



Universiteit
Leiden
The Netherlands



LEIDEN UNIVERSITY MEDICAL CENTER

DEPARTMENT OF MATHEMATICS

MASTER THESIS

STATISTICAL SCIENCE FOR THE LIFE AND BEHAVIOURAL SCIENCES

Modes of Analysis for a set of N-of-1 Trials

Author:

Giulio Flore

Thesis Advisor:

Erik van Zwet, Department of Medical Statistics and
Bioinformatics Leiden University Medical Center

Supervisor:

Marta Fiocco,
Department of Medical Statistics and Bioinformatics
Leiden University Medical Center and Department of Mathematics, LU

February 2015

Abstract

The N-of-1 trial is a single subject, cross-over, double blinded, experimental design developed in the 1980s to aid routine clinical practice. The N-of-1 trials were used to select therapies for patients when there was uncertainty on the best course of action. During the 1990s several researchers proposed the use of N-of-1 trials in clinical research for rare diseases by pooling them in a single Randomized Controlled Trial (RCT). Several types of analysis have been applied to the data from these trials. This thesis will review some of these types of analysis and will outline the advantages and disadvantages in each case.

The focus will be on the assessment of three modes of analysis for the estimation of three parameters: the Individual Treatment Effect (ITE); the Trial Treatment Effect (TTE); and the Population Treatment Effect (PTE). The ITE is used to assess if the treatment is effective for the single patient. The TTE is used to verify whether the treatment is effective, on average, across the participants to the trial. On this basis the TTE is defined as the average of the ITEs. The TTE is also used as a premise for further investigations concerning the PTE estimate. The PTE is used to verify if the treatment can be assumed to be effective, on average, across the entire population of patients.

The three modes of analysis are assessed in order of complexity. We start with a Fixed Effects Linear Model for the estimation of ITEs and of the TTE. This model requires limited assumptions and restrictions. We then review a Random Effects Linear Mixed Model. This model imposes additional restrictions as we assume a normal distribution for the random effects due to baseline heterogeneity and treatment heterogeneity. Finally we consider a Bayesian Model that replicates the specifications of the Linear Mixed Model. This latter model requires more assumptions and restrictions because it treats the parameters in the model as random variables with known distributions.

We assess the models using two different types of data. Firstly, we analyze a set of actual patient records from an RCT based on N-of-1 trials, using different models. Subsequently, we run simulations in which a virtual trial is set to be open ended with continuous enrollment of new patients. The simulations mimic the selection of subjects from different populations of patients with varying levels of baseline and treatment heterogeneity. We apply a Bayesian Model in each simulation to see how quickly the estimates of the model parameters converge towards the parameters of the virtual populations. On the basis of the analysis of real and simulated data we propose a systematic approach to the analysis of data from RCTs based on pooled N-of-1 trials.

First, ITEs and the TTE should be estimated with a simple Fixed Effect Linear Model. ITEs and the TTE can be derived and tested as contrasts of a model based on the interaction between the individual subjects and the treatment. The ITEs and TTE estimates are unbiased and are strictly objective measurements. Therefore this form of analysis should be considered as the preferred initial mode of analysis. If this first step provides evidence for a significant TTE, it is reasonable to proceed with the specification of Linear Mixed Model or a Bayesian Model analysis to estimate the PTE.

Second, Linear Mixed Models and Bayesian Models with matching specifications can be both used to estimate the PTE. We have found that Bayesian Models have some advantages if we need to estimate the variance for random effects. As a general rule we suggest that both the PTE and the Treatment random effects variances should be estimated using a Bayesian Model, provided that appropriate informative priors can be defined.

Third, Bayesian Models can be used to produce ITEs estimates under some restrictions. Bayesian ITEs estimates can be more precise than Fixed Effect Linear Model ITEs estimates, but they are biased due to shrinkage. If the analysis shows that the variance of the treatment random effects is low or negligible, then it is reasonable to use Bayesian ITEs estimates because they will be less affected by shrinkage bias. As noted before, appropriate informative priors need to be defined.

Fourth, Linear Mixed Models should not be used for ITEs estimation. If we treat the FE and RE estimates as random variables following the Student T-distribution – which is the standard mode of testing for these values – we would find that the definition of the distribution of their sum is a non-trivial task. The simplest alternative would be to use the Normal distribution for a Z-test and that would not be an appropriate assumption.

Finally, open ended trials can be analyzed with Bayesian Models in an effective manner. The PTE estimates updated by the inclusion of new subjects are, on average, affected by only a small amount of bias. However, if the variance of treatment random effects in a population is not very small, and if non-informative priors are used for the initial analysis, then the Bayesian estimates may become unstable as new patients are added to the analysis. In some instances the estimate for the variance of random treatment effects will become biased. In this scenario it is important to apply properly selected informative priors to minimize the possibility of unstable and biased estimates.

Table of Contents

1. Introduction	1
1.1 Motivation	1
1.2 Aim of the Thesis	1
1.3 Structure of the Thesis	1
1.4 Analytical Software	2
2 Literature review	3
2.1 The development of N-of-1 trials	3
2.2 Features of N-of-1 trials	4
2.2.1 Therapy Selection and Clinical Equipoise	4
2.2.2 Clinical Research	5
2.3 Modes of Analysis	5
2.3.1 Frequentist Approaches	6
2.3.2 Bayesian Approaches	8
2.3.3 Reporting Results	10
3 Modes of Evaluation of Evidence and Hypothesis Testing	11
3.1 Testing Hypotheses to Evaluate Results	11
3.2 Frequentist (Neyman-Pearson) Hypothesis Testing	12
3.3 P-Values and Confidence Intervals (CIs)	12
3.3.1 P-Values	12
3.3.2 Confidence Intervals	13
3.4 Bayesian Hypothesis Testing	14
3.5 Posterior Probabilities, Credible Intervals, Bayesian Factors (BFs)	15
3.5.1 Posterior Probabilities	15
3.5.2 Credible Intervals	15
3.5.3 Bayes Factors	16
4 The Data	18
4.1 Source	18
4.2 The Dataset	18
4.3 The Variables	18
4.3.1 Indicator Variables	18
4.3.2 The FACIT-F Score	18
4.3.3 The dataset	19
5 Analysis of the data: Individual, Trial and Population Perspectives	20
5.1 TTE and ITEs: The FE Linear Model Approach	20

5.1.1	ITEs Estimates	21
5.1.2	TTE Estimate	24
5.2	PTE and the RE Linear Mixed Model Approach	24
5.2.1	LMM Methodological Basis	26
5.2.2	The Analysis	28
5.2.3	ITEs	29
5.2.4	Estimates for the PTE and Variance Components	31
5.3	The Bayesian Model Approach	32
5.3.1	The Analysis	33
5.3.2	ITEs	37
5.3.3	Estimates for the PTE and Variance Components	38
5.4	Comparison of Modes of Estimation	39
5.4.1	Comparability of the ITEs	39
5.4.2	Comparability of the PTE and TTE	41
5.4.3	Comparability of the Variance Components	41
5.4.4	Summary	42
6	Simulations	43
6.1	Bayesian Analysis by Updating Priors	43
6.1.1	PTE simulation trends	44
6.1.2	BF for $H_0: PTE \leq 0$ Simulation Trends	45
6.1.3	Random Treatment τ_β Simulation Trends	46
6.1.4	Predicted RE Precision Simulation Trends	47
6.1.5	ITE Precision Simulation Trends	48
6.2	Stability of Bayesian Analysis: Random Sampling Sequences	49
6.2.1	Non Informative Prior – Small Treatment RE SD	49
6.2.2	Non Informative Prior – Increased Treatment RE Variance	51
6.2.3	Non-Informative Prior for Treatment RE SD	52
6.2.4	Implications of Simulations	53
7	Conclusions	54
7.1	ITEs – FE Linear Model, RE Mixed Model and Bayesian Model	54
7.2	TTE Estimation – FE Linear Model	54
7.3	PTE Estimation – Mixed and Bayesian Model Analysis	54
7.4	Use of SAS PROC MIXED and SAS PROC MCMC	55
Appendix	56
A.1	Chapter 5 Codes	56
A.1.1	Section 5.1 - SAS PROC MIXED for Fixed Effects Linear Model	56
A.1.2	Section 5.2 - SAS PROC MIXED for Random Effects Linear Mixed Model	61

A.1.3 Section 5.3 - SAS PROC MCMC for Bayesian Model	67
A.2 Chapter 6 Codes.....	72
A.2.1 Section 6.1 - SAS PROC MCMC for Simulation of Different Populations	72
A.2.2 Sections 6.2 and 6.3 - SAS PROC MCMC for Multiple Simulations	93
A.2.3 Section 6.4 - SAS PROC MCMC for Simulation with Different Prior	105
References	118

1. Introduction

1.1 Motivation

The topic of this thesis has its origin in a Randomized Control Trial (RCT) being conducted at Leiden University Medical Center (LUMC). The objective of this RCT is to determine the effect of add-on treatment with ephedrine for patients affected by Myasthenia Gravis (MG). MG is an auto-immune chronic condition that attacks the neuromuscular system. At present, MG affects approximately 2,000 subjects in the Netherlands. The RCT is based on a set of 4 patients, each undertaking an N-of-1 trial.

