University of Glasgow
Faculty of Information and Mathematical Sciences
Department of Statistics

Research Project Thesis

Simple Strategies for Variance Uncertainty in Meta-Analysis

MScR Student:
Giorgio Di Gessa

Supervisor:
Prof. Stephen Senn

Academic Year 2006-2007

Simple Strategies for Variance Uncertainty in Meta-Analysis

by Giorgio Di Gessa, MScR

A dissertation Submitted to the University of Glasgow
for the degree of Master of Science

Department of Statistics

Academic Year 2006-2007

Acknowledgments

I would like to thank Chiesi Farmaceutici SpA for providing me with funding to carry out this thesis. I hope that my work will prove useful and will prompt helpful discussion in the future. A special thanks to my supervisor Professor Stephen Senn (University of Glasgow, Department of Statistics), whose encouragement, invaluable support, advice and patience in clarifying my problems were crucial in the production of this thesis. I would also like to thank Stephen for his help on a personal level in the initial difficult times I had. Your support was much appreciated.

Finally I would like to thank both my family and my friends for trusting me just the way I am. Thanks for pushing me and supporting my projects and ideas with an enormous dose of impudent optimism and uncontainable passion. You all know how to regenerate and rejuvenate me. Thanks for your disarming patience and both your cynicism and encouragement that keep my feet "almost" on the ground. You all taught me that there is not much of a hurry to become 'mature and wise'. "There's a lot to be said about the freshness of youth ... Great strength is often born of vulnerability, new vision from unwearied eyes". Words cannot express my gratitude to my friends: you contributed your time, energy, vision, passion and support. Thanks for keeping my spirits up, sharing my vision, daring to dreams and pushing me to learn the extent of my wings! I love you all. You are all family to me!

Abstract

Reliability of the measure of precision of the estimate is crucial; a correct value of the standard error of the point estimate entails that the resulting significance of the analysis is correctly stated and that confidence intervals have correct coverage probabilities. Stating an incorrect precision, on the contrary, can often result in biased and misleading results. In particular, in fixed-effects meta-analysis the overall estimator usually used in practice tends to have a variance higher than the optimal one even though this appears to be lower, just by chance.

In performing a fixed-effects meta-analysis, individual treatment estimates are weighted proportionately to the precision of the study. Such weighting is optimal only under the assumption that variances are known, which is never the case in practice. As a consequence, the estimator is sub-optimal and the resulting meta-analysis overstates the significance of the results: in particular, overstatements are dramatic when we summarise studies with small number of patients. Focusing the attention to the fixed-effects model, the main aim of this thesis is to investigate the behaviour of the precision of the overall estimator under different circumstances in order to assess how biased and incorrectly reported the overall variance of the commonly used estimator is and also to highlight in which circumstances improved estimates are deemed necessary.

In fixed-effects meta-analysis, problems are related to poor estimates of the individual variances σ_i^2 since these values are imprecise and both θ (the point estimator) and V (the overall variance estimator) depend upon them. Poorly estimated study variances can lead to the overall estimate of the variance of the treatment effect being badly underestimated. In order to evaluate the circumstances in which the imprecision in the estimates of σ_i^2 badly affects V, a number of simulations in different settings were performed. Under both the assumption of common and uncommon variance of the observations at the patients level, the average total number of patients per study plays an important role and this appears to be more important than the total number of each single study. Moreover, the allocation of patients per arm does not seem to be decisive for the estimated overall variance of the estimator even though balanced allocation as well as having roughly the same amount of patients per study yields better results. Furthermore, true to form, the higher the average number of patients per arm,

the closer the estimator is to the optimal one, i.e. the fewer the number of patients, the less precise the estimates of σ_i^2 are and the greater the impact is on the results. Given the imprecision in the estimate of σ_i^2 , we may severely overstate the precision of $\hat{\theta}$. Better estimation of the variances are therefore investigated. Are there ways to account for the imprecise estimates of the within-studies variances?

Shrunk variances were considered in order to assess whether borrowing information across variances would produce an overall variance estimate whose 'real' and 'average' dispersion were both closer to the optimal value. Combining measurements minimises the total '*Mean Squared Error*'. Therefore, particularly when the nature of the problem is not to estimate each expected return separately but rather to minimise the total impact, shrinkage estimators represent a reasonable alternative to the classical estimators. This approach seems reasonable since the goal of this thesis is to minimise the real dispersion of the overall variance estimator. Moreover, shrinkage approaches (that combine variance information across studies and are study-specific at the same time) usually perform well under a wide range of assumptions about variance heterogeneity, behaving well both when the variances were truly constant as well as when they varied extensively from study to study. In particular, in this thesis the 'modified CHQBC estimator' suggested by Tong and Wang is used (where CHQBC stands for the James-type shrinkage estimator for variances initially proposed by Cui,Hwang,Qiu,Blades and Churchill).

Results obtained via simulations (with different patterns for various variance schemes and diverse average amounts of patients per study), emphasise that the estimator based on the 'shrunk variances' performs better than the one based on the estimated sample variances. Regardless of the variance structure across studies (homoscedasticity or uncommon variances), the estimator based on the shrunk variances performs optimally, even with an average small number of patients per trial, achieving almost optimal results even when the variances are strongly heterogenous and without relying on computational expensive procedures. Chapter 3 shows the results obtained if shrunk variances are used instead of the declared ones; moreover, this new approach is applied to some real data-sets showing how the declared variance tends to be higher in all cases and presumably closer to the 'real' optimal value.

Finally, chapter 4 highlights the merits of this new approach to the problem of imprecise precision estimates in fixed-effects methods and also looks at the further work that needs to be done in order to improve results for this and other meta-analytical settings; this thesis, in fact, only considers the case of continuous normally distributed data ignoring binary, ordinal or survival data meta-analyses. Moreover, despite the fact that the problem of estimating σ_i^2 is particularly urgent and dramatic in the fixed-effects model, the estimation of σ_i^2 might also be expected to influence random effects coverage

probabilities especially when all studies in the meta-analysis are small (Brockwell & Gordon, 2001).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Meta-analysis is intended to provide the statistical summary of a collection of results from individual studies for the purpose of integrating the findings. Data-analysis is only the last step of a long and complicated research synthesis procedure that involves a problem formulation stage, a data collection stage and a data evaluation step all of them necessary to evaluate and decide what reported studies to include in the analysis. The outcome of a meta-analysis may therefore be a long awaited process.
During the last 20 years or so, literally thousands of meta-analytic papers primarily covering applications in health and medical sciences have been published, making meta-analysis have a very important role in diverse fields of applications. Moreover, the essential character of meta-analysis is quantitative in nature and since this statistical summary is applied to numerous diverse applications in many fields, it is essential that misleading results are not produced. In fact, often, final decisions are based upon the conclusions obtained through such a quantitative research synthesis; small variations in the outputs can have an important impact and substantial consequences in public and health planning policies, for instance. *Given the length of a meta-analysis and the impact results may have, precision of the estimates is a crucial point.*

There is no empirical nor theoretical basis for preferring the fixed-effects model over the random-effects model or viceversa. There are some arguments in favour of each approach which depend on the purpose as well as conceptual difficulties linked to both points of view. This thesis, however, will focus on *fixed-effects models*, trying to investigate how reliable and accurate results are. The reason why we will concentrate attention on fixed-effects meta-analysis is that weights, overall point estimate θ as well as the precision of the estimate depend entirely and solely upon the findings of many empirical studies which are usually assumed to be known. The simplifying assumption is that the sampling variances of the effect size estimates are known; however, this is approximately true only when sample sizes are large. It is therefore essential to allow

for the imprecision of these estimates.

The ordinary method is too sensitive to individual study variances and is negatively biased when sample sizes are too small. In particular, the precision of σ_i^2 has a dramatic influence on weights and therefore on V, the overall precision of the estimator that describes how uncertain we are about the point estimate. Treating σ_i^2's as known underestimates V and can lead to a loss of efficiency, especially for small trials. When we consider small sizes, standard errors should not be considered as if they were known, because this would overestimate precision, leading to unreliable results.

1.1 Aim

Fixed-effects models do not account nor allow for the sampling error in $\hat{\sigma}_i^2$; however, it is known that sampling errors are present in practice. When we ignore this problem - as it happens with the ordinary method – the usual variance estimator performs very poorly in detecting the true variance of θ and underestimates the true value. Moreover, the actual variability of the variance estimator is always higher than both the declared and optimal ones, with a consequent overstatement of the precision of the estimator and misleading results in the form of too liberal significance tests and Confidence Intervals without correct coverage properties, in particular with small size studies.

The aim of this thesis is therefore to illustrate via simulations - and calculations where possible - what circumstances (variance structure across studies at the patient level, number of studies, allocation per arm, study size) worsen the estimate of the variance of the overall estimator. Moreover, and more importantly, it will be investigated whether a different method, able to be accurate and flexible at the same time, exists. In particular, an estimator whose variance does not diverge substantially from the optimal value both on average and in practice is highly wanted and warmly recommended. This would guarantee both more accurate statements about the precision of the point estimate and confidence intervals more likely to have the correct nominal coverage probabilities.

Chapter 2

Meta-Analysis is biased

2.1 Meta-Analysis

"Meta-analysis is a quantitative approach for systematically combining the results of previous studies in order to arrive at summary conclusions about the body of research" (Petitti, 1994, pg. 4,15). The need for such a quantitative review and synthesis of results of related but independent studies became particularly acute in the social sciences in the mid-70s, when the narrative literature reviews were perceived *selective* in the inclusion of studies and *subjective* in their weighting (Petitti, 1994). Since then, utilization of meta-analytic techniques to combine results and information from separate quantitative investigations has become increasingly common, and statistical methods for its application have been further explored and developed. "Over the past 20 years the number of published meta-analyses and discussions on meta-analysis methodology has dramatically increased. This has occurred particularly in the areas of medical and epidemiological research" (Brockwell & Gordon, 2001, pg. 825). The popularity of meta-analysis is due to its overall goal: integrated analysis has "more statistical power to detect a treatment effect than an analysis based only on one study" (Normand, 1999, pg. 321). Furthermore, "when several studies have conflicting conclusions a meta-analysis can be used to estimate an *average* effect or to identify a subset of studies associated with a beneficial effect" (Normand, 1999, pg.322). Meta-analysis can be of great advantage in situations for which individual outcomes are difficult to interpret or when treatment effects are small or not significant in each study alone. "Owing to this rapid rise in the popularity of meta-analysis, it is becoming increasingly important that the methodology and statistics used are sound" (Brockwell & Gordon, 2001, pg. 825).

Consider k separate studies looking at the same clinical question (as, for example, a comparison between a new medication and placebo) in which each trial treatment is

estimated in terms of a difference in means of a quantitative variable. Meta-analysis can be based on a fixed-effects model (where the inference is conditional on the studies actually done) or on a random-effects model (where studies are considered a random sample of some hypothetical population of studies). The two different assumptions address to two different theoretical questions. "The random-effects model is appropriate if the question is whether the treatment *will*, on average, have an effect. If the question is whether the treatment *has caused* an effect *in the studies that have been done*, then the fixed-effects model is more suitable" (Petitti, 1994, pg. 93). Evidently, these distinct assumptions entail distinct statistical methods; "the random-effects model uses a two-stage sampling idea, as if we sampled from a superpopulation of studies that might be carried out and then sampled patients within the studies. Of course, the real situation is more like a selection of studies that can be carried out" (Mosteller & Chalmers, 1992, pg. 232).
"The random-effects model in meta-analysis has actually been suggested as a way to model known differences between studies such as study-design, different within-study matching protocols, different treatment protocols" (treatment doses, lengths, exposures or intensities, for example), interventions, outcomes studied "or perhaps even gender or cultural differences between study participants" (Biggerstaff & Tweedie, 1997, pg. 753). In practice, there are so many different approaches to conducting a study that there are many different potential treatment effects that could arise. "Such diversity is commonly referred to as (methodological or clinical) heterogeneity (τ^2) and may or may not be responsible for observed discrepancies in the results of the studies. Addressing such heterogeneity has been and still is one of the most troublesome aspects of many systematic reviews" (Higgins & Thompson, 2002, pg. 1539,1540) as its magnitude can influence the conclusions of the meta-analysis. Quantifying the amount of heterogeneity is therefore one of the most important aspects of systematic reviews.

Whether fixed-effects or random-effects models are more appropriate, the choice of model is very important as this "can lead to noticeably different conclusions" (Mengersen et al., 1995, pg. 38). The impact of the choice of method can be significant. Even small absolute variations can have an important impact and "they may have substantial consequences in arenas such as public policy, health planning and litigation" (Mengersen et al., 1995, pg. 39). There are conceptual difficulties linked to both the fixed-effects and random-effects points of view: "in both models, it may be difficult to characterize precisely the universe to which we are inferring" (Normand, 1999, pg. 326). In particular, random-effects model assumes that the results from the trials are representative of the results which would be obtained from the total population of centres while, in reality, centres are not chosen at random. On the other hand, the fixed-effects model makes the assumption that the characteristics of patients in meta-analytical studies are

the same as those in the total patient population.
There is no empirical nor theoretical basis for preferring the fixed-effects model over the random-effects model or viceversa. Nonetheless, despite the long controversial debate as to the choice of the appropriate model, statisticians' attention has focused mainly on the random-effects model that incorporates a parameter explicitly accounting for the between-trial variability, producing results which can be considered more generalisable. Mosteller and Chalmers, for instance, "fear that some investigators prefer the fixed-effects approach because it gives narrower confidence limits rather than because they want to apply their inferences to the particular population sampled" (Mosteller & Chalmers, 1992, pg. 232). Biggerstaff and Tweedie remark that "the application of the fixed-effects model in meta-analytic contexts has been called into question" (Biggerstaff & Tweedie, 1997, pg. 753). Moreover, it is believed that, although random-effects models are generally conservative since they typically widen confidence intervals and lead to a lower chance of calling a difference 'statistically' different, they give a "much truer picture of variability both in individual studies and across a set of studies and consequently enable more informed inference" (Mengersen et al., 1995, pg. 41). Normand notes that "it is almost *always* reasonable to believe that there is some between-study variation and few reasons to believe it is zero". Especially when studies conflict, "it is difficult to ignore the between-study variation" (Normand, 1999, pg. 326). Furthermore, "the test for the heterogeneity for assessing the validity of the fixed effect model is of limited use, particularly when the total information is low, or when the amount of information available in each trial is very variable" (Hardy & Thompson, 1998, pg. 853). Hardy and Thompson believe that "in practical medical research, clinical homogeneity is rare owing to the nature of the studies and the many variables involved, and a degree of a statistical heterogeneity may be anticipated" (Hardy & Thompson, 1996, pg. 620). The fixed-effects approach is "open to criticism and is generally discouraged. A truly random effects approach estimating τ^2, which simplifies to a fixed effects model only if τ^2=0, may therefore be preferable" (Jackson, 2006, pg. 2689).

2.2 The models

In light of the above, attention and energies have focused mainly on random-effects models and on the quantification of the heterogeneity τ^2. This thesis, nonetheless, will focus on fixed-effects meta-analysis. The problem of estimating correct within-study variances (σ_i^2) is important for both models: it is crucial for fixed-effects models but it is also expected to have consequences on random-effect models as well. Fixed-effects models will be preferred in order to simplify the presentation of the problem but the potential strategies to handle with the imprecision of σ_i^2's could be applied to both

models.

Fixed-effects and random-effects statistical methods are outlined briefly below. We consider the problem of combining information from a series of k comparative clinical trials, where the data from each trial consists of the number of patients in treatment and control groups, n_T and n_C. For simplicity, we assume a series of parallel group trials. When means, $\overline{X}$, in each treatment arm are known, the mean difference and the associated measure of precision for each primary study can be calculated. Letting i index the trials, a potential summary measure is the difference in means, $Y_{\mathrm{i}} = \overline{X}_{T_i}$ - $\overline{X}_{C_i}$ with standard error $\hat{\sigma}_i$, calculated (under the assumption that the variances in both groups are identical in each study) by

$$var(Y_i) = \sigma_i^2 = \left(\frac{\sigma_{Ti}^2}{n_{Ti}} + \frac{\sigma_{Ci}^2}{n_{Ci}}\right) = S_i^2 \left(\frac{1}{n_{Ti}} + \frac{1}{n_{Ci}}\right) \tag{2.1}$$

where a common estimate of S_i^2 based on both σ_{Ti}^2 and σ_{Ci}^2 is given by

$$\hat{S}_i^2 = \frac{(n_{Ti}-1)\hat{s}_{Ti}^2 + (n_{Ci}-1)\hat{s}_{Ci}^2}{n_{Ti}+n_{Ci}-2} \tag{2.2}$$

where $\hat{s}_{Ti}^2$ and $\hat{s}_{Ci}^2$ are the treatment and control group sample variances, respectively, for the ith study and S_i^2 is the so called 'pooled variance'.

2.2.1 The fixed-effects model

The *fixed-effects model* assumes that each study summary statistic, Y_i, is a realization from a population of study estimates with common mean θ, i.e. every study evaluates a common treatment effect. This means that the effect of treatment, allowing for the play of chance, was the same in all studies and if all the studies were infinitely large they would give identical results.

Let θ - the average effect - be the central parameter of interest and assume there are i= 1,2,...,k studies. Assume that Y_i is such that E(Y_i) = θ (implying that each study has the same underlying effect) and let σ_i^2 = var(Y_i) be the variance of the summary statistic in the ith study. Even under a fixed effect model, in order to calculate confidence intervals for the overall estimate of treatment effect, it is assumed that the observed effects in each trial are normally distributed and approximately unbiased (which, for moderately large study sizes, is guaranteed by the central limit theorem). Thus,

$$Y_i \sim N(\theta, \sigma_i^2) \quad for \quad i = 1, 2, ..., k$$

where σ_i^2 is assumed known and equal to $\hat{\sigma}_i^2$. Making these additional assumptions, then $\hat{\theta} \sim N(\theta, 1/\sum_{i=1}^{k} w_i)$ where $w_i = 1/\hat{\sigma}_i^2$ which allows the calculation of confidence intervals for θ (Hardy & Thompson, 1998).

2.2.2 The random-effects model

The *random-effects model* is an alternative approach to meta-analysis that does not assume that a common ('fixed') treatment effect exists; on the contrary, the true treatment effects in the individual studies may be different from each other. This means there is no single number to estimate in the meta-analysis, but a distribution of numbers. The random-effects framework postulates that each study statistic, Y_i, is a draw from a distribution with a specific mean, θ_i, and variance σ_i^2 :

$$Y_i \mid \theta_i, s_i^2 \sim N(\theta_i, \sigma_i^2) \quad for \quad i = 1, 2, ..., k$$

where $\sigma_i^2 = \hat{\sigma}_i^2$. Furthermore, each study-specific mean, θ_i, is assumed to be a draw from some superpopulation of effects with mean θ and variance τ^2, under the assumption that these different true effects are normally distributed, i.e. with

$$\theta_i \mid \theta, \tau^2 \sim N(\theta, \tau^2)$$

This gives a two stage model:

$$\begin{cases} Y_i = \theta_i + e_i \\ \theta_i = \theta + \epsilon_i \end{cases}$$

where $e_i \sim N(0, \sigma_i^2)$ and $\epsilon_i \sim N(0, \tau^2)$. The error terms are assumed to be independent. In this case, the true effect for study i is centred around the overall effect, allowing individual studies to vary both in estimated effect and true effect. θ and τ^2 are referred to as *hyperparameters* and represent, respectively, the *average treatment effect* and *inter-study variation*. Given the hyperparameters, the distribution of each study summary measure, Y_i, after averaging over the study-specific effects, is Normal with mean θ and variance $(\sigma_i^2 + \tau^2)$. As in the fixed-effects model, θ is the parameter of central interest as this represents the overall treatment effect (i.e. the average effect size in the population); however, the between-study variation, τ^2 (often referred to as the heterogeneity variance) plays an important role. The special case where $\tau^2 = 0$ implies that the effect sizes are homogeneous ($\theta_i = \theta, i = 1, 2, ..., k$) and the resulting model is the fixed-effects one. The σ_i^2 values (the variance of the difference in means for the ith study) are estimated by the sample variances $\hat{\sigma}_i^2$ (see equation 2.1) usually calculated from the data of the ith observed sample and are treated as *known* constants. "In practice the variances are not known so estimated variances $\hat{\sigma}_i^2$ are used to estimate both θ and its variance. Any effect of this is generally ignored in practice" (Brockwell & Gordon, 2001, pg. 826).

2.3 The Usual Meta-Analytical Estimators

In order to account for differences in sample size and study-level characteristics, a weighted average differences of the estimates from each study is taken into account. The parameter of interest θ is estimated by

$$\hat{\theta} = \frac{\sum_{i=1}^{k} Y_i \hat{w}_i}{\sum_{i=1}^{k} \hat{w}_i} \qquad with \quad \hat{w}_i = (\hat{\tau}^2 + \hat{\sigma}_i^2)^{-1} \tag{2.3}$$

where $\hat{\tau}^2$ is a suitable estimator of the heterogeneity parameter τ^2. When a fixed effect model is considered, weights are equal to the reciprocal of the within-variability, i.e. $\hat{w}_i = 1/\hat{\sigma}_i^2$. "Any choice of weight will lead to an unbiased estimate of the common treatment effect, but w_i is generally taken to be the reciprocal of the variance for the study i. These particular weights provide the most precise estimate of the treatment effect, that is they minimise the variance of $\hat{\theta}$" (Hardy & Thompson, 1996, pg. 619,620), $\hat{V}$. Furthermore, assuming σ_i^2 known and equal to $\hat{\sigma}_i^2$ for all i implies that

$$\hat{V} = var(\hat{\theta}) = \frac{1}{\sum_{i=1}^{k} \hat{w}_i} \tag{2.4}$$

where $\hat{w}_i = 1/\hat{\sigma}_i^2$.

2.4 The Heterogeneity Parameter and the different methods of estimation

In light of the above considerations, attention has been paid particularly to estimation of the heterogeneity parameter. There exists an extensive literature about the estimation of τ^2. This parameter can be estimated using different methods of estimation: namely, the method of moments estimator by DerSimonian and Laird (DSL – DerSimonian & Laird (1986)), the variance-component type estimator by Hedges (VC – Hedges (1983)), the simple heterogeneity variance estimator by Sidik and Jonkman (SH – Sidik & Jonkman (2005)), the maximum likelihood estimator by Hardy and Thompson (ML – Hardy & Thompson (1996)) and the approximate restricted maximum likelihood estimator (REML – Thompson & Sharp (1999)).
The DerSimonian and Laird method of moments estimator is based on the test statistic of homogeneity originally proposed by Cochran (Cochran, 1937) in 1937. Using the test statistic

$$Q_C = \sum_{i=1}^{k} \hat{w}_i (Y_i - \hat{\theta}_{fix})^2$$

where $\hat{w}_i = (\hat{\sigma}_i^2)^{-1}$ for $i = 1, 2, ...k$ and where $\hat{\theta}_{fix}$ is the estimator of θ when τ^2 is set equal to zero in equation 2.3, the DerSimonian and Laird estimator has the explicit form

$$\hat{\tau}^2_{DSL} = max\left\{0; \frac{Q_C - (k-1)}{\left[\sum_{i=1}^k \hat{w}_i - (\sum_{i=1}^k \hat{w}_i^2 / \sum_{i=1}^k \hat{w}_i)\right]}\right\}$$

The DSL estimator is unbiased if the study-specific σ_i^2 are assumed known and equal to $\hat{\sigma}_i^2$. The VC estimator is

$$\tau^2_{VC} = max\left\{0, \frac{1}{k-1}\sum_{i=1}^k (Y_i - \overline{Y})^2 - \frac{1}{k}\sum_{i=1}^k \hat{\sigma}_i^2\right\}$$

where $\overline{Y} = \sum_{i=1}^k Y_i/k$. The SH estimator is based on weighted least squares, it is simple to compute and always yields a non-negative estimate of τ^2. This is given by

$$\hat{\tau}^2_{SH} = \frac{1}{k-1}\sum_{i=1}^k \hat{v}_i^{-1}(Y_i - \hat{\theta}_{\hat{v}})^2 \quad ,$$

where $\hat{v}_i = \hat{r}_i + 1$, $\hat{r}_i = \hat{\sigma}_i^2 / \left[\sum_{i=1}^k (Y_i - \overline{Y})^2/k\right]$ and $\hat{\theta}_{\hat{v}} = \sum_{i=1}^k \hat{v}_i^{-1} Y_i / \sum_{i=1}^k \hat{v}_i^{-1}$. The ML and REML estimators are less simple computationally and require iterative solutions. The ML estimator can be calculated by iterating the equation

$$\tau^2_{ML} = max\left\{0, \frac{\sum_{i=1}^k \hat{w}_i^2\{(Y_i - \hat{\theta})^2 - \hat{\sigma}_i^2\}}{\sum_{i=1}^k \hat{w}_i^2}\right\}$$

until it converges, where $\hat{\theta} = \sum_{i=1}^k \hat{w}_i Y_i / \sum_{i=1}^k \hat{w}_i$ and $\hat{w}_i = 1/(\hat{\sigma}_i^2 + \hat{\tau}^2_{ML})$, given an initial estimate of τ^2. Similarly, the REML estimator is computed using the iterative equation

$$\hat{\tau}^2_{REML} = max\left\{0, \frac{\sum_{i=1}^k \hat{w}_i^2[(k/(k-1))(Y_i - \hat{\theta})^2 - \hat{\sigma}_i^2]}{\sum_{i=1}^k \hat{w}_i^2}\right\} \quad ,$$

where $\hat{\theta} = \sum_{i=1}^k \hat{w}_i Y_i / \sum_{i=1}^k \hat{w}_i$ with $\hat{w}_i = 1/(\hat{\sigma}_i^2 + \hat{\tau}^2_{REML})$.

Despite the number of methods available to estimate τ^2, they usually yield similar estimates of θ; "this may not be surprising because the weighted mean estimator $\hat{\theta}$ that is given in equation 2.3 for an overall effect is not particularly sensitive to the estimated weights" (Sidik & Jonkman, 2005, pg. 374). It is believed that results "are nearly invariant with respect to the choice of the between-study variance estimator" (Hartung & Knapp, 2001, pg. 3876). Moreover, in practice, the point estimates from the 2 methods (fixed or random-effects) can even vary only slightly from each other. Estimates of τ^2 are important for the calculation of $\hat{V}$, the variance of the overall estimate $\hat{\theta}$. In fact, in addition to point estimates, reporting the overall variance of the estimator and a confidence interval is usually considered a useful habit in order to indicate the precision of the overall effect estimate and therefore to stress the level of uncertainty about the point estimate.

2.5 Estimating the variance of the overall effect estimate $\hat{V}$: an ignored problem

In the random effects model for meta-analysis, an overall effect is usually estimated with a weighted average of the single effect measurements. Weights are given by the precision, i.e. by the inverse of the sum of the within-study and between-study variances. Such weighting is 'optimal' provided that the correct variances are used. Nevertheless, these values are unknown and weights used in practice are obtained by substituting estimated variances in place of the true ones. "Because of sampling error, however, the precision will be estimated with some inaccuracy" (Senn, 2000, pg. 546). Hence such a weighting is not optimal anymore. "Although such weights based on estimated values are incorrect and stochastic, and may have large errors in some cases, approximate inference about an overall effect typically ignores completely the errors associated with estimation of the marginal variances" (Sidik & Jonkman, 2006, pg. 3682). Moreover, the variance of $\hat{\theta}$, V, is often estimated by equation 2.4 obtained by using the estimated weights $\hat{w}_i^{-1} = \hat{\sigma}_i^2 + \hat{\tau}^2$ in place of the original 'correct' marginal variances, a practice which fails to account for the error associated with estimated weights. Clearly, if $\hat{\sigma}_i^2$ and $\hat{\tau}^2$ have substantial errors, then $\hat{w}_i^{-1}$ would be a poor estimate of the variance of each study summary measure; as a consequence, $\hat{V}$ could be unreliable as an estimator for the variance of θ. The accuracy of the estimated values of σ_i^2 and τ^2 is therefore decisive: both the variance and the confidence intervals of θ may be considerably affected by using different methods of estimating τ^2.

In practice, the point estimates from the two methods (i.e. fixed and random effects models) can even vary only slightly from each other, but the random-effects model leads to wider confidence intervals for the overall treatment effect. However, in the calculation of $\hat{\theta}$ and $var(\hat{\theta})$, since both τ^2 and σ_i^2 are assumed known when in practice they both are estimated, the confidence interval is still too narrow. The imprecision of the estimates of both τ^2 and σ_i^2 should be considered. In general, random-effects estimators tend to weight studies more equally, because of the presence of a common variance τ^2 contributing to the weights. In the case where the relative weight of each single trial is determined more by the value of τ^2, it may be acceptable to treat the standard errors as if they were known. Nonetheless, when σ_i^2's have a consistent influence on the weights, it is essential to allow for the imprecision of these estimates, whether fixed or random-effects analyses are used. In particular, in fixed-effects models both the weights and the variance of the overall estimator depend *solely* upon the within-study variances σ_i^2. Hence, the precision with which σ_i^2's are estimated will have a dramatic influence on weights and therefore on the overall precision of the estimator. Treating σ_i^2's as known particularly in fixed-effects analysis overestimates the precision and can lead to a loss of efficiency, especially for small trials. This should not be simply

ignored. As a consequence, this project will focus on the precision of estimates of σ_i^2 in fixed-effects meta-analysis.

Problems related to poor estimates of σ_i^2 have been addressed several times in the literature and better estimates of $\text{var}(Y_i)$ have been advocated by a number of authors. DerSimonian and Laird themselves warned the reader that in their work sampling variances were "assumed known even though in reality these were estimated from the data" and exhorted to do "further research" in this area and to investigate different methods of calculating the variances (DerSimonian & Laird, 1986, pg. 187). Viechtbauer recalls that all the methods briefly cited above concentrate on the study of τ^2 given the simplifying assumption that the sampling variances of the effect size estimates are known. "This is only approximately true when the within-study sample sizes are large (in this case, $\hat{\sigma}_i^2 \approx \sigma_i^2$). On the other hand, when the within-study sample sizes are small, then the error in the $\hat{\sigma}_i^2$ values cannot be simply ignored. A meta-analysis of a large number of studies with small sample sizes yields coverage probabilities that deviate quite substantially from the nominal level" (Viechtbauer, 2007, pg. 46, 47). Both random and fixed-effects models do not account nor allow for the sampling error in $\hat{\sigma}_i^2$ which is present in practice. "Inference is carried out ignoring the sampling errors in the individual study variances. Estimated values $\hat{\sigma}_i^2$ are used without modification to the form of $\hat{\theta}$, its variance or distribution" (Brockwell & Gordon, 2001, pg. 837). Given the imprecision in the estimate of σ_i^2, we may be severely overstating the precision of the estimated overall effect size. Confidence intervals for σ_i^2 could facilitate such sensitivity analyses by suggesting a possible range of σ_i^2 values one should consider. Confidence intervals may become anticonservative especially with increasing number of trials and small sample sizes (Knapp et al., 2006). In particular, the fewer the number of patients the less precise will be the estimate of σ_i^2, and this additional uncertainty would therefore be expected to have a great impact on the results (Hardy & Thompson, 1996). The problem of estimating σ_i^2's is particularly urgent and dramatic in the fixed-effects model, even though "the estimation of σ_i^2 might also be expected to influence random effects coverage probabilities" especially when all studies in the meta-analysis are small (Brockwell & Gordon, 2001, pg. 837).

Consider the case where there are many but small, equally sized trials and homoscedasticity applies (i.e. the variances of sampling errors are identical in each trial). The optimal approach is to weight every single trial equally. Fixed-effects meta-analysis will de facto weight inversely proportional to the observed variance. In so doing it will produce an estimator whose true variance is higher than that produced by equal weighting (the 'correct one') but which will appear to be lower, "claiming to have produced a lower standard error for what is, in fact, a less precise estimate" (Senn, 2000, pg. 547) . The

real observed variances will vary and weighting by these values will produce a variance estimate that is lower than that for the optimal estimator *just by chance*. However, the estimator is sub-optimal because its 'true' variance is higher. Unfortunately, the resulting meta-analysis overstates the significance of the results (Senn, 2000). As a consequence, significance tests associated with it are too liberal and confidence intervals do not have correct coverage properties (cf. the simulation).
"The problem is not severe if individual trials are not small" (Senn, 2000, pg. 547). On the other hand, when we consider a number of small trials, standard errors should not be considered as if they were known because this would overestimate precision and could also lead to unreliable results. Especially in these cases, investigation of better estimation of the variances is highly recommended and warmly supported.

2.6 Simulations with common variance

2.6.1 Number of Patients per Arm Equal

"Simulation studies use computer intensive procedures to test particular hypotheses and assess the appropriateness and accuracy of a variety of statistical methods in relation of the known truth. These techniques provide empirical estimation of the sampling distribution of the parameters of interest that could not be achieved from a single study and enable estimation of accuracy measures, such as the bias in the estimates of interest, as the truth is known" (Burton et al., 2006, pg. 4279).
Consider a meta-analysis of k similar but independent studies. The observations consist of two sets of independent random variables $X_{Ti1}, X_{Ti2}, ...X_{Tin_{Ti}}$ and X_{Ci1}, X_{Ci2}, ... $X_{Cin_{Ci}}$ for i=1, 2, ..., k from the treatment and the control groups, respectively. Note that n_{Ti} and n_{Ci} are respectively the study specific sample sizes for the treatment and the control groups in the ith study, so the total sample size is $N_i = n_{Ti} + n_{Ci}$. Suppose that these two sets of variables have independent normal distributions with different means and equal variances as follows

$$\begin{array}{l} X_{Ti1}, X_{Ti2}, \ldots, X_{Tin_{Ti}} \sim N(\mu_{Ti}, \sigma^2_{Ti}) \\ X_{Ci1}, X_{Ci2}, \ldots, X_{Cin_{Ci}} \sim N(\mu_{Ci}, \sigma^2_{Ci}) \end{array} \qquad \text{for} \quad i = 1, \ldots, k \quad \text{where} \quad \sigma^2_{Ti} = \sigma^2_{Ci}$$

The parameter of interest is the overall mean difference, denoted by θ. The study specific mean difference is defined as $Y_i = (\mu_{Ti} - \mu_{Ci})$ and is estimated by $Y_i = \overline{X}_{Ti} - \overline{X}_{Ci}$, where $\overline{X}_{Ti} = \sum_{j=1}^{n_{Ti}} X_{Tij}/n_{Ti}$ and $\overline{X}_{Ci} = \sum_{j=1}^{n_{Ci}} X_{Cij}/n_{Ci}$. We assume that Y_i is such that $E(Y_i) = \theta$ (each study has the same underlying effect) and that the variance of the difference between two independent means based on n_{Ti} and n_{Ci} observations respectively is equal to $var(Y_i) = \sigma_i^2 = \sigma^2_{Ti}/n_{Ti} + \sigma^2_{Ci}/n_{Ci} = S_i^2(1/n_{Ti} + 1/n_{Ci})$ given the assumption that $S_i^2 = \sigma^2_{Ti} = \sigma^2_{Ci}$ (i.e. the two groups in the treatment and control arms have the same variance). For moderately large study sizes, each Y_i should be

asymptotically normal distributed. Thus,

$$Y_i \sim N(\theta, \sigma_i^2) \quad for \quad i = 1, 2, ..., k$$

For the purpose of the simulation a *fixed-effects model* is considered, that is each study summary statistic Y_i is thought as a realization from a population of study estimates with common mean θ.

