4 (22.7)	0.39 (0.12)
Donohue 2002 [56]	Tiotropium; 18mg; OD	209	74	64.5 (7.9)	NR	NR	66	9.2 (7.8)	47.0 (25.0)	0.41 (NR)
	Placebo	201	75	65.6 (7.8)	NR	NR	66	9.7 (7.9)	46.0 (24.0)	0.41 (NR)

Continued on next page

Table 4.2 – *Continued from previous page*

Study	Treatments	ITT	Male (%)	Age (sd)	Current smokers (%)	Severe or very severe (%)	ICS use (%)	COPD Duration mean (sd)	Pack years (sd)	FEV_1 % predicted
Donohue 2010 [53]	Tiotropium; 18mg; OD	420	65	64 (8.8)	NR	NR	35	NR	50.0 (25.1)	0.54 (0.16)
	Placebo	425	61	63.6 (8.9)	NR	NR	40	NR	49.7 (23.9)	0.56 (0.14)
GLOW2 [52]	Tiotropium 18mg OD	267	63	63.9 (8.2)	44	NR	52	7.5 (6.6)	50.2 (28.0)	0.56 (0.13)
	Placebo	268	65	63.6 (9.1)	46	NR	51	7.4 (6.6)	48.0 (24.0)	0.56 (0.14)
Verkindre 2006 [58]	Tiotropium; 18mg; OD	46	94	61.0 (9.5)	24	NR	NR	9.7 (6.9)	45.6 (23.1)	0.35 (0.09)
	Placebo	54	94	60.0 (10.2)	33	NR	NR	8.8 (6.6)	41.8 (18.0)	0.36 (0.09)
Casaburi 2000 [57]	Tiotropium; 18mg; OD	276	67	65.0 (8.6)	NR	NR	NR	9.3 (8.0)	64.5 (33.1)	0.39 (0.14)
	Placebo	188	63	65.0 (9.0)	NR	NR	NR	8.6 (6.9)	60.5 (30.2)	0.38 (0.14)
Covelli 2005 [59]	Tiotropium; 18mg; OD	94	66	66.0 (8.9)	40	NR	54	10.1 (8.1)	66 (35.6)	0.40 (0.13)
	Placebo	84	49	63.0 (9.2)	37	NR	58	10.4 (7.7)	65 (31.2)	0.39 (0.14)
Moita 2008 [60]	Tiotropium; 18mg; OD	147	NR	NR	28	NR	NR	NR	NR	NR
	Placebo	164	NR	NR	25	NR	NR	NR	NR	NR

Table 4.3: Summary statistics for the main patient characteristics at baseline for studies included in NMA, reporting FEV_1 value.

Variables	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
N	45											
Male (percentage)	43	0.71	0.10	0.70	0.70	0.07	0.49	0.99	0.50	0.90	1.16	0.02
Age (years)	43	64.18	1.80	63.90	64.09	1.63	60.00	68.10	8.10	0.43	0.02	0.28
Current smoker (percentage)	37	0.39	0.09	0.40	0.39	0.10	0.24	0.59	0.35	0.08	-0.87	0.01
ICS (percentage)	37	0.52	0.12	0.53	0.53	0.10	0.25	0.76	0.51	-0.35	-0.26	0.02
Duration COPD (percentage)	29	8.43	1.84	8.80	8.38	2.08	5.00	12.20	7.20	0.09	-0.97	0.34
Packyears	38	49.52	7.89	47.20	48.84	4.15	36.00	69.40	33.40	1.10	0.54	1.28
FEV_1 mean(mL)	29	1165.21	149.95	1110.00	1156.04	130.47	960.00	1510.00	550.00	0.57	-0.80	27.85
BDI	14	6.19	2.43	6.10	6.08	0.62	2.40	11.30	8.90	0.57	0.09	0.65

Table 4.4: Difference in CFB for trough FEV_1(SE) in mL at 12 and 24 weeks.

Study	Treatments	Weeks	Difference in CFB of trough FEV_1 in mL (SE)
DB2113360 [43]	UMEC/VI 62.5/25 vs TIO 18	12	80 (24.49)
		24	90 (26.02)
DB2113374[42]	UMEC/VI 62.5/25 vs TIO 18	12	95 (21.94)
		24	60 (25.26)
DB2113373 [45]	UMEC/VI 62.5/25 vs Placebo	12	195 (17.86)
		24	167 (20.15)
ZEP117115 [44]	UMEC/VI 62.5/25 vs TIO 18	12	109 (15.82)
		24	112 (16.07)
INTRUST 1 [47]	TIO 18 + IND 150	12	80 (12.76)
INTRUST 2 [47]	TIO 18 + IND 150	12	70 (10.20)
Aaron 2007 [48]	TIO 18 + SAL 50 vs TIO 18	24	18.49 (45.46)
ENLIGHTEN [49]	QVA 149 vs Placebo	12	163 (32.02)
	QVA 149 vs Placebo	24	152 (35.36)
SPARK [41]	QVA 149 vs TIO 18	12	70 (13.79)
	QVA 149 vs TIO 18	24	70 (13.79)
SHINE [40]	QVA 149 vs Placebo	12	230 (17.86)
		24	200 (17.86)
	TIO 18 vs Placebo	12	130(17.86)
		24	130(17.86)
	QVA 149 vs TIO 18	12	100 (17.86)
		24	70 (17.86)
Tashkin 2009 [46]	TIO 18 + FOR 12	12	90(28.06)
Chan 2007 [50]	TIO 18 vs Placebo	12	100(15.00)
UPLIFT [51]	TIO 18 vs Placebo	24	100 (7.00)
Niewoehner 2005 [54]	TIO 18 vs Placebo	12	100 (10.00)
	TIO 18 vs Placebo	24	100 (13.00)
Brusasco 2003 [55]	TIO 18 vs Placebo	24	120 (100.00)
Donohue 2002 [56]	TIO 18 vs Placebo	24	137 (20.00)

Continued on next page

Table 4.4 – *Continued from previous page*

Study	Treatments	Weeks	Difference in CFB of trough FEV_1 in mL (SE)
Donohue 2010	[53] TIO 18 vs Placebo	12	140 (20.41)
	TIO 18 vs Placebo	24	140 (20.41)
GLOW 2 [52]	TIO 18 vs Placebo	12	83(19.00)
	TIO 18 vs Placebo	24	84(21.60)
Verkindre 2006 [58]	TIO 18 vs Placebo	12	110(40.00)
Casaburi 2000 [57]	TIO 18 vs Placebo	12	150 (14.00)
Covelli 2005 [59]	TIO 18 vs Placebo	12	184 (37.00)
Moita 2008 [60]	TIO 18 vs Placebo	12	102 (31.38)

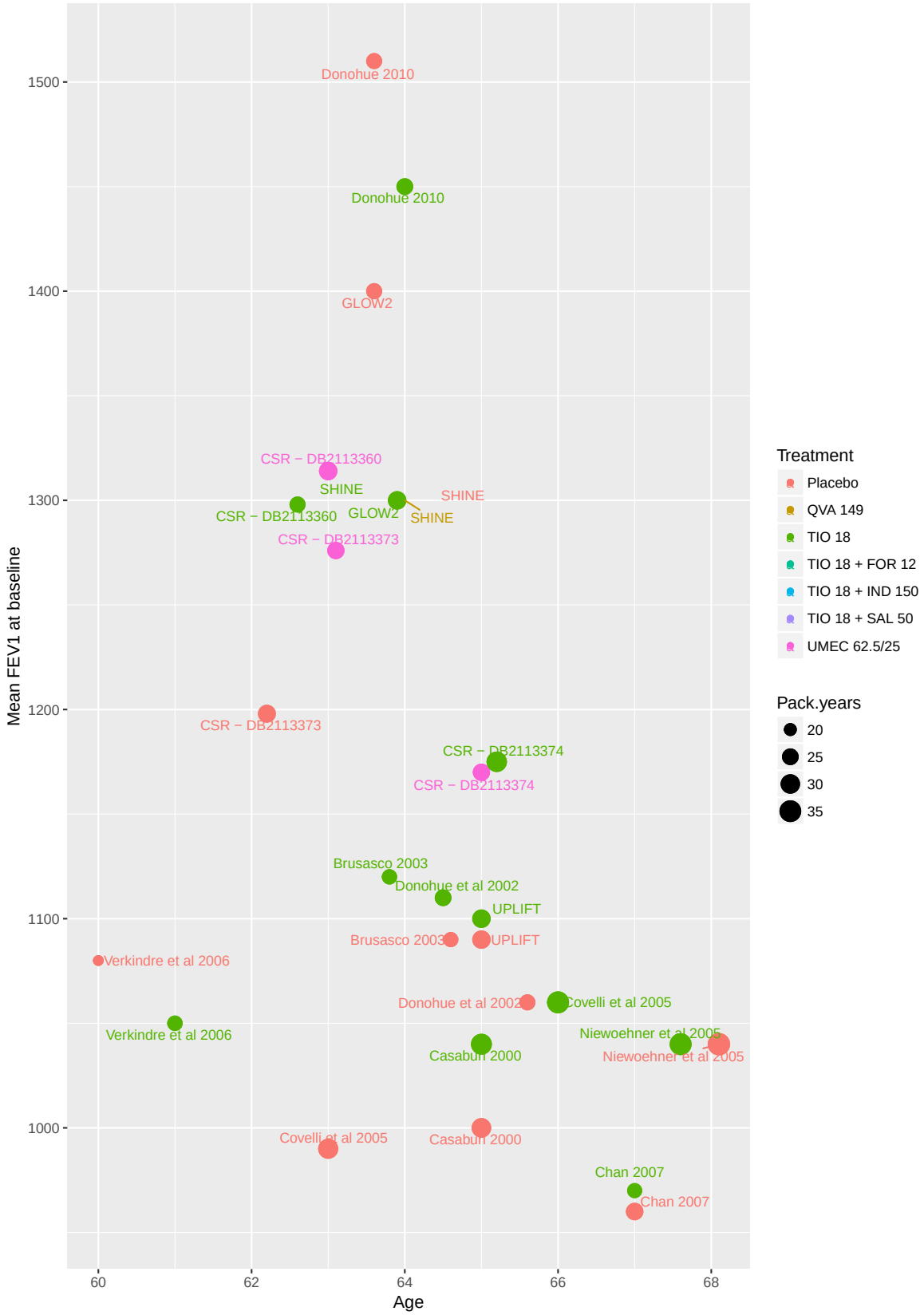


Figure 4.3: Plot of the main patient characteristics at baseline, mean FEV_1 in mL, age, treatment and pack years in each arm of the studies reporting these characteristics. Out of 45 values, 26 were reported.