The N-of-1 trial is a randomized controlled trial on a single subject, with cross-over design, a randomized sequence of treatments over a number of cycles, and double-blinding to treatment. This is a distinct procedure from the Single Subject Design or Single Patient Trial which has been developed in experimental psychology. In this latter approach double blinding is not considered for a variety of reasons, nor is cross-over randomization formalized.

Different methods have been developed to analyze the data from N-of-1 trials. This thesis will review some of these methods and outline the constraints and compromises in each case.

1.2 Aim of the Thesis

The original context for the development of the N-of-1 trial was the assessment of experimental treatments for subjects with rare and chronic illnesses, either because of the lack of established treatments (for example, the so called orphan diseases) or because standard treatments were not effective. Chronic condition is here taken to mean a condition which is relatively stable over time, so that the cross-over design is not influenced by a relatively fast change in the conditions of the patient.

As such, N-of-1 trials were not developed as a basis for the testing of a treatment for general use in the general population. More recently, researchers have designed RCTs for patients affected by orphan and chronic conditions as a collection of N-of-1 trials in order to estimate:

- the Individual Treatment Effect (ITE) for each subject participating in the RCT; and
- the Trial Treatment Effect (TTE), i.e. the mean treatment effect across the subjects participating to the RCT; and
- the Population Treatment Effect (PTE), i.e. the mean treatment effect across the population of subjects affected by the condition and eligible for the treatment.

The aim of this thesis is to discuss three different method of analysis for RCTs based on a small collection of N-of-1 trials, in order to estimate the ITEs, TTEs and PTEs of experimental treatments for patients, and to highlight the advantages and potential disadvantages of each method. These methods are:

- A Fixed Effects (FE) Linear Model, for the estimation of ITEs and the TTE;
- A Linear Mixed Model with Random Effects (RE), for the estimation of ITEs and the PTE;
- A Bayesian Model, for the estimation of ITEs and the PTE;

This thesis will also assess the implications of an open ended trial. In this approach the statistical analysis of the data from an initial small scale RCT will be used to evaluate a subsequent N-of-1 trial for new patients using Bayesian methods, and to make inferences about the validity of the treatment for the general population of patients.

1.3 Structure of the Thesis

Chapter 2 consists of a literature review on the application of N-of-1 Trials (both as individual experiments and as part of an RCT) and the modes of analysis used in these trials. Chapter 3 briefly

reviews the nature and implications of using p-values, Confidence Intervals (CIs), Posterior Probabilities, Credible Intervals and Bayes Factors to assess the evidence provided by different modes of analysis. Chapter 4 describes a subset of the data from a RCT on palliative care. This RCT has been carried out as a collection of N-of-1 trials. This subset of data will be analyzed with the different approaches set out in the previous paragraph.

In Chapter 5 the three forms of analysis are applied to the RCT Data. The results are discussed and compared. Conclusions are drawn on the relative advantages and disadvantages of the three approaches. Chapter 6 addresses an Open ended scenario, in which a virtual RCT keeps including, new patients, one at the time. A virtual population is simulated using the analysis from the RCT data. Every new patient is analyzed with a Bayesian Model, with priors based on the analysis of the previous patients, starting with the initial RCT analysis. The implications of this approach for an assessment of the ITEs and the PTE are discussed and evaluated.

Chapter 7 presents the conclusions, summarizes the findings of Chapters 5 and 6 and outlines a possible general approach to the analysis of RCTs based on a collection of N-of-1 Trials.

1.4 Analytical Software

We have selected the SAS statistical package as this is the preferred tool for the formal analysis of RCT data aimed at the evaluation of new treatments. The analysis has been carried out using SAS PROC UNIVARIATE, SAS PROC MIXED and SAS PROC MCMC. R is mostly used for graphical displays and occasionally for some sanity checks on results based on SAS. The SAS code used for the analysis is provided in the Appendix according to each chapter.

2 Literature review

In this section we will provide a short overview of the development of the N-of-1 trial as a clinical procedure, the context in which it can be meaningfully applied, the practical implications of the use of N-of-1 trials in a clinical setting. We also briefly discuss developments in the approaches to the analysis and reporting of the outcome of the analysis of the data. This review is neither meant to be systematic nor exhaustive, but it summarises key papers that address directly the methodological problems surrounding the execution of N-of-1 trials. The ethical and funding aspects of the implementation of the N-of-1 trial, while important, will not be discussed, except to mention some limitations to the use of N-of-1 trials for clinical research purposes.

The search of papers has been carried out mainly on the *PUBMED* search engine. Some papers have been selected using Google Scholar. The search terms were “N-of-1” for title and abstract in Google Scholar. For *PUBMED* the search term were for *intitle*: “N-of-1”, “Single Subject Design” and for *intitle abstract*: “Methodology”, “Bayes”, “Empirical”, “Hierarchical”, “Mixed” and “Longitudinal”.

2.1 The development of N-of-1 trials

As noted by Kravitz et al. [15], the Randomized Control Trial (RCT) has long been regarded as the gold standard for assessing which type of treatment has general – that is at population level – efficacy for patients with a given condition. Implicit in this view is the assumption that patients will respond more or less homogeneously to a tested treatment. A RCT aims to estimate the average treatment effect across patients. Thus the efficacy of a therapy as tested in RCTs does not necessarily indicate how effective the therapy is for any given individual.

According to Kravitz et al. [15], the first proposals for a personalized cross-over trial were first advanced and formalized in Psychological Research, and occasionally, in psychological therapy. There were no suggestions that such a design could be used in clinical practice and research. Janosky [13] points out that R.A. Fisher outlined the concept of single subject design as early as early as 1945.

The first researchers to consider N-of-1 trials in clinical practice were a team lead by Guyatt in the first half of the 1980s. This team developed the idea of individualized, double blinded, cross-over trials, e.g. N-of-1 trials, as the preferred method to address the effectiveness of a therapy for a specific patient. According to Guyatt et al. [12] the N-of-1 trial was in fact conceived not as a research method but as a formalized, bias-controlled testing procedure for the selection of therapies for a specific individual. Tsapas and Matthews [26] note that the N-of-1 trial was in effect developed as a form of Therapy Audit.

In the 1990s, Lilford and Braunholtz [17] and Zucker et al. [27] noted that N-of-1 trials can be pooled to study rare diseases when conventional RCTs cannot be set up. Thus small sample trials can be run as a series of concurrent (or ongoing) individual trials for research purposes. Alternatively, the results of several independent trials run according to a common template (including formal RCTs) can be aggregated to provide treatment estimates applicable to the general population.

Lilford and Braunholtz [17] noted that data pooled in this manner can provide information on the Heterogeneity of Treatment Effects (HTE) across the general population. There is a potential for a widespread application of N-of-1 trials both for the testing of therapies for rare diseases, where there are no established treatments, and for evidence based medicine practice in general to find the more appropriate treatment for a specific patient. More recently Facey et Al. [7] has noted that although rare diseases taken one by one affect small sections of the general population, when aggregated they affect a sizable proportion of the patient population. Therefore N-of-1 trials may have a very important role to play in supporting research for treatments for these conditions.

2.2 Features of N-of-1 trials

The N-of-1 trial consists of a cross-over design. The patient is subjected to a randomised sequence of different treatments. This sequence is then repeated for at least two or more times depending on the required level of precision. In its simplest form an N-of-1 trial will be represented by a randomised sequence of treatments A and B, such as ABBA, or ABAB, etc. Occasionally more than two treatments have been considered. The number of repeats may be varied to increase estimation precision and/or to account for the patient willingness to undertake repeated treatment sessions.