In addition to the point estimate, Confidence Intervals (CIs) for the overall mean difference (constructed based on the standard normal distribution) are calculated.
" The coverage probability of a random interval (A,B) for θ is defined as $Pr(\theta \in (A, B))$ which –for a nominal 95 per cent confidence interval– should be close to 0.95. The exact coverage can actually only be found if the distribution of the interval is known" (Brockwell & Gordon, 2001, pg. 831). However, as in this case, the distribution is unknown; this implies that the coverage probability must be estimated using simulation. "This is done by simulating a large number of meta-analyses and for each meta-analysis calculating the appropriate confidence interval" (Brockwell & Gordon, 2001, pg. 831). The *estimated coverage probability* is then the proportion of times that the obtained confidence interval contains the true specified parameter value θ. "The coverage should be approximately equal to the nominal coverage rate, e.g. 95 per cent of samples for the 95 per cent confidence intervals, to properly control the type I error rate for testing a null hypothesis of no effect. Over-coverage suggests that the results are too conservative as more simulations will not find a significant result when there is a true effect thus leading to a loss of statistical power with too many type II errors. In contrast, under-coverage (where the coverage rates are lower than 95%) is unacceptable as it indicates over-confidence in the estimates since more simulations will incorrectly detect a significant result, which leads to higher than expected type I errors"(Burton et al., 2006, pg. 4287).
The coverage probability is usually dependent on the parameters of the model and so the coverages presented are estimated for a range of values of S_i^2 and N_i. The value of θ is nevertheless irrelevant as "the procedure is invariant with respect to a location shift" (Brockwell & Gordon, 2001, pg. 831). For all simulations we use $\theta = 3$. The data for each meta-analysis is simulated using the fixed-effects model described above (i.e. $Y_i = \theta + e_i$), assuming normal errors e_i with zero mean and variances σ_i^2. The coverage probability is then estimated by simulating 10000 meta-analyses. The number of runs was set to 10000 in order to reduce the standard error of the simulation process for the nominal 95% coverage probability (cp) to 0.002179 (SE(cp)= $\sqrt{cp(1-cp)/M}$) without being computationally expensive. "A possible criterion for acceptability of the coverage is that the coverage should not fall outside of approximately two SEs of the nominal coverage probability (cp)" (Burton et al., 2006, pg. 4287); therefore, in our simulations between 9457 and 9543 of the 10000 confidence intervals should include the true value.

The simulations are implemented using a programme in **R**, with each simulation generating n_{Ti} and n_{Ci} observations from normal distributions with mean μ_{Ti} and $\mu_{Ci} = \mu_{Ti} + \theta$ respectively and variance S_i^2. This procedure is repeated k times and the data is then used to calculate the fixed-effects estimates for θ and the corresponding confidence interval.

In order to simplify the situation, let us assume that homoscedasticity applies, i.e. $S_i^2 = S^2$ for all i. Secondly, even if it is far from reality, we assume that all the studies have exactly the same size ($n = n_{Ti} = n_{Ci}$). This is just to give an indication of what happens in the case we consider a number of studies all of them with few patients involved; such assumption should therefore only be used as a rough guide of what would happen in an unlikely but still possible situation (large numbers of big studies is not a common occurrence in meta-analysis either).

Each estimated significance level is based on 10000 independent replications of the same model and the significance level is $\alpha = 0.05$. We discuss the meta-analytical combination of the results of k= 10, 15, 20, 35 clinical trials and N_i = 10, 16, 20, 30, 40, 60, 100 patients (i.e. as sample sizes we examine $(n_{Ti}, n_{Ci}) = (5,5), (8,8)$, (10,10), (15, 15), (20,20), (30,30) and (50, 50)). As regards the variances, we consider $S^2 = 1$. With this choice of these patterns we are able to give an impression about the general attitude of the fixed-effects meta-analysis when both the number of studies and the sample size change. We will summarise the estimates once all simulations have been performed. As in many published simulation studies, the average estimate of interest (i.e. the overall variance of the estimator) over the M simulations performed will be reported as a measure of the 'declared' estimate of interest. Similarly, as an assessment of the uncertainty in the estimate of interest between simulations, the variance of the estimates of the overall variance of the estimator from all simulations will be calculated. Moreover, in order to evaluate the performance of the obtained results from the different scenarios and approaches being studied, the coverage of the confidence intervals will be be considered as a measure of the performance and precision of the methods. When judging the performance of different methods, some argue than having less bias is more important than producing a valid estimate of sampling variance (Burton et al., 2006). In our case, not only has the empirical estimated coverage probability to correspond to the nominal value, but also- and in particular- the 'declared' precision as well as the 'real' dispersion of the overall variance of the estimator should be close to the theoretical 'optimal' one. The dispersion of the variances around the optimal value will be a good way to assess the goodness of the methods used: both the average declared variances and in particular the actual variances should be close to the optimal 'real' value.

In particular, when we perform meta-analysis, we consider the estimator

$$\theta = \frac{\sum_{i=1}^{k} w_i Y_i}{\sum_{i=1}^{k} w_i} \quad where \quad w_i = (\sigma_i^2)^{-1}$$

The variance of such estimator, V, is equal to

$$V = var(\theta) = \frac{1}{\sum_{i=1}^{k} w_i}$$

The demonstration is as follows

$$\begin{aligned} V &= var(\theta) = var\left(\frac{\sum_{i=1}^{k} w_i Y_i}{\sum_{i=1}^{k} w_i}\right) = \frac{\sum_{i=1}^{k} w_i^2 Var(Y_i)}{(\sum_{i=1}^{k} w_i)^2} = \frac{\sum_{i=1}^{k} w_i^2 \sigma_i^2}{(\sum_{i=1}^{k} w_i)^2} \\ &= \frac{\sum_{i=1}^{k} (\frac{1}{\sigma_i^2})^2 \sigma_i^2}{(\sum_{i=1}^{k} w_i)^2} = \frac{\sum_{i=1}^{k} w_i}{(\sum_{i=1}^{k} w_i)^2} = \frac{1}{\sum_{i=1}^{k} w_i} \end{aligned}$$

given that $var(Y_i) = \sigma_i^2$ and that $w_i = 1/\sigma_i^2$. However, strictly speaking, $var(\theta)$ is the 'true' variance of the 'correct' estimator ***only*** when σ_i^2's are known. When estimated weights are used both to determine θ and its variance, the equality is not valid anymore. Furthermore, if we assume that weights are fixed constants as they should be in this simulation scheme (i.e. $w_i = 1/k$), we obtain

$$V = var(\theta) = var\left(\frac{\sum_{i=1}^{k} w_i Y_i}{\sum_{i=1}^{k} w_i}\right) = \frac{\sum_{i=1}^{k} w_i^2 Var(Y_i)}{(\sum_{i=1}^{k} w_i)^2} = \frac{1}{k^2} \sum_{i=1}^{k} var(Y_i)$$

and in our simulation, as both the sample sizes for the treatment and the control groups are identical, it develops into

$$var(\theta) = \frac{2S^2}{kn} \qquad \text{where} \quad n = n_{Ti} = n_{Ci}$$

where S^2 is the variance of each arm (i.e. $S^2 = \sigma_{Ti}^2 = \sigma_{Ci}^2$), k is the number of trials considered and n is the number of patients per arm. This is the '*optimal*' value of the variance of θ provided that

$$\begin{aligned} \sigma_i^2 = var(Y_i) &= \left(\frac{1}{n_{Ti}} + \frac{1}{n_{Ci}}\right) \frac{(n_{Ti}-1)\sigma_{Ti}^2 + (n_{Ci}-1)\sigma_{Ci}^2}{(n_{Ti}-1)+(n_{Ci}-1)} = \\ &= \left(\frac{2}{n}\right) \frac{2(n-1)S^2}{2(n-1)} = \frac{2S^2}{n} \qquad \text{given} \quad n = n_{Ti} = n_{Ci}, \quad S^2 = \sigma_{Ti}^2 = \sigma_{Ci}^2 \end{aligned}$$

In practice, we use $\hat{S}_i^2$ (see equation 2.2) and therefore $\hat{\sigma}_i^2$ instead of the 'true' σ_i^2. Therefore, all the relationships described above hold ***if and only if*** σ_i^2 is perfectly estimated by $\hat{\sigma}_i^2$. This is hardly the case. As a consequence we usually end up considering the estimator

$$\hat{\theta} = \frac{\sum_{i=1}^{k} \hat{w}_i Y_i}{\sum_{i=1}^{k} \hat{w}_i}$$

and the variance of $\hat{\theta}$, $\hat{V}$, is estimated by

$$\hat{V} = \frac{1}{\sum_{i=1}^{k} \hat{w}_i} \qquad \text{with} \quad \hat{w}_i = 1/\hat{\sigma}_i^2$$

where

$$\hat{\sigma}_i^2 \;=\; \left(\frac{1}{n_{Ti}} + \frac{1}{n_{Ci}}\right)\hat{S}_i^2 = \left(\frac{1}{n_{Ti}} + \frac{1}{n_{Ci}}\right)\frac{(n_{Ti}-1)\hat{s}_{Ti}^2 + (n_{Ci}-1)\hat{s}_{Ci}^2}{(n_{Ti}+n_{Ci}-2)}$$

where $\hat{s}_{Ti}^2$ and $\hat{s}_{Ci}^2$ are sample variances.

Thanks to the simulations, it is possible in practice to calculate the ***'actual'* variance** of the estimates $\hat{\theta}$ obtained with $\hat{w}_i$ as well as the ***'declared'* variance** of the estimator. As at the end of 10000 runs we have $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_M$ with the respective variances $\hat{V}_1$, $\hat{V}_2$, $\ldots$, $\hat{V}_M$ where $M = 1, 2, \ldots, 10000$, we can calculate the real dispersion of the estimates as

$$V^{real}(\hat{\theta}) = \frac{\sum_{j=1}^{M}(\hat{\theta}_j - \overline{\hat{\theta}})^2}{(M-1)} \qquad \text{where} \quad \overline{\hat{\theta}} = \sum_{j=1}^{M} \frac{\hat{\theta}_j}{M} \tag{2.5}$$

and the expected variance $E[\hat{V}]$ as

$$E[\hat{V}] = \sum_{j=1}^{M} \frac{\hat{V}_j}{M} \tag{2.6}$$

Finally, since when simulating data we have the privilege to know the correct values for the treatment and control group variances, we can calculate the so called ***'optimal'* variance**, that is given by average of optimal variances calculated with the true variances instead of with the sampled values, that is

$$V^{opt} = \sum_{j=1}^{M} \frac{V_j}{M} \qquad \text{where} \quad V_j = \frac{1}{\sum_{i=1}^{k} w_i} \qquad \text{with} \quad w_i = 1/\sigma_i^2 \tag{2.7}$$

We give the results of the simulations if the fixed-effects model is the theoretically correct one. In Table 2.1, the estimated actual percentage of CIs is given, that is the proportion of intervals containing the true overall effect θ out of 10000 runs. We can observe, not surprisingly, that for small sample sizes the fixed effect model is rather liberal and that for increasing sample sizes in the studies the estimated coverage probability get closer to the nominal significance level (Böckenhoff & Hartung, 1998; Li et al., 1994). The proportion of intervals which contains θ drops to 90% in the balanced case $n_{Ti} = n_{Ci} = 8$ regardless of the number of trials considered. It is worth to note that with increasing sample sizes one observes a stabilization of the actual coverage probability. Moreover, from Table 2.1, it can be seen that the number of

trials involved in the meta-analysis does not have a huge impact on the results. In fact, turning to the coverage probabilities of the confidence intervals (Figure 2.1), we observe that the empirical coverage probabilities are below 95% for all k, even for a large number of studies taken into account. The empirical coverage probabilities are closer to the nominal values of 95% when a total number of 30 patients per study is considered. The four graphs show in fact the same pattern; with $n < 8$ the empirical coverage probability is far below the nominal value while with more than 15 patients per arm the figures tend to get closer to the nominal value of 95%. Therefore, the number of patients seems to be the most relevant aspect (at least in the case where only trials with the same dimension and with the same number of patients per arm are considered). The smaller the number of patients, the lower the coverage probability.

Moreover, Table 2.1 clearly shows that

$$V^{real}(\hat{\theta}) \geq V^{opt} \geq E[\hat{V}]$$

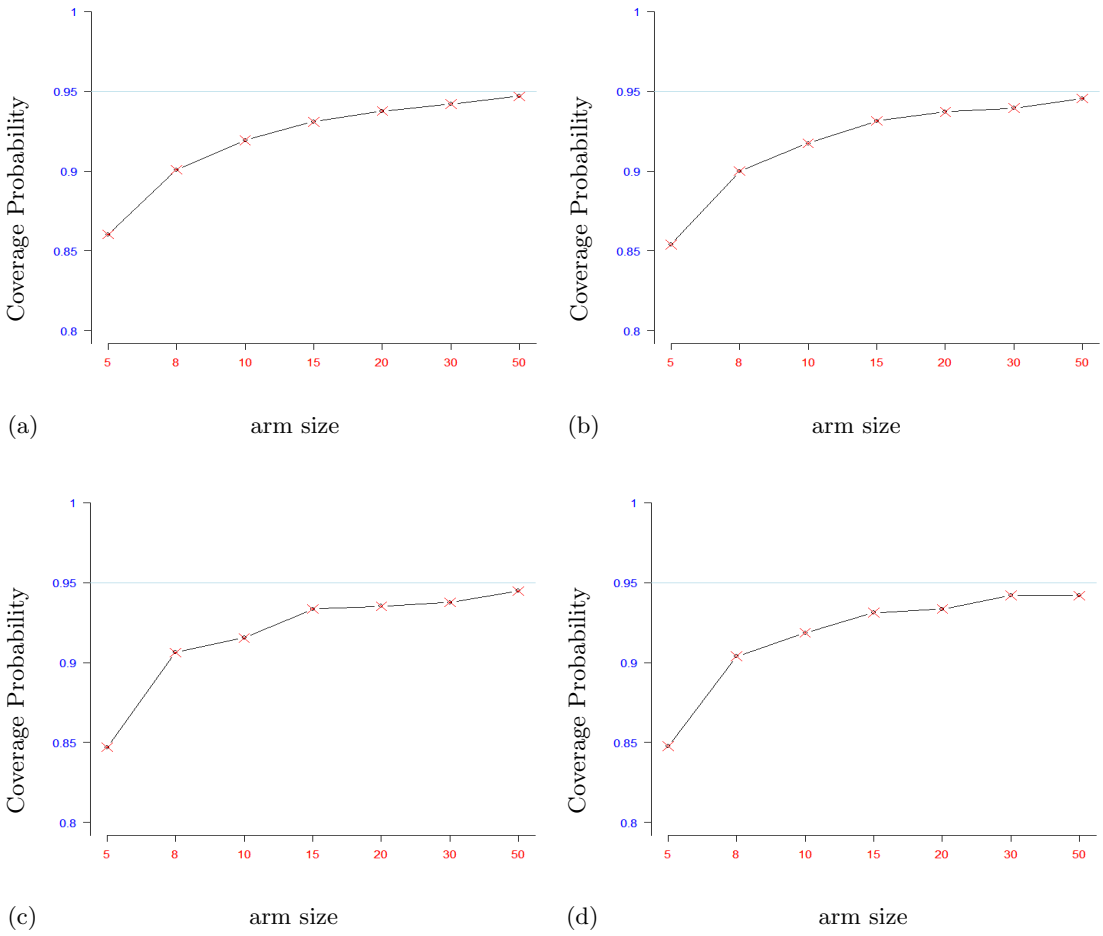
These relationships have a number of consequences. As we can only run *1* meta-analysis, *on average* we tend to assert that the variance of the estimate obtained is smaller that its 'true' optimal value. De facto, instead, the variability of such an estimate is much larger than the true one. In practice, as the Table 2.1 shows, there is the tendency to produce a variance estimate that is lower than that for the optimal estimator. Nevertheless, the estimator is sub-optimal since its 'actual' variance $V^{real}(\hat{\theta})$ is higher than V^{opt}. We claim that we are performing better than the optimal estimator while, if we could run a number of meta-analyses, we would notice that the variability of the estimate considered is larger than the 'true' one. Fixed-effects meta-analyses produce an estimator whose true variance is higher than that produced by equal weighting (the 'correct one') but which will appear to be lower. As a consequence, we tend to overstate the significance of the results.

Table 2.1: The Results of Simulations for different values of studies (k) under the assumption of equal variances and equal study sizes

k = 10	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	0.052044	0.040000	0.031227	0.8603
	8	0.029666	0.025000	0.021792	0.9008
	10	0.022188	0.020000	0.018039	0.9193
	15	0.014338	0.013333	0.012490	0.9309
	20	0.010618	0.010000	0.009534	0.9376
	30	0.006858	0.006667	0.006457	0.9420
	50	0.004087	0.004000	0.003928	0.9468
k = 15	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	0.037335	0.026667	0.020577	0.8539
	8	0.020113	0.016667	0.014495	0.9000
	10	0.015149	0.013333	0.011969	0.9173
	15	0.009435	0.008889	0.008297	0.9314
	20	0.007051	0.006667	0.006343	0.9371
	30	0.004589	0.004444	0.004302	0.9395
	50	0.002692	0.002667	0.002618	0.9454
k = 20	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	0.028918	0.020000	0.015329	0.8470
	8	0.014785	0.012500	0.010826	0.9063
	10	0.011415	0.010000	0.008959	0.9154
	15	0.007203	0.006667	0.006224	0.9334
	20	0.005288	0.005000	0.004746	0.9352
	30	0.003516	0.003333	0.003223	0.9376
	50	0.002057	0.002000	0.001961	0.9447
k = 35	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	0.016244	0.011429	0.008684	0.8476
	8	0.008535	0.007143	0.006152	0.9039
	10	0.006467	0.005714	0.005098	0.9186
	15	0.004160	0.003810	0.003546	0.9312
	20	0.003104	0.002857	0.002710	0.9334
	30	0.001997	0.001905	0.001841	0.9419
	50	0.001190	0.001143	0.001120	0.9418

This Table shows the results of simulation for $\theta = 3$ and different values of n and k, given $S^2 = \sigma^2_{Ti} = \sigma^2_{Ci} = 1$ for all i. This simulation scheme considers for each simulation k parallel group clinical trials each of whom with the same number of patients per arm ($n = n_T = n_C$). Empirical Statistics for $E(\hat{V})$ and $V^{real}(\hat{\theta})$ are based on 10000 simulation replicates as well as the Empirical Coverage Probability.

Figure 2.1: Estimated Coverage Probabilities for the Fixed-Effects method under the assumptions of Common Variances and Equal Number of Patients per Arm



The figures show Estimated Coverage Probabilities of the Confidence Intervals based on 10000 simulation replicates. Different values of k –the number of trials – and n –the number of patients per arm, with $n = n_C = n_T$– are considered: (a) $k = 10$, (b) $k = 15$, (c) $k = 25$, (d) $k = 35$.

2.6.2 Number of Patients per Arm not Equal

So far we have considered the case where the number of patients per each arm is equal. We now take into account the cases where $n_T \neq n_C$. In particular, we consider the case where on average all clinical trials have the same amount of patients per arm. This implies that $E(n_T) = E(n_C)$. Therefore, in order to consider clinical trial where each arm has on average the same amount of patients, we sample the dimension of each arm from a negative binomial distribution. The negative binomial distribution is a discrete probability distribution,commonly parameterized by two real-valued parameters p and r with $0 < p < 1$ and $r > 0$. Under this parameterization, the probability mass function of a random variable with a NegBin(r, p) distribution takes the following form:

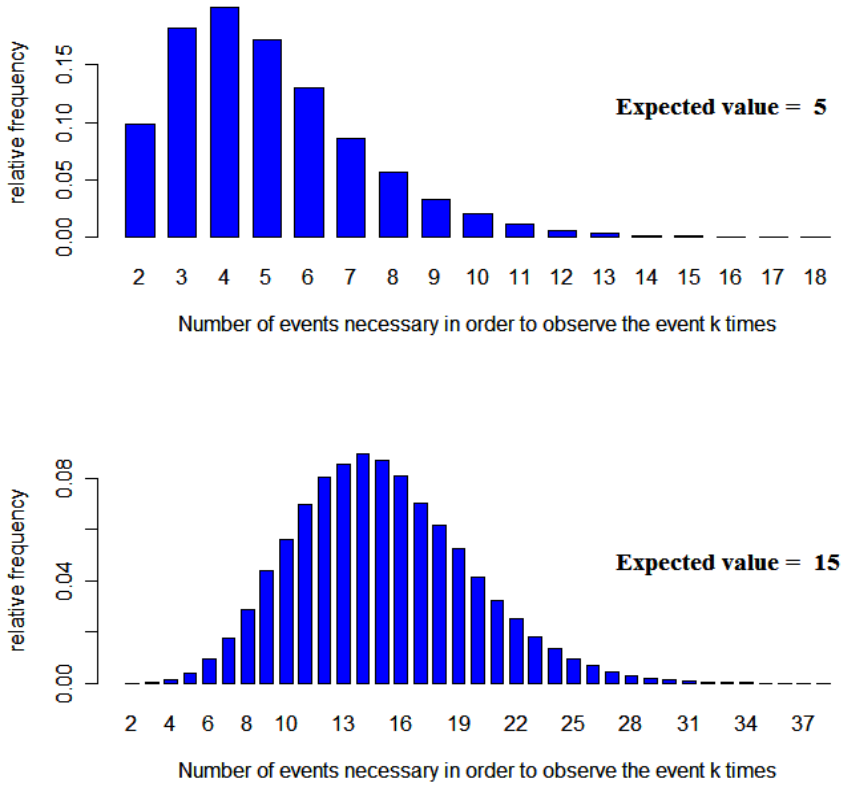
$$f(k; r, p) = \frac{\Gamma(r+k)}{k!\Gamma(r)} p^r (1-p)^k$$

for $k = 1, 2, 3, ...$ and where Γ is the Gamma Function. The NegBin(r, p) distribution is the probability distribution of a certain number of failures (r) in a series of independent and identically distributed Bernoulli trials given p as the probability of success (Piccolo, 2000). Specifically, this is the probability distribution of the number of failures before the k^{th} success in a Bernoulli process, with probability p of success on each trial. Formulae for the expectation and the variance for the negative binomial distribution are given by

$$E(X) = \frac{r(1-p)}{p} \qquad Var(X) = \frac{r(1-p)}{p^2}$$

Hence, if we impose the average number of patients per arm (with at least 2 patients per arm), we can obtain various negative binomial distributions each of which with diverse variances. In our study, we decided to consider the cases where the variances of the negative binomial distributions are equal to 5. Higher values of the variance were not taken into account as this would have implied extremely high number of patients per arm which are quite unrealistic. Given a variance equal to 5, the negative binomial distribution whose mean is 5 assumes values ranging from 2 to 18. Similarly, as shown in the following graphs, given a variance equal to 5 and an average value equal to 15, the barplot of such a probability distribution has values varying from 2 to 37.

Figure 2.2: Bar Plots of a Probability Distribution of a negative Binomial

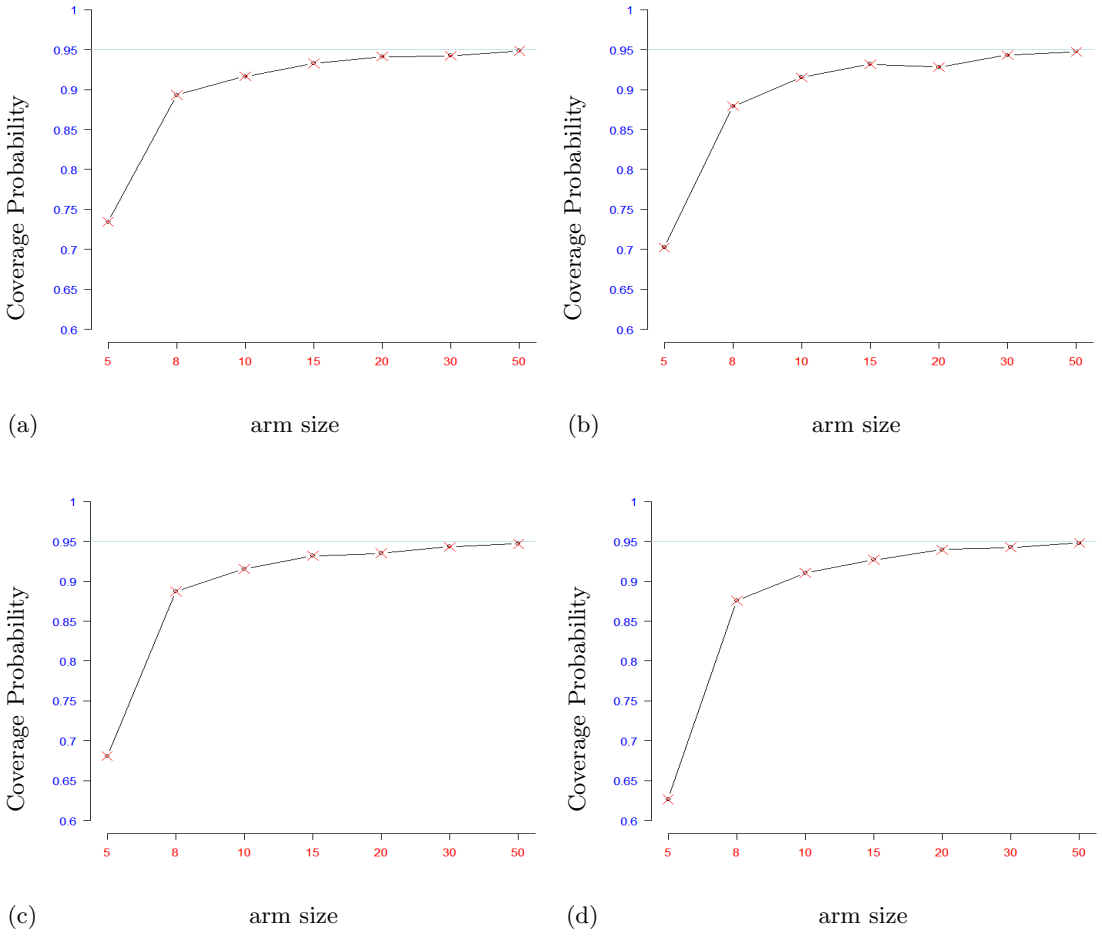


Bar Plots of a Probability Distribution of a negative Binomial with different expected values (5 and 15 respectively) and same variances in both cases equal to 5

We run 10000 meta-analyses assuming the fixed-effects model as the correct one. Table 2.2 shows the proportion of intervals containing the true overall effect θ as well as the 'real', the 'optimal' and the 'declared' variances of the estimates.
Once again, despite the fact we are now considering the case where n_T and n_C are the same only *on average*, we can observe (Figure 2.3) that the estimated coverage probability gets closer to the nominal level (i.e. 95 %) only when the sample sizes increases. The proportion of intervals which does not contain the true θ rises to 20% when we consider only 5 patients per arm on average. Interestingly, under the assumption of common variance, the number of clinical trials taken into account does not seem to have an important impact on the output. The empirical coverage probabilities assume roughly the same values regardless of k, the number of clinical trials. Again, looking at both Figure 2.3 and Table 2.2, we can observe that the coverage probability is closer to the nominal level when, on average, there are more than 15 people per arm, regardless of the number of studies. In fact, if we have a look at the ratio between the real dispersion of the estimates and the mean of the variances of these estimates we can

observe that, as the number of the average patients per arm increases, the ratio itself tends to be roughly the same for all the four scenarios considered (k = 10, 15, 20 and 35). What really matters is the average total number of patients per trial: the less the average amount of patients the lower is the coverage probability.

Figure 2.3: Estimated Coverage Probabilities for the Fixed-Effects method under the assumptions of Common Variances and Different Number of Patients per Arm



The figures show Estimated Coverage Probabilities of the Confidence Intervals based on 10000 simulation replicates. Different values of k and n (where $n = n_C = n_T$ only *on average*) are taken into account: (a) $k = 10$, (b) $k = 15$, (c) $k = 25$, (d) $k = 35$.

Considering the same average number of patients per arm entails that not necessarily every single trial has got the same allocation per arm and the same total number of patients. For instance, in a meta-analysis where each study has an average total number of 20 people it is likely to have clinical trials with more or less than 20 people allocated in a more or less extreme unbalanced way. This has 2 consequences. First, under the assumption of common variance, the average total number of patients appears to be more important than the amount of every single clinical trial. Second, the allocation per arm of these people does not seem to be significant.

In the following paragraphs the irrelevance of both the allocation and the amount of people per study is proven investigating the effect of random variation in variances on meta-analysis. In order to simplify the calculation, in a first instance we will consider only two clinical trials under the assumptions of (i) equal number of patients per trial when proving the insignificancy of the allocation and of (ii) balanced allocation when verifying the importance of the average total amount of patients considered. Subsequently, both the allocation and the total amount of patients will be taken into consideration at the same time.

Table 2.2: The Results of Simulations for different values of studies (k) under the assumption of equal variances and equal study sizes *on average*

n	**k**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	$\frac{V^{real}(\hat{\theta})}{E[\hat{V}]}$	**Coverage Probability**
5	**10**	0.13624	0.05532	0.03594	3.79055	0.7345
	15	0.11109	0.03660	0.02307	4.81575	0.7026
	20	0.09523	0.02726	0.01674	5.68843	0.6808
	35	0.08067	0.01545	0.00911	8.85376	0.6267
8	**10**	0.03897	0.03024	0.02572	1.515	0.8929
	15	0.02753	0.02000	0.01692	1.627	0.8793
	20	0.02031	0.01495	0.01264	1.607	0.8874
	35	0.01256	0.00850	0.00714	1.759	0.8756
10	**10**	0.02570	0.02280	0.02039	1.260	0.9166
	15	0.01738	0.01513	0.01345	1.292	0.9151
	20	0.01278	0.01130	0.01003	1.274	0.9153
	35	0.00763	0.00645	0.00571	1.335	0.9104
15	**10**	0.01536	0.01417	0.01327	1.158	0.9329
	15	0.01005	0.00941	0.00878	1.145	0.9315
	20	0.00752	0.00707	0.00658	1.142	0.9317
	35	0.00444	0.00403	0.00375	1.185	0.9268
20	**10**	0.01059	0.01034	0.00986	1.074	0.9413
	15	0.00748	0.00690	0.00655	1.141	0.9280
	20	0.00545	0.00517	0.00491	1.111	0.9351
	35	0.00307	0.00295	0.00280	1.095	0.9395
30	**10**	0.00702	0.00680	0.00659	1.064	0.9423
	15	0.00460	0.00453	0.00438	1.049	0.9430
	20	0.00351	0.00339	0.00328	1.070	0.9432
	35	0.00200	0.00194	0.00187	1.065	0.9425
50	**10**	0.00407	0.00406	0.00398	1.021	0.9484
	15	0.00271	0.00271	0.00266	1.021	0.9473
	20	0.00206	0.00203	0.00199	1.038	0.9468
	35	0.00117	0.00116	0.00114	1.030	0.9479

This Table shows the results of simulation for $\theta = 3$ and different values of n and k, given $S^2 = 1$. This simulation scheme considers for each simulation k parallel group clinical trials each of whom with the same number of patients per arm *on average*($E(n) = E(n_T) = E(n_C)$). Empirical Statistics for $E(\hat{V})$ and $V^{real}(\hat{\theta})$ are based on 10000 simulation replicates as well as the Empirical Coverage Probability.