4.4 Networks of evidence

According to the evidence available a network of connected treatments for trough FEV_1 at 12 weeks, Figure 4.4, and at 24 weeks Figure 4.5 is presented below. The size of its treatment's circle indicates its precision, and the width of the lines connecting treatments represents the amount of studies comparing them, as the line is thicker more studies had examined these treatments, strengthening the precision of the pooled effect estimation between them. At 12 weeks the network of evidence include 18 studies(4.4), 17 were 2-arm study and only one was a 3-arm study (SHINE [40]). Treatments connected was the QVA149, TIO 18 + IND 150, TIO 18 + FOR 12, and the treatment of interest UMEC/VI 62.5/25, placebo and TIO 18 monotherapy were included to the network in order to make the comparison of dual therapies LABA/LAMA feasible. The network of evidence for trough FEV_1 at 24 weeks contains 14 studies, again only SHINE study had three arms and the rest were 2-arm studies. The treatments included are the UMEC/VI 62.5/25 as the treatment of interest, placebo and TIO 18 monotherapy were included to make the comparison with the two LABA/LAMA treatments TIO 18 + SAL 50 and QVA 149 feasible.

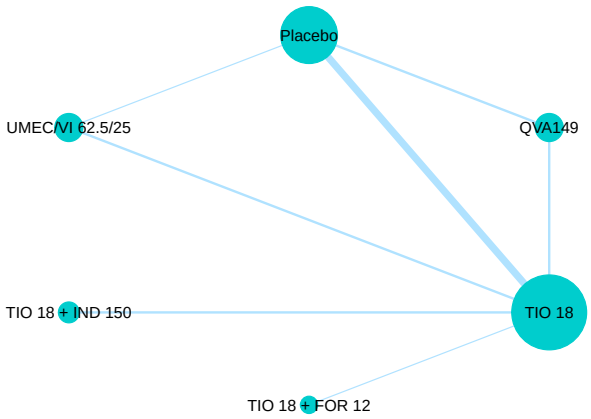


Figure 4.4: Network of evidence for trough FEV_1 at 12 weeks.

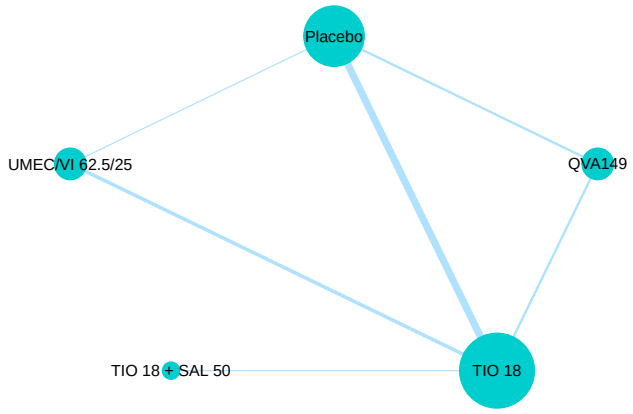


Figure 4.5: Network of evidence for trough FEV_1 at 24 weeks.

4.5 Statistical Analysis

This dissertation is part of an internship at Mapi Group, therefore the actual codes are not possible to be published as they are confidential. The frequentist meta-analyses and the corresponding figures for the examples were implemented in R using the labralies meta [61] and metafor [62]. The Bayesian meta-analyses were performed using R2OpenBUGS [63] and coda [64] libraries. The frequentist network meta-analysis was based on the graph theoretical approach [65], applied in netmeta package [66]. The Bayesian network meta-analyses were implemented in R using the models performed by NICE DSU via openBUGS [64], the only case where the code related to the implementation had been provided by Mapi. All figures for the examples and application were performed in R by the ggplot2 package [67].

4.6 Frequentist Approach

A frequentist NMA was performed for the network of evidence of the continuous outcome trough FEV_1, at 12 weeks and 24 weeks as described in the section 4.4. Both fixed and random effects models were performed duo to competence of the application. Two forest plots are given, depicting the comparison of treatments refereed to placebo, and UMEC/VI 62.5/25 for each model respectively. The forest plots were decided to be presented instead of the tables as the diversification among the included treatments is directly detected. The models were in general in line, as it was expected, with the random effects estimations having in all cases

more uncertainty around them, duo to the one additional source of variability. The analyses where conducted in R with the use of the netmeta package [66], from which the evidence networks plots were derived (Fugure 4.4 and Figure 4.5).

All therapies in the network of trough FEV_1 at 12 weeks proven to be more efficacious than placebo in both models(Figure 4.6 and Figure 4.10). The UMEC/VI 62.5/25 therapy has stricken the best performance among the other treatments in both fixed and random effects models, contrary to the TIO 18 therapy which performed worst. The comparison of UMEC/VI 62.5/25 therapy with the other treatments showed that it was comparable to LABA+/LAMA therapies (QVA149, TIO 18 + FOR 12, and TIO 18 + IND 150) for both models (Figure 4.7 and Figure 4.11). More analytically, the UMEC/VI 62.5/25 therapy compared to LABA+/LAMA interventions for both fixed and random effects showed numerically higher values of trough FEV_1 values at 12 weeks, however the results were not consider to be statistically significant. Nevertheless, the treatment of interest, UMEC/VI 62.5/25, had statistically better performance compared with the placebo and TIO 18 monotherapy.

The results for trough FEV_1 at 24 weeks of fixed and random effects were approximately the same, while again the random effects estimations give larger CI as always. The QVA149, TIO 18 and UMEC/VI 62.5/25 were proven to be statistically efficacious compared with the placebo, TIO 18 + SAL 50 was numerically grater, however not statistically significant (Figure 4.8 and Figure 4.12). The TIO 18 + SAL 50 showed the largest CI as only one study ([48]) was identified to report the outcome of interest at 24 weeks, comparing it with the TIO 18. Despite the results at 12 weeks, the trough FEV_1 values at 24 weeks in both models shown that only the QVA149 dual therapy was comparable to UMEC/VI 62.5/25 therapy, for the rest of the treatments UMEC/VI 62.5/25 performed statistically better (Figure 4.9 and Figure 4.13). The comparison of the TIO 18 + SAL 50 with the UMEC/VI 62.5/50 reported the largest CIs as they are compared in only Aaron study [48], which had the highest SE (Table 4.4) of all the studies included in the analysis.

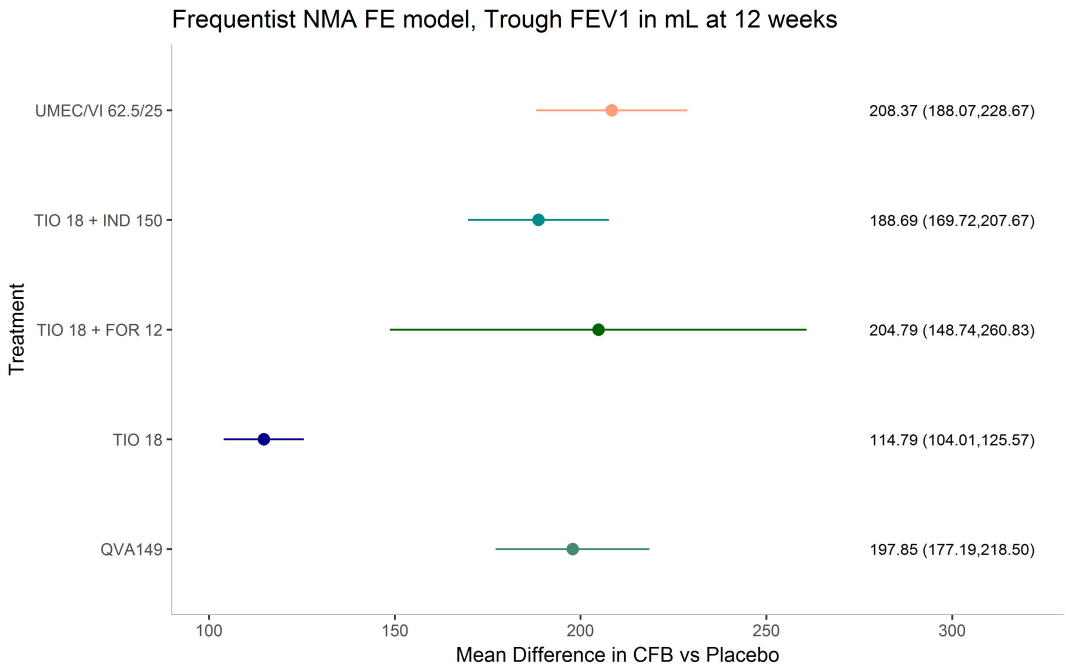


Figure 4.6: Frequentist fixed effect NMA forest plot of trough FEV_1 at 12 weeks using placebo as reference treatment.