The Null Hypothesis of the N-of-1 trial is that the average difference in outcomes for each pair A, B is not significantly different from a set value (0 or otherwise). The cross-over aspect implies that an N-of-1 trial should be applied only to patients with a chronic disease, with stable symptoms (at least in the short to medium term) and with a reversible intervention. Ideally, the treatment should also exhibit fast take-up and short washout periods, to avoid carryover effects that may distort the analysis of the data.

The randomization is usually a block randomization within each set of treatments. Duan et al. [6] report that adaptive designs procedures have been suggested to accommodate patients' feedback and minimize exposure to inferior treatments. Lillie et al. [18] note that for certain conditions a randomization over the entire trial may enable the discovery of cyclical patterns of symptoms, insofar that an AAA/BBB sequence may be generated. Cushing et al. [5] indicate that in some cases Latin Squares or counterbalanced (that is back to back, such as AB BA) designs may be a more appropriate choice.

Latin Squares can be used when the experimenter may wish to control for order effects for more than two treatment levels. For example, a researcher may set a 3x3 matrix to represent two treatment conditions compared against one control condition. The first row and column of the matrix would each be ordered ABC; the other cells would be ordered such that each condition appears only once per row and only once per column, so that no order effect may be generated.

The counterbalanced design is suggested as a way to overcome confounding time effects. Schmid et al. [24] note that if there is a steady degradation in the condition of the patient, a counter balanced design may control effectively for such a time trend. These are not ideal conditions for an N-of-1 trial, but there may be cases where the benefits of an N-of-1 trial justify this special experimental design.

The use of N-of-1 trials raises different ethical issues depending on whether they are used as a clinical research tool or in clinical routine decision making. In a clinical research context - where N-of-1 trials are used as an experimental design integrated in a formal RCT experiment- the approval and funding process would be identical to a more conventional research experiment. So the ethics of approving a set of N-of-1 trials is the same as for a conventional RCT. The situation is different when N-of-1 trials are used for clinical routine decision making and the data collected is not to be used for medical research.

2.2.1 Therapy Selection and Clinical Equipoise

Kravitz et al. [15] quote Vohra and Janoski, who claim that N-of-1 trials are mistakenly perceived as a research activity. To overcome this misconception, Kravitz et al. [15] describe N-of-1 trials as a means to achieve therapeutic precision.

Tapas and Matthew [36] also define N-of-1 trials as a therapy audit. These authors suggest that a N-of-1 trial does not need to undergo a formal Ethical Approval procedure, while still observing normal patients' rights such as confidentiality, free and informed consent to participate, and withdrawal rights without any consequences on rights for care. Tapas and Matthew [36] also suggest that no further assessment is required because clinical research is not the objective of the trial and there is no request to the patient to share the relevant data.

It is not clear to what extent any experimental treatment may be covered by this provision. If the scenario is restricted only to a choice between alternative and established forms of treatment, and the N-of-1 trial is only used to select the best available treatment, the approach proposed by these authors has a valid rationale. However if the scenario is based on a patient not treatable by conventional procedures and the N-of-1 trial is used as an experimental treatment, the burden for approval must necessarily include an ethical assessment. It also must be noted that the authors' view may be influenced by the legal and administrative framework of the country of origin.

Lillie et al. [18] make a case for restricting the use of N-of-1 trials only to cases where the clinician is in a state of Clinical Equipoise, meaning that she is neutral with respect to the actual suitability of a treatment (or competing treatments) in relation to her patient. In other words: if there is strong evidence that a treatment is generally suitable for a varied population of patients, it would be unethical to put a patient through a trial.

If however, a patient does not respond well to a therapy, or uncertainty arises about the effectiveness of the treatment of a specific patient, then an N-of-1 trial can be proposed as an ethical and clinically effective procedure to select the most suitable treatment for that specific patient.

2.2.2 Clinical Research

More recently, Zucker et al. [28] have suggested that a series of N-of-1 trials can be important in clinical research if pooled together to provide greater precision for the assessment of the general validity of a treatment. N-of-1 trials can be conducted within a more conventional RCT to reduce patient drop-out and provide important information on the degree of the HTE in routine clinical practice.

It follows that in the case of rare diseases, a RCT may be best designed as a collection of standardised N-of-1 trials, using an approved and publicly available protocol, that can be pooled together retrospectively to test the validity of treatment in the general population. Therefore, regardless of the original objectives of the trial, a case may be made for having each trial submitted by default for approval to the relevant Ethics Committee.

The basis for such an approach would be the intention to treat any N-of-1 trial as part of a generalized, ongoing process of Clinical Research, and as such should be approved by the relevant committee. In this situation, data collection and sharing would always be considered as an integral part of the N-of-1 protocol.

2.3 Modes of Analysis

In the late 1980s and early 1990s the data of single N-of-1 trials was analyzed using basic procedures, such as visual inspection of graphs and simple statistical tests (t-tests of Patient Reported Outcomes (PRO) scores). In the first systematic application of N-of-1 trials described by Guyatt et al. [12], the mode of analysis was a t-test applied to PRO scores. A systematic review of papers covering N-of-1 trials from 1985 to 2010 by Gabler et al. [8] finds that of the 108 trials selected in the review, a quarter reported only using graphical comparisons without any statistical assessment.

From the mid 1990s onwards, it was appreciated that data from multiple N-of-1 trials (either within a RCT or from reported routine clinical activities) could be pooled and analyzed with more complex techniques to provide more precise analysis of the trials' data and, potentially, to generalize the trials' outcomes to a wider population.

The techniques, as reported in the literature, can be broadly divided in four areas: t-tests, Meta-Analysis (when only mean and S.E. estimates from the trials are reported), Mixed Models, and Bayesian Models.

Table 1 shows a basic classification of the techniques reported in the literature according to the objectives of the trial.

	Frequentist	Bayesian
Clinical Routine / Therapy Selection	<i>t-test</i> <i>Non-Parametric Tests</i> <i>Graphical Inspection</i>	<i>Bayesian analysis (with previous RCT, meta-analysis, expert opinion based priors).</i>
Clinical Research	<i>t-test on subsets of treatment period outcomes</i> <i>Meta-analysis</i> <i>Mixed Models</i>	<i>Bayesian Model</i>

Table 1: Statistical Methodologies by Objective in N-of-1 trials

It must be noted that although meta-analysis is listed here in the Frequentist column, as a technique it may be applied using FEs, REs and Bayesian specifications.

The key difference between Meta-Analysis and other modes of analysis is that the Meta-Analysis aggregates other RCT mean and S.E. estimates to provide a pooled estimation. The other modes pool the individual records.

The two statistical methodologies presented in Table 1, e.g. Frequentist approaches and Bayesian approaches are briefly discussed below. More detailed theoretical assessments are carried out in the following chapters.

In the following sections we report on the different techniques used by clinicians and researchers for use in a Therapy Selection context (when only one N-of-1 trial is considered and it is used to select a therapy) and in a Clinical Research context (when multiple N-of-1 Trials are considered and they are used to test a treatment to be extended to other patients).

Recalling our stated goals in section 1.2, we note that techniques listed for clinical routine / therapy selection analysis focus on the estimation of ITEs. Clinical research analysis instead focuses on the estimation of TTEs and PTEs. This difference in focus implies quite different possibilities and restrictions for the analysis of the data, which will be discussed below.

2.3.1 Frequentist Approaches

A key feature of the statistical analysis of N-of-1 trial data is that the patient acts as his own control, removing confounding effects, thus making a more accurate assessment of the treatment effect possible. The repetition of pairs of alternative treatments enables the estimation of the variability of results and hence an indication of the precision of the estimated individual treatment effect.

2.3.1.1 Clinical Routine / Therapy Selection

T-tests and Non-Parametric Test. In its simplest form the analysis is carried out as a t-test of the differences between alternated treatments. Gabler et al. [8] report also the use of non-parametric tests such as Wilcoxon. The randomization of the treatment sequence, the use of wash-outs, and the selection of appropriate outcome measurements are the essential conditions for the validity of the analysis.

The t-tests and Wilcoxon test are not appropriate when there is missing and/or unbalanced data caused by poor reporting or shortened intervention periods due to adverse side-effects or lack of effect, making the

drawing of firm conclusions from the trial quite difficult. It is worth noting, that in the first prolonged experimentation of this method reported by Guyatt et al. [12], only 19 out of 44 trials reached a definite statistical conclusion. t-test of patient reported scores were used for the analysis in all cases.