2.7 Effect of Allocation of Patients per Arm on Meta-Analysis. 2 trials

We consider two clinical trials, each of which with the same total number of patients ($N = N_1 = N_2$). Moreover, we suppose that the variances are equal to

$$\sigma_1^2 = var(Y_1) = \left(\frac{1}{\alpha N} + \frac{1}{(1-\alpha)N}\right) S_1^2 = \left(\frac{1}{\alpha(1-\alpha)N}\right) S_1^2 = \gamma S_1^2$$

$$\sigma_2^2 = var(Y_2) = \left(\frac{1}{\beta N} + \frac{1}{(1-\beta)N}\right) S_2^2 = \left(\frac{1}{\beta(1-\beta)N}\right) S_2^2 = \delta S_2^2$$

where α and β represent the proportion of patients allocated to each arm of the trial, N is the total number of patient per clinical trial and $S_1^2 = S_2^2 = 1$. The *optimal* weight for trial 1 would be equal to

$$w_1^{opt} = \frac{\frac{1}{\sigma_1^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} = \frac{\frac{1}{\gamma S_1^2}}{\frac{1}{\gamma S_1^2} + \frac{1}{\delta S_2^2}} = \frac{\frac{1}{\gamma}}{\frac{\gamma+\delta}{\gamma\delta}} = \frac{\delta}{\gamma+\delta} \quad \text{given} \quad S_1^2 = S_2^2 = 1$$

while the *empirical* weight for trial number 1 would be

$$\begin{aligned}
\hat{w}_1^{emp} &= \frac{\frac{1}{\left(\frac{1}{\alpha N} + \frac{1}{(1-\alpha)N}\right)\hat{S}_1^2}}{\frac{1}{\left(\frac{1}{\alpha N} + \frac{1}{(1-\alpha)N}\right)\hat{S}_1^2} + \frac{1}{\left(\frac{1}{\beta N} + \frac{1}{(1-\beta)N}\right)\hat{S}_2^2}} = \frac{\frac{1}{\gamma\hat{S}_1^2}}{\frac{1}{\gamma\hat{S}_1^2} + \frac{1}{\delta\hat{S}_2^2}} = \\
&= \frac{1}{\gamma\hat{S}_1^2} * \frac{\gamma\hat{S}_1^2\delta\hat{S}_2^2}{[\delta\hat{S}_2^2 + \gamma\hat{S}_1^2]} = \frac{\delta\hat{S}_2^2}{[\delta\hat{S}_2^2 + \gamma\hat{S}_1^2]} \\
&= \frac{\delta}{\delta + \gamma\boldsymbol{r}} \qquad \text{given} \quad \frac{\hat{S}_1^2}{\hat{S}_2^2} = \boldsymbol{r}
\end{aligned}$$

Therefore the *optimal* variance of the estimator will be equal to

$$\begin{aligned}
V^{opt} &= var(\theta^{opt}) = \left[\frac{\delta S_2^2}{\delta S_2^2 + \gamma S_1^2}\right]^2 \gamma S_1^2 + \left[\frac{\delta S_2^2}{\delta S_2^2 + \gamma S_1^2}\right]^2 \delta S_2^2 = \\
&= \frac{1}{[\gamma+\delta]^2}\left(\gamma\delta^2 + \gamma^2\delta\right) = \gamma\delta\frac{[\gamma+\delta]}{[\gamma+\delta]^2} = \frac{\gamma\delta}{\gamma+\delta} = \frac{1}{\frac{1}{\gamma}+\frac{1}{\delta}}
\end{aligned}$$

Similarly, the *empirical* overall variance will become.

$$\begin{aligned}
\hat{V}^{emp} &= var(\theta^{emp}) = \left[\frac{\delta\hat{S}_2^2}{\gamma\hat{S}_1^2 + \delta\hat{S}_2^2}\right]^2 \gamma S_1^2 + \left[\frac{\gamma\hat{S}_1^2}{\gamma\hat{S}_1^2 + \delta\hat{S}_2^2}\right]^2 \delta S_2^2 = \\
&= \left[\frac{\delta}{\delta + \gamma\boldsymbol{r}}\right]^2 \gamma + \left[\frac{\gamma\boldsymbol{r}}{\delta + \gamma\boldsymbol{r}}\right]^2 \delta = \frac{1}{[\delta + \gamma\boldsymbol{r}]^2}\left(\delta^2\gamma + \gamma^2\boldsymbol{r}^2\delta\right) = \\
&= \delta\gamma\frac{(\delta + \gamma\boldsymbol{r}^2)}{(\delta + \gamma\boldsymbol{r})^2}
\end{aligned}$$

And the ratio of this to the optimal will be

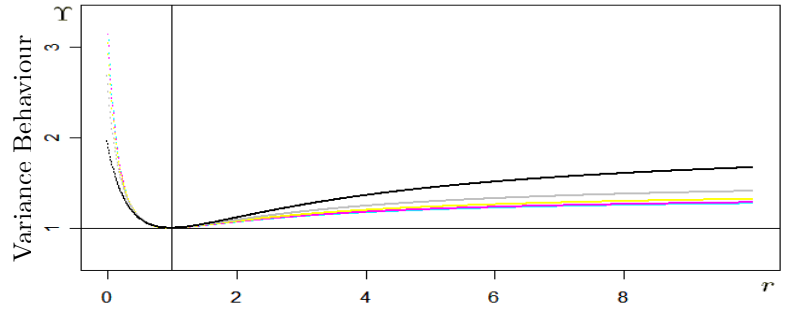
$$\Upsilon = \frac{var(\theta^{emp})}{var(\theta^{opt})} = \frac{\delta\gamma\frac{(\delta+\gamma\boldsymbol{r}^2)}{(\delta+\gamma\boldsymbol{r})^2}}{\frac{\gamma\delta}{\gamma+\delta}} = \frac{(\delta+\gamma\boldsymbol{r}^2)}{(\delta+\gamma\boldsymbol{r})^2}(\delta+\gamma) \tag{2.8}$$

where γ and δ depend on the proportion of patients per arm and where $\boldsymbol{r}$ can assume values theoretically in the range $(0, \infty)$. Actually, such a proportion does not depend on the total number of patients per clinical trials if we assume that both trials have the same dimension. In fact,

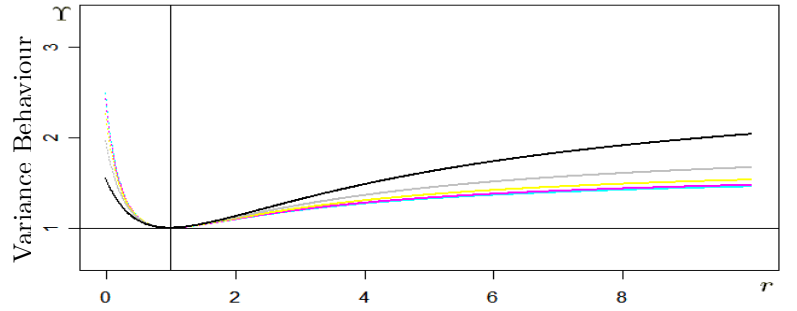
$$\begin{aligned}
\frac{(\delta+\gamma\boldsymbol{r}^2)}{(\delta+\gamma\boldsymbol{r})^2}(\delta+\gamma) &= \frac{\left(\frac{1}{\alpha(1-\alpha)N}+\frac{\boldsymbol{r}^2}{\beta(1-\beta)N}\right)\left(\frac{1}{\alpha(1-\alpha)N}+\frac{1}{\beta(1-\beta)N}\right)}{\left(\frac{1}{\alpha(1-\alpha)N}+\frac{\boldsymbol{r}}{\beta(1-\beta)N}\right)^2} = \\
&= \frac{\frac{1}{N^2}\left(\frac{1}{\alpha(1-\alpha)}+\frac{\boldsymbol{r}^2}{\beta(1-\beta)}\right)\left(\frac{1}{\alpha(1-\alpha)}+\frac{1}{\beta(1-\beta)}\right)}{\frac{1}{N^2}\left(\frac{1}{\alpha(1-\alpha)}+\frac{\boldsymbol{r}}{\beta(1-\beta)}\right)^2} = \\
&= \frac{\left(\frac{1}{\alpha(1-\alpha)}+\frac{\boldsymbol{r}^2}{\beta(1-\beta)}\right)\left(\frac{1}{\alpha(1-\alpha)}+\frac{1}{\beta(1-\beta)}\right)}{\left(\frac{1}{\alpha(1-\alpha)}+\frac{\boldsymbol{r}}{\beta(1-\beta)}\right)^2}
\end{aligned}$$

This means that the ratio Υ only depends on α, β and on $\boldsymbol{r}$. In general, to give an idea of the behaviour of the ratio we can plot it with different values of α and β. The following graphs (Figure 2.4) give us a rough idea in the cases where α assumes the values (0.1, 0.2, 0.3 and 0.5) while β ranges from 0.1 to 0.9.

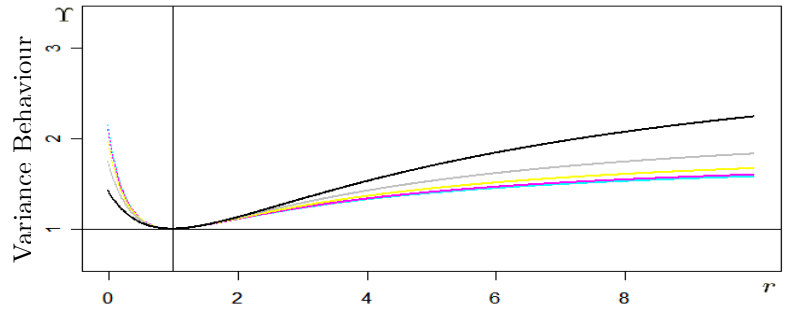
Figure 2.4: Random Variation in Variances on a Meta-Analysis with 2 trials. Effect of Allocation of patients per arm.



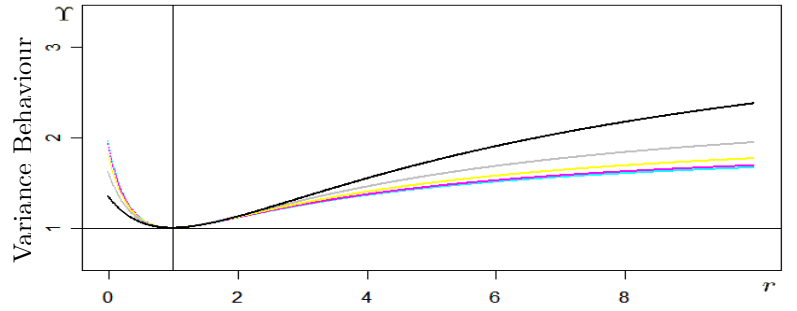
(a) Variance Ratio



(b) Variance Ratio



(c) Variance Ratio

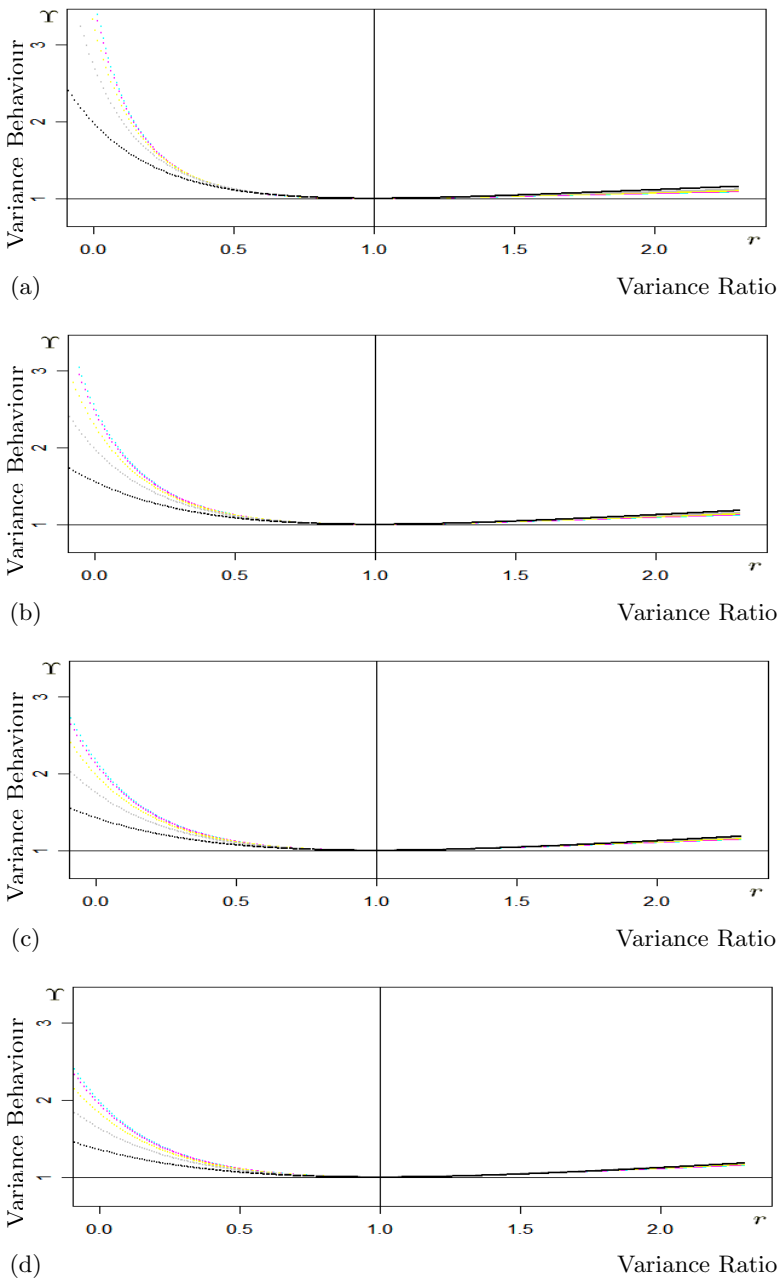


(d) Variance Ratio

These Figures show the variance behaviour of the ratio of the real overall variance to the optimal value as $\boldsymbol{r}$ increases and under the assumption of 2 studies each of which with the same total number of patients N. The 5 lines represent different values of β (– = 0.9, 0.1, – = 0.8, 0.2, – = 0.7, 0.3, – = 0.6, 0.4, – = 0.5) while α is set equal to 0.1 (a), 0.2 (b), 0.3 (c) and 0.5(d).

To have a better idea we can even plot them on a logarithmic base (Figure 2.5).

Figure 2.5: Random Variation in Variances on a Meta-Analysis with 2 trials. Effect of Allocation of patients per arm - Logarithm Scale.



These Figures show the variance behaviour of the ratio of the real overall variance to the optimal value as $\boldsymbol{r}$ increases and under the assumption of 2 studies each of which with the same total number of patients N. Again, the 5 lines represent different values of β (... = 0.9, 0.1, ... = 0.8, 0.2, ... = 0.7, 0.3, ... = 0.6, 0.4, ... = 0.5) while α is set equal to 0.1 (a), 0.2 (b), 0.3 (c) and 0.5(d). The ***logarithm*** scale is considered.

As $\boldsymbol{r}$ ranges from simulation to simulation, we can see the average behaviour of the ratio of the empirical and the optimal variances Υ calculating the mean. As we have assumed an equal total number of patients per each clinical trial, i.e. $N_1 = N_2 = N$, it follows that $\boldsymbol{r} = \frac{\hat{S}_1^2}{\hat{S}_2^2} \sim F_{(N-2),(N-2)}$, where F is the F distribution with both degrees of freedom equal to N-2. Given α, β and N, the expected value of the ratio is equal to

$$E(\alpha, \beta, N) = \int_0^\infty \Upsilon(\alpha, \beta, \boldsymbol{r}) * dF(\boldsymbol{r}, _{N-2}, _{N-2}) d\boldsymbol{r}$$

where

$$dF(\boldsymbol{r}, _{\nu_1}, _{\nu_2}) = \frac{1}{Beta(\frac{\nu_1}{2}, \frac{\nu_2}{2})} \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} r^{\frac{\nu_1}{2}-1}}{\left(1 + \frac{\nu_1 r}{\nu_2}\right)^{\frac{\nu_1+\nu_2}{2}}}$$

given $F_{(\nu_1,\nu_2)}$ with ν_1 and ν_2 representing the degrees of freedom. The following tables show the expected value of the ratio between the variance of the optimal and empirical estimator. Each table shows the expected values for the main possible combination of the proportion of total number of patients per arm, i.e. α and β.

Table 2.3: Expected values of the ratio of the empirical overall variance to the optimal. Effect of Allocation for N=30

	0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
0.1	1.0345	1.0325	1.0304	1.0291	1.0287	1.0291	1.0304	1.0325	1.0345
0.2	1.0325	1.0345	1.034	1.0335	1.0333	1.0335	1.034	1.0345	1.0325
0.3	1.0304	1.034	1.0345	1.0344	1.0343	1.0344	1.0345	1.034	1.0304
0.4	1.0291	1.0335	1.0344	1.0345	1.0345	1.0345	1.0344	1.0335	1.0291
0.5	1.0287	1.0333	1.0343	1.0345	1.0345	1.0345	1.0343	1.0333	1.0287
0.6	1.0291	1.0335	1.0344	1.0345	1.0345	1.0345	1.0344	1.0335	1.0291
0.7	1.0304	1.034	1.0345	1.0344	1.0343	1.0344	1.0345	1.034	1.0304
0.8	1.0325	1.0345	1.034	1.0335	1.0333	1.0335	1.034	1.0345	1.0325
0.9	1.0345	1.0325	1.0304	1.0291	1.0287	1.0291	1.0304	1.0325	1.0345

This Table shows the expected means of the ratio of the real overall variance to the optimal value when different allocations per arm per study are considered. The rows represent the allocation in study 1 (i.e. α), while the columns represent possible values of β, the allocation in the study 2. The total number of patients in both studies (N_1, N_2) is set to 30.

Table 2.4: Expected values of the ratio of the empirical overall variance to the optimal. Effect of Allocation for N=80

	0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
0.1	1.0127	1.0118	1.0108	1.0103	1.0101	1.0103	1.0108	1.0118	1.0127
0.2	1.0118	1.0127	1.0125	1.0122	1.0121	1.0122	1.0125	1.0127	1.0118
0.3	1.0108	1.0125	1.0127	1.0126	1.0126	1.0126	1.0127	1.0125	1.0108
0.4	1.0103	1.0122	1.0126	1.0127	1.0127	1.0127	1.0126	1.0122	1.0103
0.5	1.0101	1.0121	1.0126	1.0127	1.0127	1.0127	1.0126	1.0121	1.0101
0.6	1.0103	1.0122	1.0126	1.0127	1.0127	1.0127	1.0126	1.0122	1.0103
0.7	1.0108	1.0125	1.0127	1.0126	1.0126	1.0126	1.0127	1.0125	1.0108
0.8	1.0118	1.0127	1.0125	1.0122	1.0121	1.0122	1.0125	1.0127	1.0118
0.9	1.0127	1.0118	1.0108	1.0103	1.0101	1.0103	1.0108	1.0118	1.0127

This Table shows the expected means of the ratio of the real overall variance to the optimal value when different allocations per arm per study are considered. The rows represent the allocation in study 1 (i.e. α), while the columns represent possible values of β, the allocation in the study 2. The total number of patients in both studies (N_1, N_2) is set to 80.

Regardless of the combination of the values α and β, given N, it appears that our function Υ tends to assume roughly the same values; this implies that the allocation per arm does not affect the analysis. What really matters is the total number of patients per clinical trial and not the way patients are allocated in each arm. In other words, the crucial factor is the degrees of freedom available for estimating the within trial variance. In fact, as the total number per clinical trial increases, the expected values tend to be closer to 1 (cf. Tables 2.3 and 2.4, with N=30 and N=80 respectively).

2.8 Effect of Number of patients per Trial on meta-analysis.

2 Trials Example

In this case we consider two clinical trials with a given total number of patients equal to 2N. However, in this case, the number of patients per trial will vary. This means that we can consider $2n_1$ patients in trial 1 and $2n_2$ in trial 2, where $2n_1 + 2n_2 = 2N$. As the allocation per arm does not really affect the meta-analysis in the case where the variance is supposed to be equal in both clinical trials, we consider a balanced allocation (i.e. n_1 patients per arm in the clinical trial 1 and n_2 patients per arm in the clinical trial 2) in order to simplify the calculation. Moreover, we suppose that the variances are equal to

$$\sigma_1^2 = var(Y_1) = \left(\frac{1}{n_1} + \frac{1}{n_1}\right) S_1^2 = \left(\frac{2}{n_1}\right) S_1^2$$

$$\sigma_2^2 = var(Y_2) = \left(\frac{1}{n_2} + \frac{1}{n_2}\right) S_2^2 = \left(\frac{2}{n_2}\right) S_2^2$$

where $S_1^2 = S_2^2 = 1$.
In this case, the optimal scheme will be to weight the trials according to the numbers of patients, that is to use weights equal to n_1/N and to n_2/N respectively. In fact, if we consider only the trial 1, we have that the *optimal* weight is equal to

$$w^{opt} = \frac{\frac{n_1}{2S_1^2}}{\frac{n_1}{2S_2^2} + \frac{n_2}{2S_2^2}} = \frac{\frac{n_1}{2S_1^2}}{\frac{n_1 S_2^2 + n_2 S_1^2}{2S_1^2 S_2^2}} = \frac{n_1}{n_1 + n_2} = \frac{n_1}{N}$$

given that in the optimal scenario S_1^2 and S_2^2 are known and equal to 1 and where $n_1 + n_2 = N$. Given the optimal weights, the optimal estimator will have an overall variance equal to 2/N. In fact, recalling that when we perform meta-analysis, we consider the estimator

$$\theta = \frac{\sum_{i=1}^k w_i Y_i}{\sum_{i=1}^k w_i} = \sum_{i=1}^k q_i Y_i \quad \text{where} \quad q_i = \frac{w_i}{\sum_{i=1}^k w_i} \quad \text{with} \quad w_i = 1/\sigma_i^2$$

and that the overall variance of such estimator, V, is usually estimated by

$$V = var(\theta) = \frac{1}{\sum_{i=1}^k w_i} \quad \text{with} \quad w_i = \sigma_i^2$$

the *optimal* overall variance becomes

$$V^{opt} = \left(\frac{n_1}{N}\right)^2 \frac{2}{n_1} + \left(\frac{n_2}{N}\right)^2 \frac{2}{n_2} = \frac{2(n_1 + n_2)}{N^2} = \frac{2}{N}$$

However, the estimated weights as well as the estimated overall variance should take into account the fact that the variance is unknown. De facto, we could impose and presume to know the variance of each single trial when it comes to calculate the overall

variance of the estimator. Nonetheless, the weights will inevitably depend on the observed variances in each trial. Therefore, for trial 1 -for instance - the weight will be equal to

$$\begin{aligned} w^{emp} &= \frac{\frac{n_1}{2\hat{S}_1^2}}{\frac{n_1}{2\hat{S}_2^2}+\frac{n_2}{2\hat{S}_2^2}} = \frac{\frac{n_1}{2\hat{S}_1^2}}{\frac{n_1\hat{S}_2^2+n_2\hat{S}_1^2}{2\hat{S}_1^2\hat{S}_2^2}} = \frac{n_1}{2\hat{S}_1^2}\left(\frac{2\hat{S}_1^2\hat{S}_2^2}{n_1\hat{S}_2^2+n_2\hat{S}_1^2}\right) = \\ &= \frac{n_1\hat{S}_2^2}{n_1\hat{S}_2^2+n_2\hat{S}_1^2} = \frac{n_1}{n_1+n_2\boldsymbol{r}} \end{aligned}$$

given $\boldsymbol{r}$ equal to the observed ratio of within trial variances for the trial 1 compared to the trial 2 (i.e. $\boldsymbol{r} = \hat{S}_1^2/\hat{S}_2^2$). As a consequence, the overall observed variance, in the case where we take into account the observed variances only when calculating the weights of the trials, becomes

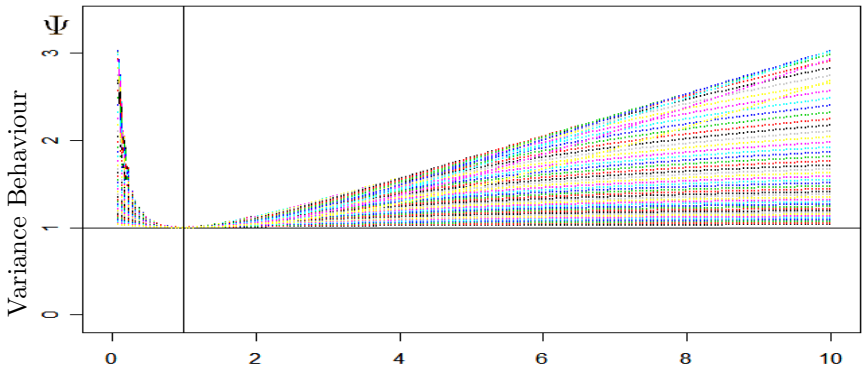
$$\hat{V}^{emp} = \left(\frac{n_1}{n_1+n_2\boldsymbol{r}}\right)^2\frac{2}{n_1} + \left(\frac{n_2\boldsymbol{r}}{n_1+n_2\boldsymbol{r}}\right)^2\frac{2}{n_2} = 2\frac{(n_1+n_2\boldsymbol{r}^2)}{(n_1+n_2\boldsymbol{r})^2}$$

The ratio of this variance to the optimal one will be equal

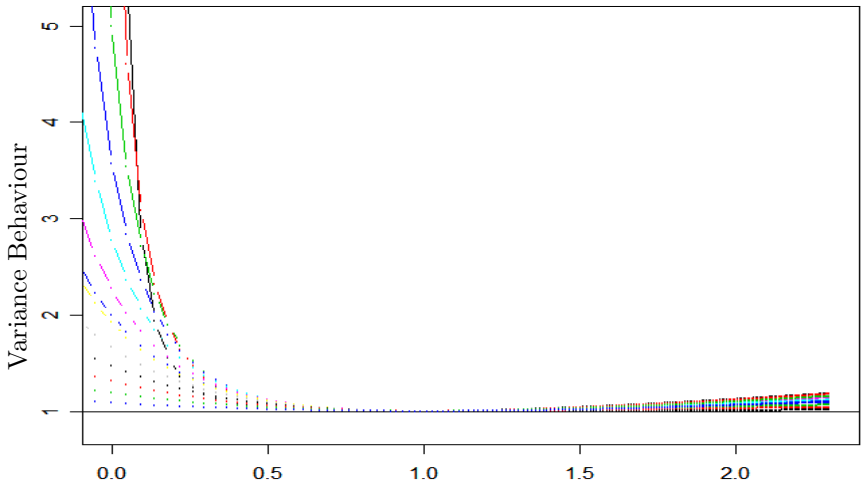
$$\Psi = \frac{var(\theta^{emp})}{var(\theta^{opt})} = \frac{2\frac{(n_1+n_2\boldsymbol{r}^2)}{(n_1+n_2\boldsymbol{r})^2}}{\frac{2}{N}} = \frac{(n_1+n_2\boldsymbol{r}^2)}{(n_1+n_2\boldsymbol{r})^2}(n_1+n_2) \tag{2.9}$$

where n_1 and n_2 represent the total number of patients per arm in each clinical trial and $\boldsymbol{r}$ (the ratio of the observed variances) assumes theoretically values $\in (0,\infty)$. This means that the ratio depends only on the *average total number of patients* and on $\boldsymbol{r}$. In general, to give an idea of the behaviour of the ratio we can plot it with different values of n_1 and n_2. The following graphs (Figure 2.6) give us a rough idea in the cases where N assumes the value 50 with at least 5 people per clinical trial on a Normal (a) and a Logarithmic scale (b).

Figure 2.6: Random Variation in Variances on a Meta-analysis with 2 trials. Effect of Number of patients per trial



(a)



(b)

These figures show the variance inflation of the ratio of the real overall variance to the optimal overall variance (Ψ) as $\boldsymbol{r}$ –the observed ratio of within trial variances for trial 1 compared to trial 2– increases. Each line represents a possible combination of $2n_1$ and $2n_2$ under the constraint of a total number of patients $2n_1 + 2n_2 = 50$ and under the assumption of a balanced allocation in each trial. Both the **Normal Scale (a)** and the **Logarithmic Scale (b)** are considered

As $\boldsymbol{r}$ ranges from simulation to simulation, we can see the average behaviour of the ratio of the empirical and the optimal variances calculating the mean. As we have assumed an average total number of patients equal to 2N, it follows that $\boldsymbol{r} = \frac{\hat{S}_1^2}{\hat{S}_2^2} \sim F_{(2n_1-2),(2N-2n_1-2)}$, where F is the F distribution with degrees of freedom equal to $2n_1 - 2$ and to $2n_2 - 2$. The expected value of the ratio, given N, is then equal to

$$E(n_1, n_2) = \int_0^\infty \Psi(n_1, n_2, \boldsymbol{r}) * dF(\boldsymbol{r}, 2n_1-2, 2n_2-2)d\boldsymbol{r}$$

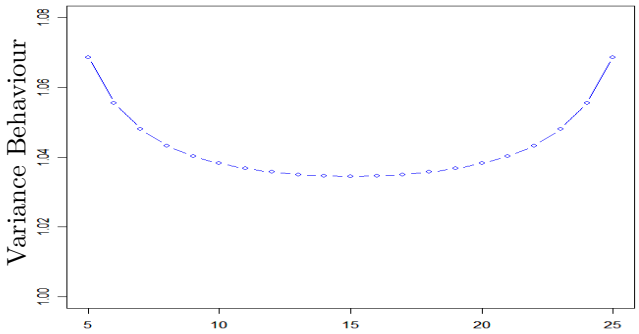
where

$$dF(\boldsymbol{r}, \nu_1, \nu_2) = \frac{1}{Beta(\frac{\nu_1}{2}, \frac{\nu_2}{2})} \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} r^{\frac{\nu_1}{2}-1}}{\left(1 + \frac{\nu_1 r}{\nu_2}\right)^{\frac{\nu_1+\nu_2}{2}}}$$

given $F_{(\nu_1,\nu_2)}$ with ν_1 and ν_2 representing the degrees of freedom.

As a consequence, if we give the total number of patients 2N we can compute the average of the ratio between the observed and the optimal overall variances for the estimator for all possible combinations of n_1 and n_2. Imposing that each single trial has at least 5 patients in total, we can see that the the ratio has on average values close to 1. Moreover, as the average total number of patients increases, the figures tend to be closer to 1 (i.e. the overall empirical and optimal variances of the estimator tend to be the same).

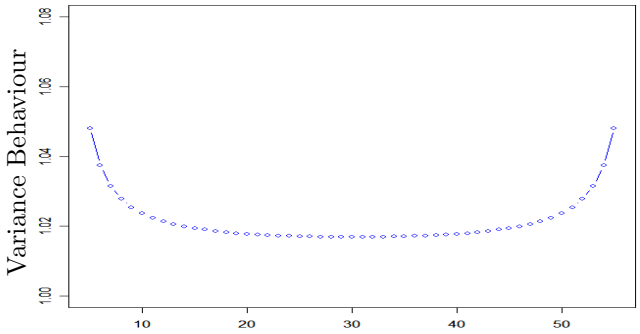
Figure 2.7: Expected Variance Inflation for N=30 on a Meta-Analysis with 2 trials



(a)

This figure shows the Expected Variance Inflation on a Meta-Analysis with only 2 trials for different number of patients (perfectly balanced among arms) in each of the 2 studies; the x axis represents the patients allocated in trial 1, given a total of 30 patients in the 2 trials

Figure 2.8: Expected Variance Inflation for N=60 on a Meta-Analysis with 2 trials



This figure shows the Expected Variance Inflation on a Meta-Analysis with only 2 trials for different numbers of patients (perfectly balanced in the 2 arms) in 2 studies. The x axis represents the patients allocated in trial 1, given that a total number of 60 patients in the 2 trials is considered

2.9 Effect of Patients and Allocation per Arm. 2 Trials Example

Let the variances of responses for trial 1 and 2 be both equal to 1. However, in this case, both the number of patients per trial and the allocation per arm in each trial will vary. Suppose n_1 patients in trial 1 and n_2 in trial two for a total number of N patients. Moreover, let's denote the allocation of patients in one arm as α and β for trial 1 and 2 respectively. Hence,

$$\sigma_1^2 = var(Y_1) = \left(\frac{1}{\alpha n_1} + \frac{1}{(1-\alpha)n_1}\right) S_1^2 = \left(\frac{1}{\alpha(1-\alpha)n_1}\right) S_1^2$$

$$\sigma_2^2 = var(Y_2) = \left(\frac{1}{\beta n_2} + \frac{1}{(1-\beta)n_2}\right) S_2^2 = \left(\frac{1}{\beta(1-\beta)n_2}\right) S_2^2$$

The optimal scheme to weight the trials will therefore depend on both the allocation and the number of people involved in each study. In fact, the *optimal* weight is equal to

$$w_1^{opt} = \frac{\frac{1}{var(Y_1)}}{\frac{1}{var(Y_1)} + \frac{1}{var(Y_2)}} \overset{S_1^2=S_2^2=1}{=} \frac{\alpha(1-\alpha)n_1}{\alpha(1-\alpha)n_1 + \beta(1-\beta)n_2}$$

These weights yields an estimator with an *optimal* overall variance given by

$$\begin{aligned} V^{opt} = var(\theta) &= (w_1^{opt})^2 var(Y_1) + (w_2^{opt})^2 var(Y_2) = \\ &= \left[\frac{\alpha(1-\alpha)n_1}{\alpha(1-\alpha)n_1 + \beta(1-\beta)n_2}\right]^2 \frac{S_1^2}{\alpha(1-\alpha)n_1} \quad + \\ &\qquad \left[\frac{\beta(1-\beta)n_2}{\alpha(1-\alpha)n_1 + \beta(1-\beta)n_2}\right]^2 \frac{S_2^2}{\beta(1-\beta)n_2} = \\ &= \frac{1}{\alpha(1-\alpha)n_1 + \beta(1-\beta)n_2} \qquad \text{given} \quad S_1^2 = S_2^2 = 1 \end{aligned}$$

Now, let us consider the *observed* variances. The *empirical* weight for trial 1 becomes

$$w_1^{emp} = \frac{\frac{\alpha(1-\alpha)n_1}{\hat{S}_1^2}}{\frac{\alpha(1-\alpha)n_1}{\hat{S}_1^2} + \frac{\beta(1-\beta)n_2}{\hat{S}_2^2}} = \frac{\alpha(1-\alpha)n_1}{\hat{S}_1^2} \times \frac{\hat{S}_1^2\hat{S}_2^2}{\hat{S}_2^2\alpha(1-\alpha)n_1 + \hat{S}_1^2\beta(1-\beta)n_2} =$$

$$= \frac{\alpha(1-\alpha)n_1}{\alpha(1-\alpha)n_1 + \beta(1-\beta)n_2\boldsymbol{r}} \qquad \text{given} \quad \hat{S}_1^2/\hat{S}_2^2 = \boldsymbol{r}$$

and the overall empirical variance of the estimator becomes

$$\hat{V}^{emp} = var(\hat{\theta}) = \left(\frac{\alpha(1-\alpha)n_1}{\alpha(1-\alpha)n_1 + \beta(1-\beta)n_2\boldsymbol{r}}\right)^2 \frac{1}{\alpha(1-\alpha)n_1} +$$

$$\left(\frac{\beta(1-\beta)n_2\boldsymbol{r}}{\alpha(1-\alpha)n_1 + \beta(1-\beta)n_2\boldsymbol{r}}\right)^2 \frac{1}{\beta(1-\beta)n_2} =$$

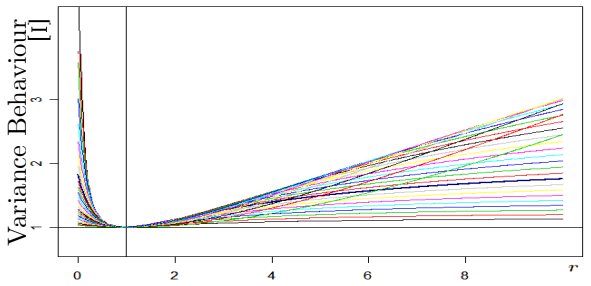
$$= \frac{\alpha(1-\alpha)n_1 + \beta(1-\beta)n_2\boldsymbol{r}^2}{[\alpha(1-\alpha)n_1 + \beta(1-\beta)n_2\boldsymbol{r}]^2}$$

The ratio of this variance to the optimal one will be equal to

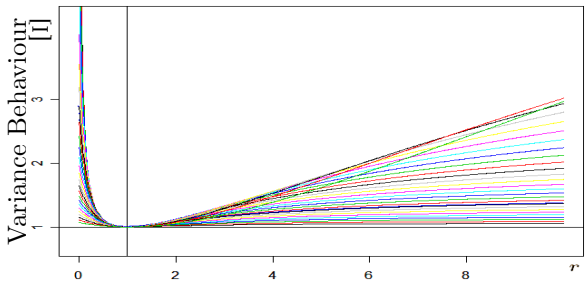
$$\Xi = \frac{var(\hat{\theta})}{var(\theta)} = \frac{\alpha(1-\alpha)n_1 + \beta(1-\beta)n_2\boldsymbol{r}^2}{[\alpha(1-\alpha)n_1 + \beta(1-\beta)n_2\boldsymbol{r}]^2} \times [\alpha(1-\alpha)n_1 + \beta(1-\beta)n_2]$$

Ξ depends on α, β, n_1 and n_2. Just to give an idea of the behaviour of Ξ when $\boldsymbol{r}$ ranges from 0 to 10, we can plot this function for different values of n_1 and n_2, given α and allowing for different values of β or viceversa. The following four plots (Figure 2.9), for example, show the behaviour of Ξ for all the possible combinations of n_1 and n_2 given N=30 and for four different values of α (0.1, 0.3, 0.6, 0.8) given β equal to 0.3 in all four cases.