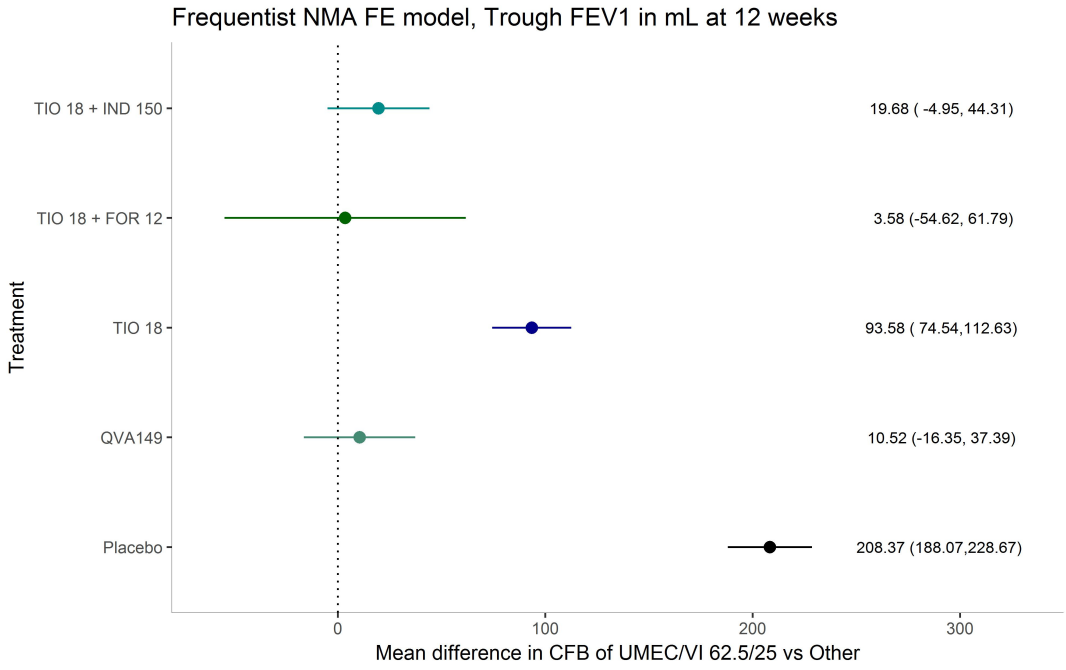


Figure 4.7: Frequentist fixed effect NMA forest plot of trough FEV_1 at 12 weeks using UMEC/VI 62.5/25 as reference treatment.

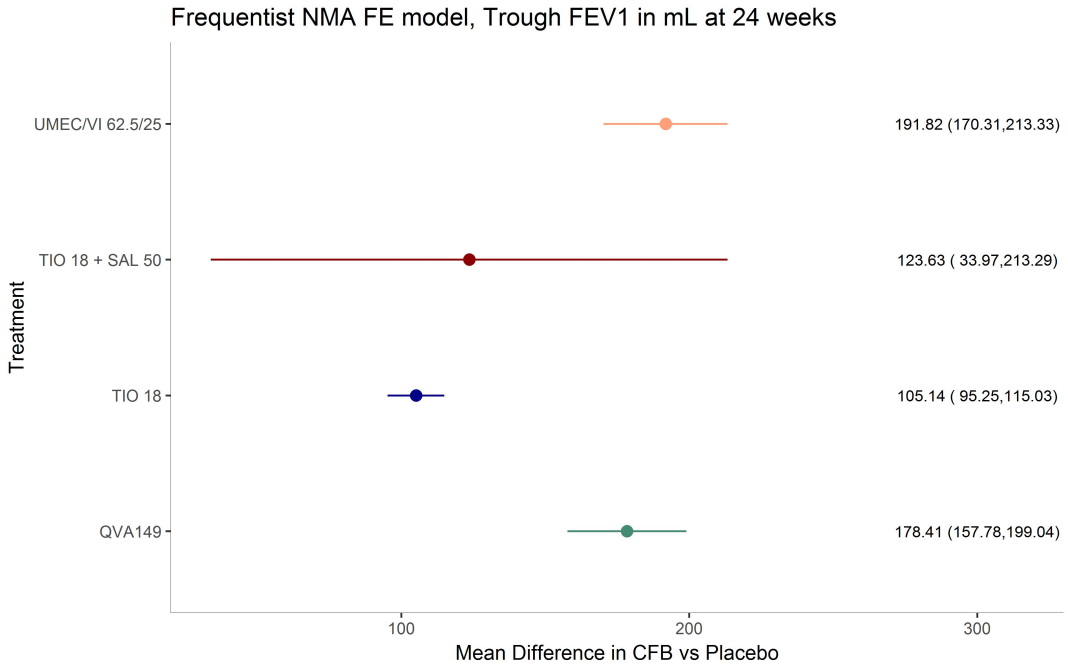


Figure 4.8: Frequentist fixed effect NMA forest plot of trough FEV_1 at 24 weeks using placebo as reference treatment.

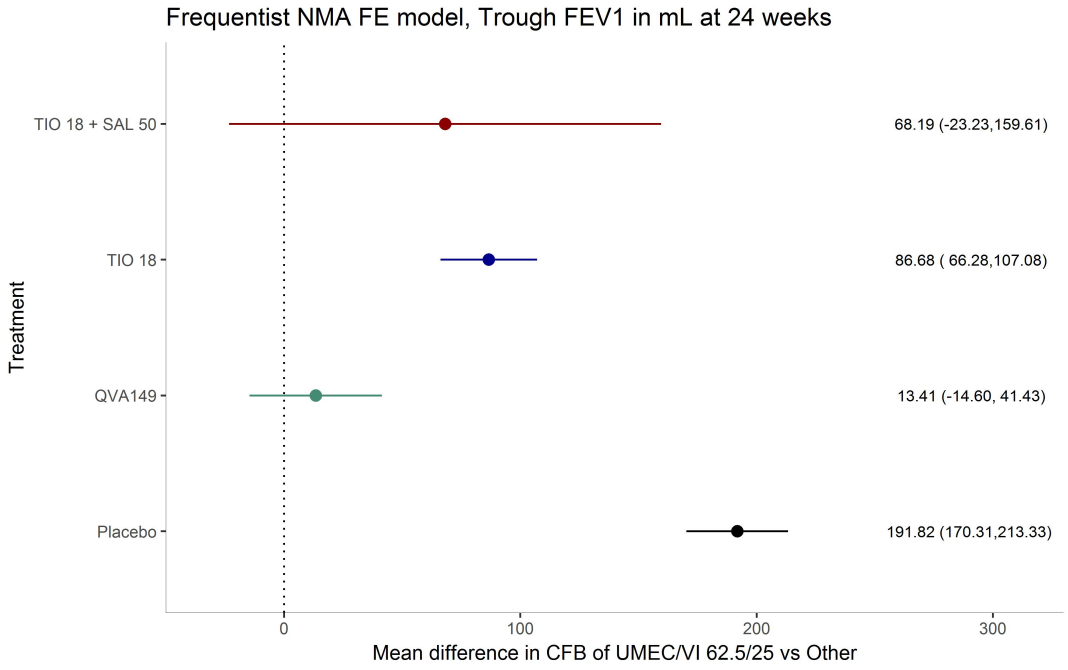


Figure 4.9: Frequentist fixed effect NMA forest plot of trough FEV_1 at 24 weeks using UMEC/VI 62.5/25 as reference treatment

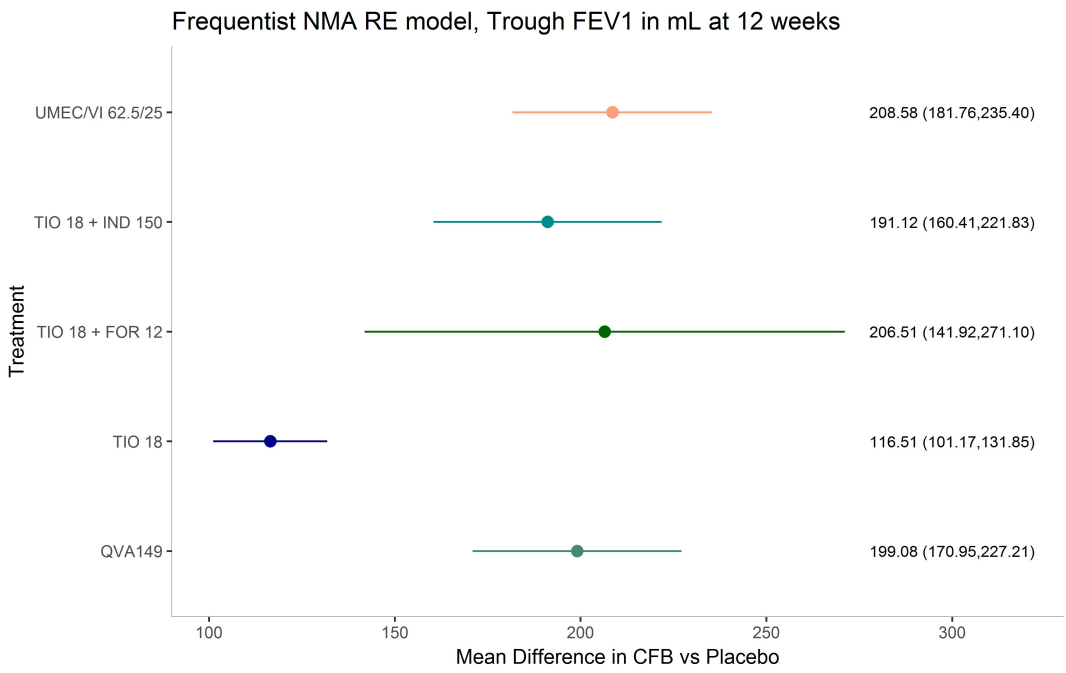


Figure 4.10: Frequentist random effects NMA forest plot of trough FEV_1 at 12 weeks using placebo as reference treatment.