Mixed Models. Mixed Models are more robust insofar that they are robust against missing data and unequal variances. The literature generally discusses this approach in the context of pooled N-of-1 trials, in which the combined information is used to provide an Average Treatment Estimate. However Rochon [22] and Avins et al. [1] propose a longitudinal model also for the analysis of a single N-of-1 trial. Cushing et al. [5] remark that a Mixed Model enables the modelling of carry-over effects when wash-out procedures are ineffective.

2.3.1.2 Clinical Research

Meta Analysis. A discussion of this approach is found in Zucker et al. [28]. The key advantage of meta-analysis procedures lies in the fact that variations in trial designs can be reconciled by the pooled weighted estimation of the mean results and their respective variances. One disadvantage is that the quality of the sources of data may be uneven.

Another important issue is the chance that there may be different values of residual variance in different sources of data, especially in the case of rare diseases. In that regard, Zucker et al. [28] also note that it is in practice more effective to assume that all the sources have the same degree of residual variance, suggesting that in a Bayesian Meta Analysis approach this is equivalent to setting a highly informative prior for the σ^2 .

T-tests (FE Models). Another approach, suggested by Chen and Chen [4] consists of t-tests on outcomes of multiple N-of-1 trials in which a cycle consist of a pair of treatment periods (placebo vs. drug). For example, given N cycles of N-of-1 trials with K subjects enrolled, a total $K \times N$ pairs of data are generated, each subject providing N pairs of data. Paired t-test are used to analyze the pooled pairs of data, without accounting for between-subject variance. This approach has been tested using data simulation and the authors claim that t-tests provide the best results in terms of predictive precision, provided that normality and independence of observations is proven. Otherwise Mixed Models are found to be a better approach.

Zucker et al. [28] mention two alternative t-test approaches derived from meta-analysis literature. In one approach, all first period treatments outcomes are pooled together as from a conventional RCT (where patients are randomized in two groups). In the other approach, all first two treatment periods outcomes are collected and analyzed as in a conventional RCT AB / BA cross-over design.

Both Chen and Chen [4] and Zucker et al. [28] apply in practice FE Linear models, under different sets of assumptions on the homogeneity of effects across patients and treatment cycles.

Mixed Models. These models can be effectively applied to the analysis of pooled N-of-1 trials. The main advantages of this approach are:

- model appropriate for the hierarchical and longitudinal structure of the data;
- resilience against missing data;
- control of nuisance factors;
- simultaneous evaluation of the population and individual mean treatment effects; and
- the ability to specify covariance structures that reflect autocorrelation, heterogeneous response variance, hierarchical (clustered and nested) data structures.

A possible disadvantage is that, by definition, these RCTs are based on small numbers of pooled N-of-1 trials, each trial reporting a small number of treatment cycles. The amount of data available – including

the possibility of missing data and drop-outs – may not be able to support analysis based on complex mixed models.

2.3.2 Bayesian Approaches

Lilford and Braunholtz [17] make a case that Bayesian analysis is a natural fit to N-of-1 trials. These authors list a number of advantages, such as the ability to apply different clinical judgement scenarios as priors and the integration of both quantitative and qualitative sources of information on the trial data. This position has been supported by several other authors¹. The main benefits can be summarized as follows:

- the ability to incorporate heterogeneous information concerning the patient response;
- the ease of specification of distributional and covariance assumptions, and related hierarchical structures, in the form of priors;
- the ability to produce relevant information from small samples;
- the model output as probabilistic statements on the effectiveness of the treatments in an intuitive and self-explanatory way; and
- the updating of the data as additional trials are made available.

These advantages are counterbalanced by several disadvantages, such as:

- lack of familiarity with the Bayesian framework;
- the subjective nature, and lack of objectivity, of the Bayesian analytical framework;
- individual subjects are still assumed to be essentially homogeneous: otherwise some form of data stratification should be considered;
- sensitivity to the ratio of within and between subject variance;
- sensitivity to the choice of priors; and
- the difficulty of defining appropriate non-informative priors.

An assessment of the balance between advantages and disadvantages is not a straightforward one and it is contingent on each research context. We can however note that the Bayesian approach can be considered as a counterpart to the Mixed Model approach, insofar that both address the fact that the data is grouped at individual level. The Bayesian approach enjoys some degree of advantage over other methods because it is possible to generate posterior estimates even with limited information, provided that suitable priors can be used.

The choice of suitable priors is nonetheless a non-trivial problem. We will see in the following chapters that the Bayesian approach is constrained by the sensitivity of Bayesian models to chosen prior assumptions. For example:

- Priors assumptions that are perfectly suitable for parameter estimation may create problems in some forms of hypothesis testing;
- If a Bayesian model does not produce a clear shift from the prior assumptions after the incorporation of the data, it is not possible to tell whether this is due to the fact that the data is in agreement with the prior assumptions, or it is due to a prior assumption that dominates the data represented by a small sample;
- The Bayesian approach is also bound to introduce a certain degree of bias in certain circumstances, in relation to prior assumptions for the variance of random effects for small samples.

This dependency on prior assumptions characterizes the subjective nature of the Bayesian approach.

¹ Amongst others N Duan, R Kravitz, P Schluter, C Schmid, R Ware, D Zucker.

2.3.2.1 Therapy Selection Context

Bayesian Analysis is extensively discussed as a suitable tool to select therapies by Lilford and Braunholtz [17]. The justification for this approach is twofold:

- borrowing strength from prior knowledge, so that the individual results are ‘weighted’ against the information available for other subjects, in the form of empirical priors and, related to that;
- the posterior distribution for the single individual is subject to shrinkage, e.g. extreme results are ‘damped’ by evidence provided by other subjects.

The same paper makes implicitly a more general case for the use of Bayesian Analysis to interpret trial data for rare diseases. This paper presents four cases of initial clinical judgement in terms of priors:

- Absolute Clinical Equipoise, as a Non-informative Prior giving equal weight to competing hypothesis on the effect of the treatment;
- Belief in greater likelihood of benefits, as Informative Prior with stated beliefs on the expected magnitude of effects;
- Belief in absence of effects, as Informative Prior with stated belief of absence of (zero mean) effect; and
- Belief of positive effects only in the sense of expected positive results, as vague prior centred on a positive mean effect result.

Lilford and Braunholtz [17] state that in all cases the posterior mean is expressed as a more balanced view of the original clinical statement, in so far that it is adjusted by the information provided by the available data. Hence, the clinical decision that results from the analysis of the data is always a more realistic version of the original clinicians’ views. The authors include a final additional example which shows that incorporating information from previous trials (not necessarily with the same outcome end points) as empirical priors still provides useful information for a clinical decision.

2.3.2.2 Clinical Research Context

Zucker et al. [27] make a case for the use of Bayesian Models for clinical research on account of the following advantages:

- Population Estimates are equivalent to the average effects estimated in Mixed Models, so that comparisons can be made across different studies;
- Individual observed extreme results will generate posterior estimates that are shrunk towards the population estimate if their within variance estimate is large reflecting so to speak the possibility that these are outlier results;
- Covariate predictors – for example dosage stratification – can be easily added to the model.

Zucker et al. [27] have also noted that N-of-1 trials estimates that show a strong shrinkage towards the population mean may be used as a criterion to either extend the trial (more paired periods to yield more informative within-subject variance) or modify subsequent trials designed to yield stronger evidence, in a sense providing an objective base for adaptive design. It must be noted that this is by no means a commonly accepted view.

This recommendation by Zucker et al. [27] raises the question of what threshold a statistician should adopt to trigger a change in design. The same consideration can be made in terms of the stopping rule, i.e. when additional trials are considered to have a too low marginal gain with respect to what is already known.

Schluter and Ware [23] propose a Bayesian Model in which the primary end point is a binomial (or a ranked categorical) variable. This variable expresses the preference of the subject for one treatment (or placebo) over another treatment. According to the authors, this approach provides several advantages:

- the patient has a straightforward manner to report the treatment outcome;
- results from other trials (N-of-1 and or RCTs) may be reasonably converted on the same preference scale of the new trials;
- outcome of analysis can be provided either in terms of a posterior probability range or a mean posterior probability; and
- posterior probabilities thresholds can be used to classify individual subjects as responsive to the treatment.