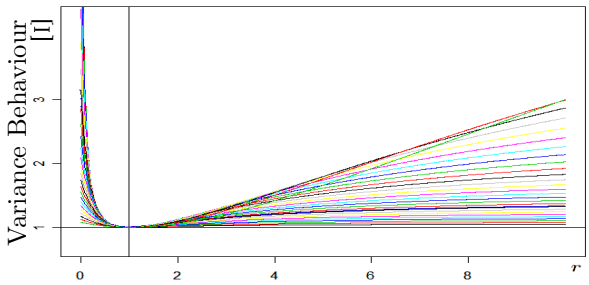
Figure 2.9: Random Variation in Variances on a Meta-analysis with 2 trials. Effect of both *Allocation* and *Number of patients per trial*



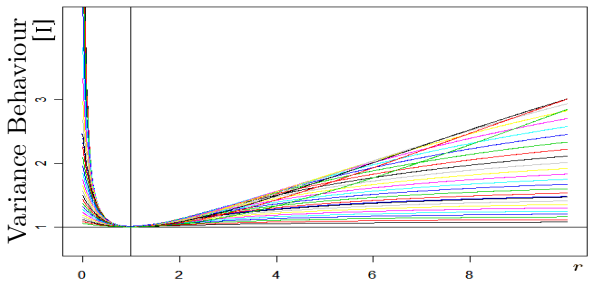
(a) Variance Ratio



(b) Variance Ratio



(c) Variance Ratio

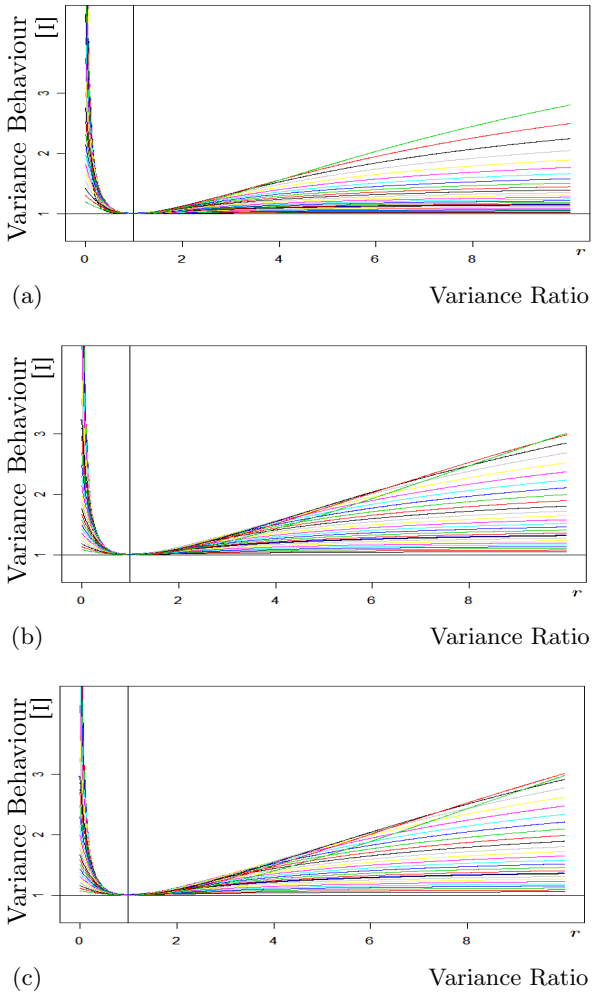


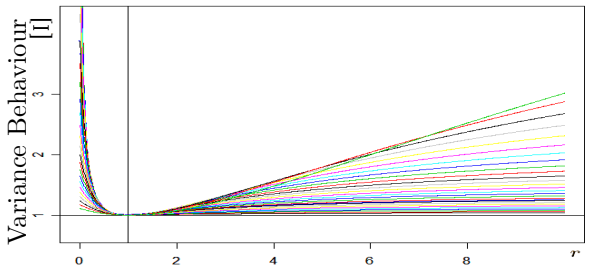
(d) Variance Ratio

These 4 figures show the variance inflation of the ratio of the real overall variance to the optimal overall variance as $\boldsymbol{r}$ –the observed ratio of within trial variances for trial 1 compared to trial 2– increases. Each line represents the variance behaviour for a possible combination of n_1 and n_2, given $n_1 + n_2 = 30$. Unbalanced allocation is assumed: the allocation per arm in the trial 2 (β) is fixed and set to 0.3 while the allocation per arm in trial 1 (α) changes for every figure (0.1, 0.3, 0.6 and 0.8 for figures from (a) to (d) respectively).

Similarly, Figure 2.10 shows the same situation in the case where α is fixed equal to 0.5 while β varies and assumes the values 0.1, 0.3, 0.6 and 0.8.

Figure 2.10: Random Variation in Variances on a Meta-analysis with 2 trials. Effect of both *Allocation* and *Number of patients per trial*





These 4 figures show the variance inflation of the ratio of the real overall variance to the optimal overall variance as the observed ratio of within trial variances for trial 1 compared to trial 2 increases. Each line represents the variance behaviour for each possible combination of n_1 and n_2, given $n_1 + n_2 = 30$. While the allocation per arm in the trial 1 is balanced ($\alpha = 0.5$), β changes for every figure (0.1, 0.3, 0.6 and 0.8 for figures from (a) to (d) respectively).

In order to have a better idea of the average behaviour of the function, we can calculate the expected value of Ξ under the assumption that $\boldsymbol{r} \sim F_{n_1-2,n_2-2}$. Results are given in Table 2.5 for N=30, β fixed and equal to 0.8 (in Table 2.5.a), 0.4 (in Table 2.5.b) and 0.1 (in Table 2.5.c) while α assumes values ranging from 0.1 to 0.9. Each row represents different values of α (from 0.1 to 0.9) while each column represents the number of patients in trial 1 given a total number of patients of 30 patients and under the assumption that there are at least 5 patients in each study (i.e., if N=30, n_1 ranges from 5 to 25).

Table 2.5: Expected value of the ratio between the real variance and the optimal variance of the estimator on a Meta-Analysis with only 2 trials with different sizes and allocations

(a)	n=5	n=6	n=7	n=8	n=9	n=10	n=11	n=12	n=13	n=14	n=15	n=16	n=17	n=18	n=19	n=20	n=21	n=22	n=23	n=24	n=25
$\alpha = 0.1$	1.304	1.188	1.136	1.109	1.093	1.083	1.077	1.073	1.070	1.069	1.069	1.070	1.072	1.074	1.079	1.085	1.095	1.109	1.131	1.170	1.246
$\alpha = 0.2$	1.282	1.185	1.139	1.113	1.098	1.088	1.081	1.076	1.074	1.072	1.071	1.072	1.074	1.076	1.081	1.088	1.098	1.113	1.139	1.185	1.282
$\alpha = 0.3$	1.266	1.179	1.136	1.112	1.097	1.087	1.081	1.076	1.073	1.072	1.071	1.071	1.073	1.075	1.080	1.086	1.097	1.113	1.139	1.188	1.295
$\alpha = 0.4$	1.257	1.175	1.134	1.111	1.096	1.087	1.080	1.076	1.073	1.071	1.070	1.070	1.072	1.075	1.079	1.085	1.095	1.111	1.138	1.189	1.299
$\alpha = 0.5$	1.255	1.174	1.134	1.110	1.096	1.086	1.080	1.075	1.072	1.071	1.070	1.070	1.072	1.074	1.078	1.085	1.095	1.111	1.138	1.189	1.301
$\alpha = 0.6$	1.257	1.175	1.134	1.111	1.096	1.087	1.080	1.076	1.073	1.071	1.070	1.070	1.072	1.075	1.079	1.085	1.095	1.111	1.138	1.189	1.299
$\alpha = 0.7$	1.266	1.179	1.136	1.112	1.097	1.087	1.081	1.076	1.073	1.072	1.071	1.071	1.073	1.075	1.080	1.086	1.097	1.113	1.139	1.188	1.295
$\alpha = 0.8$	1.282	1.185	1.139	1.113	1.098	1.088	1.081	1.076	1.074	1.072	1.071	1.072	1.074	1.076	1.081	1.088	1.098	1.113	1.139	1.185	1.282
$\alpha = 0.9$	1.304	1.188	1.136	1.109	1.093	1.083	1.077	1.073	1.070	1.069	1.069	1.070	1.072	1.074	1.079	1.085	1.095	1.109	1.131	1.170	1.246

(b)	n=5	n=6	n=7	n=8	n=9	n=10	n=11	n=12	n=13	n=14	n=15	n=16	n=17	n=18	n=19	n=20	n=21	n=22	n=23	n=24	n=25
$\alpha = 0.1$	1.308	1.181	1.127	1.100	1.085	1.076	1.070	1.067	1.065	1.064	1.064	1.065	1.067	1.070	1.074	1.081	1.089	1.102	1.121	1.154	1.216
$\alpha = 0.2$	1.299	1.189	1.138	1.111	1.095	1.085	1.079	1.075	1.072	1.070	1.070	1.071	1.073	1.076	1.080	1.087	1.096	1.111	1.134	1.175	1.257
$\alpha = 0.3$	1.289	1.187	1.139	1.113	1.097	1.087	1.081	1.076	1.073	1.072	1.071	1.072	1.074	1.076	1.081	1.088	1.098	1.113	1.138	1.182	1.275
$\alpha = 0.4$	1.282	1.185	1.139	1.113	1.098	1.088	1.081	1.076	1.074	1.072	1.071	1.072	1.074	1.076	1.081	1.088	1.098	1.113	1.139	1.185	1.282
$\alpha = 0.5$	1.280	1.184	1.139	1.113	1.098	1.088	1.081	1.077	1.074	1.072	1.071	1.072	1.074	1.076	1.081	1.088	1.098	1.113	1.139	1.186	1.284
$\alpha = 0.6$	1.282	1.185	1.139	1.113	1.098	1.088	1.081	1.076	1.074	1.072	1.071	1.072	1.074	1.076	1.081	1.088	1.098	1.113	1.139	1.185	1.282
$\alpha = 0.7$	1.289	1.187	1.139	1.113	1.097	1.087	1.081	1.076	1.073	1.072	1.071	1.072	1.074	1.076	1.081	1.088	1.098	1.113	1.138	1.182	1.275
$\alpha = 0.8$	1.299	1.189	1.138	1.111	1.095	1.085	1.079	1.075	1.072	1.070	1.070	1.071	1.073	1.076	1.080	1.087	1.096	1.111	1.134	1.175	1.257
$\alpha = 0.9$	1.308	1.181	1.127	1.100	1.085	1.076	1.070	1.067	1.065	1.064	1.064	1.065	1.067	1.070	1.074	1.081	1.089	1.102	1.121	1.154	1.216

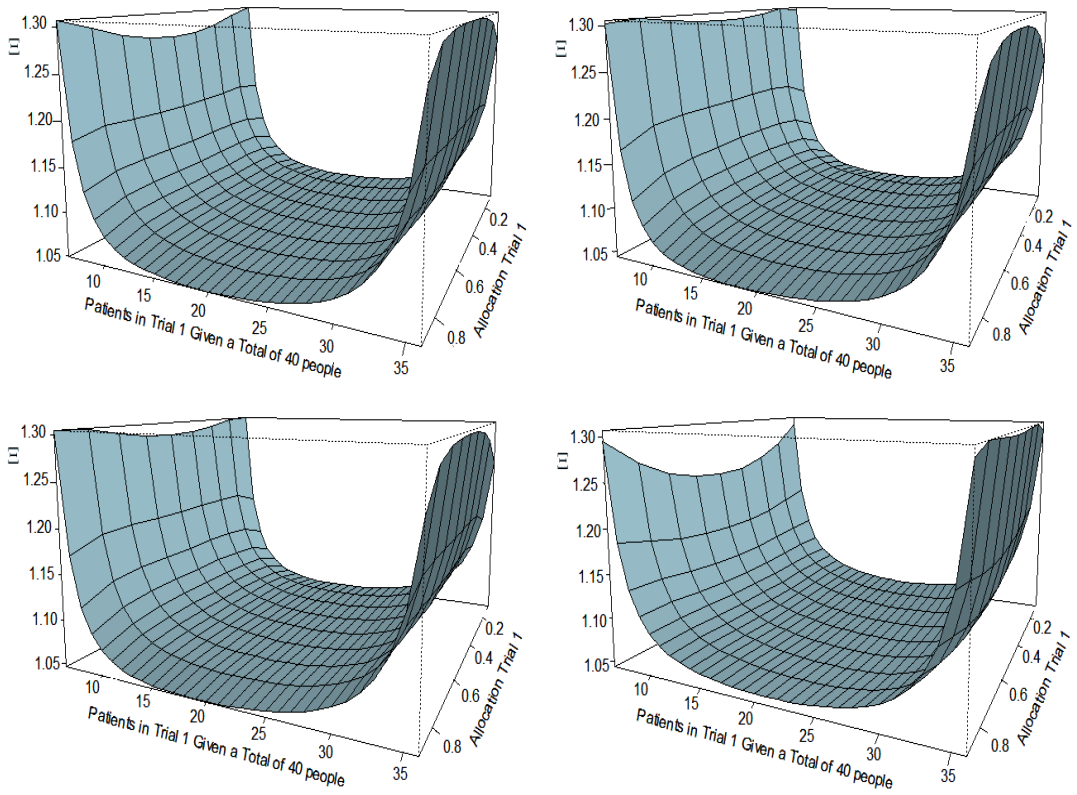
(c)	n=5	n=6	n=7	n=8	n=9	n=10	n=11	n=12	n=13	n=14	n=15	n=16	n=17	n=18	n=19	n=20	n=21	n=22	n=23	n=24	n=25
$\alpha = 0.1$	1.308	1.184	1.131	1.104	1.088	1.079	1.073	1.069	1.067	1.066	1.066	1.067	1.069	1.072	1.076	1.082	1.091	1.104	1.125	1.159	1.226
$\alpha = 0.2$	1.295	1.188	1.139	1.113	1.097	1.086	1.080	1.075	1.073	1.071	1.071	1.072	1.073	1.076	1.081	1.087	1.097	1.112	1.136	1.179	1.266
$\alpha = 0.3$	1.282	1.185	1.139	1.113	1.098	1.088	1.081	1.076	1.074	1.072	1.071	1.072	1.074	1.076	1.081	1.088	1.098	1.113	1.139	1.185	1.282
$\alpha = 0.4$	1.275	1.182	1.138	1.113	1.098	1.088	1.081	1.076	1.074	1.072	1.071	1.072	1.073	1.076	1.081	1.087	1.097	1.113	1.139	1.187	1.289
$\alpha = 0.5$	1.272	1.182	1.137	1.113	1.098	1.088	1.081	1.076	1.074	1.072	1.071	1.072	1.073	1.076	1.080	1.087	1.097	1.113	1.139	1.188	1.291
$\alpha = 0.6$	1.275	1.182	1.138	1.113	1.098	1.088	1.081	1.076	1.074	1.072	1.071	1.072	1.073	1.076	1.081	1.087	1.097	1.113	1.139	1.187	1.289
$\alpha = 0.7$	1.282	1.185	1.139	1.113	1.098	1.088	1.081	1.076	1.074	1.072	1.071	1.072	1.074	1.076	1.081	1.088	1.098	1.113	1.139	1.185	1.282
$\alpha = 0.8$	1.295	1.188	1.139	1.113	1.097	1.086	1.080	1.075	1.073	1.071	1.071	1.072	1.073	1.076	1.081	1.087	1.097	1.112	1.136	1.179	1.266
$\alpha = 0.9$	1.308	1.184	1.131	1.104	1.088	1.079	1.073	1.069	1.067	1.066	1.066	1.067	1.069	1.072	1.076	1.082	1.091	1.104	1.125	1.159	1.226

The columns represent the total number of subjects in trial 1 given a combined number of 30 subjects in the 2 trials, while each row, denoted α, represents the value of the allocation in each arm in trial 1 (ranging from 0.1 to 0.9 by a pace of 0.1). The three matrices consider 3 different values of β, the allocation in arms in trial 2; (a) 0.8, (b) 0.4 and (c) 0.1

All these matrices may not be easy to interpret, especially when N increases. A 3D visualization may better show the pattern of the calculated expected values of the ratio of the variances. Once again, we fix the value β for each single 3D representation (0.8, 0.6, 0.3 and 0.1) and we set N, the combined number of subjects in the 2 trials, equal to 40. The x axis represents the number of patients in trial 1 (and consequently we can determine the patients in the second trial) while the y axis shows the different values of α, i.e. the allocation per arm in trial 1 (while patients in trials 2 will always have the same allocation β). The 4 graphs have roughly the same pattern and the same values for every possible combination of the four variables taken into account. It can be seen from the three-dimensional plots that the distribution of the ratio between the empirical and the optimal variances is, as expected, symmetric (as we only considered 2 trials). Patterns and values change slightly with different values of β. The ratio Ξ tends to assume values distant from 1 when we consider the extreme cases with only few patients allocated in one trial, regardless of the allocation per arm. In all other cases, differences are, on average, undetectable; the ratio calculations change by only 0.03 in the region covering most of the cases and combinations (from 10 to 30 subjects in trial 1). Therefore, the distribution of the combined number of patients (i.e. the total subjects in each single study) as well as the allocation per arm per study does not have much influence on $\hat{V}$.

Nonetheless, if we observe the profiles of the expected values it can be seen that the minimum of the calculated ratio is reached when both the trials have the same number of patients as well as when the allocation is balanced (both arm in the trial have the same number of patients). Under these circumstances, the ratio is closer to 1; this means that, on average, the overall variance of the estimator is almost perfectly estimated. Just so almost, though. The expected ratio never reaches the value 1. Therefore, the overall variance of the estimator is on average higher than the optimal one. As proven via simulations, the 'real' dispersion of the estimator is *de facto* higher than that of the optimal one even though we tend to declare that the estimate has a smaller overall variance. On the average, the estimate of the variance of the overall effect by using the variance weighted method in meta-analysis underestimates the true pooled variance (i.e. the significance level is overestimated). There is a tendency to produce variance estimate lower than the optimal true one even though the 'real' variance of the estimator used is higher. The estimate of the overall variance should then consider both the effect of the sample size and the variation of sample variances in order to produce more reliable figures.

Figure 2.11: 3-D Visualisation of the Expected Variance Inflation for N=40 on a Meta-Analysis with 2 trials



These 4 figures show the Expected variance inflation of the ratio of the real overall variance to the optimal overall variance for different values of patients and allocation in both trials 1 and 2. Each figure represents the expected values for a fixed value of β while n_1 and α varies. In particular, α ranges from 0.1 to 0.9 while the number of patients in trial 1 ranges from 5 to 35 given a total number of 40 patients among the 2 studies. β for each single 3D representation is set to (a) 0.8, (b) 0.6, (c) 0.3 and (d) 0.1

2.10 Precision of the overall estimator: a recap

Even though "the simulated data sets should have some resemblance to reality for the results to be generalizable to real situations and have any credibility" (Burton et al., 2006, pg. 4283), so far we have investigated the behaviour of the precision of the overall estimator (paying attention particularly to the individual within-study variances) under the assumption of homoscedasticity where all studies have exactly the same nominal value of the internal variance (i.e. $S^2 = \sigma^2_{Ti} = \sigma^2_{Ci}$ for all i). We simulated different scenarios in order to evaluate both the importance of the total patients present in each study and the significance of the allocation of patients in each arm of each study. At least under the assumption of common variance, the *average number of patients per trial* is more important than the total number of each single study. Moreover, the allocation of patients per arm does not seem to be decisive for the estimated overall

variance of the estimator. Nonetheless, having a perfect balanced allocation as well as having roughly the same amount of patients per study yield better results. Furthermore, true to form, the higher the average number of patients considered in each arm, the closer the estimator is to the optimal one, i.e. the fewer the number of patients the less precise the estimates of S_i^2's are and the greater the impact is on the results. These conclusions were obtained not only via simulations but they even were mathematically demonstrated, at least for the case including only two studies. Moreover, coverage probabilities (obtained by simulating 10000 meta-analyses and considering different patterns and scenarios) show that wrongly reported values of the variance can badly underestimate the overall variance of the estimator. This happens especially when trials with few patients are taken into consideration. Unsurprisingly, when we consider a number of small trials we generally overestimate precision and this leads to unreliable results, as we tend to overstate the significance of the results. In particular when the meta-analysis is dominated by very small studies, caution needs to be exercised. In fact, "the estimated weight $\hat{w}_i$ has expectation $E(\hat{w}_i) = (n_i)w_i/(n_i - 3)$, where $n_i = n_{Ti}n_{Ci}/(n_{Ti} + n_{Ci})$ rather than $E(\hat{w}_i) = w_i$. This bias in the estimation of the weights not only affects the power of the test, but also the estimate of the variance of the overall effect and the estimate of the between-study variance τ. Hence, when numbers are very small, results should be interpreted cautiously" (Hardy & Thompson, 1998, pg. 853).

"The bias of the estimate of the variance of the overall effect synthesised from individual studies by using the variance weighted method is proven to be negative" (Li et al., 1994, pg. 1063). Furthermore, such an estimate of the variance of the estimator is also *too sensitive* to the minimum of the estimates of the variances in the k studies. "If the $\hat{\sigma}^2_{min}$ happens to be wrongly reported to have a very small value, the influence on $var(\hat{\theta})$ by this estimate would be over-emphasized" (Li et al., 1994, pg. 1065) and $\hat{V}$ would badly underestimate the true overall variance.

As we have shown via calculation (for two trials, without loss of generality) and simulations, on average, the estimate of the overall variance ($\hat{V}$) underestimates the true value. "If there is an outlier or measurement error which gives an extremely small sample variance, the pooled variance will badly underestimate the true variance by the ordinary method and the weight of this individual sample for the combined mean will be too *high*" (Li et al., 1994, pg. 1083). "The ordinary method is too sensitive to the variation of the sample variances and biased if the sample size is not large" (Li et al., 1994, pg. 1082). As a consequence, an adjusted method that considers both the effect of *small sample* and the *variation of the variances* and which is *not too sensitive to any individual result* is necessary.

Especially when we consider small trials, are there better estimations of the variances? As the estimation of the variance and the number of patients are highly correlated (the larger the sample size the closer the estimate of the the variance is to the real value), are there methods capable to shrink the variances and account for their random variation in order to have an estimator which does not depend so badly on the sizes of clinical trials?

How can we produce less biased estimate? Can we adjust the weight method in order to consider the effect of small sample sizes? Is there a robust weight method with regard to the variation of the sample variances and sample sizes not too sensitive to any individual result?

Chapter 3

Shrinkage Estimators and a more Realistic Scenario

3.1 Acting in a more Realistic Scenario

So far we have considered the assumption of common variances, however, such an assumption may be unrealistic. Therefore, in this chapter we consider the case where the population variances of the observation at the patient level differ from study to study, by analysing two different simulation schemes. The first one will arbitrarily impose the values of the variances, and the second one will draw the values from a Gamma distribution. In both designs we will consider a perfect balance because in most of the trials the sample sizes n_i for the control and the treatment groups are generally similar. Nonetheless, instead of considering k studies all of which have the same size, we will impose the sample size to be equal only *on average*.

First Simulation Design

Following the example of the simulation study conducted by Knapp et al. (2006), the first simulation study considers equal allocation (i.e. $n_i = n_{Ci} = n_{Ti}$) and different values for the S_i^2 (where $S_i^2 = \sigma_{Ti}^2 = \sigma_{Ci}^2$). As in the paper by Knapp et al., we "arbitrarily choose the base value of S_i^2 to be equal to 100 and deviations from this value for a few of the S_i^2 are made to reflect patterns of imbalance possible in application." The first pattern has roughly half of the $S_i^2 = 100$ while the second one - the most imbalanced - has roughly 80-90% of the $S_i^2 = 100$. When S_i^2 is not equal to 100, the value chosen is $S_i^2 = 10$. As mentioned above, the sample sizes, varying from 5 to 50 people per arm, are balanced. This means that for $S_i^2 = 100$ the within-trial variability ranges (on average) from 40 to 4 as n increases from 5 to 50, respectively.
The aim of such a simulation design is to evaluate how important the effect on the overall

variance is when a few estimates from a few clinical trials carry a disproportionate amount of weight, that is when a few of the within trial variances $\sigma_i^2 = var(Y_i) = (1/n_{Ci} + 1/n_{Ti})\, S_i^2$ are much smaller than the others.

Second Simulation Design

In this case, as suggested by Tong and Wang (2007), S_i^2's for i=1,...,k, are simulated from a Gamma distribution with shape parameter γ and scale parameter β. β is set at 1 "because it has little impact on the comparative performance" (Tong & Wang, 2007, pg. 116). In order to evaluate the performance of the estimator under different levels of variance heterogeneity, three different shape parameters are considered, $\gamma = 0.25$, 1 and 4, which correspond to three different coefficients of variation ($CV = \sqrt{\gamma\beta^2}/(\gamma\beta) = \sqrt{\gamma}/\gamma$) at levels 2, 1, and 0.5 respectively. These three different settings represent commonly encountered cases; for example, γ=0.25 corresponds to the case with different variances across studies whereas $\gamma = 4$ corresponds to highly similar values of S_i^2.

3.2 Simulations with uncommon variances

We design the simulation study to roughly follow the characteristic of a more realistic meta-analysis. We consider two different schemes for the variance of the observations at the patients level and for each pattern we discuss the meta-analytical combination of the results of k clinical trials, where k = 10, 15, 20, 35. The true overall effect θ is set at 3. The error probability α is restricted to the common value 0.05 in constructing the approximate $100(1-\alpha)$ confidence interval for θ.

As in the previous simulation designs, we consider a meta-analysis of k similar but independent studies. The observations consist of two sets of independent random variables $X_{Ti1}, X_{Ti2}, ...X_{Tin_{Ti}}$ and X_{Ci1}, X_{Ci2}, ... $X_{Cin_{Ci}}$ for i=1, 2, ..., k from the treatment and the control groups, respectively. These two sets of variables have independent normal distributions with different means and equal variances, S_i^2, as follows

$$\begin{array}{l} X_{Ti1}, X_{Ti2}, \ldots, X_{Tin_{Ti}} \sim N(\mu_{Ti}, \sigma_{Ti}^2) \\ X_{Ci1}, X_{Ci2}, \ldots, X_{Cin_{Ci}} \sim N(\mu_{Ci}, \sigma_{Ci}^2) \end{array} \quad \text{for} \quad i = 1, \ldots, k \quad \text{where} \quad \sigma_{Ti}^2 = \sigma_{Ci}^2 = S_i^2$$

The parameter of interest, denoted by θ, is the overall mean difference. The study specific mean difference is defined as $Y_i = (\mu_{Ti} - \mu_{Ci})$ and is estimated by $Y_i = \overline{X}_{Ti} - \overline{X}_{Ci}$, where $\overline{X}_{Ti} = \sum_{j=1}^{n_{Ti}} X_{Tij}/n_{Ti}$ and $\overline{X}_{Ci} = \sum_{j=1}^{n_{Ci}} X_{Cij}/n_{Ci}$. We assume that Y_i is such that $E(Y_i) = \theta$ and that the *variance of the difference* between two independent means based on n_{Ti} and n_{Ci} observations respectively is equal to $var(Y_i) = \sigma_i^2 = \sigma_{Ti}^2/n_{Ti} + \sigma_{Ci}^2/n_{Ci} = S_i^2(1/n_{Ti} + 1/n_{Ci})$ given that the two groups in the treatment and control arms have the same variance. For moderately large study sizes, each Y_i

should be asymptotically normally distributed. Thus,

$$Y_i \sim N(\theta, \sigma_i^2) \quad \text{for} \quad i = 1, 2, ..., k$$

where σ_i^2 varies from study to study accordingly to the values assumed by S_i^2. Once S_i^2 is set for each study, the study specific mean difference is computed. For each single study, we use equal sample sizes ($n_i = n_{Ci} = n_{Ti}$) for the control and the treatment groups since these values tend to be similar in parallel trials; n_i's are sampled from a Negative Binomial Distribution in order for the k clinical trials to have *on average* the same amount of patients per arm and per study, where E(n_i)'s are equal to 5, 8, 10, 15, 20, 30 and 50. The simulations are implemented using the software package **R**, with each simulation generating n_{Ti} and n_{Ci} observations from normal distributions with mean μ_{Ti} and $\mu_{Ci} = \mu_{Ti} + \theta$ respectively and variance S_i^2. This procedure is replicated k times for each of the 10000 independent simulations run. At each replicate, the study specific mean differences (Y_i) as well as their variances $\hat{\sigma}_i^2$ are computed. These estimates allow us to obtain the estimated weights $\hat{w}_i$ for each study and therefore to calculate the overall effect estimate $\hat{\theta}$ and its overall variance estimate $\hat{V}$ for each replicate.

As done in the previous simulations, we summarise the estimates once all simulations have been performed. In particular, we report the 'declared' estimate of the overall variance as well as the 'optimal' and the 'real' variance of the point estimates (see equations 2.5 - 2.7). Furthermore, in order to evaluate the performance and the precision of the results obtained from the different scenarios and approaches being studied, the coverage of the confidence intervals is shown. Again, a possible criterion for acceptability of the coverage is that the coverage should not fall outside of approximately two SEs of the nominal coverage probability.

In general our goal is to obtain not only an empirical estimated coverage probability corresponding to the nominal value but also, and more importantly, both the 'declared' precision and the 'actual' dispersion of the overall variance of the estimator close to the theoretical 'optimal' one.

3.3 The simulation results

We give the results if no heterogeneity in the treatment effect is present, that is assuming the *fixed-effects model* is the theoretically correct one. The 'actual', 'optimal' and 'declared' variances are given in Tables 3.1 - 3.5 for the 5 different scenarios (i.e. different values of k and S_i^2). Also, the empirical coverage probabilities of the approximate 95% confidence intervals based on the 'empirical' weights and variances are shown in Figs 3.1 - 3.2. From the Tables 3.1 - 3.5, it should be first noted that the 'empirical' variances are reasonably close to the true 'optimal' value *only* for high values of n

(where $n = E(n_i)$ for all i), regardless of the number of trials and of the S_i^2 pattern considered. As n decreases, the 'declared' variance is less accurate and precise. In all cases, however, the 'optimal' variance is badly underestimated. In particular, as the average number of patients per arm decreases, the estimated 'declared' overall variances start to deviate from the 'optimal' value with negative and increasingly large bias. Thus, we can assume that the method currently used leads to large error in the estimated weights, both when common or uncommon S_i^2's are assumed. Moreover, considering the column of the 'actual' variance of the estimator, $V^{real}(\hat{(\theta)})$, we note that, once again, these values are always observed to be higher than both the 'optimal' and the 'declared' ones, in particular when n decreases. Again, even without the assumption of common S_i^2 for all studies, there is the tendency to produce an overall variance estimate $\hat{V}$ that is lower than that produced by the 'optimal' estimator, whereas the 'actual' variance of the estimator is higher. Fixed-effects meta-analysis uses an estimator whose true variance is higher than the 'correct' one but which will appear to be lower. As a consequence, the ordinary method tends to overstate the significance and the precision of the results; in particular, this tendency is even more marked under the assumption of uncommon S_i^2.
As regards the Empirical Coverage Probabilities for the Confidence Intervals, the proportion of intervals containing θ falls for decreasing sample sizes. Note that the coverage probabilities for the confidence intervals in Figs. 3.1 - 3.2 are also very similar, exception made for small values of n. Under different levels of variance heterogeneity and under different schemes of simulation, patterns are about the same regardless the number of studies taken into account. The coverage probabilities based on the 'declared' overall variances are generally below 95%, although they do increase when n increases. However, for an average number of patients per arm per study less than 10, the coverage probability falls to 55%. For small sample sizes, the coverage probability is far from the nominal level whether we consider the slightly imbalanced scenario or the most imbalanced one. The different choices of within-trial variances (S_i^2) do not have an impact on the simulation results. As regards the number of trials, if the latter are "large enough" then no matter how many studies we include we get roughly the same results. On the contrary, if sizes are small, the more studies we consider the lower the coverage probability is.

In short, the usual variance estimator performs very poorly in detecting the true variance of θ and underestimates the true value for all values of n and k. Moreover, the 'actual' variability of the variance estimator is always higher than both the 'declared' and 'optimal' ones, with a consequent overstatement of the precision of the estimator and misleading results in the form of too liberal significance tests and Confidence Intervals without correct coverage properties. In particular, these problems arise when

a small average number of patients per arm per study is considered: the smaller the number of patients, the lower the coverage probability and the higher the overestimation of the precision of the estimate.

When we perform fixed-effects meta-analysis, inference is based on the assumption that weights are perfectly estimated, whereas since the variances are poorly estimated the inferences drawn may be in error. In practice, no matter whether the true variances are assumed to be equal or whether they vary from trial to trial, in both cases there is the tendency to badly estimate V, the overall variance of the point estimate. Moreover, regardless of the true values of S_i^2, the fewer the average number of subjects per arm the higher the underestimation of V. In fact, under the assumption of both equal S_i^2's (see Fig 2.3 and Table 2.2) and unequal S_i^2's (see Tables 3.1 and 3.2), the empirical coverage probabilities are generally below 95% and figures get closer to the nominal level only for high values of n. For instance, referring to the empirical coverage probabilities, we see that if n=8 and S_i^2's are assumed to be equal, the coverage probabilities are roughly between 87 and 89% (depending on the number of trials involved) whereas under the assumption of unequal S_i^2's these values may vary between 82 and 88%.
Since when performing a single meta-analysis the real values of the study variances are unknown, little can be said about the assumptions on the variances. However, regardless of the assumption made, a lot can be said about the expected outcome: if small-sized studies are considered, the overall variance will be underestimated quite badly and inferences based on these values are likely to be wrong.