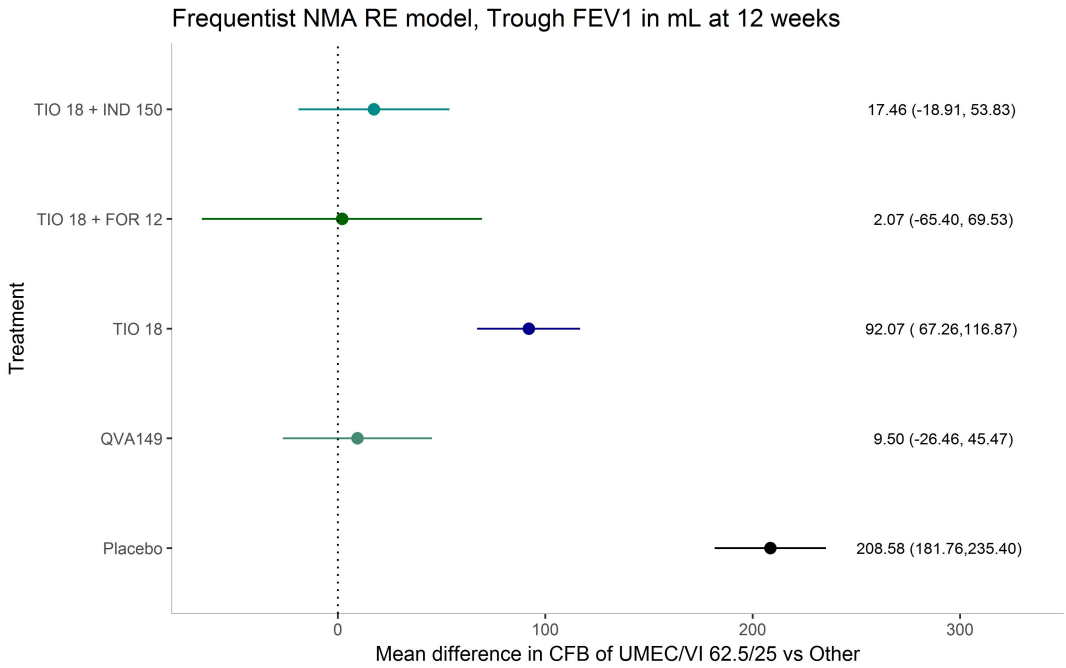


Figure 4.11: Frequentist random effects NMA forest plot of trough FEV_1 at 12 weeks using UMEC/VI 62.5/25 as reference treatment.

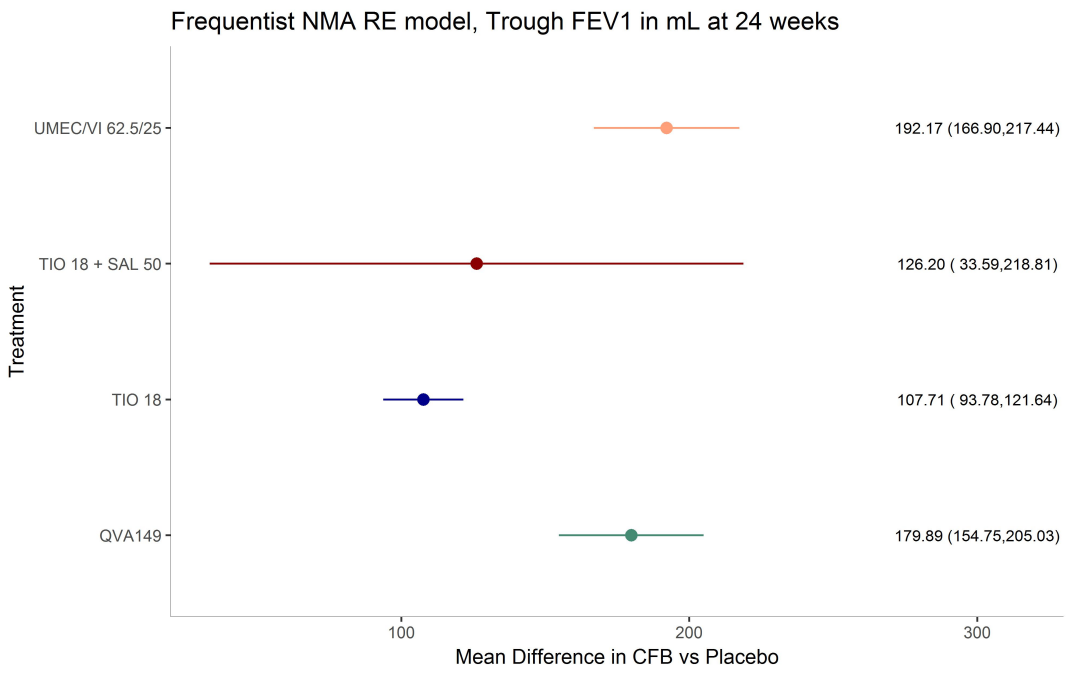


Figure 4.12: Frequentist randon effect NMA forest plot of trough FEV_1 at 24 weeks using placebo as reference treatment.

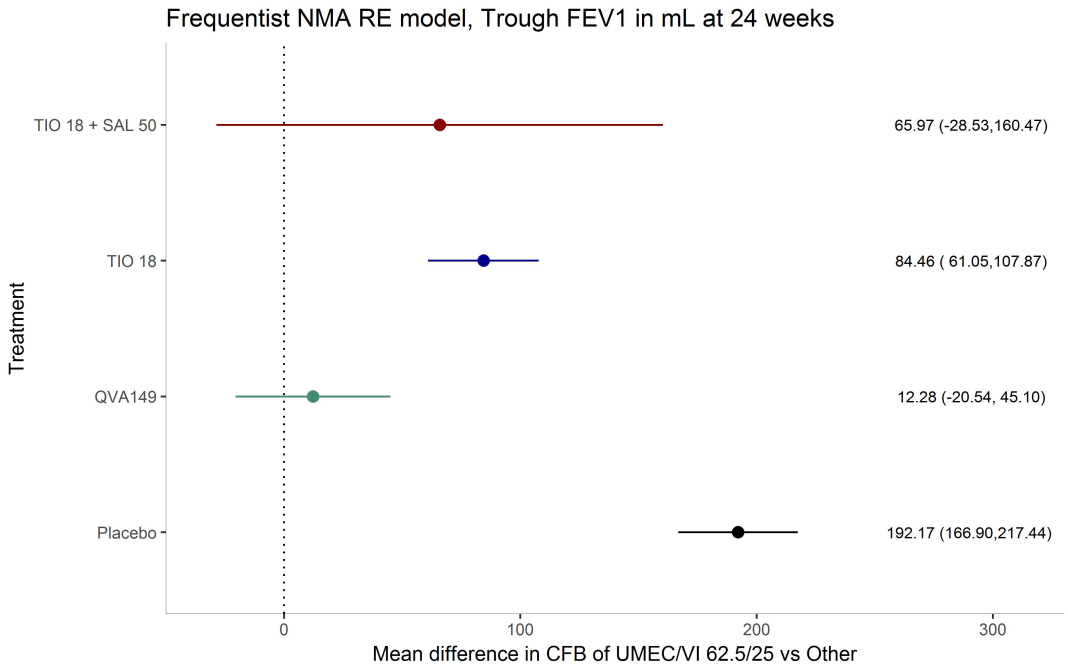


Figure 4.13: Frequentist random effects NMA forest plot of trough FEV_1 at 24 weeks using UMEC/VI 62.5/25 as reference treatment

4.7 Bayesian Approach

The analyses of trough FEV_1 at 12 and 24 weeks have been performed from the Bayesian perspective to serve the completeness of the analysis. As a general observation, the two frameworks, frequentist and Bayesian, present approximately same results, which was the main question to answer of this application. All estimations at two time points had same direction, leading eventually to the same decisions as in the frequentist perspective. Due to consistence, the presentation of the results is the same as in the frequentist approach. Two forest plots have been presented for each time point for each model, fixed and random effects, the first illustrates the comparison of the treatments with the placebo and the second one compares the treatment of interest, UMEC/VI 62.5/25, with the rest of therapies included in the network. The fixed and random effects models were in line with the corresponding models given from the frequentist framework, the CrIs were approximately the same and even identical in some estimations e.g. at 12 weeks the effect of UMEC 62.5/25 versus placebo (CrI: 187.70, 228.20). The analyses were performed with the use of R and openBUGS, as discribed in section 4.5.

All active treatments compared with the placebo by means of trough FEV_1 at 12 weeks for both models were proven to be statistically efficacious (Figure 4.16 and Figure 4.20). In that case, the estimation of TIO 18 + FOR 12 vs placebo showed the greatest uncertainty among the other treatments, as it is reported in only one study, [46], reporting quite large SE. As it is expected the copmarison of the TIO 18 monotherapy with the placebo had the smallest CrI, being compared in 8 studies (Table 4.4). All dual therapies included in the network of evidence for the trough FEV_1 at 12 weeks showed numerically lower performance compared with the UMEC/VI 62.5/25 dual therapy, however none of these estimations was consider statistically significant (Figure 4.15 and Figure 4.19)

At 24 weeks the NMA results of the outcome of interest comparing all active treatments to placebo were in line with the results at 12 weeks, thus the two dual therapies (UMEC/VI 62.5/25, QVA149, and TIO 18 + SAL 50) and the one monotherapy (TIO 18) were proven to be statistically significant more efficacious than placebo (Figure 4.16 and Figure 4.20). The comparison of the treatment of interest compared the other treatments for both models, illustrated that the dual therapies, QVA149 and TIO 18 + SAl 50, where comparable by means of trough FEV_1 at 24 weeks (Figure 4.17 and Figure 4.21).