2.3.3 Reporting Results

The reviewed literature does not examine in depth the issue of how to best report the results to clinicians and patients. However it is possible to summarise some main themes:

- Graphical methods are useful in reporting outcomes to patients as a visual aid for the clinical decision process;
- Bayesian methods provide more intuitive feedback to patients and clinicians alike in the form of probabilistic statements on posterior means;
- No article discusses in any length how best to report frequentist forms analysis to patients and clinicians.

3 Modes of Evaluation of Evidence and Hypothesis Testing

In this chapter we briefly discuss the basic principles underpinning the evaluation of evidence using test statistics, as well as the test statistics that will be used in this thesis. We will discuss two paradigms: The Frequentist Paradigm and the Bayesian Paradigm.

The rationale for this discussion is to clarify that the adoption of a mode of analysis implies also a different approach to assess the evidence. The FE and RE Linear Models are Frequentist modes of analysis and the test statistics that are used to assess the evidence are also Frequentist. The Bayesian Hierarchical Model provides a different set of test statistics that need to be interpreted in a different manner. In essence, depending on how we analyze the data we must adopt different approaches to assess and interpret the evidence. The following section addresses the basic features and justifications for the use of these test statistics.

3.1 Testing Hypotheses to Evaluate Results

We analyse the data in a sample to draw inferences about the behaviour of the data in a population. We need to test the validity of those inferences on the basis of the information available. We can do this by setting up statistical hypothesis and testing them against the available evidence.

From a formal point of view, we can use the definition of a statistical hypothesis by Casella and Berger [3] whereby a hypothesis is a statement on a population parameter. For a given hypothesis H_0 it is possible to formulate a complementary hypothesis H_1 , so that if H_0 is true, then H_1 is false and *vice versa*. The test consists of using the evidence from the sample to select either hypothesis.

The complementary hypotheses are conventionally defined as Null (H_0) and Alternative (H_1) hypotheses. If we denote the parameter to be tested as θ , the Null and Alternative hypotheses can be stated as:

$$H_0: \theta \in \Theta_0 \text{ and } H_1: \theta \in \Theta_0^c \quad (1)$$

where Θ_0 is the subset of the parameter space that validates the H_0 and Θ_0^c is its complement. The Null Hypothesis is so defined because it is used as the hypothesis that negates the experimental hypothesis. It is important to note that the term parameter is a generic and potentially broad one. It could represent a vector of population parameters. It could also be a parameter evaluated after controlling for confounding effects, such as those defined by a regression analysis.

We now turn to the concept of testing. Casella and Berger [3] state that we can define a hypothesis testing procedure or hypothesis test as a rule. The rule defines the values derived from the sample for which we accept H_0 and the values for which we reject H_0 and accept H_1 . The subset of the sample space for which we accept H_0 is called the *acceptance region* and the complementary subset is called the *rejection region*.

By convention the hypothesis testing is formulated using a test statistic, which is a function of the sample data, so that we can define a test statistic $W(x_1, \dots, x_n) = W(\mathbf{X})$. For this test statistic we can therefore set acceptance and rejection regions which are the basis for our decisions. Traditionally certain test statistics (such as the t-test for example) were used because they had pre-calculated Null Distribution tables, which provided acceptance and rejection regions. More recently, computer based methods have been used to calculate a null distribution for any chosen statistic directly from the sample data.

The theory concerning hypothesis testing covers a number of important subjects, which are outside the scope of this thesis. In the next section we will look at two different approaches to hypothesis testing.

3.2 Frequentist (Neyman-Pearson) Hypothesis Testing

Following Rice [21], we note that the decision problem approach proposed by Neyman and Pearson is characterized by the fact that the parameter θ is considered a fixed quantity that cannot be treated as a random variable, and the assessment of the Null and Alternative hypotheses as either true or false is based only on the data. Consider the following rule:

Reject H_0 if a chosen test statistic $W(\mathbf{X})$ takes values in some set R .

Now, we have the probability of rejecting the Null Hypothesis when it is true (Type I error). This is also known as the test size α , or significance level.

Second, we can consider the probability of rejecting the Null Hypothesis when it is false as a function of the parameter θ and the sample data \mathbf{X} . This function is called the power of the test. The set R must be chosen so that if H_0 is true, the probability that $W(\mathbf{X}) \in R$ is not greater than the chosen level of significance α . Given requirement, $W(\mathbf{X})$ and R should be chosen in such a way that the probability that $W(\mathbf{X}) \in R$, is as large as possible when H_0 is not true, i.e. it has the greatest possible power.

We will now look at an important difference between this approach and the Bayesian approach. In the Frequentist approach it is not necessary to specify in advance the probability that either hypothesis is true. The decision is entirely based on the data observed and it is, in that sense, strictly objective. We note that the consequences of this procedure is that it is not possible to state that given the data, the H_0 has a probability to be true. Objectivity is gained at the expense of defining a probability for H_0 being true.

There is also a methodological problem in the definition of Null Hypothesis for parameters describing continuous random variables. Let us consider a sample \mathbf{x} for a random variable r.v. $x \sim N(\mu, \sigma^2)$, for which we need to estimate the mean μ . If we set the simple hypotheses for μ as $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$, we are assuming that the sample mean is exactly μ_0 and the data is perfectly distributed according to the Normal distribution.

If the sample is large enough the H_0 will always be rejected. A composite hypothesis $H_0: \mu \leq \mu_0, H_1: \mu > \mu_0$ would not suffer from this potential contradiction, and it is in fact the form that is usually adopted in Bayesian testing. In the next sections we will give a description of statistics used in these two alternative approaches and their interpretations.

3.3 P-Values and Confidence Intervals (CIs)

We now discuss two statistics that are used in the Frequentist paradigm as the basis to either accept or reject the Null Hypothesis: P-Values and CIs. They have been developed in different contexts and with different objectives and they are the most commonly reported test statistics.

3.3.1 P-Values

This test-statistic has been originally developed by R. A. Fisher. The purpose of this statistic is to provide a measure of the strength of the evidence on a continuous scale, so that ‘small’ p-values are evidence in favour of the alternative hypothesis. According to the definition by Casella and Berger [3], a p-value $p(\mathbf{x})$ is a test statistic such that $0 \leq p(\mathbf{x}) \leq 1$ for any possible sample \mathbf{x} and it is a valid statistic if, for every $\theta \in \Theta_0$ and every test size $0 \leq \alpha \leq 1$, $P_\theta(p(\mathbf{X}) \leq \alpha) \leq \alpha$

Thus according to Rice [21], the p-value may be interpreted as the smallest possible test size that could be imposed on the data sample so that H_0 is rejected. It follows that we can define a decision rule by setting *a priori* a level for α , e.g. the significance level, so that if $p(\mathbf{X}) \leq \alpha$ we reject the H_0 . This decision rule is then a valid alpha level test.

The guidelines for clinical research aimed to the approval of medical treatments recommend that we should consider a two-sided test with $\alpha = 0.05$. This is, to an extent, an arbitrary choice and should be open to revision in different contexts. For example, if the $p - value = 0.06$, we may not say that there is no evidence for the treatment to be effective. We should say that the evidence is not strong enough according to the stated requirements of evidence.

The actual calculation of the p-value changes depending on the type of inference – and test – being carried out. The p-values associated to tests statistics such as the t-test, Z-test, F-test, etc, apply certain distributional assumption and derive the p-value accordingly.

3.3.2 Confidence Intervals

The Confidence Interval (CI) is an interval estimate and its function is to define a range of values that has a set probability to contain the ‘true’ parameter value. It can also be said that the function of a CI is to associate a measure of accuracy to a point estimate. The CI is not a test statistic. However it does depend on a choice of test size, and we can use the CI to set a decision rule.

Following Casella and Berger [3] we define an interval estimate for a real valued parameter θ as any pair of functions $L(x_1, \dots, x_n), U(x_1, \dots, x_n)$ of a sample \mathbf{x} that satisfies the inequality $L(\mathbf{x}) < U(\mathbf{x})$ for all $\mathbf{x} \in \mathfrak{X}$, \mathfrak{X} being the space of possible samples. If a given sample $\mathbf{X} = \mathbf{x}$, is observed, then it is possible to make the inference $L(\mathbf{x}) \leq \theta \leq U(\mathbf{x})$. The random interval $[L(\mathbf{x}), U(\mathbf{x})]$ is then defined as an interval estimator.

The notation used so far implies a closed interval. It is in fact possible to state the CI as a one-sided interval if we consider either $L(\mathbf{x}) = -\infty$ or $U(\mathbf{x}) = +\infty$, so that the CIs are represented by the intervals $[-\infty, U(\mathbf{x})], (L(\mathbf{x}), +\infty]$. In these cases the definition would be accordingly adjusted as $\theta \leq U(\mathbf{x}), L(\mathbf{x}) \leq \theta$.