Table 3.1: The Results of Simulation for different values of studies (k) under the assumption of Unequal S_i^2 and Study sizes - First Simulation Scheme (A)

k = 10	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	5.787174	1.724529	1.090686	0.7069
	8	1.716196	0.985787	0.822931	0.8691
	10	0.951359	0.763757	0.681596	0.9041
	15	0.537592	0.490154	0.462380	0.9274
	20	0.380268	0.363268	0.348901	0.9400
	30	0.245991	0.240309	0.233714	0.9439
	50	0.144902	0.143970	0.141717	0.9471
k = 15	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	4.947632	1.087694	0.656184	0.6625
	8	1.418933	0.631846	0.522482	0.8620
	10	0.631725	0.496492	0.440474	0.8984
	15	0.334990	0.323049	0.302369	0.9361
	20	0.251068	0.240756	0.230016	0.9377
	30	0.165133	0.159585	0.155199	0.9418
	50	0.095679	0.095648	0.094113	0.9499
k = 20	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	4.892815	0.796330	0.467671	0.6373
	8	1.258493	0.467765	0.386032	0.8506
	10	0.492900	0.368996	0.326740	0.8989
	15	0.257675	0.241634	0.225962	0.9319
	20	0.191488	0.179803	0.171074	0.9346
	30	0.122043	0.119628	0.116054	0.9473
	50	0.073591	0.071738	0.070481	0.9428
k = 35	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	4.098463	0.436752	0.245839	0.5698
	8	0.845240	0.262924	0.212878	0.8319
	10	0.403127	0.207927	0.182103	0.8869
	15	0.146905	0.137120	0.127746	0.9319
	20	0.111066	0.102539	0.097383	0.9344
	30	0.069944	0.068164	0.065977	0.9443
	50	0.042369	0.040884	0.040116	0.9444

This Table shows the results of simulation for $\theta = 3$ and different values of n, k and S_i^2. This simulation scheme considers for each simulation k parallel group clinical trials each of whom with the same number of patients per arm only *on average* ($n = E[n_i] = n_{Ti} = n_{Ci}$). As regards the within-study variances, the most imbalanced scenario is shown, i.e. 80% of the studies has S_i^2 set to 100 while the remaining 20% are set equal to 10. Empirical Statistics for $\mathrm{E}[\hat{V}]$ and $V^{real}(\hat{\theta})$ are based on 10000 simulation replicates as well as the Empirical Coverage Probability.

Table 3.2: The Results of Simulation for different values of studies (k) under the assumption of Unequal S_i^2 and Study sizes - First Simulation Scheme (B)

k = 10	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	2.917085	0.844070	0.519157	0.6988
	8	0.822160	0.486116	0.401863	0.8652
	10	0.496175	0.378410	0.335633	0.9046
	15	0.261349	0.247350	0.231493	0.9344
	20	0.197728	0.183608	0.175550	0.9333
	30	0.125123	0.121904	0.118438	0.9419
	50	0.075708	0.073157	0.071895	0.9423
k = 15	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	2.515061	0.567343	0.341732	0.6644
	8	0.608366	0.335919	0.275871	0.8487
	10	0.361150	0.263301	0.232215	0.8970
	15	0.187226	0.172940	0.161681	0.9307
	20	0.134901	0.129138	0.123035	0.9367
	30	0.087392	0.085815	0.083122	0.9449
	50	0.051520	0.051463	0.050523	0.9479
k = 20	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	2.050539	0.393981	0.231733	0.6381
	8	0.599061	0.234913	0.191897	0.8494
	10	0.271898	0.185648	0.162671	0.8943
	15	0.132802	0.122398	0.114006	0.9269
	20	0.097786	0.091385	0.086845	0.9353
	30	0.062851	0.060831	0.058860	0.9422
	50	0.036432	0.036428	0.035755	0.9480
k = 35	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	2.019225	0.222689	0.124061	0.5592
	8	0.441426	0.135820	0.109607	0.8233
	10	0.163837	0.107798	0.094237	0.8911
	15	0.076543	0.071281	0.066329	0.9329
	20	0.055763	0.053325	0.050581	0.9398
	30	0.037028	0.035547	0.034353	0.9416
	50	0.021329	0.021306	0.020883	0.9484

This Table shows the results of simulation for $\theta = 3$ and different values of n, k and S_i^2. This simulation scheme considers for each simulation k parallel group clinical trials each of whom with the same number of patients per arm only *on average* ($n = E[n_i] = n_{Ti} = n_{Ci}$). As regards the within-study variances, a slight imbalanced scenario is shown, i.e. 50% of the studies had S_i^2 set to 100 while the remaining 50% were set equal to 10. Empirical Statistics for $E[\hat{V}]$ and $V^{real}(\hat{\theta})$ are based on 10000 simulation replicates as well as the Empirical Coverage Probability.

Table 3.3: The Results of Simulation for different values of studies (k) under the assumption of Unequal S_i^2 and Study sizes - Second Simulation Scheme (A)

k = 10	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	4.81074E-04	1.53915E-04	1.08749E-04	0.8079
	8	1.62910E-04	9.29367E-05	7.92131E-05	0.8982
	10	8.81519E-05	7.32187E-05	6.63829E-05	0.9166
	15	4.73614E-05	4.41435E-05	4.21259E-05	0.9340
	20	3.37472E-05	3.30902E-05	3.18899E-05	0.9401
	30	2.45493E-05	2.23964E-05	2.19227E-05	0.9460
	50	1.37721E-05	1.37271E-05	1.35344E-05	0.9467
k = 15	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	1.55977E-04	3.99302E-05	2.74604E-05	0.8110
	8	4.02419E-05	2.20131E-05	1.87943E-05	0.8924
	10	2.36055E-05	1.80732E-05	1.64116E-05	0.9186
	15	1.09118E-05	1.08852E-05	1.04165E-05	0.9363
	20	8.53535E-06	8.35436E-06	8.01270E-06	0.9409
	30	5.53268E-06	5.49154E-06	5.34444E-06	0.9469
	50	3.13219E-06	3.05720E-06	3.02327E-06	0.9435
k = 20	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	7.25774E-05	1.43276E-05	9.27051E-06	0.8038
	8	1.81251E-05	8.17075E-06	7.04632E-06	0.8961
	10	8.29189E-06	6.33262E-06	5.79714E-06	0.9168
	15	4.08453E-06	4.01735E-06	3.81904E-06	0.9387
	20	3.49810E-06	2.87931E-06	2.81890E-06	0.9416
	30	1.98370E-06	1.96991E-06	1.92363E-06	0.9424
	50	1.26418E-06	1.25725E-06	1.23345E-06	0.9469
k = 35	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	1.25921E-05	1.79404E-06	1.17625E-06	0.8040
	8	1.51794E-06	1.07902E-06	9.32548E-07	0.8943
	10	1.19789E-06	7.97212E-07	7.20237E-07	0.9135
	15	5.48354E-07	5.05667E-07	4.79389E-07	0.9365
	20	3.33613E-07	3.52159E-07	3.37293E-07	0.9423
	30	2.92712E-07	2.67102E-07	2.61131E-07	0.9437
	50	1.57857E-07	1.57786E-07	1.55980E-07	0.9486

This Table shows the results of simulation for $\theta = 3$ and different values of n, k and S_i^2. This simulation scheme considers for each simulation k parallel group clinical trials each of whom with the same number of patients per arm only *on average* ($n = E[n_i] = n_{Ti} = n_{Ci}$). In this simulation scheme, within-study variances S_i^2 are drawn from a Γ distribution with shape parameter $\gamma = 0.25$. Empirical Statistics for $E[\hat{V}]$ and $V^{real}(\hat{\theta})$ are based on 10000 simulation replicates as well as the Empirical Coverage Probability.

Table 3.4: The Results of Simulation for different values of studies (k) under the assumption of Unequal S_i^2 and Study sizes - Second Simulation Scheme (B)

k = 10	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	0.0515724	0.0155874	0.0098259	0.7165
	8	0.0154709	0.0089368	0.0074546	0.8733
	10	0.0085677	0.0069199	0.0061864	0.9128
	15	0.0047821	0.0045230	0.0042589	0.9328
	20	0.0034570	0.0032938	0.0031601	0.9386
	30	0.0022969	0.0022170	0.0021583	0.9419
	50	0.0013495	0.0013065	0.0012880	0.9441
k = 15	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	0.0409151	0.0088253	0.0053688	0.6775
	8	0.0090774	0.0050160	0.0041460	0.8650
	10	0.0059114	0.0039948	0.0035261	0.9031
	15	0.0028288	0.0026268	0.0024566	0.9294
	20	0.0020033	0.0019289	0.0018477	0.9389
	30	0.0013050	0.0012869	0.0012500	0.9456
	50	0.0007744	0.0007619	0.0007493	0.9491
k = 20	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	0.0357162	0.0059225	0.0035595	0.6558
	8	0.0076422	0.0034974	0.0028850	0.8514
	10	0.0036417	0.0027604	0.0024472	0.9059
	15	0.0019467	0.0017814	0.0016727	0.9335
	20	0.0013540	0.0013215	0.0012625	0.9374
	30	0.0008975	0.0008707	0.0008459	0.9413
	50	0.0005269	0.0005250	0.0005160	0.9447
k = 35	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	0.0242387	0.0028335	0.0016546	0.6036
	8	0.0062553	0.0016724	0.0013653	0.8411
	10	0.0019950	0.0013180	0.0011652	0.9047
	15	0.0009610	0.0008661	0.0008112	0.9285
	20	0.0006909	0.0006430	0.0006125	0.9327
	30	0.0004454	0.0004323	0.0004194	0.9430
	50	0.0002613	0.0002531	0.0002485	0.9468

This Table shows the results of simulation for $\theta = 3$ and different values of n, k and S_i^2. This simulation scheme considers for each simulation k parallel group clinical trials each of whom with the same number of patients per arm *on average* ($n = E[n_i] = n_{Ti} = n_{Ci}$). In this simulation scheme, within-study variances S_i^2 are drawn from a Γ distribution with shape parameter $\gamma = 1$. Empirical Statistics for $\mathrm{E}[\hat{V}]$ and $V^{real}(\hat{\theta})$ are based on 10000 simulation replicates as well as the Empirical Coverage Probability.

Table 3.5: The Results of Simulation for different values of studies (k) under the assumption of Unequal S_i^2 and Study sizes - Second Simulation Scheme (C)

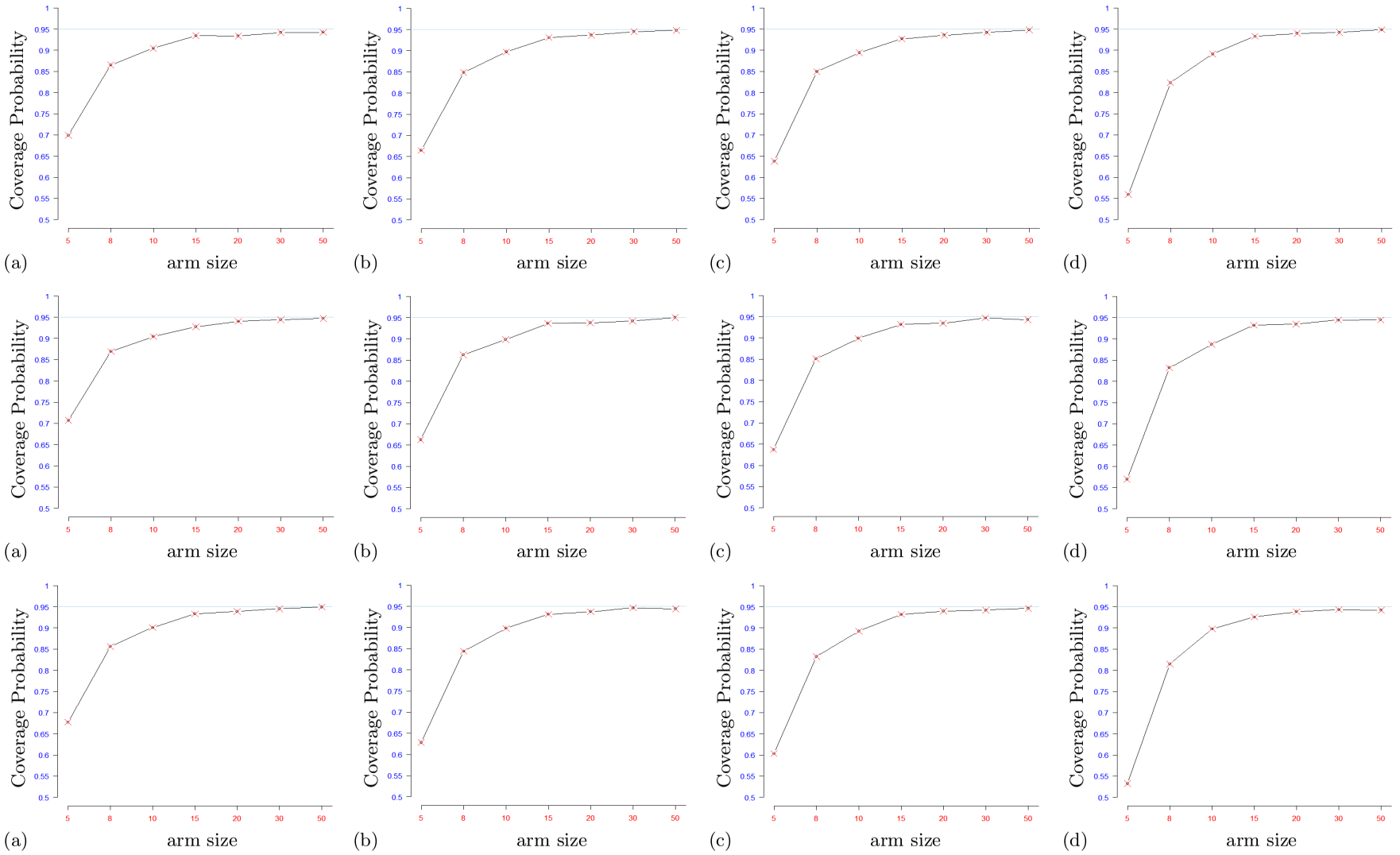
k = 10	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	0.4561987	0.1399444	0.0849630	0.6901
	8	0.1471794	0.0825180	0.0680434	0.8585
	10	0.0846646	0.0644753	0.0569307	0.9009
	15	0.0469084	0.0425057	0.0397063	0.9262
	20	0.0333541	0.0316287	0.0301946	0.9319
	30	0.0220455	0.0210269	0.0203818	0.9401
	50	0.0128987	0.0125241	0.0123001	0.9414
k = 15	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	0.3793491	0.0894609	0.0525835	0.6501
	8	0.1129277	0.0532958	0.0435367	0.8497
	10	0.0592798	0.0422414	0.0371362	0.8974
	15	0.0303040	0.0276316	0.0257761	0.9297
	20	0.0222863	0.0208020	0.0198107	0.9350
	30	0.0142411	0.0138214	0.0133774	0.9419
	50	0.0086080	0.0082729	0.0081168	0.9428
k = 20	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	0.3520310	0.0653238	0.0374189	0.6112
	8	0.0927032	0.0393147	0.0319142	0.8361
	10	0.0445698	0.0312078	0.0274058	0.8933
	15	0.0225143	0.0206696	0.0192208	0.9306
	20	0.0162632	0.0153991	0.0146369	0.9386
	30	0.0104935	0.0102330	0.0098997	0.9436
	50	0.0063069	0.0061575	0.0060413	0.9475
k = 35	**n**	$V^{real}(\hat{\theta})$	V^{opt}	$\mathbf{E}[\hat{V}]$	**Coverage Probability**
	5	0.3044838	0.0361784	0.0197527	0.5428
	8	0.0714990	0.0220422	0.0178129	0.8166
	10	0.0264298	0.0175457	0.0153523	0.8938
	15	0.0125205	0.0116370	0.0108208	0.9316
	20	0.0092827	0.0087110	0.0082593	0.9357
	30	0.0059134	0.0058103	0.0056140	0.9399
	50	0.0035171	0.0034782	0.0034116	0.9478

This Table shows the results of simulation for $\theta = 3$ and different values of n, k and S_i^2. This simulation scheme considers for each simulation k parallel group clinical trials each of whom with the same number of patients per arm only *on average* ($n = E[n_i] = n_{Ti} = n_{Ci}$). In this simulation scheme, S_i^2 are drawn from a Γ distribution with shape parameter $\gamma = 4$. Empirical Statistics for $\mathrm{E}[\hat{V}]$ and $V^{real}(\hat{\theta})$ are based on 10000 simulation replicates as well as the Empirical Coverage Probability.

Under the assumptions of both common and uncommon S_i^2, the average number of patients per arm per clinical trial plays an important role since for small values of n there is the tendency to badly underestimate V, the overall variance of the estimator. Specifically, we have pointed out that the usual variance estimator commonly used in meta-analytical inference is not robust to the estimated weights, and that in fact it may not be a good estimator for the correct variance of an overall effect estimate when estimated weights are used, as in practice. Hence, the weights used in practice are not the correct ones and a new method to better estimate the variances $\hat{\sigma}_i^2$ (and the weights) in order to have a more precise and accurate overall variance is desperately needed. For instance, should there be some random variation in the treatment or control group sample variances that causes an extremely small pooled variance, the variance of the mean difference $\hat{\sigma}_i^2$ will badly underestimate the true one and the weight of this individual sample will be too *high*. Since the ordinary method is too sensitive to the variation of the sample variances and biased if the sample size is not large "enough" (being small variances more suspicious for small studies), is there a robust method to adjust weights with regard to the variation of S_i^2 and sample sizes and not too sensitive to any individual result?

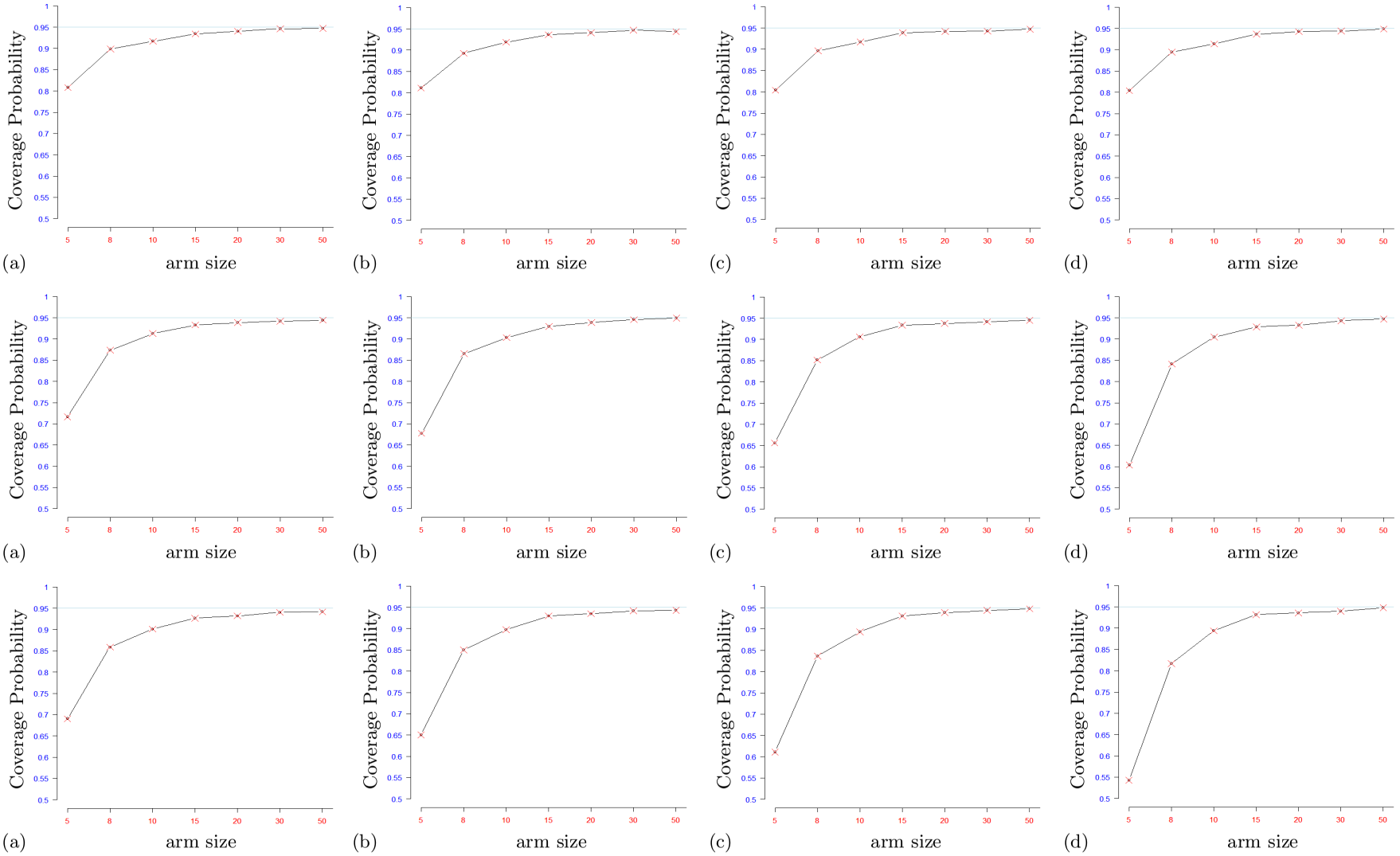
In order to minimise the overall variance estimation error and to have better weights a shrinkage estimator for within trials variances S_i^2 will be taken into account. This should guarantee enough robustness in order to provide protection against errors in the estimated weights (i.e. random variation in sample variances); this way, the 'declared' estimated variance should be closer to the optimal one, more importantly, the 'actual' variability of the overall effect estimator computed with the new weights should be closer to both the 'optimal' and the 'declared' values. The dispersion of the variances around the optimal value will be an indicator of the goodness of the method used.

Figure 3.1: 95% Empirical Coverage Probabilities of the CIs based on $\hat{V}$ – First Simulation Design, with S_i^2's arbitrarily imposed



Each row corresponds to the number of studies included in the meta analysis [10 (a), 15 (b), 20 (c) and 35 (d)] whereas lines show the results for each different S_i^2 pattern. In these cases, S_i^2 follow the first variance simulation scheme. In particular, the first line shows a slight imbalance with half of S_i^2=10 and the other half equal to 100; the second corresponds to the most imbalanced case where 20% of S_i^2 are set to 10 and the remaining to 100 while the third line corresponds to the common variance case (100% of S_i^2 have values equal to 100). Results from 10000 simulations.

Figure 3.2: 95% Empirical Coverage Probabilities of the CIs based on $\hat{V}$ – Second Simulation Design, with S_i^2's drawn from a Γ Distribution



Each line shows the results for each different variance pattern whereas rows correspond to the number of studies included in the meta analysis [10 (a), 15 (b), 20 (c) and 35 (d)]. S_i^2's follow the second simulation scheme. In particular, for the first line of figures, $\gamma = 0.25$, for the second $\gamma = 1$ whereas for the third $\gamma = 4$ (i.e. highly similar values of S_i^2). Results from 10000 simulations.

3.4 Shrinkage Estimators

3.4.1 Basic Logic

The 'Shrinkage Estimators' are commonly considered consistent with Bayesian logic since their main idea is that, when estimating a parameter, one should not simply use the information coming from the sample, but also some 'extra-sample' information. In fact, combining measurements (i.e. estimating single parameters using some sort of overall information) minimises the total '*Mean Squared Error*' (MSE). When the nature of the problem is not to estimate each expected return separately but rather minimise the total impact, shrinkage estimators represent an efficient and reasonable alternative to the classical estimators. In this study it is therefore reasonable to combine variance measurements since the goal is to minimise the total variance estimation error.
Considering both the informative prior and the information obtained through the sample, the shrinkage estimators compress the general values of each single study towards an identical common value (usually referred to as the 'common mean' or informative prior) (Braga, 2004). In fact, the general logic of a shrinkage estimator is similar to the weighted mean of a 'common value' and a 'sample mean' where weights determine how close the expected value is to the common one that functions as a target. There are several approaches to shrinking least squares estimators towards a common mean; all of which suggest that in general shrinking "produces estimators with greater predictive power than classical pooling techniques" (Smith, 1997, pg. 359).
The 'shrinking factor' is the element that determines the intensity of the 'shrinking' towards the 'mean value' and therefore this is the element that tells us the proximity of the informative prior to the sample information at disposal. The shrinking factor is quantified accordingly to the informative prior used, that is the value of the 'common mean' assumed. Usually the 'common mean' depends on the sampled values; that is the reason why the shrinkage estimators have similarities with the *empirical* Bayes approach. In general, the 'shrinking factor' is influenced by the dimension of the study, the total number of studies included as well as the dispersion of the single values around the common mean (Braga, 2004). The effect of the empirical Bayes approach is to smooth estimates based on small numbers of events more heavily than estimates based on large numbers of events (Cox & Solomon, 1997).

3.4.2 An introduction to Stein-Estimators

"Stein (1956) obtained the surprising result that for estimating p independent normal means simultaneously, the sample mean was inadmissible under squared error loss when $p \geq 3$"(Ghosh et al., 1983, pg. 351). In this case ($p \geq 3$), "the cost of estimating the shrinkage intensity is already (and always!) offset by the savings in total risk" (Opgen-

Rhein & Strimmer, 2007, pg. 4).
Stein's paradox (or phenomenon or problem) mainly demonstrated that when 3 or more parameters are estimated simultaneously, their combined estimator is more accurate than any other method which handles the parameters separately, even when the measurements and the parameters are totally unrelated. In fact, the combined estimator achieves a lower MSE and – even if not necessarily better estimates for the single variable alone are obtained – a better estimate (which has a reduced total risk) for the means of all of the random variables is produced. *Surprisingly, the cost of a bad estimate in one component can be compensated by a better estimate in another component.* When multiple observations are present (no matter whether those observations are statistically independent), the simultaneous measurement of several parameters reduces the *total* error of the parameters. Such a correction to the reduced mean squared error can be obtained by shrinking the ordinary estimator. For example the MSE of the MLE of the variance of the normal distribution can be reduced by shrinking the estimates toward zero. (Hedges & Olkin, 1985). Furthermore, recent studies have demonstrated that in many cases "shrinking towards a data-based point yields more reduction in risk than shrinking towards the origin" (Ghosh et al., 1983, pg. 353). In general, it is sensible to use a shrinkage estimator when it is reasonable to expect the values to be quite close together and the possibility of an overall improvement in the estimates at the expense of a worsening of individual ones is considered acceptable or desirable (Cox & Hinkley, 1974). Shrinkage technique has been used in different problems and under different assumptions and settings, both in simultaneous estimation problems for normal, exponential or non-normal distributions (Ghosh et al., 1983). In practice, in fact, there are several situations where it is a requirement to shrink the usual plain estimators in order to obtain a uniformly smaller risk than the usual plain estimator. Surprising results have been shown in the estimation problems of the variance; in particular, in small sample problems the concept of shrinkage has been recognised as beneficial. (Kubokawa, 1999).

3.4.3 Properties of the Shrinkage Estimators

While the Bayes estimators make use of the prior knowledge, the usual procedures such as UMVUE and MLE neglect such a knowledge. The empirical Bayes estimator, on the contrary, can be interpreted as an intermediate of the Bayes and usual ones as this incorporates parts of the prior information (guessed or taken from the sample) even if one cannot suppose any exact prior information.
The empirical Bayesian approach to shrinking naturally allows for *information sharing across studies*, which can be important especially when the number and sizes of studies considered are small. In fact, even though "analytic shrinkage estimators combine properties that render them attractive for analyzing large-dimensional studies" (Opgen-

Rhein & Strimmer, 2007, pg. 2), these can be used for small-dimensional studies as well. Moreover, shrinkage estimators are generally *fully analytic* and usually require no computer-intensive procedures and only *little distributional assumptions*.

3.5 Why draw on Bioinformatics?

In this research, we use an estimator of the error variance that can borrow information across studies using the James-Stein shrinkage concept. Tong et al. (Tong & Wang, 2007), in particular, employed James-Stein Estimation, further developing the estimator by Cui et al. (2005), to obtain shrinkage estimates of the gene-specific variances, making only weak prior assumptions about the distribution of the variance components (the sampling distribution of the logarithm of the variance estimators is assumed to be normal) and achieving an estimator with an explicit expression that is computationally simple. Such an estimator was originally used for **microarray experiments**. In the original paper, a new test statistic was developed based and constructed on this shrinkage estimator and this provided "a powerful and robust approach to test the differential expressions of genes that utilises information not available in individual gene testing approaches and does not suffer from biases of the pooled variance approach" (Cui et al., 2005, pg. 59).

Similarly to the meta-analysis, combining information across genes in the statistical analysis of microarray data is desirable because of the relatively small number of data points obtained for each individual gene. *Small number of freedom* due to few replicates is a common situation for microarray experiments (Lin et al., n.d.). In fact, since microarrays are expensive, experiments are typically performed with a limited number of replicates. When this is the case, the use of within-gene estimates of variability provides unreliable results (Jain et al., 2003). Specifically, if variance heterogeneity is assumed, individual gene-specific tests are used even though the standard gene-specific estimators of variances are unreliable due to the relatively small number of replications (Cui et al., 2005). On the other hand, more powerful tests can be used assuming common variance; nevertheless, this assumption is unlikely to be true. "Thus, tests based on a pooled common variance estimator for all genes are at the risk of generating misleading results" (Tong & Wang, 2007, pg. 113).

"A number of approaches to improving estimates for variability and statistical tests of differential expression have thus recently emerged. Several variance function methods have been proposed" (Jain et al., 2003, 1945). In particular, over the last few years, shrinkage approaches that combine variance information across genes have been developed. Tests based on variance estimates that are gene specific but combine in-

formation across many genes are nowadays considered better approaches in microarray experiments in order to increase power by utilizing more information in the data and also to avoid bias (Cui et al., 2005). These approaches were usually proposed to handle better "*the situation where a gene with low expression may have very low variance by chance*" (Jain et al., 2003, pg. 1946).

Meta-analysis, i.e. a combination of results and information from independent quantitative investigations, shares somehow the same problems with microarray experiments. Also in meta-analysis, both the sample size within individual studies and the number of studies are typically relatively small. This is the reason why the treatment and control group sample variances as well as the $\hat{\sigma}_i^2$ may be unreliable and, as a consequence, the reported overall variance is less than the optimal value. Tests based on the common overall variance estimator for all studies are at risk of generating misleading results in meta-analyses as well (too liberal significance tests and Confidence Intervals without correct coverage properties). As in meta-analysis one is usually interested in the possibility of an overall gain (i.e. a better estimate of the overall variance of the estimator rather than merely better individual estimates $\hat{\sigma}_i^2$) and as the setting and problems faced are similar to those which crop up in Bioinformatics, we decided to borrow the shrinkage estimator (originally developed for microarray problems) to evaluate whether this would be useful in a meta-analysis context as well. Moreover, the specific modified shrinkage estimator used in our simulations usually require little assumptions and is therefore easily adaptable to diverse and numerous frameworks. "Even though motivated and applied to microarray data, the optimal shrinkage variance estimator [...] can have a wide range of applications". Tong and Wang "methodology and theory extend Stein's landmark results from shrinkage estimation of means to shrinkage estimation of variance, and from shrinkage estimation of a single variance to the shrinkage estimation of multiple variances" (Tong & Wang, 2007, pg. 121).

3.6 Shrinkage Statistic of Variance Vector: method used

A shrinkage estimator for gene-specific variance components based on the James-Stein estimator was proposed by Cui et al in 2005 (Cui et al., 2005). Their estimator made no prior assumptions about the distribution of variances across genes. The test based on such an estimator performed well under a wide range of assumptions about variance heterogeneity, behaving well both when the variances were truly constant as well as when they varied extensively from gene to gene. How did they obtain a shrinkage estimator of variance components that provided a gene-specific variance also using information across all of the genes in the data in order to improve estimation?
Stein discovered that the standard sample variance is improved by a shrinkage esti-

mator using information contained in the sample mean. Much research has been done since then. Nevertheless, most research concerned with a *single variance*, assumption not applicable to microarray data analysis since the homogeneity of the variances is unlikely to be true. Cui et al. focused on heterogenous variances and, instead of using information in the sample mean, extended Stein's theory for multiple means to *multiple variances*. This way they obtained variance estimates that were gene-specific but combined at the same time information across many genes, improving power but also avoiding bias. Cui et al. method was recently further developed and improved by Tong and Wang who presented their methods in the framework of microarray data analysis, stressing however that both their methods and theory are general may be implemented in a much wider range of scenarios (Tong & Wang, 2007).

3.6.1 CHQBC Estimator

As initially suggested by Cui et al. (2005), an improved estimator of variance from an ensemble of individual variance estimators (herein referred to as the CHQBC estimator) can be constructed by shrinking them towards their common *corrected geometric mean*. "The amount of shrinkage depends on the variability of the individual variance estimators. When individual variance estimates are similar, indicating homogeneity, the shrinkage estimator effectively pools these estimates. When individual variance estimates are widely dispersed, indicating heterogeneity, the shrinkage estimator gives greater weight" to the study specific contributions (Cui et al., 2005, pg. 61).
For g = 1, ..., G ($G \geq 3$ with G equal to the number of studies), let X_g be the residual sum of squared errors and σ_g^2 be the true variance of g. Assuming that X_g/σ_g^2's are mutually independent (each having a Chi-squared distribution χ_ν^2 with ν degrees of freedom) we have $X_g \sim \sigma_g^2 \chi_\nu^2$. Considering the natural logarithm transformation of X_g we then have

$$ln\frac{X_g}{\nu} \sim ln\sigma_g^2 + ln\frac{\chi_\nu^2}{\nu}$$

Therefore, if we denote the mean of $\ln\frac{\chi_g^2}{\nu}$ as m, by subtracting m from both sides we could write the following equation

$$X_g' \sim ln\sigma_g^2 + \epsilon_g'$$

where $X_g' = ln(X_g/\nu) - m$ and $\epsilon_g' = ln(\chi_\nu^2/\nu) - m$. Applying the James-Stein shrinkage method to X_g' and then transforming back to the original scale gives the shrinkage estimator for σ_g^2,

$$\tilde{\sigma}_g^2 = \left(\prod_{g=1}^{G}(X_g/\nu)^{1/G}\right) B * exp\left[\left(1 - \frac{(G-3)V}{\sum(lnX_g - \overline{lnX_g})^2}\right)_+ * (lnX_g - \overline{lnX_g})\right] \quad (3.1)$$

where V is the variance of ϵ_g', $\overline{lnX_g} = \frac{1}{G}\sum_{g=1}^{G} ln(X_g)$ and $B = exp(-m)$ is a bias correction (*all the details are provided by Cui et al. (2005)*). "Note that multiplying

the geometric mean $\left(\prod_{g=1}^{G}(X_g/\nu)^{1/G}\right)$ by B gives an unbiased estimator of σ^2 when $\sigma_g^2 = \sigma^2$ for all g. The values of B (and also V) depend on ν. They can be simulated easily and values are given in Table 3.6. Note that B is always larger than 1, hence, the geometric mean without B underestimates σ^2 when all σ_g^2 are equal to σ^2" (Cui et al., 2005, pg. 61).

Table 3.6: Values of B (bias correction) and V/(2/ν) as a function of ν.