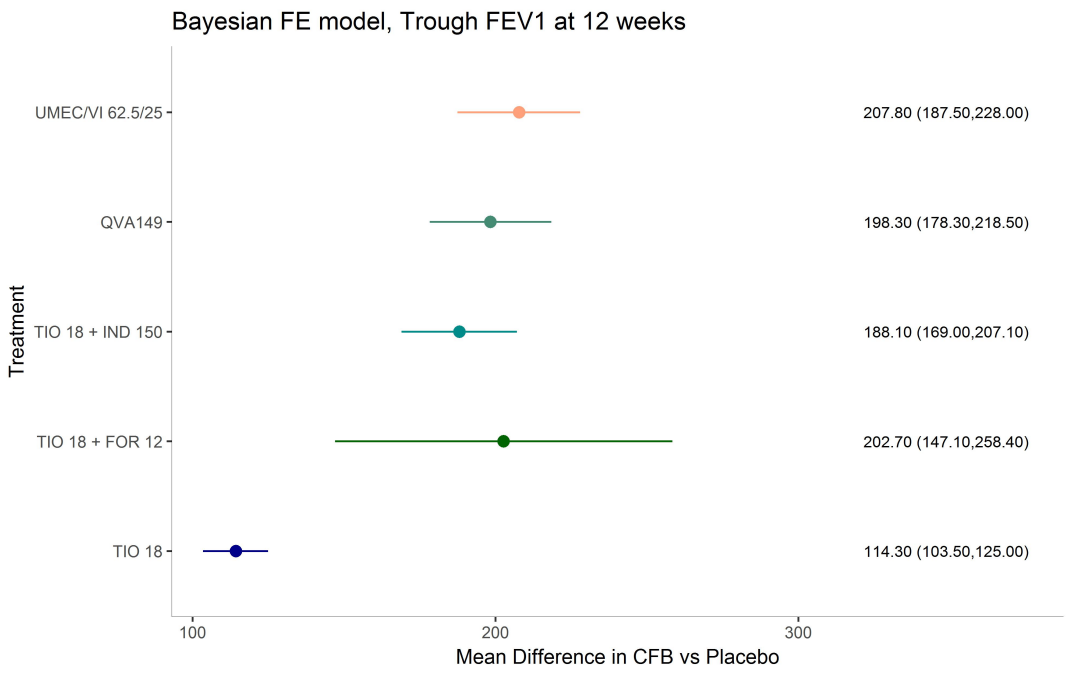


Figure 4.14: Bayesian fixed effect NMA forest plot of trough FEV_1 at 12 weeks using placebo as reference treatment

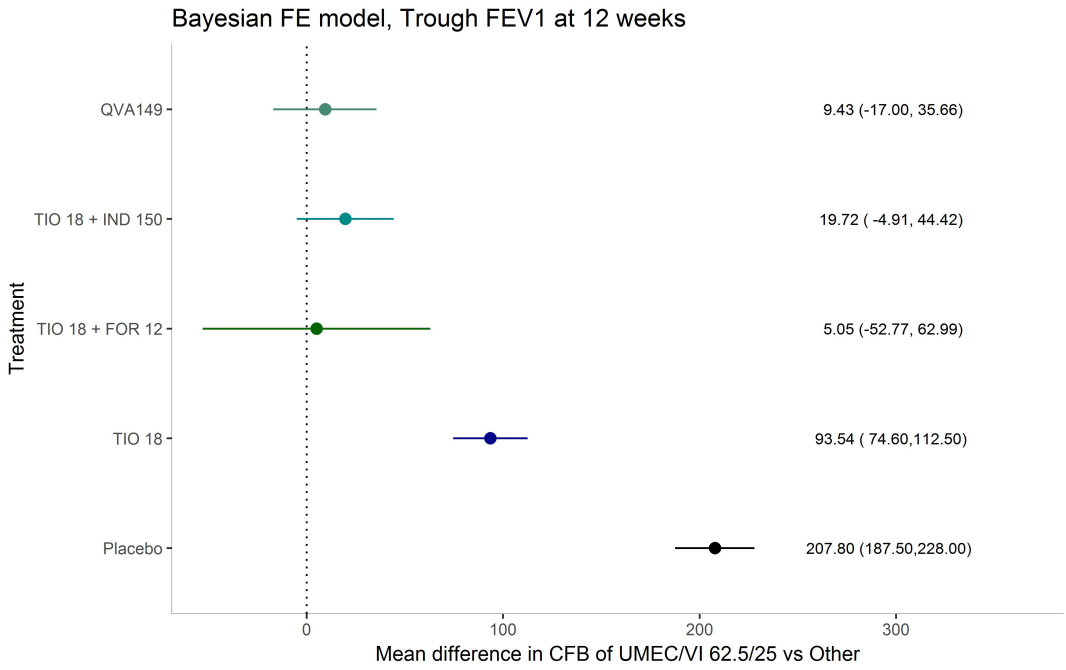


Figure 4.15: Bayesian fixed effect NMA forest plot of trough FEV_1 at 12 weeks using UMEC/VI 62.5/25 as reference treatment

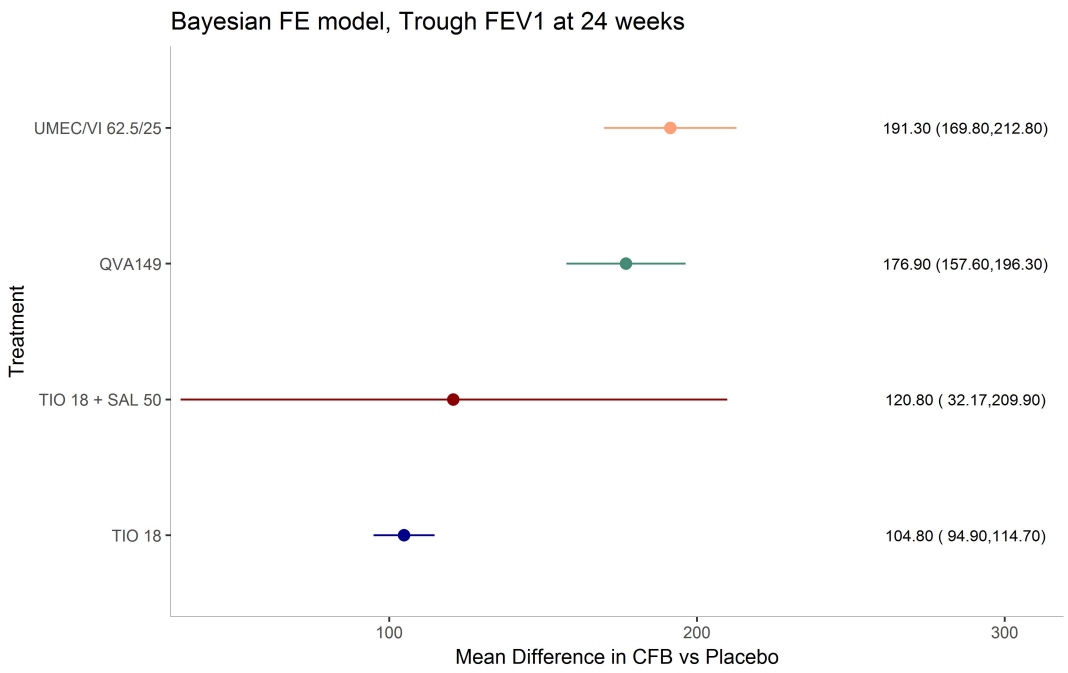


Figure 4.16: Bayesian fixed effect NMA forest plot of trough FEV_1 at 24 weeks using placebo as reference treatment

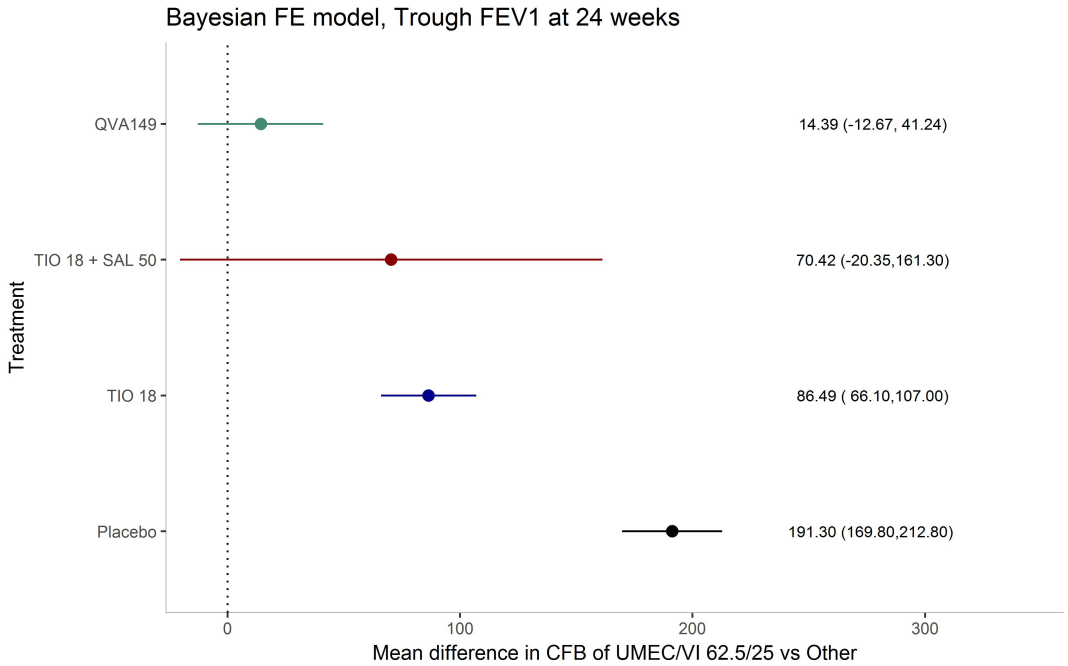


Figure 4.17: Bayesian fixed effect NMA forest plot of trough FEV_1 at 24 weeks using UMEC/VI 62.5/25 as reference treatment