One of the purposes of defining an interval estimate is to have some degree of certainty that the parameter θ is included in the interval. To provide a formal basis for this concept we supply these additional definitions:

1) The coverage probability for an interval estimator $[L(\mathbf{x}), U(\mathbf{x})]$ is the probability that the random interval $[L(\mathbf{x}), U(\mathbf{x})]$ includes the parameter θ . The formal definition for the coverage probability is $P_\theta(\theta \in [L(\mathbf{x}), U(\mathbf{x})])$ or $P(\theta \in [L(\mathbf{x}), U(\mathbf{x})]|\theta)$.

2) Furthermore, we define the confidence coefficient for an interval estimator $[L(\mathbf{x}), U(\mathbf{x})]$, as the infimum of the coverage probabilities, e.g. $\inf_\theta P_\theta(\theta \in [L(\mathbf{x}), U(\mathbf{x})])$.

The interval estimator and a confidence coefficient represent a Confidence Interval. To the extent that we can establish a confidence coefficient at a given value α , we can talk of a Confidence Interval (or Confidence Set) $1 - \alpha$. Now we discuss a common mode of calculating a CI: as inversion of a test statistic.

Following Rice [21], let us assume that for every $\theta_0 \in \Theta$ there is a test of size α for the hypothesis $H_0: \theta = \theta_0$. Define the acceptance region of this test as $A(\theta_0)$. Then the set $C(\mathbf{X}) = \{\theta: \mathbf{X} \in A(\theta)\}$ is the $100(1 - \alpha)\%$ confidence set for a given θ . This means that $100(1 - \alpha)\%$ confidence set for θ consists of all θ_0 for which the $H_0: \theta = \theta_0$ will not be rejected at the significance level α .

This statement can be proved as follows. If $A(\theta)$ is the acceptance region of a test of size α , then the probability that the sample \mathbf{X} belongs to the acceptance region given the Null Hypothesis is $1 - \alpha$, e.g. $P(\mathbf{X} \in A(\theta_0)|\theta = \theta_0) = 1 - \alpha$. However by the definition of $C(\mathbf{X})$ this probability is also equal to the probability that θ_0 belongs to the confidence set $C(\mathbf{X})$ so that:

$$P(\theta_0 \in C(\mathbf{X})|\theta = \theta_0) = P(\mathbf{X} \in A(\theta_0)|\theta = \theta_0) = 1 - \alpha \quad (2)$$

This duality between hypothesis testing and CI can be shown also starting from the definition of confidence set. Let $C(\mathbf{X})$ be a $100(1 - \alpha)\%$ confidence set for θ . Then, for every θ_0 , the probability of θ_0 to belong to the confidence set is $P(\theta_0 \in C(\mathbf{X})|\theta = \theta_0) = 1 - \alpha$. It follows that an acceptance region for a test of size α for the $H_0: \theta = \theta_0$ is $A(\theta_0) = \{\mathbf{X}|\theta_0 \in C(\mathbf{X})\}$.

This means that the $H_0: \theta = \theta_0$ is not rejected if θ_0 belongs to the confidence set. This is proven again by definition. As $A(\theta)$ is defined by all samples \mathbf{X} such that $\theta_0 \in C(\mathbf{X})$, then:

$$P(\mathbf{X} \in A(\theta_0)|\theta = \theta_0) = P(\theta_0 \in C(\mathbf{X})|\theta = \theta_0) = 1 - \alpha \quad (3)$$

The duality between a CI and the acceptance region means that we can start from any test, such as a Z-Test or t-test. Then, by inverting the test, e.g. defining an interval in terms of the T or Z statistic and the chosen test size α , we can derive the equivalent Confidence Interval.

Thus a decision rule can be set to reject the Null Hypothesis $H_0: \theta = \theta_0$ simply by checking whether the random CI $C(\mathbf{X})$ is not inclusive of the θ_0 .

3.4 Bayesian Hypothesis Testing

In the Bayesian approach the key consideration is that θ is a random variable with its own probability distribution. Once we have observed the data, we can compute the (posterior) probability that the Null Hypothesis is true. Thus the Bayesian approach requires the specification of the probability for each hypothesis to be true before analysing the data. These are the Prior Probabilities $P(H_0), P(H_1)$. Then we re-evaluate these probabilities given the analysis of the data x , to produce the Posterior Probabilities, applying Probability Theory, as follows:

$$P(H_1|x) = \frac{P(H_1, x)}{P(x)} = \frac{P(x|H_1)P(H_1)}{P(x)} \quad (4)$$

$$P(H_0|x) = \frac{P(H_0, x)}{P(x)} = \frac{P(x|H_0)P(H_0)}{P(x)} \quad (5)$$

where $P(H_1, x), P(H_0, x)$ are defined with respect to the parameter space for continuous r.v.:

$$P(H_0, x) = \int_{\theta_0} p(x|\theta)dp(\theta) \quad (6)$$

$$P(H_1, x) = \int_{\theta_1} p(x|\theta)dp(\theta) \quad (7)$$

The integrals would be replaced by summation for discrete r.v. These probability distributions are based on the parameter θ , and, in the case of the prior distribution $p(\theta)$ they express beliefs of the researchers (with different degrees of evidence) on the distribution of θ . The conditional probabilities can be re-written as:

$$P(H_1|x) = P(\theta \in \Theta_1^c|x), P(H_0|x) = P(\theta \in \Theta_0|x) \quad (8)$$

The degree by which the posterior probability for the alternative hypothesis H_1 is greater than the posterior probability for the Null Hypothesis H_0 is used to decide in favour of one hypothesis rather than the other. For example, we could use a very simple rule to reject the Null Hypothesis:

$$\text{Reject } H_0 \text{ if } P(\theta \in \Theta_0^c | x) > P(\theta \in \Theta_0 | x) \Rightarrow \frac{P(\theta \in \Theta_0^c | x)}{P(\theta \in \Theta_0 | x)} > 1 \quad (9)$$

We could make the test more stringent by choosing a $c > 1$, as follows:

$$\frac{P(\theta \in \Theta_0^c | x)}{P(\theta \in \Theta_0 | x)} > c \quad (10)$$

where c is a threshold agreed in advance.

The main advantage of the Bayesian approach is that we can associate to the parameter estimates a probability that they are a true expression of the statistical process that we analyse. This advantage comes at a cost. Probability theory requires the adoption of prior probability. This implies a subjective testing procedure. The result of the test depends on the choice of prior probabilities for the hypothesis, and such choice may be more or less properly justified.

3.5 Posterior Probabilities, Credible Intervals, Bayesian Factors (BFs)

In section 3.3 we have already discussed the idea of using posterior probabilities as a possible test statistic to be used for a decision rule. We now briefly discuss these probabilities and other derived test statistics. Also, we use the terms Null and Alternative simply to state two competing hypothesis concerning θ , without adopting the Frequentist view of θ as a fixed parameter value.

3.5.1 Posterior Probabilities

Bayesian Analysis produces a posterior density function (p.d.f.) (or posterior mass function p.m.f.) of the parameter θ :

$$p(\theta | x) = \frac{p(x | \theta)p(\theta)}{p(x)} \quad (11)$$

The information contained in the posterior distributions makes it possible to assess the probability that the true parameter is higher or lower than an assigned value θ_0 . Thus we can define for each θ also a $P(\theta > \theta_0 | x)$. If the researcher requires the probability $P(\theta > \theta_0 | x)$ to be greater than a pre-set value c , then we have a decision rule:

$$\text{Reject } H_0 \text{ if } P(\theta > \theta_0 | x) > c$$

For example, a researcher may require $P(\theta > \theta_0 | x) > 95\%$. Or researchers may take a stance in which no decision rule is adopted and the researcher simply reports the $P(\theta > \theta_0 | x)$, just as a researcher may report the p-values as a sliding scale measure of strength of evidence.

The advantage of a Bayesian Posterior Probability over a p-value is that the former is a probabilistic statement about θ while the p-value is not. It must also be noted that setting a Null Hypothesis as $H_0: \theta = \theta_0, H_1: \theta \neq \theta_0$ is not meaningful in a Bayesian Perspective for a parameter with a continuous prior. Only composite Null and Alternative Hypothesis are possible in a Bayesian framework for this type of parameter.