ν	**B**	$V/(2/\nu)$		ν	**B**	$V/(2/\nu)$
1	3.53	2.45		13	1.08	1.08
2	1.77	1.64		14	1.08	1.08
3	1.44	1.39		15	1.07	1.07
4	1.31	1.27		16	1.07	1.06
5	1.24	1.22		17	1.06	1.06
6	1.19	1.18		18	1.06	1.06
7	1.16	1.15		19	1.06	1.05
8	1.14	1.13		20	1.05	1.05
9	1.12	1.12		25	1.04	1.04
10	1.11	1.11		30	1.04	1.03
11	1.10	1.10		40	1.03	1.03
12	1.09	1.09		50+	1.02	1.02

These values are used in equation 3.1 to construct the estimates that shrink the unbiased estimators of variances to their corrected geometric mean. When ν is greater than 50, B and V/(2/ν) are effectively 1.

3.6.2 Improvements on the CHQBC Estimator

Even though the CHQBC estimator may work well as an estimator of variance, Tong and Wang (2007) suggested an improvement to such an estimator (for full details, refer to the article). Let $Z_g = X_g/\nu$, $Z_{pool} = \prod_{g=1}^{G} Z_g^{1/G}$ and $\hat{\alpha}_0 = 1 - (1 - (G - 3)V/\sum(lnX_g - \overline{lnX_g})^2)_+$. It is easy to check that the CHQBC estimator can be rewritten as

$$\tilde{\sigma}_g^2 = B(Z_{pool})^{\alpha}(Z_g)^{1-\alpha} \qquad \text{with} \quad \alpha = \hat{\alpha}_0.$$

Note that when $\sigma_g^2 = \sigma^2$ for all g, $E(Z_{pool}) = \sigma^2/B$. That is, BZ_{pool} is an unbiased estimator of σ^2 when $\sigma_g^2 = \sigma^2$ for all g. On the other hand, Z_g is an unbiased estimator of σ_g^2. Therefore, it is reasonable to consider the following combination of two unbiased estimators

$$\sigma_g^{2+} = (BZ_{pool})^{\alpha}(Z_g)^{(1-\alpha)}, \qquad 0 \leq \alpha \leq 1. \tag{3.2}$$

Referring to $\sigma_g^{2+}(\hat{\alpha_0})$ as the *modified CHQBC estimator*, $\sigma_g^{2+}(\hat{\alpha_0})$ in the simulations

shown always performs better than the original CHQBC estimator $\tilde{\sigma}_g^2(\hat{\alpha_0})$ for estimating σ_g^2. "The estimator σ_g^{2+} has a very simple structure; it borrows information across studies by shrinking each specific variance towards the bias corrected geometric mean of variances for all studies. The amount of shrinkage depends on the variability of the individual variances. In particular, the shrinkage parameter $\hat{\alpha}$ was obtained by applying the James-Stein method to the logarithm of sample variances which do not follow the normal distribution" (Tong & Wang, 2007, pg. 114). "Although the James-Stein shrinkage estimator was developed in a context of a normal model, it is the sampling distribution of the logarithm of the variance estimators, not the values themselves, that are assumed to be normal" (Cui et al., 2005, pg. 73). On the logarithm scale, the *modified CHQBC estimator* is a weighted average of the study-specific variance and the bias corrected geometric mean. If the empirical variances can be reliably determined from the data, and consequently exhibit only a small variance themselves, there will be little shrinkage, whereas if the empirical variance is comparatively large pooling across studies will take place.

According to Tong and Wang simulations, "the modified CHQBC estimator $\sigma_g^{2+}(\hat{\alpha_0})$ has smaller risk than the original CHQBC estimator $\tilde{\sigma}_g^2(\hat{\alpha_0})$ in *all* settings" (Tong & Wang, 2007, pg. 117), in particular when the variance heterogeneity and ν are both small.

3.7 Comparison of methods by Simulations

We perform simulation studies to compare $\hat{V}$, the usual overall variance estimator which uses the estimated within variances, with $\hat{V}^*$, the estimator that takes into account the 'shrunk' variances. We then consider the ordinary estimators $\hat{\theta}$ and $\hat{V}$ as in equations 2.3 and 2.4 with $\hat{w}_i = 1/\hat{\sigma}_i^2$ where

$$\hat{\sigma}_i^2 = \hat{S}_i^2 \left(\frac{1}{n_{Ti}} + \frac{1}{n_{Ci}} \right) \qquad \text{with} \quad \hat{S}_i^2 = \frac{(n_{Ti}-1)\hat{s}_{Ti}^2 + (n_{Ci}-1)\hat{s}_{Ci}^2}{n_{Ti}+n_{Ci}-2}$$

When 'shrunk' variances are considered, we denote the 'new' point and overall variance estimators as follows

$$\hat{\theta}^* = \frac{\sum_{i=1}^k \hat{w}_i^* Y_i}{\sum_{i=1}^k \hat{w}_i^*} \quad \text{and} \quad \hat{V}^* = \frac{1}{\sum_{i=1}^k \hat{w}_i^*} \tag{3.3}$$

where

$$\hat{w}_i^* = \frac{1}{\hat{\sigma}_{si}^2} \qquad \text{with} \quad \hat{\sigma}_{is}^2 = \hat{S}_{shr.i}^2 \left(\frac{1}{n_{Ti}} + \frac{1}{n_{Ci}} \right) \tag{3.4}$$

where $\hat{S}_{shr.i}^2$'s are obtained with the shrinkage estimator (see eqn 3.2) applied on $\hat{S}_i^2$'s. The empirical coverage probabilities of the two Confidence Intervals based on $\hat{V}$ and $\hat{V}^*$ are computed and compared as well.

Since it is known that errors are present in the estimated sample variances, the objective of this study is to evaluate whether the 'improved' CHQBC estimator improves the behaviour of the overall variance estimates, i.e. to establish whether $\hat{V}^*$ performs in general better than $\hat{V}$. More specifically, we want to evaluate the robustness, precision and accuracy of the overall variance estimator used in fixed-effects meta-analysis when 'shrunk' estimates of the pooled variances are used. Theoretically, the shrinkage estimators for the pooled variances should better handle the situation when a single study carry a disproportionate amount of weight and has few subjects and a very low sample variances just by chance; moreover, 'shrunk' variances should take into consideration the effect of study sizes as well. $\hat{V}^*$ should therefore be less sensitive to any individual results.
In order to assess and evaluate the goodness of the estimator whose weights are based on the 'shrunk' variances, the dispersion around the optimal value will be computed. In particular, non only $\hat{V}^*$ should yield values on average closer to the optimal levels, but also the 'real' dispersion should not be too imprecise nor be too far from both the 'declared' and the 'optimal' values.

Simulation settings are as specified in section 3.2. In addition, the common variance scenario is considered, i.e. the case where $S_1^2 = 100$ for all i. Results are given in tables 3.7 - 3.12 for the different S_i^2 settings. Each table shows the Confidence Intervals based on both $\hat{V}$ and $\hat{V}^*$. Moreover, since when we simulate studies we have the privilege to know the 'real' values of each single within study variances, the average (over 10000 replicates) 'optimal' overall variance of the estimator is calculated. The 'declared' and 'actual' dispersion for both methods are also given. Furthermore, in order to make comparisons between the two methods easier, ratio index numbers are shown in columns 10-15. Ratio Index numbers measure changes or differences and are used in a variety of contexts to compare series of numbers of different size in a standardised and directly comparable way. An index number is generally formed by the ratio between the 'current value' of an indicator and its 'base value', against which all the observations are measured and compared. The ratio itself has no units and expresses the changes around the base. In Tables 3.7 - 3.12, columns *10* and *11* are the ratio between the 'real' dispersion of the estimator and its 'declared' value, for $\hat{V}$ and $\hat{V}^*$ respectively. Values in these columns indicate how much larger the 'real' dispersion is compared to the 'declared' one; the bigger the value, the wider the absolute difference between the two numbers. Columns *12 - 15* show the ratio between either the 'real' or the 'declared' variances (for both $\hat{V}$, $\hat{V}^*$) and the 'optimal' value which is the base for all four columns. Values less than 1 mean that there is a negative bias in the estimate while values greater than 1 indicate the opposite. The ideal situation would be to have both the ratio of the 'declared' and the 'real' dispersion to the 'optimal' equal to 1,

indicating perfect estimation of the overall variance of the estimator.

From the results in Tables 3.7 - 3.12, we note that $\hat{V}^*$ always perform better than $\hat{V}$. Observing the absolute values, the estimator that uses weights based on the variance shrinkage estimator has values closer to the 'optimal' ones. In particular, not only is the 'declared' variance closer to the 'optimal', but the 'real' dispersion is also not badly estimated. This means that the new method tends to underestimate the 'optimal' value less severely, on average and in all cases. In practice, when performing a single meta-analysis, with the new method we tend to declare on average a variability of the point estimate closer to the correct one; moreover, the 'real' dispersion of the new method is smaller than the one obtained with the ordinary estimator, meaning that the new method yields less liberal results because of the slight difference between the 'real' and the 'declared' dispersion of $\hat{V}^*$. As a consequence, in general we still tend to overstate the precision of the estimator, but less badly. In fact, observing the absolute values, we can note that the following relationship always hold

$$V^{real}(\hat{\theta}) \geq V^{real}(\hat{\theta}^*) \geq V^{opt}$$

As regards the 'declared' variances, in general we have '

$$V^{opt} \geq E[\hat{V}^*] \geq E[\hat{V}]$$

Nevertheless, it may happen that the 'average' declared variance obtained using the new method is slightly larger than the 'optimal' value. Again, if we consider the absolute difference between the 'optimal' and the 'declared' variances obtained with the usual and the new methods, such a difference is always smaller when the new method is taken into account. For instance, when $n > 10$, the new method may declare a variance 1% greater than the 'optimal' instead of underestimating it by between 3 and 12%. In the worse scenario (Table 3.10), when n=5 and k=35, the new method overstates the variance by 10% whereas the usual method underestimates it by 35%.
In general, however, the following relationship holds

$$V^{real}(\hat{\theta}) \geq V^{real}(\hat{\theta}^*) \geq V^{opt} \geq E[\hat{V}^*] \geq E[\hat{V}]$$

This relationship is confirmed even when ratio index numbers are taken into consideration. The ratio $V^{real}(\hat{\theta}^*)/E[\hat{V}^*]$ is always smaller than $V^{real}(\hat{\theta})/E[\hat{V}]$, entailing that the variability (and range) of the new method is smaller than the one obtained with the usual method. In addition, from columns *12* to *15* we note that, given 1 as the optimal value, both $V^{real}(\hat{\theta}^*)$ and $E[\hat{V}^*]$ are closer to the target 1. If the difference between the two methods is almost imperceptible when n is large, this becomes dramatically important when small studies are combined. For instance, Table 3.12 shows that for $n = 5$ the 'optimal' variance is underestimated on average by only around 15% with

the new method instead of 40% or more with the ordinary method. It is worth to noting that even when n increases, both $V^{real}(\hat{\theta}^*)$ and $E[\hat{V}^*]$ are more accurate and less dispersed around the 'optimal' value.

As a consequence, turning to the empirical coverage probabilities of the the 95% Confidence Intervals for the two methods, we note that with the new method values are always closer to the nominal value (and only in few occasions above the nominal level). Especially when n is small, the coverage probabilities for the interval based on $\hat{V}^*$ are much better than the usual ones. They still suffer from inadequate coverage; nevertheless, the estimated Coverage probabilities of the CIs based on $\hat{V}$ fall well below the ones obtained with the new method whose weights were shrunk.

These results generally emphasise that the estimator based on the 'shrunk variances' rather than the estimated ones performs better. Regardless of the variance structure across studies (homoscedasticity or uncommon variances), the ordinary method shows values close to the optimal only if large sized studies are summarised. The new method, on the other hand, performs optimally even with an average small number of patients per trial. $\hat{V}^*$ shows a certain accuracy and flexibility since better results are achieved even when variances are strongly heterogenous. It is quite remarkable that the new method based on the shrinkage estimators proposed by Tong and Wang performs well, providing highly accurate overall variances for simulated data for all considered scenarios without relying on computational expensive procedures.

Table 3.7: Comparison of the Results of Simulations Considering the 'estimated' and the 'shrunk' variances for different k under the assumption of Unequal S_i^2 and Study sizes - First Simulation Scheme (A)

	n	Coverage Prob (CP)	Improved CP	$V^{real}(\hat{\theta})$	$V^{real}(\hat{\theta}^*)$	V^{opt}	$\mathbf{E}[\hat{V}^*]$	$\mathbf{E}[\hat{V}]$	$\frac{V^{real}(\hat{\theta})}{\mathrm{E}[\hat{V}]}$	$\frac{V^{real}(\hat{\theta}^*)}{\mathrm{E}[\hat{V}^*]}$	$\frac{V^{real}(\hat{\theta})}{V^{opt}}$	$\frac{V^{real}(\hat{\theta}^*)}{V^{opt}}$	$\frac{E[\hat{V}^*]}{V^{opt}}$	$\frac{E[\hat{V}]}{V^{opt}}$
k = 10	**5**	0.6988	0.8084	2.917085	1.876222	0.844070	0.710398	0.519157	5.6189	2.6411	3.4560	2.2228	0.8416	0.6151
	8	0.8652	0.9018	0.822160	0.650713	0.486116	0.445866	0.401863	2.0459	1.4594	1.6913	1.3386	0.9172	0.8267
	10	0.9046	0.9209	0.496175	0.454169	0.378410	0.358564	0.335633	1.4783	1.2666	1.3112	1.2002	0.9476	0.8870
	15	0.9344	0.9399	0.261349	0.260267	0.247350	0.240428	0.231493	1.1290	1.0825	1.0566	1.0522	0.9720	0.9359
	20	0.9333	0.9375	0.197728	0.197418	0.183608	0.180408	0.175550	1.1263	1.0943	1.0769	1.0752	0.9826	0.9561
	30	0.9419	0.9440	0.125123	0.125059	0.121904	0.120502	0.118438	1.0565	1.0378	1.0264	1.0259	0.9885	0.9716
	50	0.9423	0.9434	0.075708	0.075690	0.073157	0.072625	0.071895	1.0530	1.0422	1.0349	1.0346	0.9927	0.9828
k = 15	**5**	0.6644	0.8172	2.515061	1.377154	0.567343	0.483562	0.341732	7.3598	2.8479	4.4331	2.4274	0.8523	0.6023
	8	0.8487	0.8984	0.608366	0.447773	0.335919	0.311271	0.275871	2.2053	1.4385	1.8110	1.3330	0.9266	0.8212
	10	0.8970	0.9192	0.361150	0.320755	0.263301	0.250819	0.232215	1.5552	1.2788	1.3716	1.2182	0.9526	0.8819
	15	0.9307	0.9375	0.187226	0.185702	0.172940	0.169038	0.161681	1.1580	1.0986	1.0826	1.0738	0.9774	0.9349
	20	0.9367	0.9417	0.134901	0.134427	0.129138	0.127098	0.123035	1.0964	1.0577	1.0446	1.0410	0.9842	0.9527
	30	0.9449	0.9476	0.087392	0.087334	0.085815	0.084860	0.083122	1.0514	1.0292	1.0184	1.0177	0.9889	0.9686
	50	0.9479	0.9493	0.051520	0.051503	0.051463	0.051139	0.050523	1.0197	1.0071	1.0011	1.0008	0.9937	0.9817
k = 20	**5**	0.6381	0.8242	2.050539	0.944725	0.393981	0.334267	0.231733	8.8487	2.8263	5.2047	2.3979	0.8484	0.5882
	8	0.8494	0.9079	0.599061	0.332034	0.234913	0.217217	0.191897	3.1218	1.5286	2.5501	1.4134	0.9247	0.8169
	10	0.8943	0.9188	0.271898	0.231411	0.185648	0.175938	0.162671	1.6715	1.3153	1.4646	1.2465	0.9477	0.8762
	15	0.9269	0.9333	0.132802	0.131815	0.122398	0.119218	0.114006	1.1649	1.1057	1.0850	1.0769	0.9740	0.9314
	20	0.9353	0.9389	0.097786	0.097418	0.091385	0.089732	0.086845	1.1260	1.0857	1.0701	1.0660	0.9819	0.9503
	30	0.9422	0.9455	0.062851	0.062755	0.060831	0.060099	0.058860	1.0678	1.0442	1.0332	1.0316	0.9880	0.9676
	50	0.9480	0.9498	0.036432	0.036431	0.036428	0.036193	0.035755	1.0187	1.0066	1.0001	1.0001	0.9936	0.9815
k = 35	**5**	0.5592	0.8171	2.019225	0.593985	0.222689	0.188896	0.124061	16.2761	3.1445	9.0675	2.6673	0.8483	0.5571
	8	0.8233	0.9051	0.441426	0.181561	0.135820	0.126070	0.109607	4.0273	1.4402	3.2501	1.3368	0.9282	0.8070
	10	0.8911	0.9260	0.163837	0.125479	0.107798	0.102658	0.094237	1.7386	1.2223	1.5198	1.1640	0.9523	0.8742
	15	0.9329	0.9405	0.076543	0.075427	0.071281	0.069626	0.066329	1.1540	1.0833	1.0738	1.0582	0.9768	0.9305
	20	0.9398	0.9441	0.055763	0.055544	0.053325	0.052417	0.050581	1.1025	1.0597	1.0457	1.0416	0.9830	0.9485
	30	0.9416	0.9450	0.037028	0.037003	0.035547	0.035146	0.034353	1.0779	1.0528	1.0416	1.0410	0.9887	0.9664
	50	0.9484	0.9497	0.021329	0.021328	0.021306	0.021164	0.020883	1.0214	1.0078	1.0011	1.0010	0.9933	0.9801

This Table shows the results of simulation for $\theta = 3$ and different values of n, k and S_i^2. This simulation scheme considers for each simulation k parallel group clinical trials each of whom with the same number of patients per arm only *on average* ($n = E[n_i] = n_{Ti} = n_{Ci}$). As regards the within-study variances, half of the $S_i^2 = 100$ and the other half is equal to 10. Empirical Statistics for $E[\hat{V}]$, $E[\hat{V}^*]$, V^{opt}, $V^{real}(\hat{\theta})$ and $V^{real}(\hat{\theta}^*)$ are based on 10000 simulation replicates as well as the Empirical Coverage Probability.

Table 3.8: Comparison of the Results of Simulations Considering the 'estimated' and the 'shrunk' variances for different k under the assumption of Unequal S_i^2 and Study sizes - First Simulation Scheme (B)

	n	Coverage Prob (CP)	Improved CP	$V^{real}(\hat{\theta})$	$V^{real}(\hat{\theta}^*)$	V^{opt}	$\mathbf{E}[\hat{V}^*]$	$\mathbf{E}[\hat{V}]$	$\frac{V^{real}(\hat{\theta})}{\mathrm{E}[\hat{V}]}$	$\frac{V^{real}(\hat{\theta}^*)}{\mathrm{E}[\hat{V}^*]}$	$\frac{V^{real}(\hat{\theta})}{V^{opt}}$	$\frac{V^{real}(\hat{\theta}^*)}{V^{opt}}$	$\frac{E[\hat{V}^*]}{V^{opt}}$	$\frac{E[\hat{V}]}{V^{opt}}$
k = 10	**5**	0.7069	0.8296	5.787174	3.965697	1.724529	1.588877	1.090686	5.3060	2.4959	3.3558	2.2996	0.9213	0.6325
	8	0.8691	0.9074	1.716196	1.363653	0.985787	0.950408	0.822931	2.0855	1.4348	1.7409	1.3833	0.9641	0.8348
	10	0.9041	0.9262	0.951359	0.899337	0.763757	0.754291	0.681596	1.3958	1.1923	1.2456	1.1775	0.9876	0.8924
	15	0.9274	0.9368	0.537592	0.539245	0.490154	0.493515	0.462380	1.1627	1.0927	1.0968	1.1002	1.0069	0.9433
	20	0.9400	0.9459	0.380268	0.380441	0.363268	0.366476	0.348901	1.0899	1.0381	1.0468	1.0473	1.0088	0.9605
	30	0.9439	0.9469	0.245991	0.246253	0.240309	0.241378	0.233714	1.0525	1.0202	1.0236	1.0247	1.0044	0.9726
	50	0.9471	0.9498	0.144902	0.144957	0.143970	0.144463	0.141717	1.0225	1.0034	1.0065	1.0069	1.0034	0.9844
k = 15	**5**	0.6625	0.8299	4.947632	2.529562	1.087694	0.987782	0.656184	7.5400	2.5609	4.5487	2.3256	0.9081	0.6033
	8	0.8620	0.9147	1.418933	0.988629	0.631846	0.610570	0.522482	2.7158	1.6192	2.2457	1.5647	0.9663	0.8269
	10	0.8984	0.9230	0.631725	0.577880	0.496492	0.491923	0.440474	1.4342	1.1747	1.2724	1.1639	0.9908	0.8872
	15	0.9361	0.9456	0.334990	0.333555	0.323049	0.325176	0.302369	1.1079	1.0258	1.0370	1.0325	1.0066	0.9360
	20	0.9377	0.9437	0.251068	0.250842	0.240756	0.243159	0.230016	1.0915	1.0316	1.0428	1.0419	1.0100	0.9554
	30	0.9418	0.9478	0.165133	0.164979	0.159585	0.160991	0.155199	1.0640	1.0248	1.0348	1.0338	1.0088	0.9725
	50	0.9499	0.9518	0.095679	0.095675	0.095648	0.095189	0.094113	1.0166	1.0051	1.0003	1.0003	0.9952	0.9839
k = 20	**5**	0.6373	0.8279	4.892815	2.162006	0.796330	0.719145	0.467671	10.4621	3.0064	6.1442	2.7150	0.9031	0.5873
	8	0.8506	0.9123	1.258493	0.723114	0.467765	0.454560	0.386032	3.2601	1.5908	2.6904	1.5459	0.9718	0.8253
	10	0.8989	0.9290	0.492900	0.430310	0.368996	0.366807	0.326740	1.5085	1.1731	1.3358	1.1662	0.9941	0.8855
	15	0.9319	0.9448	0.257675	0.257210	0.241634	0.243951	0.225962	1.1403	1.0544	1.0664	1.0645	1.0096	0.9351
	20	0.9346	0.9421	0.191488	0.191426	0.179803	0.181414	0.171074	1.1193	1.0552	1.0650	1.0646	1.0090	0.9515
	30	0.9473	0.9521	0.122043	0.121898	0.119628	0.120647	0.116054	1.0516	1.0104	1.0202	1.0190	1.0085	0.9701
	50	0.9428	0.9465	0.073591	0.073552	0.071738	0.072131	0.070481	1.0441	1.0197	1.0258	1.0253	1.0055	0.9825
k = 35	**5**	0.5698	0.8400	4.098463	1.092604	0.436752	0.397013	0.245839	16.6713	2.7521	9.3840	2.5017	0.9090	0.5629
	8	0.8319	0.9145	0.845240	0.349540	0.262924	0.254336	0.212878	3.9705	1.3743	3.2148	1.3294	0.9673	0.8097
	10	0.8869	0.9240	0.403127	0.254163	0.207927	0.205927	0.182103	2.2137	1.2342	1.9388	1.2224	0.9904	0.8758
	15	0.9319	0.9438	0.146905	0.146200	0.137120	0.138575	0.127746	1.1500	1.0550	1.0714	1.0662	1.0106	0.9316
	20	0.9344	0.9421	0.111066	0.110946	0.102539	0.103688	0.097383	1.1405	1.0700	1.0832	1.0820	1.0112	0.9497
	30	0.9443	0.9497	0.069944	0.069873	0.068164	0.068779	0.065977	1.0601	1.0159	1.0261	1.0251	1.0090	0.9679
	50	0.9444	0.9479	0.042369	0.042352	0.040884	0.041125	0.040116	1.0562	1.0298	1.0363	1.0359	1.0059	0.9812

This Table shows the results of simulation for $\theta = 3$ and different values of n, k and S_i^2. This simulation scheme considers for each simulation k parallel group clinical trials each of whom with the same number of patients per arm only *on average* ($n = E[n_i] = n_{Ti} = n_{Ci}$). As regards the within-study variances, the most imbalanced pattern is considered; roughly 80% of S_i^2 are set to 100 with the remaining are set equal to 10. Empirical Statistics for $E[\hat{V}]$, $E[\hat{V}^*]$, V^{opt}, $V^{real}(\hat{\theta})$ and $V^{real}(\hat{\theta}^*)$ are based on 10000 simulation replicates as well as the Empirical Coverage Probability.

Table 3.9: Comparison of the Results of Simulations Considering the 'estimated' and the 'shrunk' variances for different k under the assumption of Equal S_i^2 and Study sizes - First Simulation Scheme (C)

	n	Coverage Prob (CP)	Improved CP	$V^{real}(\hat{\theta})$	$V^{real}(\hat{\theta}^*)$	V^{opt}	$\mathbf{E}[\hat{V}^*]$	$\mathbf{E}[\hat{V}]$	$\frac{V^{real}(\hat{\theta})}{\mathrm{E}[\hat{V}]}$	$\frac{V^{real}(\hat{\theta}^*)}{\mathrm{E}[\hat{V}^*]}$	$\frac{V^{real}(\hat{\theta})}{V^{opt}}$	$\frac{V^{real}(\hat{\theta}^*)}{V^{opt}}$	$\frac{E[\hat{V}^*]}{V^{opt}}$	$\frac{E[\hat{V}]}{V^{opt}}$
k = 10	**5**	0.6776	0.8362	13.932203	7.761157	4.383944	3.606886	2.627364	5.3027	2.1518	3.1780	1.7704	0.8227	0.5993
	8	0.8559	0.9024	4.603915	3.440154	2.589141	2.336047	2.131950	2.1595	1.4726	1.7782	1.3287	0.9022	0.8234
	10	0.9009	0.9252	2.621340	2.296118	2.053049	1.907176	1.804959	1.4523	1.2039	1.2768	1.1184	0.9289	0.8792
	15	0.9330	0.9421	1.446309	1.387196	1.348546	1.301078	1.259245	1.1486	1.0662	1.0725	1.0287	0.9648	0.9338
	20	0.9389	0.9460	1.071095	1.041215	1.006607	0.982020	0.956923	1.1193	1.0603	1.0641	1.0344	0.9756	0.9506
	30	0.9451	0.9488	0.671534	0.669052	0.668573	0.655896	0.647553	1.0370	1.0201	1.0044	1.0007	0.9810	0.9686
	50	0.9495	0.9507	0.401515	0.401218	0.400902	0.395643	0.393509	1.0203	1.0141	1.0015	1.0008	0.9869	0.9816
k = 15	**5**	0.6286	0.8433	12.865521	5.379387	2.848081	2.350470	1.648318	7.8052	2.2886	4.5173	1.8888	0.8253	0.5787
	8	0.8440	0.9103	3.472281	2.186208	1.710592	1.543520	1.396262	2.4868	1.4164	2.0299	1.2780	0.9023	0.8162
	10	0.8980	0.9260	1.874152	1.528303	1.357033	1.261771	1.190205	1.5746	1.2112	1.3811	1.1262	0.9298	0.8771
	15	0.9311	0.9419	0.961792	0.922547	0.894816	0.861010	0.832305	1.1556	1.0715	1.0748	1.0310	0.9622	0.9301
	20	0.9372	0.9438	0.708870	0.686861	0.670588	0.654584	0.637030	1.1128	1.0493	1.0571	1.0243	0.9761	0.9500
	30	0.9463	0.9500	0.452064	0.445644	0.445236	0.437140	0.431201	1.0484	1.0195	1.0153	1.0009	0.9818	0.9685
	50	0.9438	0.9463	0.278931	0.267958	0.267025	0.263635	0.262110	1.0642	1.0164	1.0446	1.0035	0.9873	0.9816
k = 20	**5**	0.6031	0.8510	11.427807	3.868244	2.096980	1.740143	1.194155	9.5698	2.2229	5.4497	1.8447	0.8298	0.5695
	8	0.8320	0.9070	2.861575	1.582369	1.277377	1.149683	1.032548	2.7714	1.3764	2.2402	1.2388	0.9000	0.8083
	10	0.8923	0.9251	1.521242	1.128749	1.010111	0.938886	0.884435	1.7200	1.2022	1.5060	1.1175	0.9295	0.8756
	15	0.9319	0.9435	0.725269	0.691283	0.669735	0.644486	0.622684	1.1647	1.0726	1.0829	1.0322	0.9623	0.9297
	20	0.9396	0.9485	0.529465	0.511877	0.501333	0.488486	0.475182	1.1142	1.0479	1.0561	1.0210	0.9744	0.9478
	30	0.9420	0.9464	0.345586	0.337646	0.333686	0.327000	0.322393	1.0719	1.0326	1.0357	1.0119	0.9800	0.9662
	50	0.9461	0.9495	0.206331	0.203752	0.200363	0.197695	0.196521	1.0499	1.0306	1.0298	1.0169	0.9867	0.9808
k = 35	**5**	0.5323	0.8542	10.812036	2.346861	1.172893	0.981972	0.639560	16.9054	2.3899	9.2183	2.0009	0.8372	0.5453
	8	0.8146	0.9182	2.384724	0.845162	0.722304	0.651800	0.580498	4.1081	1.2967	3.3016	1.1701	0.9024	0.8037
	10	0.8976	0.9344	0.871098	0.612211	0.575768	0.535051	0.503311	1.7307	1.1442	1.5129	1.0633	0.9293	0.8742
	15	0.9256	0.9375	0.421685	0.399635	0.382415	0.367861	0.355109	1.1875	1.0864	1.1027	1.0450	0.9619	0.9286
	20	0.9383	0.9452	0.297313	0.286424	0.286198	0.279193	0.271432	1.0954	1.0259	1.0388	1.0008	0.9755	0.9484
	30	0.9433	0.9480	0.198110	0.193776	0.190646	0.186878	0.184161	1.0757	1.0369	1.0391	1.0164	0.9802	0.9660
	50	0.9416	0.9447	0.120361	0.118815	0.114415	0.112862	0.112162	1.0731	1.0527	1.0520	1.0384	0.9864	0.9803

This Table shows the results of simulation for $\theta = 3$ and different values of n, k. This simulation scheme considers for each simulation k parallel group clinical trials each of whom with the same number of patients per arm only *on average* ($n = E[n_i] = n_{Ti} = n_{Ci}$). In this simulation scheme, within-study variances S_i^2 are imposed to be all equal to 100, so this is the balanced case. Empirical Statistics for $E[\hat{V}]$, $E[\hat{V}^*]$, V^{opt}, $V^{real}(\hat{\theta})$ and $V^{real}(\hat{\theta}^*)$ are based on 10000 simulation replicates as well as the Empirical Coverage Probability.

Table 3.10: Comparison of the Results of Simulations Considering the 'estimated' and the 'shrunk' variances for different k under the assumption of Unequal S_i^2 and Study sizes - Second Simulation Scheme (A)

	n	**Coverage Prob (CP)**	**Improved CP**	$V^{real}(\hat{\theta})$	$V^{real}(\hat{\theta}^*)$	V^{opt}	$\mathbf{E}[\hat{V}^*]$	$\mathbf{E}[\hat{V}]$	$\frac{V^{real}(\hat{\theta})}{\mathrm{E}[\hat{V}]}$	$\frac{V^{real}(\hat{\theta}^*)}{\mathrm{E}[\hat{V}^*]}$	$\frac{V^{real}(\hat{\theta})}{V^{opt}}$	$\frac{V^{real}(\hat{\theta}^*)}{V^{opt}}$	$\frac{E[\hat{V}^*]}{V^{opt}}$	$\frac{E[\hat{V}]}{V^{opt}}$
k = 10	**5**	0.8079	0.8573	4.8107E-04	3.9726E-04	1.5392E-04	1.5289E-04	1.0875E-04	4.4237	2.5983	3.1256	2.5810	0.9933	0.7065
	8	0.8982	0.9158	1.6291E-04	1.3177E-04	9.2937E-05	9.0269E-05	7.9213E-05	2.0566	1.4597	1.7529	1.4178	0.9713	0.8523
	10	0.9166	0.9266	8.8152E-05	8.6180E-05	7.3219E-05	7.2405E-05	6.6383E-05	1.3279	1.1902	1.2040	1.1770	0.9889	0.9066
	15	0.9340	0.9394	4.7361E-05	4.7432E-05	4.4143E-05	4.4054E-05	4.2126E-05	1.1243	1.0767	1.0729	1.0745	0.9980	0.9543
	20	0.9401	0.9440	3.3747E-05	3.3660E-05	3.3090E-05	3.2908E-05	3.1890E-05	1.0582	1.0228	1.0199	1.0172	0.9945	0.9637
	30	0.9460	0.9474	2.4549E-05	2.4555E-05	2.2396E-05	2.2362E-05	2.1923E-05	1.1198	1.0981	1.0961	1.0964	0.9985	0.9788
	50	0.9467	0.9475	1.3772E-05	1.3757E-05	1.3727E-05	1.3692E-05	1.3534E-05	1.0176	1.0047	1.0033	1.0022	0.9975	0.9860
k = 15	**5**	0.8110	0.8703	1.5598E-04	1.0739E-04	3.9930E-05	4.1752E-05	2.7460E-05	5.6801	2.5720	3.9063	2.6893	1.0456	0.6877
	8	0.8924	0.9137	4.0242E-05	3.5682E-05	2.2013E-05	2.2078E-05	1.8794E-05	2.1412	1.6162	1.8281	1.6209	1.0029	0.8538
	10	0.9186	0.9297	2.3606E-05	2.1291E-05	1.8073E-05	1.8161E-05	1.6412E-05	1.4383	1.1724	1.3061	1.1781	1.0049	0.9081
	15	0.9363	0.9419	1.0912E-05	1.0953E-05	1.0885E-05	1.0875E-05	1.0417E-05	1.0475	1.0072	1.0024	1.0062	0.9990	0.9569
	20	0.9409	0.9446	8.5354E-06	8.5243E-06	8.3544E-06	8.3061E-06	8.0127E-06	1.0652	1.0263	1.0217	1.0203	0.9942	0.9591
	30	0.9469	0.9488	5.5327E-06	5.5316E-06	5.4915E-06	5.4679E-06	5.3444E-06	1.0352	1.0117	1.0075	1.0073	0.9957	0.9732
	50	0.9435	0.9446	3.1322E-06	3.1306E-06	3.0572E-06	3.0636E-06	3.0233E-06	1.0360	1.0219	1.0245	1.0240	1.0021	0.9889
k = 20	**5**	0.8038	0.8751	7.2577E-05	5.0415E-05	1.4328E-05	1.4919E-05	9.2705E-06	7.8288	3.3792	5.0655	3.5187	1.0413	0.6470
	8	0.8961	0.9227	1.8125E-05	1.5748E-05	8.1707E-06	8.3649E-06	7.0463E-06	2.5723	1.8826	2.2183	1.9273	1.0238	0.8624
	10	0.9168	0.9294	8.2919E-06	7.8098E-06	6.3326E-06	6.4844E-06	5.7971E-06	1.4303	1.2044	1.3094	1.2333	1.0240	0.9154
	15	0.9387	0.9438	4.0845E-06	4.0387E-06	4.0174E-06	4.0039E-06	3.8190E-06	1.0695	1.0087	1.0167	1.0053	0.9967	0.9506
	20	0.9416	0.9447	3.4981E-06	3.5015E-06	2.8793E-06	2.9304E-06	2.8189E-06	1.2409	1.1949	1.2149	1.2161	1.0177	0.9790
	30	0.9424	0.9448	1.9837E-06	1.9815E-06	1.9699E-06	1.9730E-06	1.9236E-06	1.0312	1.0043	1.0070	1.0059	1.0016	0.9765
	50	0.9469	0.9486	1.2642E-06	1.2650E-06	1.2572E-06	1.2511E-06	1.2334E-06	1.0249	1.0111	1.0055	1.0061	0.9951	0.9811
k = 35	**5**	0.8040	0.8847	1.2592E-05	7.6874E-06	1.7940E-06	1.9839E-06	1.1763E-06	10.7052	3.8748	7.0188	4.2850	1.1058	0.6556
	8	0.8943	0.9267	1.5179E-06	1.3864E-06	1.0790E-06	1.1269E-06	9.3255E-07	1.6277	1.2303	1.4068	1.2849	1.0443	0.8643
	10	0.9135	0.9318	1.1979E-06	9.9714E-07	7.9721E-07	8.1329E-07	7.2024E-07	1.6632	1.2261	1.5026	1.2508	1.0202	0.9034
	15	0.9365	0.9448	5.4835E-07	5.4808E-07	5.0567E-07	5.0936E-07	4.7939E-07	1.1439	1.0760	1.0844	1.0839	1.0073	0.9480
	20	0.9423	0.9463	3.6561E-07	3.6152E-07	3.5216E-07	3.5001E-07	3.3729E-07	1.0840	1.0329	1.0382	1.0266	0.9939	0.9578
	30	0.9437	0.9463	2.9271E-07	2.9277E-07	2.6710E-07	2.6825E-07	2.6113E-07	1.1209	1.0914	1.0959	1.0961	1.0043	0.9776
	50	0.9486	0.9498	1.5786E-07	1.5901E-07	1.5779E-07	1.5845E-07	1.5598E-07	1.0120	1.0036	1.0004	1.0078	1.0042	0.9886

This Table shows the results of simulation for $\theta = 3$ and different values of n, k and S_i^2. This simulation scheme considers for each simulation k parallel group clinical trials each of whom with the same number of patients per arm only *on average* ($n = E[n_i] = n_{Ti} = n_{Ci}$). As regards the within-study variances, S_i^2's are drawn from a Γ distribution with shape parameter $\gamma = 0.25$. Empirical Statistics for $E[\hat{V}]$, $E[\hat{V}^*]$, V^{opt}, $V^{real}(\hat{\theta})$ and $V^{real}(\hat{\theta}^*)$ are based on 10000 simulation replicates as well as the Empirical Coverage Probability.