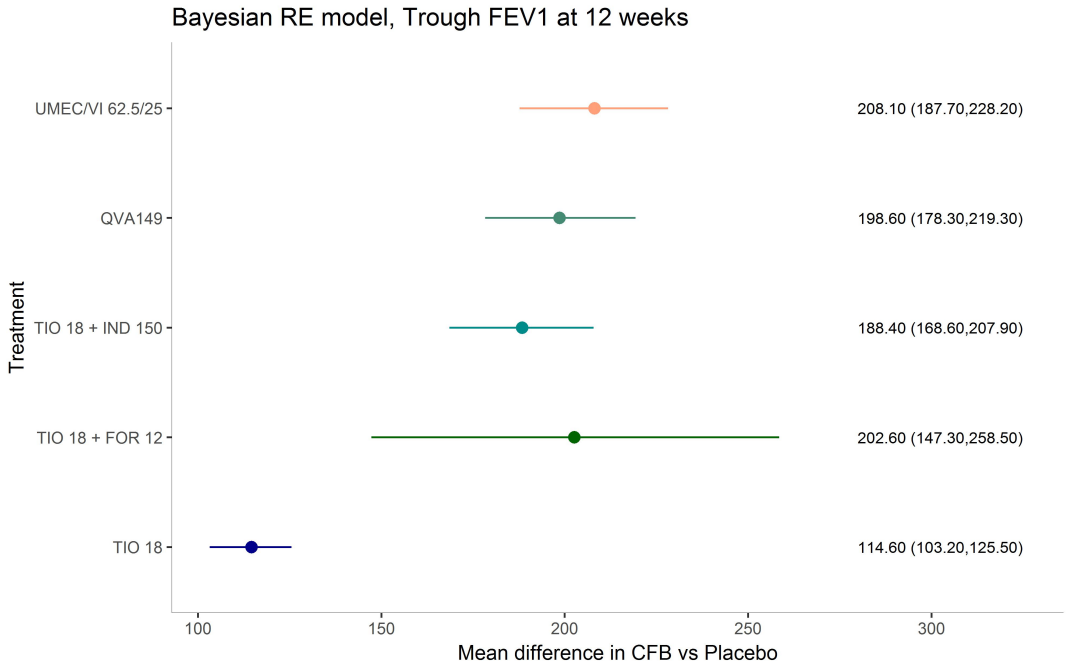


Figure 4.18: Bayesian random effects NMA forest plot of trough FEV_1 at 12 weeks using placebo as reference treatment

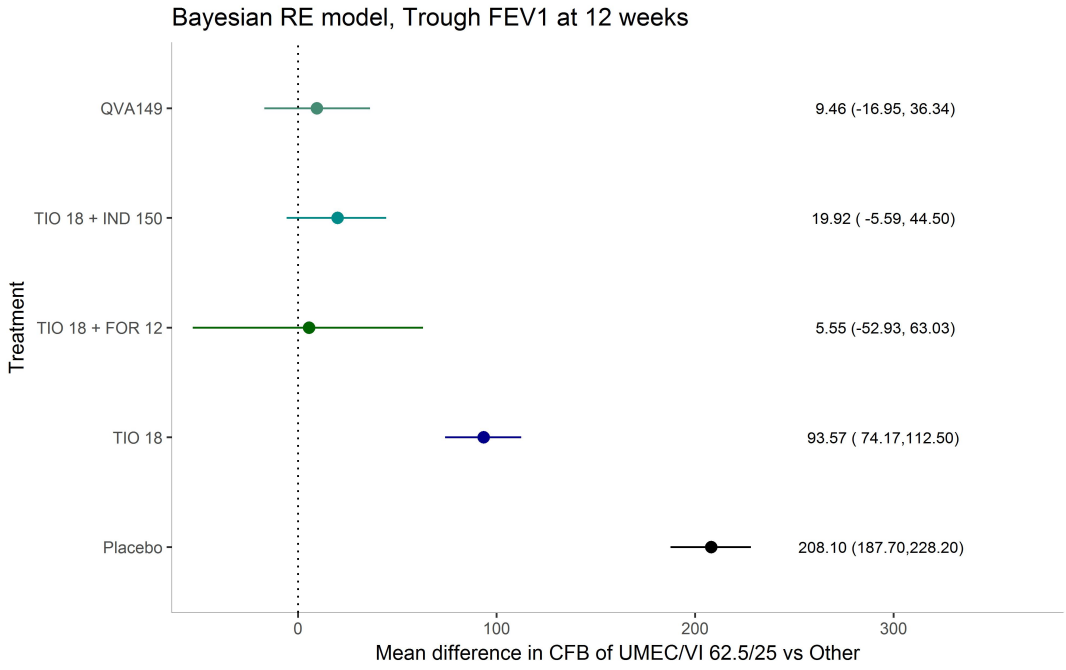


Figure 4.19: Bayesian random effects NMA forest plot of trough FEV_1 at 12 weeks using UMEC/VI 62.5/25 as reference treatment

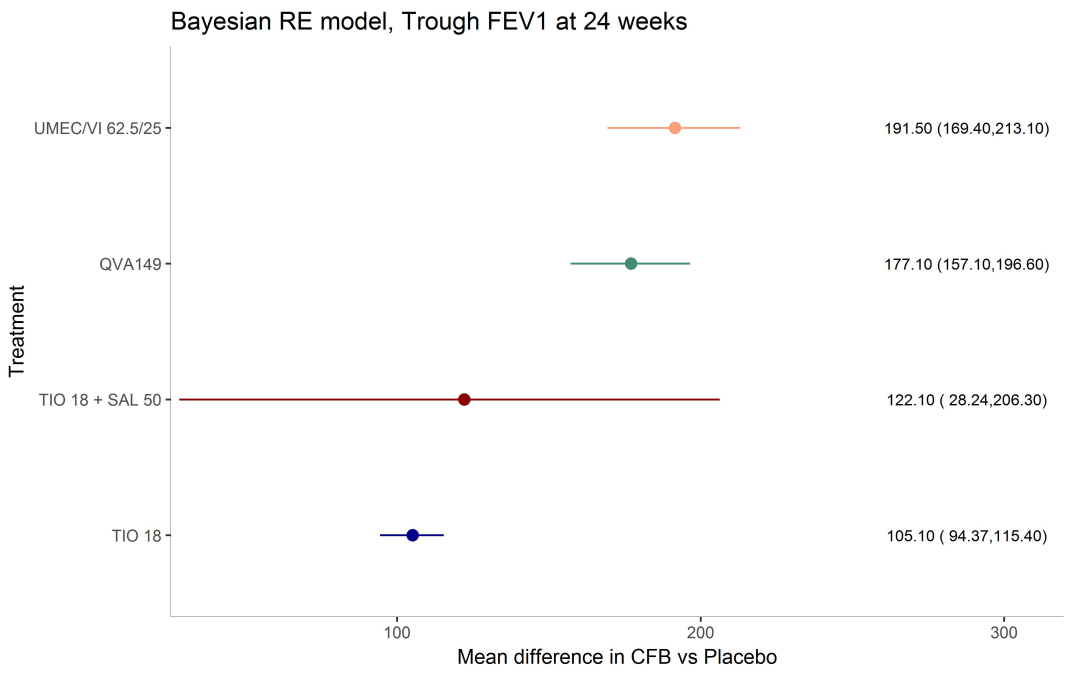


Figure 4.20: Bayesian random effects NMA forest plot of trough FEV_1 at 21 weeks using placebo as reference treatment

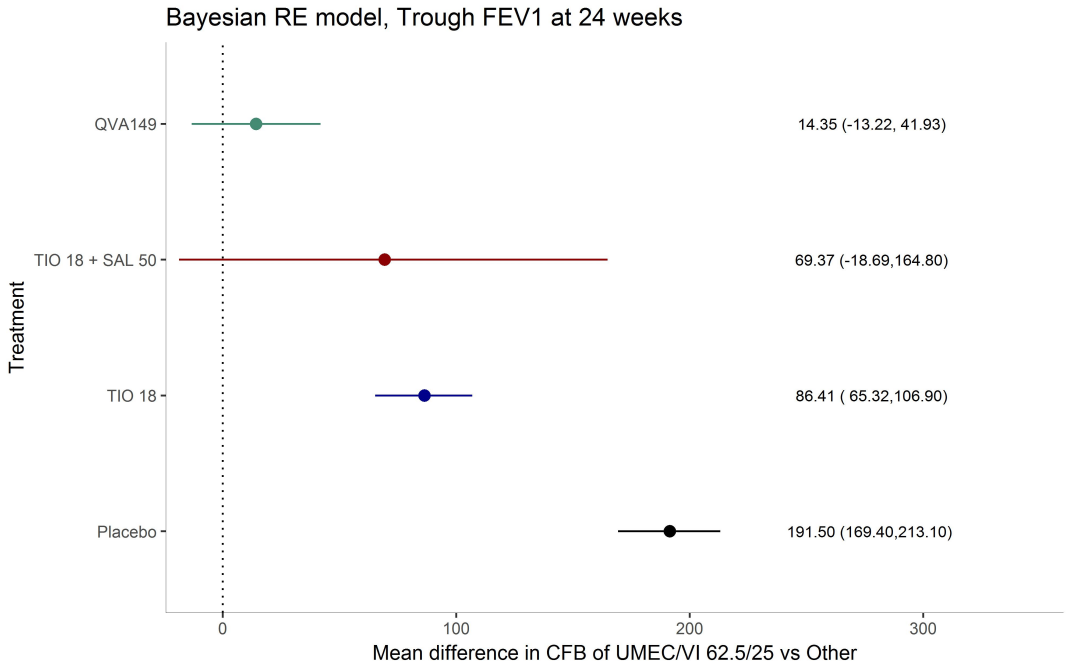


Figure 4.21: Bayesian random effects NMA forest plot of trough FEV_1 at 24 weeks using UMEC/VI 62.5/25 as reference treatment

Chapter 5

Conclusion

This dissertation has shed light into the area of evidence synthesis by into detail discussing and providing examples for the meta-analysis and network meta-analysis techniques in frequentist and Bayesian frameworks for both fixed and random effects models for continuous outcomes. The theory has been analyzed in detail in the first part, starting from the meta-analysis to the mixed treatment comparison, in which the indirect treatment comparison (ITC), pooled effect estimation, and network meta-analysis (NMA) were presented. Moreover, the dissertation's second part presented an application in COPD, using an existing project performed by Mapi [1], enhancing this NMA by providing all possible complex of frameworks and models.