3.5.2 Credible Intervals

Credible Intervals are interval estimates for the parameter θ . We have seen that Confidence Intervals are r.v. that have a probability to include within themselves the population parameter θ , which is fixed. As θ is a r.v. from a Bayesian perspective, we can state that θ has a probability to be included in the Credible Interval. This statement is however conditional on the data, and, as it is based on a posterior distribution, on the stated prior. A formal definition is provided in Casella and Berger [3]:

Given the posterior probability distribution $p(\theta|\mathbf{x})$ of θ given an observed sample $\mathbf{X} = \mathbf{x}$, then for any set $A \subset \Theta$, Θ being the parameter space, the credible probability of A is:

$$P(\theta \in A|\mathbf{x}) = \int_A p(\theta|\mathbf{x})d\theta \quad (12)$$

and A is the credible set for θ . We may set the credible interval to be a posterior interval such that we have a $100(1 - \alpha)\%$ probability for θ to be included in such interval.

In practice we build the credible intervals by taking the appropriate quantiles of the posterior probability. Therefore, we may decide to set the credible interval for a 90% (or 95%) probability by taking the 5% (2.5%) quantile for the lower bound and the 95% (97.5%) quantile for the upper bound i.e. the $\frac{\alpha}{2}$ and the $1 - \frac{\alpha}{2}$ quantiles.

A variant of the credible interval is the Highest Posterior Density (HPD) interval. This interval has an additional restriction compared to the credible interval. It imposes for the density of the distribution to be always greater than the distribution density outside the interval. In unimodal, symmetrical posterior distributions the two type of interval statistics are identical. Credible intervals are usually preferred because they are easier to compute and can be directly based to the $\frac{\alpha}{2}$ and the $1 - \frac{\alpha}{2}$ quantiles.

3.5.3 Bayes Factors

This section is broadly based on the discussion of the Bayes Factors (BF) by Lesaffre and Lawson [16]. The BFs have been developed by Jeffreys as a test statistic. It is based on the concept of prior and posterior odds. BFs can be used to either test the evidence for the rejection (or acceptance) of the Null Hypothesis, or to compare alternative models. As such they may be seen as the Bayesian equivalent of the Likelihood Ratio Test. In our case we want to use the BF as test statistic for the Null Hypothesis. Recalling the equations in section 3.4:

$$P(H_1|\mathbf{x}) = \frac{P(H_1, \mathbf{x})}{P(\mathbf{x})} = \frac{P(\mathbf{x}|H_1)P(H_1)}{P(\mathbf{x})} \quad (13)$$

$$P(H_0|\mathbf{x}) = \frac{P(H_0, \mathbf{x})}{P(\mathbf{x})} = \frac{P(\mathbf{x}|H_0)P(H_0)}{P(\mathbf{x})} \quad (14)$$

$$\frac{P(H_0|\mathbf{x})}{P(H_1|\mathbf{x})} = \frac{P(\mathbf{x}|H_0)}{P(\mathbf{x}|H_1)} \times \frac{P(H_0)}{P(H_1)} \quad (15)$$

As H_0 and H_1 are set as hypothesis that are alternative to each other, we can set $P(H_1|\mathbf{x}) = 1 - P(H_0|\mathbf{x})$ and $P(H_1) = 1 - P(H_0)$. Then:

$$\frac{P(H_0|\mathbf{x})}{1 - P(H_0|\mathbf{x})} = \frac{P(\mathbf{x}|H_0)}{P(\mathbf{x}|H_1)} \times \frac{P(H_0)}{1 - P(H_0)} \quad (16)$$

The ratio of likelihoods on the right hand side of the formula is the Bayes Factor. The equation can be rearranged as follows:

$$\frac{P(\mathbf{x}|H_0)}{P(\mathbf{x}|H_1)} = \frac{P(H_0)}{1 - P(H_0)} \times \frac{1 - P(H_0|\mathbf{x})}{P(H_0|\mathbf{x})} = \left(\frac{P(H_0|\mathbf{x})}{1 - P(H_0|\mathbf{x})} \right) / \left(\frac{P(H_0)}{1 - P(H_0)} \right) \quad (17)$$

Thus the Bayes Factor can be expressed as the odds ratio between the posterior odds for the H_0 and the prior odds for the H_0 . If we are dealing with discrete r.v., the BF can be set up for simple hypotheses for either or both the null and alternative hypothesis, such as $H_0: \theta = \theta_0, H_1: \theta = \theta_1$, or $H_0: \theta = \theta_0, H_1: \theta \neq$

θ_1 . If we deal with θ as a continuous random variable, simple hypothesis are not possible because at least the denominator prior probabilities would be equal to 0.

Gelman et al. [10] note that in principle it may be possible to use a prior density constructed as a mixture of a discrete density with mass at θ_0 and a continuous distribution elsewhere, but such a solution is judged to be “contrived”, in the sense of wanting to apply the simple hypothesis approach in a Bayesian context. If we set up the hypothesis as composite, such as $H_0: \theta \leq \theta_0, H_1: \theta > \theta_0$, then BF can be used both for discrete and continuous variables.

The BF is a very flexible tool and it is commonly used in evaluating Bayesian models. As for the other Bayesian statistics used in hypothesis testing, the BF is dependent on an appropriate selection of priors. Furthermore, the increase in sample size does not affect the prior odds. These are set in advance and only the posterior odds are modified as a consequence of using a larger sample.

Kass and Raftery [14] note that an appropriate choice of priors for estimation may not be appropriate for hypothesis testing. These authors give the example of selecting a non-informative prior for a Bayesian analysis on the basis of having a large sample. The researcher is confident that the large sample will be able to move the posterior in the direction of the data distribution.

From the perspective of the BF, however, a non-informative prior with a large spread implies increasing the evidence in favour of the H_0 , because a greater fraction of the prior density may be associated to the H_0 , raising the prior odds in its favour. Kass and Raftery [14] suggest that from a hypothesis testing perspective, the priors should be proper (e.g. the density function has a finite integral) and with a contained spread. It follows that in a situation when informative priors are not available or are not appropriate, the testing of the Null Hypothesis by means of the BF with non-informative priors may give rise to additional hurdles to the rejection of the H_0 .

4 The Data

4.1 Source

We compare the three different methods of analysis by using a set of actual N-of-1 trial records. These N-of-1 trial records have been generated as part of a single RCT that has been carried out in 2014. The RCT tested a treatment for fatigue in advanced cancer. This data was kindly provided by an anonymous source.

An experimental drug was tested against a placebo in the N-of-1 trials. The primary outcome in the trial is an improvement in the Quality of Life as a reduction in fatigue severity, as measured by the Functional Assessment of Cancer Therapy-fatigue subscale (FACIT-F). Each N-of-1 trial consists of three cycles of treatment with each cycle lasting a week. Each cycle is composed of two periods of three days each, for a total six days. The patient is not treated for one day, and then the cycle is started again. In each cycle the treatment order has been randomly assigned.

This design is a fairly common feature of N-of-1 trials. The analysis of this trial data provide a suitable example of the type of problems of analysis and interpretation that may be encountered by statisticians and clinicians.

4.2 The Dataset

In the RCT 15 subjects are reported with a complete set of outcomes. Most patients in the RCT have incomplete data but this is to be expected in the context of terminally ill cancer patients. As we use the data only for comparative purposes between different statistical techniques, we do not make any assumption on whether the data is missing at random or not. The selection of these 15 records is adequate for the purpose of this thesis and it does not require any further assessment for modelling purposes. There are 90 observations (6 reported FACIT-F scores per each subject) in total.

For the same reason, the profile of the subjects (such as age or gender) is not considered to be relevant for the purpose of the analysis. The parameters of interest are therefore restricted to the following variables:

- Subject ID
- FACIT-F score
- Period ID
- Cycle ID
- Treatment

4.3 The Variables

4.3.1 Indicator Variables

Period and Cycle take values 1,2 and 1,2,3 in sequential order respectively. Treatment is a dummy variable where 2 stands for administration of drug and 1 stands for administration of placebo. This is the opposite of the original data. The change has been carried out for ease of interpretation. In the simulation of the data for Chapter 7 the values will be set as 1 for treatment and 0 for placebo for simplicity's sake.