Table 3.11: Comparison of the Results of Simulations Considering the 'estimated' and the 'shrunk' variances for different k under the assumption of Unequal S_i^2 and Study sizes - Second Simulation Scheme (B)

	n	Coverage Prob (CP)	Improved CP	$V^{real}(\hat{\theta})$	$V^{real}(\hat{\theta}^*)$	V^{opt}	$\mathbf{E}[\hat{V}^*]$	$\mathbf{E}[\hat{V}]$	$\frac{V^{real}(\hat{\theta})}{\mathrm{E}[\hat{V}]}$	$\frac{V^{real}(\hat{\theta}^*)}{\mathrm{E}[\hat{V}^*]}$	$\frac{V^{real}(\hat{\theta})}{V^{opt}}$	$\frac{V^{real}(\hat{\theta}^*)}{V^{opt}}$	$\frac{E[\hat{V}^*]}{V^{opt}}$	$\frac{E[\hat{V}]}{V^{opt}}$
k = 10	**5**	0.7165	0.8255	0.0515724	0.0347001	0.0155874	0.0140105	0.0098259	5.2486	2.4767	3.3086	2.2262	0.8988	0.6304
	8	0.8733	0.9088	0.0154709	0.0123376	0.0089368	0.0084475	0.0074546	2.0754	1.4605	1.7311	1.3805	0.9452	0.8341
	10	0.9128	0.9295	0.0085677	0.0078429	0.0069199	0.0067256	0.0061864	1.3849	1.1661	1.2381	1.1334	0.9719	0.8940
	15	0.9328	0.9395	0.0047821	0.0047685	0.0045230	0.0044796	0.0042589	1.1228	1.0645	1.0573	1.0543	0.9904	0.9416
	20	0.9386	0.9430	0.0034570	0.0034587	0.0032938	0.0032793	0.0031601	1.0940	1.0547	1.0496	1.0501	0.9956	0.9594
	30	0.9419	0.9447	0.0022969	0.0022954	0.0022170	0.0022103	0.0021583	1.0642	1.0385	1.0361	1.0354	0.9970	0.9735
	50	0.9441	0.9461	0.0013495	0.0013498	0.0013065	0.0013064	0.0012880	1.0478	1.0332	1.0330	1.0331	0.9999	0.9858
k = 15	**5**	0.6775	0.8303	0.0409151	0.0221388	0.0088253	0.0080231	0.0053688	7.6208	2.7594	4.6361	2.5086	0.9091	0.6083
	8	0.8650	0.9125	0.0090774	0.0066936	0.0050160	0.0048151	0.0041460	2.1894	1.3901	1.8097	1.3344	0.9599	0.8265
	10	0.9031	0.9271	0.0059114	0.0049595	0.0039948	0.0039000	0.0035261	1.6765	1.2717	1.4798	1.2415	0.9763	0.8827
	15	0.9294	0.9394	0.0028288	0.0028112	0.0026268	0.0026119	0.0024566	1.1515	1.0763	1.0769	1.0702	0.9943	0.9352
	20	0.9389	0.9444	0.0020033	0.0020006	0.0019289	0.0019332	0.0018477	1.0843	1.0348	1.0386	1.0372	1.0023	0.9579
	30	0.9456	0.9486	0.0013050	0.0013055	0.0012869	0.0012873	0.0012500	1.0440	1.0141	1.0141	1.0145	1.0003	0.9714
	50	0.9491	0.9510	0.0007744	0.0007740	0.0007619	0.0007624	0.0007493	1.0336	1.0153	1.0164	1.0158	1.0006	0.9833
k = 20	**5**	0.6558	0.8351	0.0357162	0.0157924	0.0059225	0.0054913	0.0035595	10.0340	2.8759	6.0306	2.6665	0.9272	0.6010
	8	0.8514	0.9121	0.0076422	0.0047029	0.0034974	0.0034030	0.0028850	2.6489	1.3820	2.1851	1.3447	0.9730	0.8249
	10	0.9059	0.9326	0.0036417	0.0031723	0.0027604	0.0027330	0.0024472	1.4881	1.1607	1.3193	1.1492	0.9901	0.8865
	15	0.9335	0.9442	0.0019467	0.0019272	0.0017814	0.0017904	0.0016727	1.1638	1.0764	1.0928	1.0819	1.0050	0.9390
	20	0.9374	0.9446	0.0013540	0.0013523	0.0013215	0.0013280	0.0012625	1.0724	1.0183	1.0245	1.0232	1.0049	0.9553
	30	0.9413	0.9445	0.0008975	0.0008975	0.0008707	0.0008741	0.0008459	1.0611	1.0268	1.0309	1.0308	1.0039	0.9715
	50	0.9447	0.9472	0.0005269	0.0005270	0.0005250	0.0005261	0.0005160	1.0211	1.0018	1.0036	1.0039	1.0021	0.9829
k = 35	**5**	0.6036	0.8360	0.0242387	0.0077050	0.0028335	0.0027268	0.0016546	14.6495	2.8256	8.5543	2.7192	0.9623	0.5839
	8	0.8411	0.9163	0.0062553	0.0025824	0.0016724	0.0016587	0.0013653	4.5818	1.5568	3.7403	1.5441	0.9918	0.8163
	10	0.9047	0.9355	0.0019950	0.0014816	0.0013180	0.0013273	0.0011652	1.7121	1.1162	1.5137	1.1241	1.0071	0.8841
	15	0.9285	0.9415	0.0009610	0.0009491	0.0008661	0.0008793	0.0008112	1.1847	1.0794	1.1095	1.0958	1.0153	0.9366
	20	0.9327	0.9430	0.0006909	0.0006894	0.0006430	0.0006504	0.0006125	1.1280	1.0600	1.0745	1.0722	1.0115	0.9526
	30	0.9430	0.9481	0.0004454	0.0004453	0.0004323	0.0004362	0.0004194	1.0618	1.0210	1.0303	1.0303	1.0091	0.9704
	50	0.9468	0.9505	0.0002613	0.0002615	0.0002531	0.0002544	0.0002485	1.0515	1.0279	1.0326	1.0332	1.0052	0.9821

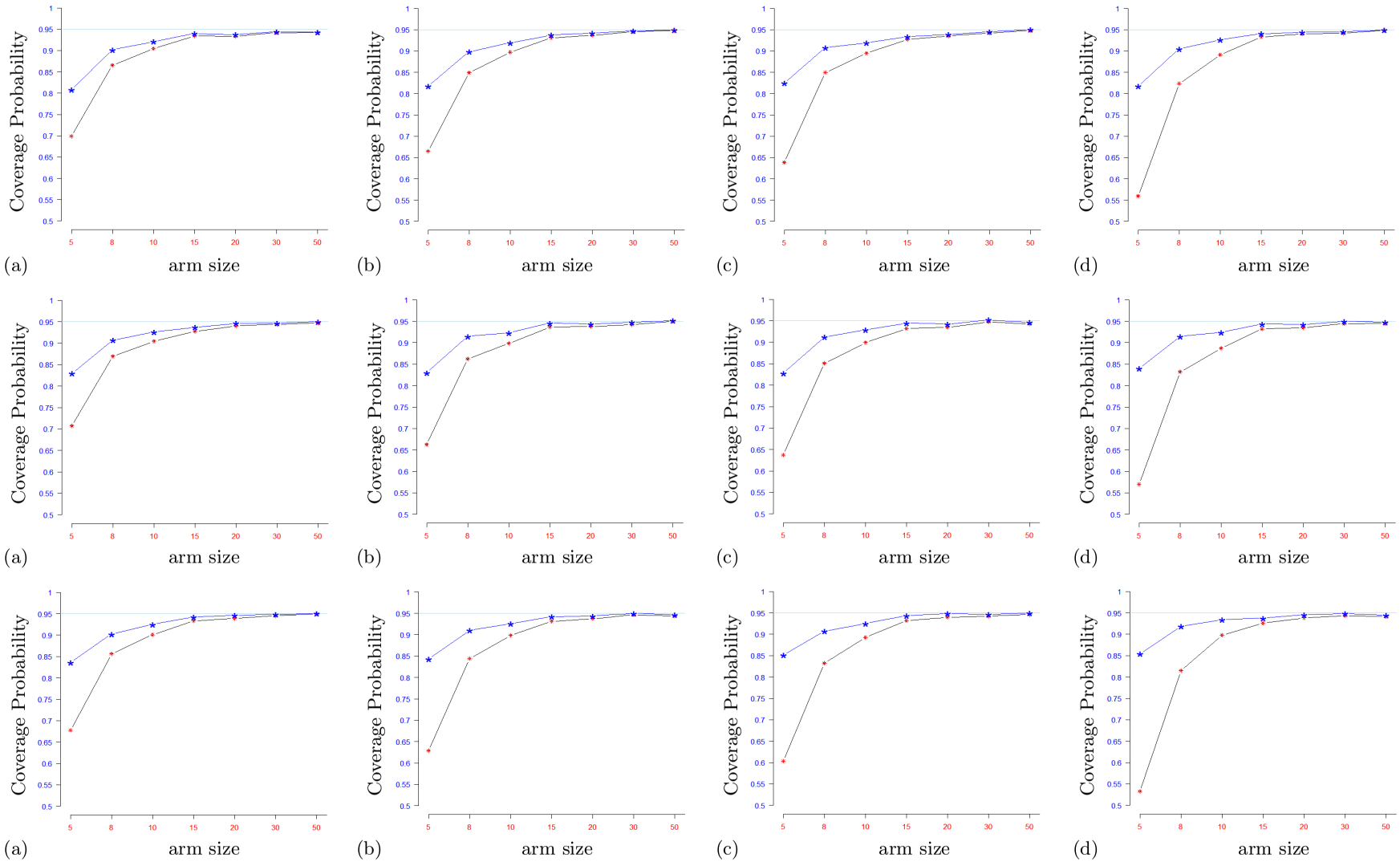
This Table shows the results of simulation for $\theta = 3$ and different values of n, k and S_i^2. This simulation scheme considers for each simulation k parallel group clinical trials each of whom with the same number of patients per arm only *on average* ($n = E[n_i] = n_{Ti} = n_{Ci}$). As regards the within-study variances, S_i^2's are drawn from a Γ distribution with shape parameter $\gamma = 1$. Empirical Statistics for $E[\hat{V}]$, $E[\hat{V}^*]$, V^{opt}, $V^{real}(\hat{\theta})$ and $V^{real}(\hat{\theta}^*)$ are based on 10000 simulation replicates as well as the Empirical Coverage Probability.

Table 3.12: Comparison of the Results of Simulations Considering the 'estimated' and the 'shrunk' variances for different k under the assumption of Unequal S_i^2 and Study sizes - Second Simulation Scheme (C)

	n	Coverage Prob (CP)	Improved CP	$V^{real}(\hat{\theta})$	$V^{real}(\hat{\theta}^*)$	V^{opt}	$\mathbf{E}[\hat{\theta}^*]$	$\mathbf{E}[\hat{\theta}]$	$\frac{V^{real}(\hat{\theta})}{\mathrm{E}[\hat{\theta}]}$	$\frac{V^{real}(\hat{\theta}^*)}{\mathrm{E}[\hat{\theta}^*]}$	$\frac{V^{real}(\hat{\theta})}{V^{opt}}$	$\frac{V^{real}(\hat{\theta}^*)}{V^{opt}}$	$\frac{E[\hat{\theta}^*]}{V^{opt}}$	$\frac{E[\hat{\theta}]}{V^{opt}}$
k = 10	**5**	0.6901	0.8276	0.4561987	0.2683907	0.1399444	0.1172602	0.0849630	5.3694	2.2888	3.2599	1.9178	0.8379	0.6071
	8	0.8585	0.9020	0.1471794	0.1147609	0.0825180	0.0754963	0.0680434	2.1630	1.5201	1.7836	1.3907	0.9149	0.8246
	10	0.9009	0.9237	0.0846646	0.0743347	0.0644753	0.0609399	0.0569307	1.4872	1.2198	1.3131	1.1529	0.9452	0.8830
	15	0.9262	0.9336	0.0469084	0.0462202	0.0425057	0.0414748	0.0397063	1.1814	1.1144	1.1036	1.0874	0.9757	0.9341
	20	0.9319	0.9379	0.0333541	0.0331702	0.0316287	0.0312389	0.0301946	1.1046	1.0618	1.0546	1.0487	0.9877	0.9547
	30	0.9401	0.9432	0.0220455	0.0220064	0.0210269	0.0208337	0.0203818	1.0816	1.0563	1.0484	1.0466	0.9908	0.9693
	50	0.9414	0.9429	0.0128987	0.0129030	0.0125241	0.0124628	0.0123001	1.0487	1.0353	1.0299	1.0303	0.9951	0.9821
k = 15	**5**	0.6501	0.8394	0.3793491	0.1757212	0.0894609	0.0753356	0.0525835	7.2142	2.3325	4.2404	1.9642	0.8421	0.5878
	8	0.8497	0.9090	0.1129277	0.0706677	0.0532958	0.0488919	0.0435367	2.5939	1.4454	2.1189	1.3260	0.9174	0.8169
	10	0.8974	0.9233	0.0592798	0.0487866	0.0422414	0.0400220	0.0371362	1.5963	1.2190	1.4034	1.1549	0.9475	0.8791
	15	0.9297	0.9360	0.0303040	0.0296294	0.0276316	0.0270534	0.0257761	1.1757	1.0952	1.0967	1.0723	0.9791	0.9328
	20	0.9350	0.9416	0.0222863	0.0221456	0.0208020	0.0205885	0.0198107	1.1250	1.0756	1.0714	1.0646	0.9897	0.9523
	30	0.9419	0.9467	0.0142411	0.0142086	0.0138214	0.0137225	0.0133774	1.0646	1.0354	1.0304	1.0280	0.9928	0.9679
	50	0.9428	0.9436	0.0086080	0.0086014	0.0082729	0.0082414	0.0081168	1.0605	1.0437	1.0405	1.0397	0.9962	0.9811
k = 20	**5**	0.6112	0.8391	0.3520310	0.1308555	0.0653238	0.0551617	0.0374189	9.4078	2.3722	5.3890	2.0032	0.8444	0.5728
	8	0.8361	0.9083	0.0927032	0.0507408	0.0393147	0.0361051	0.0319142	2.9048	1.4054	2.3580	1.2906	0.9184	0.8118
	10	0.8933	0.9275	0.0445698	0.0355504	0.0312078	0.0296446	0.0274058	1.6263	1.1992	1.4282	1.1392	0.9499	0.8782
	15	0.9306	0.9394	0.0225143	0.0220425	0.0206696	0.0202365	0.0192208	1.1713	1.0892	1.0892	1.0664	0.9790	0.9299
	20	0.9386	0.9440	0.0162632	0.0161094	0.0153991	0.0152541	0.0146369	1.1111	1.0561	1.0561	1.0461	0.9906	0.9505
	30	0.9436	0.9463	0.0104935	0.0104693	0.0102330	0.0101753	0.0098997	1.0600	1.0289	1.0255	1.0231	0.9944	0.9674
	50	0.9475	0.9486	0.0063069	0.0063001	0.0061575	0.0061414	0.0060413	1.0440	1.0259	1.0243	1.0232	0.9974	0.9811
k = 35	**5**	0.5428	0.8446	0.3044838	0.0731298	0.0361784	0.0308069	0.0197527	15.4148	2.3738	8.4162	2.0214	0.8515	0.5460
	8	0.8166	0.9080	0.0714990	0.0278685	0.0220422	0.0203844	0.0178129	4.0139	1.3672	3.2437	1.2643	0.9248	0.8081
	10	0.8938	0.9305	0.0264298	0.0193275	0.0175457	0.0166986	0.0153523	1.7216	1.1574	1.5063	1.1016	0.9517	0.8750
	15	0.9316	0.9408	0.0125205	0.0122981	0.0116370	0.0114353	0.0108208	1.1571	1.0754	1.0759	1.0568	0.9827	0.9299
	20	0.9357	0.9412	0.0092827	0.0091951	0.0087110	0.0086376	0.0082593	1.1239	1.0645	1.0656	1.0556	0.9916	0.9481
	30	0.9399	0.9436	0.0059134	0.0058993	0.0058103	0.0057854	0.0056140	1.0533	1.0197	1.0177	1.0153	0.9957	0.9662
	50	0.9478	0.9497	0.0035171	0.0035115	0.0034782	0.0034739	0.0034116	1.0309	1.0108	1.0112	1.0096	0.9988	0.9809

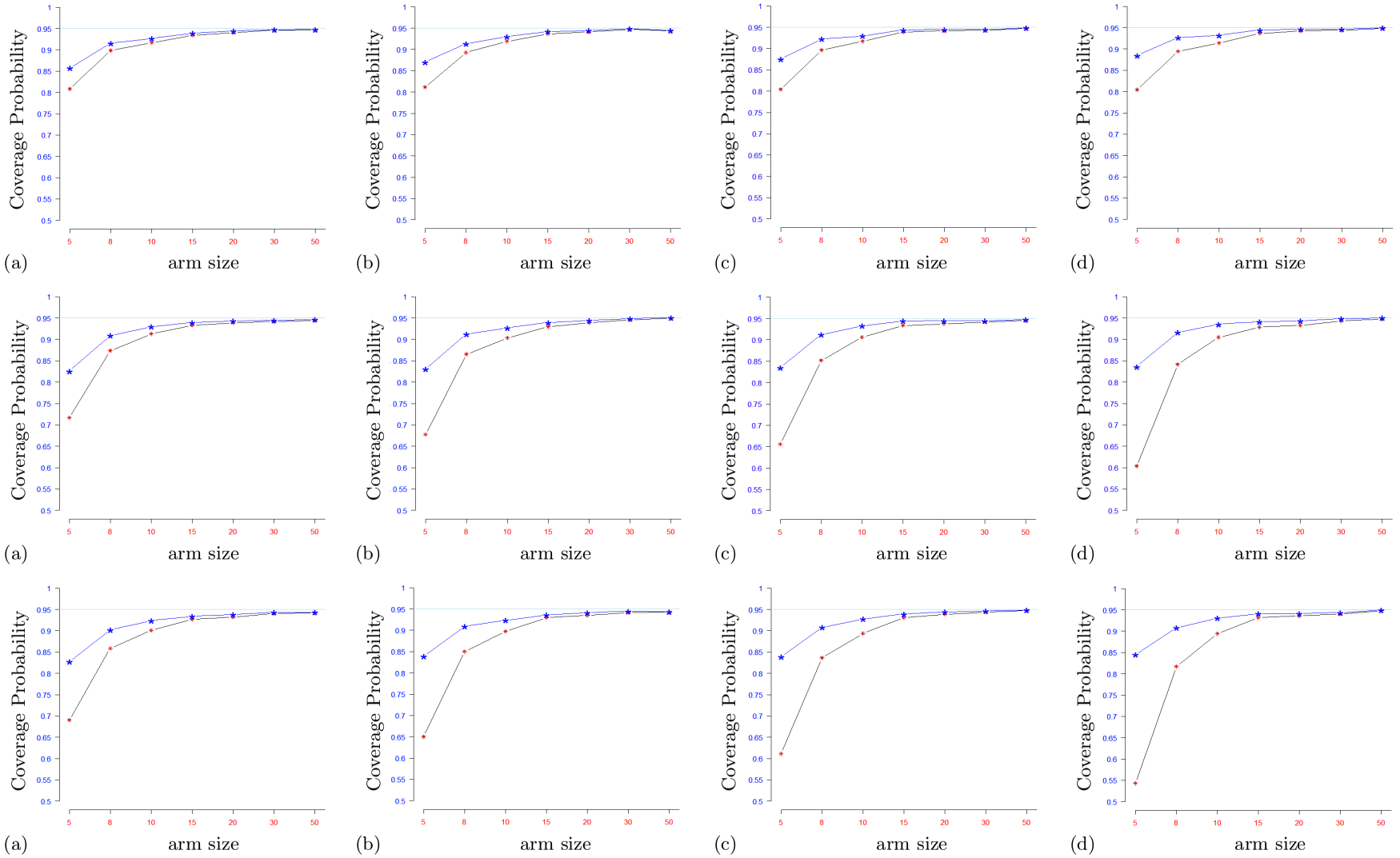
This Table shows the results of simulation for $\theta = 3$ and different values of n, k and S_i^2. This simulation scheme considers for each simulation k parallel group clinical trials each of whom with the same number of patients per arm only *on average* ($n = E[n_i] = n_{Ti} = n_{Ci}$). In this simulation scheme, within-study variances are drawn from a Γ distribution with shape parameter $\gamma = 4$. Empirical Statistics for $E[\hat{V}]$, $E[\hat{V}^*]$, V^{opt}, $V^{real}(\hat{\theta})$ and $V^{real}(\hat{\theta}^*)$ are based on 10000 simulation replicates as well as the Empirical Coverage Probability.

Figure 3.3: 95% Comparison of the Empirical Coverage Probabilities of the CIs based on both $\hat{V}$ and $\hat{V}^*$, i.e. the ordinary and the 'shrunk' estimators – First Simulation Design, with S_i^2's arbitrarily imposed



Each line shows the results for each different variance pattern whereas rows correspond to the number of studies included in the meta analysis [10 (a), 15 (b), 20 (c), 35 (d)]. S_i^2's follow the first variance simulation scheme. In particular, the first line corresponds to the slight imbalance case with half of S_i^2=10 and the other half equal to 100%; the second line corresponds to the most imbalanced case where 20% of S_i^2's are set to 10 and the remaining to 100 while the third line shows the 'common' variance case ($S_i^2 = 100$ for all i). Red dots represent values obtained with the ordinary method whereas blue stars represent values obtained with the new proposed method. Results from 10000 simulations.

Figure 3.4: 95% Comparison of the Empirical Coverage Probabilities of the CIs based on both $\hat{V}$ and $\hat{V}^*$, i.e. the ordinary and the 'shrunk' estimators – Second Simulation Design, with S_i^2's randomly drawn from a Γ distribution



Each line shows the results for each different variance pattern whereas rows correspond to the number of studies included in the meta analysis [10 (a), 15 (b), 20 (c) and 35 (d)]. S_i^2's follow the second simulation scheme. In particular, for the first line of figures, $\gamma = 0.25$, for the second $\gamma = 1$ whereas for the third $\gamma = 4$ (i.e. highly similar values of S_i^2). Red dots represent values obtained with the ordinary method whereas blue stars represent values obtained with the new proposed method. Results from 10000 simulations.

3.8 Real Data Examples

In this section, as examples to illustrate the two methods compared and discussed in the previous sections 4 data sets are taken into consideration. In particular, meta-analyses and data sets given in Rees et al. (2004), Thompson & Pope (2005), Whitehead (2002) and Salpeter et al. (2002) are presented in Figs 3.5 - 3.8 (For full details on trials, heterogeneity tests, response variables and protocols for inclusion of studies, refer to the full articles). These data-sets were considered as reasonable real examples on which to apply the meta-analytical 'shrunk' estimator since the fixed-effects method was originally used to combine these data. Moreover, the variable of interest is in all cases a continuous variable, assumed to be normally distributed, and summarised with a weighted absolute mean difference (as in our simulations). Furthermore, what made these data-sets particularly appealing to our study was the average number of patients per arm as well as the different number of studies combined together. For instance, in the exercise duration studies (Rees et al. (2004)) we can observe a total of 510 randomised participants measured in 15 studies with an average of 17 patients per arm (cf. Fig. 3.5). Fifteen trials involving 22 patients per arm on average (cf. Fig. 3.6) were selected to compare the frequency of Raynaud's Phenomenon (RP) attacks over a 1-week period in those taking calcium channel blockers vs. placebo (Thompson & Pope (2005)). A multicentre study with 9 centres (cf. Fig. 3.7) considered as being from separate studies each of which with an average of 10 patients per arm were included in a fixed-effects meta-analysis comparing two anaesthetic agents in patients undergoing short surgical procedures (Whitehead (2002)). Finally, twenty-five studies each of which including an average of 13 patients per arm were included to compare single-dose of cardioselective β-blockers with placebo (cf. Fig. 3.8). The latter data-set, however, could not be used for our purpose of illustrating and comparing results from the two methods; in fact, an error in the printed table was present and, despite the access to the original article (Chatterjee, 1986) the reproducibility of the same output as in Salpeter et al. (2002) was not possible. Therefore, analysis was performed (using again the statistical package **R**) only on the remaining data-sets.

Recall that the parameter of interest is the overall effect, denoted by θ. The fixed-effects model is assumed to be the correct one for our analysis, i.e. $\theta_i = \theta$ for $i = 1,2, \ldots k$. This implies that the estimated effect size Y_i is normally distributed with mean θ and variance σ_i^2. The estimator of θ is generally a simple weighted average of the Y_i, with the optimal weights proportional to $w_i = 1/var(Y_i)$. In practice, the variances are not known so estimated variances $\hat{\sigma}_i^2$ are used to estimate both θ and V=var(θ). Hence we define $\hat{w}_i = 1/\hat{\sigma}_i^2$ giving

$$\hat{\theta} = \frac{\sum \hat{w}_i Y_i}{\sum \hat{w}_i} \quad \text{and} \quad \hat{V} = var(\theta) = \frac{1}{\sum \hat{w}_i}$$

with

$$var(Y_i) = \hat{\sigma}_i^2 = \left(\frac{1}{n_{Ti}} + \frac{1}{n_{Ci}}\right)\hat{S}_i^2$$

where, generally, $\hat{S}_i^2$ is equal to the within-study pooled variance calculated by

$$\hat{S}_i^2 = \frac{(n_{Ti}-1)\hat{s}_{Ti}^2 + (n_{Ci}-1)\hat{s}_{Ci}^2}{n_{Ti}+n_{Ci}-2}$$

This is the ordinary method usually used in fixed-effects meta-analysis. However, results will be given also when within-study variances are estimated via the modified shrinkage CHQBC estimator (see equation 3.2). Both the formulae to compute the point estimate θ and its overall variance V are the same but $S_{shr.i}^2$ instead of S_i^2 is used to calculate $var(Y_i)$, where $S_{shr.i}^2$s are obtained with the shrinkage estimator (see eqn 3.2) applied on S_i^2s (as thoroughly described in section 3.6), i.e.

$$\hat{S}_{shr.i}^2 = (BZ_{pool})^{\alpha}(Z_g)^{(1-\alpha)}, \qquad 0 \leq \alpha \leq 1.$$

This estimator borrows information across studies by shrinking each specific variance towards the bias corrected geometric mean of variances for all studies, where the amount of shrinkage depends on the variability of the individual variances.
Moreover, these results will be given even under the assumption of a common within-treatment group variance across all studies. This entails that not every study has its own variance term and the common group variance across studies is estimated by S_p^2 where

$$\hat{S}_p^2 = \frac{\sum_{i=1}^{k}(n_{Ti}+n_{Ci}-2)\hat{S}_i^2}{\sum_{i=1}^{k}(n_{Ti}+n_{Ci}-2)}$$

Usually the decision to assume a common variance can be based upon Bartlett's test, even though strict adherence to a specific level for this test is not advisable and this test is extremely sensitive to non-normality of data (Whitehead, 2002).

Therefore, 3 different methods to calculate $var(Y_i) = \hat{S}^2\,(1/n_T + 1/n_C)$ are considered; $\hat{S}^2$ can be replaced by the overall pooled variance $\hat{S}_p^2$, by the usual within-study pooled variances $\hat{S}_i^2$ or by the shrunk variances $\hat{S}_{shr.i}^2$ leading to 3 different estimates of the variance of the mean difference, i.e. $\hat{\sigma}_{pi}^2$, $\hat{\sigma}_i^2$ and $\hat{\sigma}_{si}^2$ respectively. In general, we would expect the change in the weights due to different ways to calculate the variances to lead to a change in the overall fixed-effect estimate of treatment difference and in particular a change in the estimate of the variance of the overall effect. In particular, we expect V to be higher and more reliable when shrunk estimates are used.

Study, Year	Treatment n_T	Treatment Mean Change ± SD	Control n_C	Control Mean Change ± SD	Weight %	**WMD** (95% CI *FIXED*)
Belardelli, 1992	10	-2.30 ± 1.20	10	-0.10 ± 1.24	19.0	-2.20 (-3.27; -1.13)
Coats, 1990	10	-2.00 ± 3.65	10	0.60 ± 3.35	2.3	-2.60 (-5.67; 0.47)
Coats, 1992	17	-2.65 ± 4.80	17	-0.05 ± 4.80	2.1	-2.60 (-5.83; 0.63)
Dubach et al. studies	12	-3.52 ± 1.85	13	-1.60 ± 5.50	2.0	-1.92 (-5.09; 1.25)
Gottlieb, 1999	11	-4.90 ± 3.40	14	0.20 ± 3.40	3.0	-5.10 (-7.78; -2.42)
Hambrecht, 1995	9	-2.73 ± 2.82	9	0.28 ± 0.42	6.3	-3.01 (-4.87; -1.15)
Hambrecht, 2000	31	-3.95 ± 4.07	33	0.17 ± 3.67	6.0	-4.12 (-6.02; -2.22)
Keteyian, 1996	15	-2.80 ± 2.30	14	-0.50 ± 1.87	9.3	-2.30 (-3.82; -0.78)
Kiilavuori, 1996	12	-12.30 ± 8.40	15	-2.80 ± 9.30	0.5	-9.50 (-16.19; -2.81)
Maiorana, 2000	13	-2.50 ± 3.50	13	0.30 ± 3.20	3.3	-2.80 (-5.38; -0.22)
Oka, 2000	18	-0.60 ± 3.35	18	0.30 ± 3.30	4.6	-0.90 (-3.07; 1.27)
Ponikowski, 1997	16	-1.20 ± 2.20	13	-0.93 ± 2.95	6.2	-0.27 (-2.20; 1.66)
Quittan, 1999	11	-2.90 ± 2.20	12	-0.58 ± 1.95	7.6	-2.32 (-4.03; 0.61)
Wielenga, 1998	35	-2.16 ± 3.66	32	-0.08 ± 3.05	8.3	-2.08 (-3.69; -0.47)
Wielenga, 1999 - Change	35	-2.43 ± 2.22	32	-0.09 ± 2.18	19.5	-2.34 (-3.39; -1.29)
					100	**-2.38** [-2.85 ; -1.92]

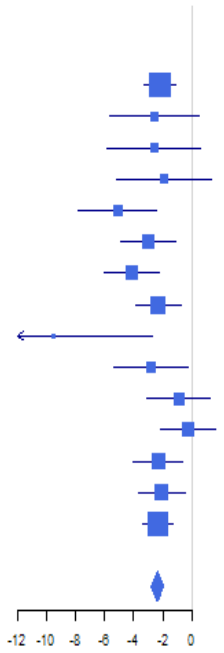


Figure 3.5: **Exercise based rehabilitation for heart failure. All exercise interventions *vs* Usual Care.** Forest Plot: Mean Difference Estimates with 95% Intervals for the data of the 15 trials - The area of squares is proportional to the study precision. Using a fixed-effects model, the meta-analysis showed that significant improvements were seen for exercise duration which increased by 2.38 minutes (95% CI 1.92, 2.85) in the training group compared to the 'usual-care group'. For all variables concerned with exercise capacity, an increase in the value from baseline indicates improvement with the exercise intervention; this is the reason why the sign of the mean change for both group has been changed for the pooled analysis so that the direction of effect is the appropriate direction on the plot.