The evidence synthesis methodologies start from the systematic literature review; a crucial process in which all the available data, referring to the question of interest should be identified avoiding any source of bias The importance of that stage should not be underestimated, as it ensures the validation of the analysis. After deciding the considerable homogeneous RCTs, able to be pooled, the techniques of meta-analysis can be applied. The decisions of the similar enough studies, discussed in section 3.2.1, can be challenging and demanding and prerequisite the collaboration of expert disease, clinicians and statisticians.

Meta-analysis has been presented in Chapter 2 synthesises the available RCTs comparing two specific treatments to provide an overall effect. There are two main models, fixed and random effects, the latter is considered as more conservative allowing two sources of variation and is preferred when the number of combined studies is reasonable large or there is heterogeneity among the combined studies. Both frameworks have been analyzed and presented in a extensive fictional example, leading to approximately the same results. In addition, in Chapter 3 the mixed treatment comparison methodologies have been presented, starting with the ITC methodology, based on a mathematical equation [24], which overcomes the meta-analysis limitation; the ability to synthesize indirectly compare treatments. A network of direct and indirect evidence is available by the pooled estimation, based on the ITC methodology, however this is deemed to be more appropriate for trivial networks containing small number of studies and treatments, as it does not take into consideration

the geometry of the network of evidence. The final step of the evidence synthesis pyramid is the network meta-analysis, which analyze a complex network of an arbitrary number of connected treatments directly and indirectly. As it is the extension of meta-analysis, it has also two possible models, fixed and random effects, and both of them have been discussed in frequentist and Bayesian framework, in the same chapter. All the methodologies have been accompanied with examples, however in order to emphasize the NMA techniques, an extensive example has been performed in form of an application.

The application of the NMA has been presented in Chapter 4, based on a network meta-analysis in patients with COPD [1]. The scope of this application was to replicated the results of the Bayesian analyses published and additionally performing the frequentist analyses as well, to serve question of interest about the differentiation of these two frameworks. Furthermore, in the publication only the analyses based on the random effects model were provided, while in this application both models have been performed for each approach. The continuous outcome of the trough FEV_1 at the time point of 12 and 24 weeks were decided to be analyzed. As a summary, it can be said that both frameworks are equivalent, as they lead to the approximately same results.

The fundamental idea of evidence synthesis by means of meta analysis and mixed treatment comparison, with emphasis to the NMA methodology, has been presented in this dissertation. Due to the limitation of this analysis only the continuous outcome has been discussed. It could have been interesting to present also binary or count data which are also common outcomes of interest, e.g. deaths or annual exacerbation rates in a disease area. Another important technique could be the meta-regression, which requires a large number of studies and can deal with more heterogeneous studies. Moreover, in cases when there is no available evidence connecting the treatments of interest, a tool for indirectly comparisons, the Matching-Adjusted Indirect Comparisons [68] can be performed, however this was beyond the scope of this dissertation.

Bibliography

[1]E. L. Huisman, S. M. Cockle, A. S. Ismaila, A. S. Ismaila, and Y. S. Punekar, "Comparative efficacy of combination bronchodilator therapies in COPD: A network meta-analysis", International Journal of COPD **10**, 1863–1881 (2015).

[2]N. J. Welton, A. J. Sutton, N. J. Cooper, K. R. Abrams, and A. Ades, *Evidence Synthesis for Decision Making in Healthcare (Statistics in Practice)* (Wiley, 2012), p. 294.

[3]K. Dwan, C. Gamble, R. Kolamunnage-Dona, S. Mohammed, C. Powell, and P. R. Williamson, "Assessing the potential for outcome reporting bias in a review: a tutorial.", Trials **11**, 52 (2010).

[4]J. S. Silva, "Maximum Likelihood", **58**, 19–20 (2011).

[5]M. Borenstein, L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein, "Introduction to Meta-Analysis", Psychotherapy research journal of the Society for Psychotherapy Research **19**, 421 (2009).

[6]R. Dersimonian and N. Laird, "Meta-Analysis in Clinical Trials*", Statistics in medicine **188**, 177–188 (1986).

[7]R. J. Hardy and S. G. Thompson, "A likelihood approach to meta-analysis with random effects", Statistics in Medicine **15**, 619–629 (1996).

[8]A. A. Veroniki, D. Jackson, W. Viechtbauer, R. Bender, J. Bowden, G. Knapp, O. Kuss, J. P. Higgins, D. Langan, and G. Salanti, "Methods to estimate the between-study variance and its uncertainty in meta-analysis", Research Synthesis Methods **7**, 55–79 (2016).

[9]W. Viechtbauer, "Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model", Journal of Educational and Behavioral Statistics **30**, 261–293 (2005).

[10]E. Kontopantelis, D. A. Springate, and D. Reeves, "A Re-Analysis of the Cochrane Library Data: The Dangers of Unobserved Heterogeneity in Meta-Analyses", PLoS ONE **8**, edited by T. Friede, e69930 (2013).

[11]R. DerSimonian, R. Kacker, F. Mosteller, R. Collins, and P. Sleight, "Random-effects model for meta-analysis of clinical trials: an update.", Contemporary clinical trials **28**, 105–14 (2007).

[12]L. V. Hedges, "A random effects model for effect sizes.", Psychological Bulletin **93**, 388–395 (1983).

[13]A. L. Rukhin, "Estimating heterogeneity variance in meta-analysis", Journal of the Royal Statistical Society: Series B (Statistical Methodology) **75**, 451–469 (2013).

[14]T. C. Smith, D. J. Spiegelhalter, and A. Thomas, "Bayesian approaches to random-effects meta-analysis: A comparative study", Statistics in Medicine **14**, 2685–2699 (1995).

[15]J. Hartung and K. H. Makambi, "Reducing the Number of Unjustified Significant Results in Meta-analysis", Communications in Statistics - Simulation and Computation **32**, 1179–1190 (2003).

[16]H. Raiffa and R. Schlaifer, *Applied statistical decision theory* (Wiley, 2000), p. 356.

[17]M. Jeffrey and E. J. G. Bradford, "Neurosis in escaped prisoners of war", British Journal of Medical Psychology **20**, 422–435 (1946).

[18]E. Lesaffre and A. B. Lawson, *Bayesian Biostatistics* (Wiley, 2012), pp. 1–516.

[19]G. L. Jones and J. P. Hobert, "Honest Exploration of Intractable Probability Distributions via Markov Chain Monte Carlo", Statistical Science **16**, 312–334 (2001).

[20]A. E. Gelfand and A. F. M. Smith, "Sampling-Based Approaches to Calculating Marginal Densities", Journal of the American Statistical Association **85**, 398 (1990).

[21]A. Gelman and D. B. Rubin, "Inference from Iterative Simulation Using Multiple Sequences", Statistical Science **7**, 457–472 (1992).

[22]T. Lumley, "Network meta-analysis for indirect treatment comparisons", Statistics in Medicine **21**, 2313–2324 (2002).

[23]J. P. A. Ioannidis, S. Evans, P. Gotzsche, f. t. C. Group, M. Zwahlen, and G. Chene, "Indirect comparisons: the mesh and mess of clinical trials.", Lancet (London, England) **368**, 1470–2 (2006).

[24]H. C. Bucher, G. H. Guyatt, L. E. Griffith, and S. D. Walter, "The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials.", Journal of clinical epidemiology **50**, 683–91 (1997).

[25]L. J. Gleser and I. Olkin, *Stochastically dependent effect sizes*, 2009.

[26]J. P. T. Higgins and A. Whitehead, "Borrowing strngth from extreral trials in a meta-analysis", Statistics in Medicine **15**, 2733–2749 (1996).

[27]G. Lu and A. E. Ades, "Combination of direct and indirect evidence in mixed treatment comparisons", Statistics in Medicine **23**, 3105–3124 (2004).

[28]G. Salanti, J. P. Higgins, A. Ades, and J. P. Ioannidis, "Evaluation of networks of randomized trials", Statistical Methods in Medical Research **17**, 279–301 (2007).

[29]T. Greenhalgh, "How to read a paper. Getting your bearings (deciding what the paper is about).", BMJ (Clinical research ed.) **315**, 243–246 (1997).

[30]J. P. Jansen, B. Crawford, G. Bergman, and W. Stam, "Bayesian Meta-Analysis of Multiple Treatment Comparisons: An Introduction to Mixed Treatment Comparisons", Value in Health **11**, 956–964 (2008).

[31]J. P. Jansen, C. H. Schmid, and G. Salanti, "Directed acyclic graphs can help understand bias in indirect and mixed treatment comparisons", Journal of Clinical Epidemiology **65**, 798–807 (2012).

[32]S. G. Thompson and J. P. Higgins, "How should meta-regression analyses be undertaken and interpreted?", Statistics in Medicine **21**, 1559–1573 (2002).