	Treatment		Control			
Study, Year	n_T	Mean ± SD, %	n_C	Mean ± SD, %	Weight (%)	**WMD** (95% CI *FIXED*)
Rodeheffer,1983	5	1.3 ± 1.4	5	5 ± 2.62	17.8	-3.7 (-6.30;-1.10)
Ettinger,1984	3	15 ± 9.54	3	15.83 ± 7.1	0.7	-0.83 (-14.29;12.63)
Kahan,1985	5	4.2 ± 5.67	5	18.6 ± 9.61	1.3	-14.4 (-24.18;-4.62)
Kahan -DILT, 1985	6	4.25 ± 3.9	6	8.15 ± 1.1	11.5	-3.9 (-7.14;-0.66)
Aldoori, 1986	9	5 ± 9.3	9	11.3 ± 9.9	1.5	-6.3 (-15.17;2.57)
Corbin, 1986	22	4.49 ± 5.51	22	8.67 ± 8.83	6.4	-4.18 (-8.53;0.17)
Gjorup, 1986a	21	6.97 ± 7.38	21	14.56 ± 8.51	5.2	-7.59 (-12.41;-2.77)
Gjorup, 1986b	19	7.77 ± 7.7	19	10.57 ± 7.91	4.9	-2.8 (-7.76;2.16)
Sarkozi, 1986	15	7.82 ± 17.82	17	10.63 ± 12.37	1.1	-2.81 (-13.58;7.96)
Waller, 1986	29	7.5 ± 8.31	29	10 ± 9.35	5.8	-2.5 (-7.05;2.05)
Challenor, 1987	36	10.29 ± 11.27	36	10.5 ± 9.17	5.3	-0.21 (-4.96;4.54)
Kahan, 1987	3	9.17 ± 9.93	3	15.17 ± 8.61	0.5	-6 (-20.87;8.87)
Challenor, 1989	22	4 ± 8.91	22	6.4 ± 7.97	4.8	-2.4 (-7.40;2.60)
French Gp, 1991	69	4.93 ± 6.2	69	5.8 ± 5.7	30.5	-0.87 (-2.86;1.12)
RTSI, 2000	68	0.84 ± 19.04	62	1.47 ± 19.04	2.8	-0.63 (-7.18;5.92)
					100	**-2.80** [-3.90 ; -1.70]

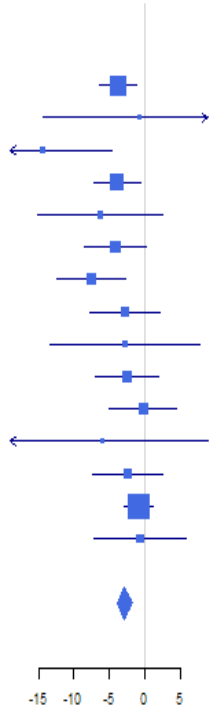


Figure 3.6: **Calcium Channel Blocker (CCB) *vs* Placebo. Frequency of Raynaud's Phenomenon (RP) attacks over a 1-week period.** Forest Plot: Mean Difference Estimates with 95% Intervals for the data of the 15 trials - The area of squares is proportional to the study precision. Using a fixed-effects model, the meta-analysis showed that CCB provided a significance reduction, compared to the placebo, in the frequency of ischaemic attacks over a 1-week period, with a weighted mean difference (WMD) of -2.8 (-3.9, -1.7) (P=0.0007), which means a reduction of about 3 attacks in a 1-week period.

	Treatment A		Treatment B			
Centre	n_{T_A}	Mean± SD	n_{T_B}	Mean ± SD	Weight %	**WMD** (95% CI *FIXED*)
1	4	1.141 ± 0.967	5	0.277 ± 0.62	3.5	0.864 (-0.172; 1.900)
2	10	2.165 ± 0.269	10	1.519 ± 0.913	10.8	0.646 (0.056; 1.235)
3	17	1.79 ± 0.795	17	1.518 ± 0.849	12.2	0.272 (-0.281; 0.825)
4	8	2.105 ± 0.387	9	1.189 ± 1.061	6.2	0.916 (0.136; 1.696)
5	7	1.324 ± 0.47	10	0.456 ± 0.619	12.6	0.868 (0.322; 1.412)
6	11	2.369 ± 0.401	10	1.55 ± 0.558	22	0.819 (0.407; 1.232)
7	10	1.074 ± 0.67	12	0.265 ± 0.502	15.6	0.809 (0.319; 1.299)
8	5	2.583 ± 0.409	4	1.37 ± 0.934	4.6	1.213 (0.312; 2.113)
9	14	1.844 ± 0.848	19	2.118 ± 0.749	12.5	-0.274 (-0.820; 0.274)
					100	**0.627** [0.433 ; 0.820]

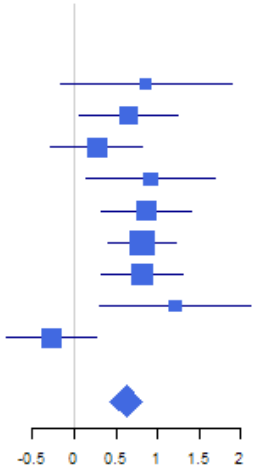


Figure 3.7: **Comparison in log-Recovery Time of two anaesthetic agents (A *vs* B).** Forest Plot: Mean Difference Estimates (treatment A - treatment B) with 95% Intervals for the data of the 9 centres; the area of squares is proportional to the study precision. Using a fixed-effects model, the meta-analysis showed that the log-recovery time is longer on anaesthetic A than on anaesthetic B ($\hat{\theta} = 0.627$, with 95% CI 0.433 ; 0.820).

		β-Blocker		Control			
	Study, Year	n_T	Mean Change in FEV_1 from Baseline ± SD, %	n_C	Mean Change in FEV_1 from Baseline ± SD, %	Weight (%)	Net Change in FEV_1 (95% CI *Fixed*)
Group 1: Cardioselective *without* intrinsic sympathomimetic activity (ISA) vs placebo	Adam et al., 1982	10	-7.1 ± 9.3	10	-5.3 ± 6.8	6.8	-1.80(-8.94 ; 5.34)
	Benson et al., 1978	12	-5.1 ± 11.3	12	1.9 ± 8.3	5.5	-7.00(-14.93 ; 0.93)
	Chatterjee, 1986	12	-5.1 ± 13.5	12	1.1 ± 10.5	3.7	-6.90 (-16.58; 2.78)
	Chodosh et al., 1988	16	-17.6 ± 16.8	16	-4.4 ± 13	3.2	-13.20(-23.61 ; -2.79)
	Doshan et al., 1986a	15	-11.8 ± 17.1	15	3.7 ± 14.2	2.7	-15.50(-26.75 ; -4.25)
	Doshan et al., 1986b	34	-8.4 ± 15	34	5.5 ± 12.3	8.2	-13.90 (-20.42 ; -7.38)
	Ellis et al., 1981	14	-13 ± 13.5	14	-2.8 ± 10.5	4.3	-10.20 (-19.16 ; -1.24)
	Falliers et al., 1986	18	-8.8 ± 15	18	-3 ± 15.5	3.5	-5.80 (-15.76 ; 4.16)
	Greefhorst et al., 1984	8	-8.9 ± 15.5	8	-0.8 ± 11.5	1.9	-8.10 (-21.47 ; 5.27)
	Johnsson et al., 1975	7	-9.2 ± 11.2	7	-3 ± 9.8	2.9	-6.20 (-17.22 ; 4.82)
	Lammers et al, 1984	8	-10.5 ± 10.7	8	-5.1 ± 8.3	3.9	-5.40(-14.78 ; 3.98)
	Lammers et al, 1986	11	-12.3 ± 14.2	11	-0.3 ± 11.2	3.0	-12.00 (-22.69 ; -1.31)
	Lammers et al, 1988	11	-12.3 ± 13.5	11	-0.3 ± 10.5	3.4	-12.00 (-22.11 ; -1.89)
	Lawrence et al., 1982	14	-9.2 ± 11.8	14	3.6 ± 9.1	5.7	-12.80 (-20.61 ; -4.99)
	Lofdahl et al., 1981	8	-10 ± 17	8	2.6 ± 7.8	2.1	-12.60 (-25.56 ; 0.36)
	Ruffin et al., 1979	12	-3.7 ± 13.5	12	0 ± 10.5	3.7	-3.70 (-13.38 ; 5.98)
	Skinner et al., 1975	10	-8.3 ± 12	10	-2 ± 7.5	4.5	-6.30(-15.70 ; 2.47)
	Tantucci et al., 1990	12	-13.5 ± 15.8	12	-1.1 ± 11.7	2.8	-12.40 (-23.52 ; -1.28)
	van den Bergh et al., 1981	8	-7.4 ± 15	8	7 ± 12	2.0	-14.40 (-27.71 ; -1.09)
Group 2: *with* ISA vs placebo	Benson et al., 1978	12	-5.9 ± 11.3	12	1.9 ± 8.3	5.5	-7.80(-15.73; 0.13)
	Doshan et al., 1986a	15	10.4 ± 17.1	15	3.7 ± 14.2	2.7	6.70 (-4.55 ; 17.95)
	Doshan et al., 1986b	34	9.3 ± 15	34	5.5 ± 12.3	8.2	3.80(-2.72 ; 10.32)
	Greefhorst et al., 1984	8	-8.6 ± 15.5	8	-0.8 ± 11.5	1.9	-7.80 (-21.17 ; 5.57)
	Lammers et al., 1986	11	-8.8 ± 14.2	11	-0.3 ± 11.2	3.0	-8.50 (-19.19 ; 2.19)
	Skinner et al., 1975	10	-9.6 ± 12	10	-2 ± 7.5	4.5	-7.60 (-16.37 ; 1.17)
						100	**-7.46** [-9.32 ; -5.59]

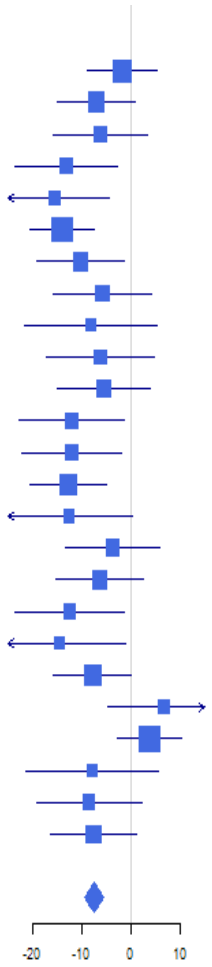


Figure 3.8: **Effects of Treatment on FEV_1 for single-dose studies.** Single doses of cardioselective β-blockers (divided in two groups; with or without intrinsic sympathomimetic activity) *vs* Placebo. Forest Plot: Mean Difference Estimates with 95% Intervals for the data of the 25 trials - The area of squares is proportional to the study precision. Fixed-Effects (test for heterogeneity, P = 0.09) meta-analysis on 25 studies; Compared with placebo, single doses of cardioselective β-blockers *as a group* were associated with a 7.43% (CI, 5.57% ,9.29%) ($P < 0.0001$) reduction in FEV_1.

Randomised controlled trials were used in the Cochrane meta-analysis (Table 3.13) to determine the effectiveness of exercise based interventions. In the meta-analysis performed with the pooled variances, significant improvements were seen for exercise duration which increased by 2.38 minutes. Roughly the same point estimate is obtained when 'shrunk' variances are taken into consideration;furthermore, as expected, the variance of the overall estimator increases with the new proposed method. As illustrated through simulations we expect 'shrunk' variances to produce an overall variance estimate whose 'real' and 'average' dispersion are both closer to the optimal value; therefore, the bigger standard error of $\hat{\theta}^*$ should be closer to and less negatively biased than the true value. Similarly, the 95% Confidence Interval [-1.912 ; -2.864] is supposed to have coverage properties more likely to be correct.

Looking at table 3.14, 15 studies were used in a fixed-effects meta-analysis to detect whether calcium channel blockers would significantly reduce, compared to the placebo, the frequency of ischemic attacks in a 1-week period of time. Again, both the point estimate and the overall variance differ slightly depending on the method we are applying to the data-set. In fact, whilst the reduction between CCB and placebo has a weighted difference mean of -2.802 if the pooled variances are considered, the reduction is equal to -2.759 when 'shrunk' variances are taken into account. Moreover, the overall variance estimate increases from 0.313 to 0.327 with the new method: this means that the confidence intervals for the new estimate obtained with 'shrunk variances' are wider (95% CI [-3.88 ; -1.64]). Again, we believe that the CIs obtained considering 'shrunk' variances are more likely to have correct coverage properties and that the declared precision of the overall estimate should underestimate less remarkably the true value. In this real-data example, it is interesting to note that Rodeheffer and Kahan-DILT studies (i.e. 1 and 4) decrease their relative weights as expected. The original control and treatment sample variances are extremely small compared to the other studies. Not necessarily one has to expect these values similar to the other studies with the same number of subjects but, at the same time, it is reasonable to suspect that these values might be unreliable. Therefore, with the shrunk variances, the relative weight of these 2 studies decreases, accounting this way for the possible imprecision of the measures due to the small number of subjects; in particular, Rodeheffer's weight drops to 15.3% from the initial 17.8% while Kahan-DILT's weight goes from 11.5% down to 10.7%.

Table 3.15 gives the estimates from the the log-recovery time meta-analysis, presenting details for different methods to calculate the within-study variances. Interestingly, we can observe that all centres have approximately the same weights regardless of the method used. Nevertheless, centres 1, 3, 4 and 9 have, with the new shrunk estimates, a

slightly higher weight in the fixed-effects meta-analysis. The change in the weights due to the use of shrunk variances has led to a decrease in the overall fixed-effects estimate of treatment difference. In particular, the new overall point estimate (i.e. difference in mean log-recovery time after anaesthesia between treatment A and treatment B) is equal to 0.611, showing that the recovery time (minutes from when the anaesthetic gases are turned off until the patient opens their eyes) is longer on anaesthetic A than on B. Not only has the point estimate changed but also has the overall variance V; in fact, the standard error of $\hat{\theta}^*$ is equal to 0.101 and this should, according to the simulations done, better represent the real variability of the estimator. This gives even wider confidence intervals which are supposed to have a coverage probability closer to the nominal value of 95%.

Table 3.13: Fixed-Effects Meta-Analysis of the absolute mean difference (All exercise interventions *vs* Usual Care) for exercise based rehabilitation for heart failure. The first column represents the total number of patients in each study; the second represents the absolute mean difference for each study. The other columns represent the different values of $\hat{S}^2$ – single pooled variances ($\hat{S}_i^2$), shrunk variances ($\hat{S}_{shr.i}^2$) and a common pooled variance parameter (S_p^2) respectively – based on which the relative $SE(Y_i)$ and weights are calculated. The last two rows show the *Estimated overall effects*, the corresponding *SE* as well as the *Confidence Intervals* for the three different methods

	N	$\hat{\theta}_i$	$\hat{S}_i^2$	$\hat{\sigma}_i$	Weight	$\hat{S}_{shr.i}^2$	$\hat{\sigma}_{si}$	Weight	$\hat{\sigma}_{pi}$	Weight
Study, Year					(%)			(%)		(%)
Belardelli, 1992	20	-2.2	1.489	0.546	19.0	1.738	0.590	17.0	1.674	3.9
Coats, 1990	20	-2.6	12.273	1.567	2.3	12.058	1.553	2.4	1.674	3.9
Coats, 1992	34	-2.6	23.040	1.646	2.1	22.156	1.615	2.3	1.284	6.7
Dubach et al.,	25	-1.92	17.419	1.671	2.0	16.791	1.640	2.2	1.498	4.9
Gottlieb, 1999	25	-5.1	11.560	1.370	3.0	11.433	1.362	3.2	1.508	4.8
Hambrecht, 1995	18	-3.01	4.064	0.950	6.3	4.415	0.990	6.0	1.764	3.5
Hambrecht, 2000	64	-4.12	14.967	0.968	6.0	14.809	0.963	6.4	0.936	12.6
Keteyian, 1996	29	-2.3	4.427	0.782	9.3	4.615	0.798	9.3	1.391	5.7
Kiilavuori, 1996	27	-9.5	79.481	3.453	0.5	70.385	3.249	0.6	1.449	5.2
Maiorana, 2000	26	-2.8	11.245	1.315	3.3	11.145	1.309	3.4	1.468	5.1
Oka, 2000	36	-0.9	11.056	1.108	4.6	10.991	1.105	4.8	1.247	7.1
Ponikowski, 1997	29	-0.27	6.557	0.956	6.2	6.694	0.966	6.3	1.397	5.6
Quittan, 1999	23	-2.32	4.297	0.865	7.6	4.544	0.890	7.4	1.562	4.5
Wielenga, 1998	67	-2.08	11.444	0.827	8.3	11.394	0.826	8.7	0.915	13.1
Wielenga, 1999 - Change	67	-2.34	4.844	0.538	19.5	4.913	0.542	20.1	0.915	13.1

	$\hat{\theta}$ $[se(\hat{\theta})]$	**-2.378** [**0.238**]	**-2.388** [**0.243**]	**-2.866** [**0.332**]
	95% CI	(-2.845 ; -1.911)	(-2.864 ; -1.912)	(-3.517 ; -2.215)

Table 3.14: Fixed-Effects Meta-Analysis of the absolute mean difference (Calcium Channel Blocker *vs* Placebo) in frequency of RP attacks over a 1-week period: Comparisons of methods. The first column represents the total number of subjects in each study while the second shows the absolute mean difference for each study. The other columns represent the different values of $\hat{S}^2$ – single pooled variances ($\hat{S}_i^2$), shrunk variances ($\hat{S}_{shr.i}^2$) and a common pooled variance parameter (S_p^2) respectively – based on which the relative $SE(Y_i)$ and weights are calculated. The last two rows show the *Estimated overall effects*, the corresponding *SE* as well as the *Confidence Intervals* for the three different methods

	N	$\hat{\theta}_i$	$\hat{S}_i^2$	$\hat{\sigma}_i$	Weight	$\hat{S}_{shr.i}^2$	$\hat{\sigma}_{si}$	Weight	$\hat{\sigma}_{pi}$	Weight
Study, Year					(%)			(%)		(%)
Rodeheffer,1983	10	-3.70	4.412	1.328	17.8	5.335	1.461	15.3	7.243	1.5
Ettinger,1984	6	-0.83	70.711	6.866	0.7	71.876	6.922	0.7	9.350	0.9
Kahan,1985	10	-14.40	62.251	4.990	1.3	62.654	5.006	1.3	7.243	1.5
Kahan -DILT, 1985	12	-3.90	8.210	1.654	11.5	9.202	1.751	10.7	6.612	1.8
Aldoori, 1986	18	-6.30	92.250	4.528	1.5	91.166	4.501	1.6	5.398	2.7
Corbin, 1986	44	-4.18	54.165	2.219	6.4	54.243	2.221	6.6	3.453	6.7
Gjorup, 1986a	42	-7.59	63.442	2.458	5.2	63.420	2.458	5.4	3.534	6.4
Gjorup, 1986b	38	-2.80	60.929	2.533	4.9	60.940	2.533	5.1	3.715	5.8
Sarkozi, 1986	32	-2.81	229.800	5.370	1.1	224.802	5.311	1.2	4.057	4.8
Waller, 1986	58	-2.50	78.239	2.323	5.8	78.057	2.320	6.1	3.007	8.8
Challenor, 1987	72	-0.21	105.551	2.422	5.3	105.133	2.417	5.6	2.699	10.9
Kahan, 1987	6	-6.00	86.369	7.588	0.5	85.097	7.532	0.6	9.350	0.9
Challenor, 1989	44	-2.40	71.455	2.549	4.8	71.322	2.546	5.0	3.453	6.7
French Gp, 1991	138	-0.87	35.465	1.014	30.5	35.532	1.015	31.8	1.950	20.9
RTSI, 2000	130	-0.63	362.522	3.343	2.8	360.028	3.332	2.9	2.011	19.7

	$\hat{\theta}$ $[se(\hat{\theta})]$	**-2.802** [**0.560**]	**-2.759** [**0.572**]	**-2.347** [**0.892**]
	95% CI	(-3.900 ; -1.704)	(-3.880 ; -1.638)	(-4.095 ; -0.599)

Table 3.15: Fixed-Effects Meta-Analysis of the absolute mean difference of two anaesthetic agents (A *vs* B) for patients undergoing short surgical procedures. The first column represents the total number of patients in each centre whereas the second shows the absolute mean difference for each study. The other columns represent the different values of $\hat{S}^2$ – single pooled variances ($\hat{S}_i^2$), shrunk variances ($\hat{S}_{shr.i}^2$) and a common pooled variance parameter (S_p^2) respectively – based on which the relative $SE(Y_i)$ and weights are calculated. The last two rows show the *Estimated overall effects*, the corresponding *SE* as well as the *Confidence Intervals* for the three different methods

	N	$\hat{\theta}_i$	$\hat{S}_i^2$	$\hat{\sigma}_i$	Weight	$\hat{S}_{shr.i}^2$	$\hat{\sigma}_{si}$	Weight	$\hat{\sigma}_{pi}$	Weight
Centre					(%)			(%)		(%)
1	9	0.864	0.620	0.528	3.49	0.572	0.507	3.94	0.477	4.93
2	20	0.646	0.453	0.301	10.76	0.460	0.303	11.02	0.318	11.10
3	34	0.272	0.676	0.282	12.25	0.651	0.277	13.23	0.244	18.87
4	17	0.916	0.670	0.398	6.16	0.623	0.384	6.89	0.346	9.40
5	17	0.868	0.318	0.278	12.61	0.353	0.293	11.82	0.350	9.14
6	21	0.819	0.232	0.211	21.99	0.266	0.225	19.94	0.311	11.63
7	22	0.809	0.341	0.250	15.61	0.362	0.258	15.26	0.304	12.11
8	9	1.213	0.469	0.460	4.61	0.504	0.476	4.47	0.477	4.93
9	33	-0.274	0.627	0.279	12.52	0.608	0.275	13.43	0.250	17.89

	$\hat{\theta}$ $[se(\hat{\theta})]$	**0.627** [**0.099**]	**0.611** [**0.101**]	**0.535** [**0.106**]
	95% CI	(0.433 ; 0.820)	(0.414 ; 0.808)	(0.327 ; 0.743)

Chapter 4

Conclusions and Discussion

Meta-analysis is the statistical summary of a collection of analytic results from individual studies for the purpose of integrating the findings. Data-analysis is only the last step of a long and complicated research synthesis procedure; the outcome of a meta-analysis may therefore be a long awaited process. Furthermore, conclusions obtained through such a quantitative research synthesis can have an important impact and substantial consequences in public and health planning policies. Clearly, "an estimate of the overall effect size should be accompanied by a confidence interval to indicate the precision with which the overall effect size has been estimated" (Viechtbauer, 2007, pg. 50). As a consequence, reliability of the output and in particular of the measure of precision of the point estimate is crucial; a correct value of the standard error of the point estimate ensures that the resulting significance of the analysis is correctly stated and that confidence intervals have correct coverage probabilities. On the contrary, stating an incorrect precision can often result in biased and misleading results.

In this thesis, reliability of the overall variance of the point estimate was investigated in *fixed-effects* meta-analyses since, in this case, the weights, the overall point estimate θ, as well as its variance V depend entirely and solely upon the within-study variances, usually assumed to be known. Nonetheless, this assumption is approximately true only when sample sizes are large enough. Imprecision of the within-study variances should not be simply ignored: in fact, when sampling errors are not taken into account, the usual variance estimator performs very poorly in detecting the true variance of θ and underestimates the true value. Additionally, the actual variability of the variance estimator is always higher than both the optimal and declared values, with a consequent overstatement of the precision of the estimator and misleading results in the form of Confidence Intervals without correct coverage properties, in particular with small sample size studies.

The aim of this thesis was not only to illustrate via simulations what circumstances

worsen the estimate of the variance of the overall estimator (variance structure across studies at the patient level, number of studies, allocation per arm, study size) but also, and more importantly, to investigate whether a different method, which can be accurate and flexible at the same time, existed. In particular, an estimator whose variance does not diverge substantially from the optimal value, both on average and in practice, was sought and found to provide both more accurate statements about the precision of the point estimate and confidence intervals which are more likely to have the correct nominal coverage probabilities.

The overall average number of patients per study plays an important role which appears to be more important than the total number of patients in each single study. Moreover, the allocation of patients per arm does not seem to be decisive for the estimated overall variance of the estimator even though balanced allocation, as well as having roughly the same amount of patients per study, yields better results. Furthermore, true to form, the higher the average number of patients per arm, the closer the variance estimator is to the optimal one. However, when small studies are combined, the σ_i^2's are less precise and this leads to severely unreliable results. The ordinary method is too sensitive to individual study variances and is negatively biased when sizes are too small.

In order to overcome this problem, we decided to shrink the individual pooled estimates towards a common value before calculating the variances. The shrinkage estimators for the pooled variances prove to be particularly advantageous compared with conventional approaches when a single study carries a disproportionate amount of weight and has few subjects, which may indicate a very low sample variance just by chance. Borrowing information across variances through the "modified CHQBC estimator" produced an overall variance estimate whose 'real' and 'average' dispersion were both *closer to the optimal value*, representing a reasonable alternative to the ordinary method. Results obtained via simulations (with different patterns for various variance schemes and diverse average numbers of patients per study), emphasised that the estimator of the overall variance based on the 'shrunk variances' ($\hat{V}^*$), performed better than the one based on the estimated sample variances ($\hat{V}$), minimising the real dispersion of the overall variance estimator. Moreover, regardless of the variance structure across studies, $\hat{V}^*$ (calculated with the new proposed weighting scheme) performs optimally even with a small average number of patients per trial, achieving almost optimal results without relying on computationally expensive procedures.

As a consequence, since the overall variance of the point estimator was better estimated with the proposed technique, the coverage probabilities of the approximate 95% confidence intervals based on $\hat{V}^*$ were found to be generally accurate, in that they were

approximately equal to the nominal coverage value. The proposed weighting scheme yields confidence intervals with higher coverage probability than the commonly used interval based on pooled variances, particularly when the number of studies is moderate and the average number of patients per arm is small. In light of these results, $\hat{V}^*$ is strongly recommended, since this appears to perform better than (or at least as well as) the usual variance estimator across the range of cases, and this provides additional protection against large errors in the estimated sample variances and hence in imprecisely estimated weights.

In this thesis we have shown the consequences of using the estimated weights in the calculation of the overall variance of the common effect estimator in combining estimates from independent studies. We have pointed out that $(\hat{V})$ is not a good estimator for the correct variance of the overall effect estimate when the weights are merely based on sample variances, as in practice. Protection against errors in the estimated weights should therefore be provided. We recommend the use of the proposed weighting scheme to achieve more reliable estimates of the overall variance and better approximations of the nominal significance level, which has the added advantage of simplicity. Due to its easy application and its good performance in the simulation study, the proposed shrinkage estimator for the pooled variances is a good alternative to the ordinary method. The use of shrunk variances for the variance of an overall effect estimate is advocated in making inference for the fixed-effects meta-analysis, particularly when the studies considered have, on average, a small number of subjects because of its more accurate estimate of the overall variance.

Finally, it should be noted that this thesis has only considered the weighted mean difference for continuous data. Many other outcome variables are possible when dealing with the comparison of two treatments, control and experimental, in an effort to find out whether there is a significant difference between the two. Standardised mean difference for continuous variables as well as odds ratio, difference or ratio of proportions in the case of qualitative attributes play an important role to detect the effect size and measure such a difference. Nevertheless, no attempt has been made to assess the precision of the estimators and to evaluate whether shrinkage estimators might improve overall variance estimates in these cases as well. Further investigation and research are needed in this area to evaluate whether shrinking the pooled variances before calculating the variances would be a superior method than the ordinary one for categorical variables as well.

Another possible step would be to try an extension of the considered method to the case of random-effects models where we could observe the same deficiencies as in the

the fixed-effects models considered above. Despite the fact that the problem of estimating σ_i^2 is particularly urgent and dramatic in the fixed-effects model, the estimation of σ_i^2 might also be expected to influence random effects coverage probabilities especially when all studies in the meta-analysis are small (Brockwell & Gordon, 2001).

Bibliography

Biggerstaff, B. & Tweedie, R. (1997), 'Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis', *Statistics in Medicine* **16**, 753–768.

Böckenhoff, A. & Hartung, J. (1998), 'Some Corrections of the Significance Level in Meta-Analysis', *Biometrical Journal* **40**(8), 937–947.

Braga, M. (2004), *Mean-Variance Efficiency: da Markowitz ... ad oggi*, L'innovazione finanziaria. Osservatorio Newfin, Bancaria Editrice, chapter 17.

Brockwell, S. & Gordon, I. (2001), 'A comparison of statistical methods for meta-analysis', *Statistics in Medicine* **20**, 825–840.

Burton, A., Altman, D., Royston, P. & Holder, R. (2006), 'The design of simulation studies in medical statistics', *Statistics in Medicine* **25**, 4279–4292.

Chatterjee, S. (1986), 'The Cardioselective and Hypotensive Effects of Bisoprolol in Hypertensive Asthmatics', *Journal of Cardiovascular Pharmacology* **8**((Suppl. 11)), S74–S77.

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

Cox, D. & Hinkley, D. (1974), *Theoretical Statistics*, London : Chapman and Hall.

Cox, D. & Solomon, P. (1997), *Statistical Iinference*, Monographs on statistics and applied probability, Boca Raton, FL; London: Chapman & Hall/CRC.

Cui, X., Hwang, J., Qiu, J., Blades, N. & Churchill, G. (2005), 'Improved statistical tests for differential gene expression by shrinking variance components estimates', *Biostatistics* **6**, 59–75.

DerSimonian, R. & Laird, N. (1986), 'Meta-Analysis in Clinical Trials', *Controlled Clinical Trials* **7**, 177–188.

Ghosh, M., Hwang, J. & Tsui, K. (1983), 'Construction of Improved Estimators in Multiparameter Estimation for Discrete Exponential Families', *The Annals of Statistics* **11**(2), 351–367.

Hardy, R. & Thompson, S. (1996), 'A likelihood approach to meta-analysis with random effects', *Statistics in Medicine* **15**, 619–629.

Hardy, R. & Thompson, S. (1998), 'Detecting and describing heterogeneity in meta-analysis', *Statistics in Medicine* **17**, 841–856.

Hartung, J. & Knapp, G. (2001), 'A refined method for the meta-analysis of controlled clinical trials with binary outcome', *Statistics in Medicine* **20**, 3875–3889.

Hedges, L. (1983), 'A random effects model for effect sizes', *Psychological Bulletin* **93**(2), 388–395.

Hedges, L. & Olkin, I. (1985), *Statistical Methods for Meta-Analysis*, Orlando ; London : Academic Press.

Higgins, J. & Thompson, S. (2002), 'Quantifying heterogeneity in a meta-analysis', *Statistics in Medicine* **21**, 1539–1558.

Jackson, D. (2006), 'The power of the standard test for the presence of heterogeneity in meta-analysis', *Statistics in Medicine* **25**, 2688–2699.

Jain, N., Thatte, J., Braciale, T., Ley, K., O'Connell, M. & Lee, J. (2003), 'Local-pooled-error test for identifying differentially expressed genes with a small number of replicated microarrays', *BioInformatics* **19**(15), 1945–1951.

Knapp, G., Biggerstaff, B. & Hartung, J. (2006), 'Assessing the Amount of Heterogeneity in Random-Effects Meta-Analysis', *Biometrical Journal* **48**(2), 271–285.

Kubokawa, T. (1999), 'Shrinkage and Modification Techniques in Estimation of Variance and the Related Problems: A Review', *Communications in Statistics: Theory and Methods* **28**(3/4), 613–650.

Li, Y., Shi, L. & Roth, H. (1994), 'The bias of the commonly-used estimate of variance in meta-analysis', *Communications in Statistics – Theory and Methods* **23**, 1063–1085.

Lin, Y., Nadler, S., Lan, H., Attie, A. & Yandell, B. (n.d.), *Adaptive Gene Picking with Microarray Data: Detecting Important Low Abundance Signals*, The Analysis of Gene Expression Data: Methods and Software,, Springer, NY.

Mengersen, K., Tweedie, R. & Biggerstaff, B. (1995), 'The impact of method choice on meta-analysis', *Australian Journal of Statistics* **37**(1), 19–44.

Mosteller, F. & Chalmers, T. (1992), 'Some Progress and Problems in Meta-Analysis of Clinical Trials', *Statistical Science* **7**(2), 227–236.

Normand, S. (1999), 'Tutorial in Biostatistics. Meta-Analysis: Formulating, Evaluating, Combining and Reporting', *Statistics in Medicine* **18**, 321–359.

Opgen-Rhein, R. & Strimmer, K. (2007), 'Accurate Ranking of Differentially Expressed Genes by a Distribution-Free Shrinkage Approach', *Statistical Applications in Genetics and Molecular Biology* **6**(1), Article 9.

Petitti, D. (1994), *Meta-analysis, decision analysis, and cost-effectiveness analysis : methods for quantitative synthesis in medicine*, New York; Oxford: Oxford University Press.

Piccolo, D. (2000), *Statistica*, Bologna; Il Mulino.

Rees, K., Taylor, R., Singh, S., Coats, A. & Ebrahim, S. (2004), 'Exercise based rehabilitation for heart failure', *Cochrane Database of Systematic Reviews* **Issue 3**, Art. No.: CD003331.

Salpeter, S., Ormiston, T. & Salpeter, E. (2002), 'Cardioselective β-Blockers in Patients with Reactive Airway Disease: A Meta-Analysis', *Annals of Internal Medicine* **137**(9), 715–725.

Senn, S. (2000), 'The many modes of meta', *Drug Information Journal* **34**, 535–549.

Sidik, K. & Jonkman, J. (2005), 'Simple heterogeneity variance estimation for meta-analysis', *Applied Statistics* **54**(2), 367–384.

Sidik, K. & Jonkman, J. (2006), 'Robust variance estimation for random effects meta-analysis', *Computational Statistics & Data Analysis* **50**, 235–246.

Smith, K. (1997), 'Explaining Variation in State-Level Homicide Rates: Does Crime Pay?', *The Journal of Politics* **59**(2), 350–367.

Thompson, A. & Pope, J. (2005), 'Calcium channel blockers for primary Raynaud's phenomenon: a meta-analysis', *Rheumatology* **44**(2), 145–150.

Thompson, S. & Sharp, S. (1999), 'Explaining heterogeneity in meta-analysis: a comparison of methods', *Statistics in Medicine* **18**, 2693–2708.

Tong, T. & Wang, Y. (2007), 'Optimal Shrinkage Estimation of Variance with Applications to Microarray data Analysis', *Journal of the American Statistical Association 102 (477):* **102 (477)**, 113–122.

Viechtbauer, W. (2007), 'Confidence intervals for the amount of heterogeneity in meta-analysis', *Statistics in Medicine* **26**, 37–52.

Whitehead, A. (2002), *Meta-Analysis of Controlled Clinical Trials – Statistics in Practice*, Chichester: John Wiley & Sons Ltd.