[33]D. M. Caldwell, A. E. Ades, and J. P. T. Higgins, "Simultaneous comparison of multiple treatments: combining direct and indirect evidence", BMJ : British Medical Journal **331**, 897–900 (2005).

[34]G. Salanti, J. P. Higgins, A. Ades, and J. P. Ioannidis, "Evaluation of networks of randomized trials", Statistical Methods in Medical Research **17**, 279–301 (2008).

[35]G. Lu, N. J. Welton, J. P. T. Higgins, I. R. White, and A. E. Ades, "Linear inference for mixed treatment comparison meta-analysis: A two-stage approach", Research Synthesis Methods **2**, 43–60 (2011).

[36]G. Rücker, "Network meta-analysis, electrical networks and graph theory", Research Synthesis Methods **3**, 312–324 (2012).

[37]G. Lu and A. E. Ades, "Assessing Evidence Inconsistency in Mixed Treatment Comparisons", Journal of the American Statistical Association **101**, 447–459 (2006).

[38]G. Lu and A. Ades, "Modeling between-trial variance structure in mixed treatment comparisons", Biostatistics **10**, 792–805 (2009).

[39]C. Mathers, G. Stevens, D. M. Fat, J. Ho, A. Mahanani, D. Hogan, W. R. Mahanani, B. Black, T. Boerma, P. Boucher, F. Bray, T. Burton, D. Chou, R. Cibulskis, S. Cousens, J. Ferlay, M. Gacic-Dobo, A. Gemmill, P. Gerland, P. Ghys, P. Glaziou, K. Iaych, R. Jakob, D. Jamison, P. Jha, J. Lawn, L. Liu, R. Lozano, C. Murray, L. Newman, M. Peden, J. Rehm, I. Rudan, L. Say, C. Sismanidis, P. Strebel, T. Vos, and D. You, "WHO methods and data sources for country-level causes of death Estimates of country-level deaths by cause for years 2000-2012 were primarily prepared by", (2014).

[40]E. D. Bateman, G. T. Ferguson, N. Barnes, N. Gallagher, Y. Green, M. Henley, D. Banerji, and E. D. Bateman, "Dual bronchodilation with QVA149 versus single bronchodilator therapy: the SHINE study", Eur Respir J **42**, 1484–1494 (2013).

[41]J. A. Wedzicha, M. Decramer, J. H. Ficker, D. E. Niewoehner, T. Sandström, A. F. Taylor, P. D 'andrea, C. Arrasate, H. Chen, and D. Banerji, "Articles Analysis of chronic obstructive pulmonary disease exacerbations with the dual bronchodilator QVA149 compared with glycopyrronium and tiotropium (SPARK): a randomised, double-blind, parallel-group study", The Lancet Respiratory **1**, 199–209 (2013).

[42]GlaxoSmithKline, *CSR DB21133747*, tech. rep. (2012).

[43]GlaxoSmithKline, *CSR DB2113360*, tech. rep. (2012).

[44]GlaxoSmithKline, *CSR ZEP117115*, tech. rep. (2013).

[45]GlaxoSmithKline, *CSR DB2113373*, tech. rep. (2012).

[46]D. P. Tashkin, J. Pearle, and D. Iezzoni, "Formoterol and Tiotropium Compared With Tiotropium Alone for Treatment of COPD", 17–25 (2009).

[47]D. A. Mahler, A. DUrzo, E. D. Bateman, S. A. Ozkan, T. White, C. Peckitt, C. Lassen, B. Kramer, I.-1. study Investigators, and INTRUST-2, "Concurrent use of indacaterol plus tiotropium in patients with COPD provides superior bronchodilation compared with tiotropium alone: a randomised, double-blind comparison.", Thorax **67**, 781–788 (2012).

[48]S. D. Aaron, K. L. Vandemheen, D. Fergusson, F. Maltais, J. Bourbeau, R. Goldstein, M. Balter, D. O'Donnell, A. McIvor, S. Sharma, G. Bishop, J. Anthony, R. Cowie, S. Field, A. Hirsch, P. Hernandez, R. Rivington, J. Road, V. Hoffstein, R. Hodder, D. Marciniuk, D. McCormack, G. Fox, G. Cox, H. B. Prins, G. Ford, D. Bleskie, S. Doucette, I. Mayers, K. Chapman, N. Zamel, M. FitzGerald, and C. T. S. R. C. R. Consortium, "Tiotropium in combination with placebo, salmeterol, or fluticasone-salmeterol for treatment of chronic obstructive pulmonary disease: a randomized trial.", Ann. Intern. Med. **146**, 545–555 (2007).

[49]R. Dahl, K. Chapman, M. Rudolf, R. Mehta, P. Kho, V. Alagappan, I. Berhane, H. Chen, and D. Banerji, "P190 QVA149 Once Daily Provides Significant Improvements in Lung Function Over 1 Year in Patients with COPD: The Enlighten Study: Abstract P190 Table 1", Thorax **67**, A146.2–A147 (2012).

[50]C. K. N. Chan, F. Maltais, C. Sigouin, and J. M. Haddon, "ORIGINAL ARTICLE A randomized controlled trial to assess the efficacy of tiotropium in Canadian patients with chronic obstructive pulmonary disease", **14**, 465–472 (2007).

[51]D. P. Tashkin, B. Celli, S. Senn, D. N. Burkhart, S. Kesten, S. Menjoge, and M. Decramer, "A 4-Year Trial of Tiotropium in Chronic Obstructive Pulmonary Disease", n engl j med **35915359**, 1543–54 (2008).

[52]E. Kerwin, J. H??bert, N. Gallagher, C. Martin, T. Overend, V. K. T. Alagappan, Y. Lu, and D. Banerji, "Efficacy and safety of NVA237 versus placebo and tiotropium in patients with COPD: The GLOW2 study", European Respiratory Journal **40**, 1106–1114 (2012).

[53]J. F. Donohue, C. Fogarty, J. Lötvall, D. A. Mahler, H. Worth, A. Yorgancioğlu, A. Iqbal, J. Swales, R. Owen, M. Higgins, and B. Kramer, "Once-daily bronchodilators for chronic obstructive pulmonary disease: Indacaterol versus tiotropium", American Journal of Respiratory and Critical Care Medicine **182**, 155–162 (2010).

[54]D. E. Niewoehner, K. Rice, C. Cote, D. Paulson, J. A. D. Cooper, L. Korducki, C. Cassino, and S. Kesten, "Prevention of exacerbations of chronic obstructive pulmonary disease with tiotropium, a once-daily inhaled anticholinergic bronchodilator: A randomized trial", Annals of Internal Medicine **143**, 317–26 (2005).

[55]V. Brusasco, R. Hodder, M. Miravitlles, L. Korducki, L. Towse, and S. Kesten, "Health outcomes following treatment for six months with once daily tiotropium compared with twice daily salmeterol in patients with COPD", Thorax **58**, 399–404 (2003).

[56]J. F. Donohue, J. a. V. Noord, E. D. Bateman, S. J, A. Lee, T. J. Witek, and S. Kesten, "Comparing Lung Function and Health Status Changes in COPD Patients Treated With Tiotropium or Salmeterol * A 6-Month , Placebo-Controlled Study Comparing Lung Function and Health Status Changes in COPD Patients Treated With Tiotropium or Salmeterol *", Chest, 47–55 (2002).

[57]R. Casaburi, W. Serby, S. S. Menjoge, and T. J. Witek, "The spirometric efficacy of once daily dosing with tiotropium in stable COPD", Chest **118**, 1294–1302 (2000).

[58]C. Verkindre, F. Bart, B. Aguilaniu, F. Fortin, J.-C. Guérin, C. Le Merre, P. Iacono, and G. Huchon, "The effect of tiotropium on hyperinflation and exercise capacity in chronic obstructive pulmonary disease.", Respiration; international review of thoracic diseases **73**, 420–7 (2006).

[59]H. Covelli, S. Bhattacharya, C. Cassino, C. Conoscenti, and S. Kesten, "Absence of Electrocardiographic Findings and Improved Function with Once-Daily Tiotropium in Patients with Chronic Obstructive Pulmonary Disease", Pharmacotherapy **25**, 1708–1718 (2005).

[60]J. Moita, C. Bárbara, J. Cardoso, R. Costa, M. Sousa, J. Ruiz, and M. Santos, "Tiotropium improves FEV1 in patients with COPD irrespective of smoking status", Pulmonary Pharmacology & Therapeutics **21**, 146–151 (2008).

[61]G. Schwarzer, "meta: An R package for meta-analysis.", R News **7(3)** (2007).

[62]W. Viechtbauer, "Conducting Meta-Analyses in R with the metafor Package", Journal of Statistical Software **36**, 1–48 (2010).

[63]S. Sturtz, U. Ligges, A. Gelman, and N. Thomas, "R2OpenBUGS: A Package for Running OpenBUGS from R", Journal of Statistical Software (2005).

[64]M. Plummer, N. Best, K. Cowles, and K. Vines, "Coda: convergence diagnosis and output analysis for mcmc", R News **6**, 7–11 (2006).

[65]G. Rücker and G. Schwarzer, "Ranking treatments in frequentist network meta-analysis works without resampling methods", 1–9 (2015).

[66]G. Rucker, G. Schwarzer, U. Krahn, and J. Konig, *Netmeta: network meta-analysis using frequentist methods*, R package version 0.9-3 (2017).

[67]H. Wickham, *Ggplot2*, Vol. 35, Use R! July (Springer International Publishing, Cham, 2016), p. 211.

[68]J. E. Signorovitch, V. Sikirica, M. H. Erder, J. Xie, M. Lu, P. S. Hodgkins, K. A. Betts, and E. Q. Wu, "Matching-Adjusted Indirect Comparisons: A New Tool for Timely Comparative Effectiveness Research", Value in Health **15**, 940–947 (2012).