4.3.2 The FACIT-F Score

The FACIT-F score is generated by a 13 items questionnaire² used to gauge chronic fatigue levels in patients affected by a variety of conditions. Each item is measured by a 5 point, Likert Scale score from 0 to 4. The rankings of the 13 items are combined together to produce the FACIT-F score. The score itself ranges from 0 to 52, so the higher the final score the better the quality of life of the subject. The design of the score is such that an increase of 3 points is considered to have Clinical Significance.

² Available at <http://www.facit.org/FACITOrg/Questionnaires>.

4.3.3 The dataset

The trial data is reported as the average FACIT-F score over three days (duration of administration period of either placebo or drug). We have inspected the data with SAS PROC UNIVARIATE. Figure 1 indicates that the placebo observations have a symmetrical distribution, whereas the treatment observations show a skewed distribution to the right for the outcome of the treatment periods, suggesting the possibility that some patients have a better response to treatment than others.

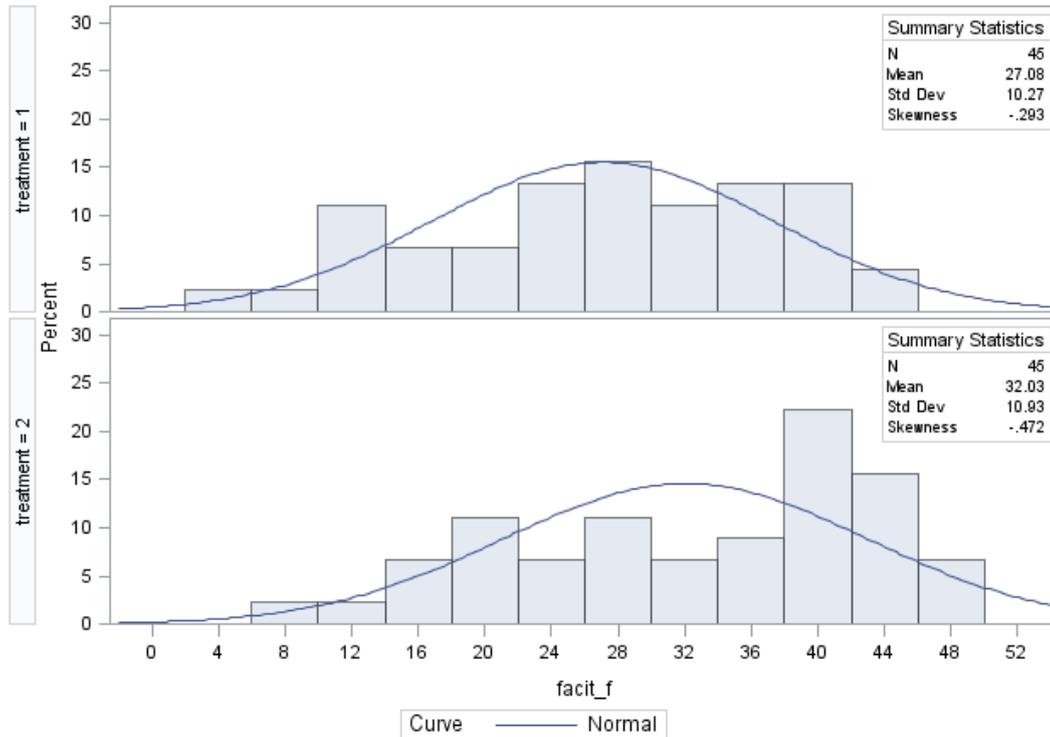


Figure 1: Distribution of FACIT-F score by Treatment (Drug = 2 – Placebo = 1)

Simple visual inspection of the data and of descriptive statistics often provides useful information. For example, the skewness of treatment data shown in Figure 1 suggests that some form of interaction between the individual patient and the treatment effect should be considered in the specification of the statistical model.

5 Analysis of the data: Individual, Trial and Population Perspectives

In chapter 2 we have briefly described the modes and objectives of the analysis which will be used in this chapter. We analyze the data according to three statistical models: FE Linear Model, RE Linear Mixed Model, and a Bayesian Model.

The FE Linear Model is the simplest. It requires minimal assumptions on the model's parameters. This simplicity comes at the expense of efficiency but it minimizes the risk of misspecification. In the FE model we can estimate both the ITEs and the TTE. In the Linear Mixed Model we impose additional assumptions about the parameters. The ITE estimates gain in precision, but this comes at the expense of some bias. The estimates of the ITEs are pulled toward the average treatment effect. An advantage of the mixed model is that it allows estimation of the PTE.

In the Bayesian model further assumptions are added to those considered for the Linear Mixed Model. Prior and hyperprior probabilities are assigned to the model parameters. All parameters may be tested by using the posterior distributions. As in Linear Mixed Model, estimates are affected by shrinkage and are biased towards the PTE. The selection of appropriate priors is not straightforward and may induce variable degrees of bias depending on the nature of the data.

5.1 TTE and ITEs: The FE Linear Model Approach

Our aim is to estimate TTEs and ITEs in an objective manner, that is, only on the basis of the observed data. We have seen in chapter 4 that there is some evidence of a differential effect of the treatment across the subjects. In general, Longford [19] notes that it is not uncommon to observe HTE across individuals for many treatments.

In addition, each subject which has been recruited for the trial may have different baseline values. It follows that the average reported outcome in absence of treatment may vary from subject to subject. This type of heterogeneity may be caused by a variety of factors, which, in our case, cannot be modelled explicitly. We therefore start with a FE linear model that will support the assessment of TTE, ITEs and HTE.

$$y_{i,j} = \alpha_i + \beta_i x_{i,j} + \varepsilon_{i,j} \quad (18)$$

with:

$i = 1, \dots, I = 15$ subjects and $j = 1, \dots, J = 6$ observations for subject

α_i : Individual FE Intercepts

β_i : FE Individual Treatment Effect (ITE)

$x_{i,j}$: Dummy Variable, with $x_{i,j} = 0$ for placebo and $x_{i,j} = 1$ for treatment intervention

$\varepsilon_{i,j}$: Random Error term

This model assumes that:

$\varepsilon_{i,j}$ are i. i. d. r. v., with $\varepsilon_{i,j} \sim N(0, \sigma^2)$

We define the Trial Treatment Effect (TTE) as the average of the ITEs. Each ITE can be defined as the average difference between the effect of the drug and the effect of the placebo in the FACIT-F score of each individual. The TTE is the average, across individuals, of those individual differences.

We ignore period and cycle effects assuming that the wash-out procedures are effective. Also, we assume that the random disturbances are normally distributed, all with equal variance. We can define two

experimental hypotheses. We define the Null Hypothesis for the TTE as a test for an average of the contrasts between intercepts and treatment effects in each individual being equal to 0. In formal terms:

$$TTE H_0: \frac{1}{n} \sum_{i=1}^n \beta_i = 0 \text{ and } H_1: \frac{1}{n} \sum_{i=1}^n \beta_i \neq 0 \quad (19)$$

For each ITE we define the Null Hypothesis as the contrast between the individual intercept and individual treatment effect being equal to 0, so that:

$$ITE H_0: \beta_i = 0 \text{ and } H_1: \beta_i \neq 0 \text{ for a given } i \quad (20)$$

There are four rationales for the use of this model. Firstly, these ITE estimates are unbiased. Secondly, this model can be used in the analysis of the data to provide individual (ITE) as well as group (TTE) estimates, using appropriate sets of contrasts.

Thirdly, this approach is related to Linear Mixed Model Theory. McCulloch et al. [20] note that heterogeneity of effects, at the unit level for which repeated observations are available, can be modelled in two ways: by assuming that those units represent a sampling from a wider population (RE Model), and by treating the same units as a finite experimental set (FE Model) in which each unit is treated as a FE interacting with the Treatment FE.

If we wish to report estimates that are specific only to the trial participants, the model with FE individual treatment effects should be considered as an experimental design in which the subjects participating to the RCT are not a sample from a wider population. Therefore, we can report the estimates for the ITEs and their average, the TTE, to be specific only to the trial subjects.

Fourthly, we use the testing of the Null Hypothesis for TTE as a premise to the estimation of the PTE. Weak evidence in favour of the Null Hypothesis for the TTE necessarily implies weak evidence for PTE (whose estimate has by definition lower precision). It is also possible to consider a simpler model:

$$y_{i,j} = \alpha_i + \beta x_{i,j} + \varepsilon_{i,j} \quad (21)$$

in which we assume that the ITEs are constant across subjects and equal to a common treatment effect (CTE) as β . A H_0 for HTE among subjects can be set as a contrast so that:

$$HTE H_0: \beta_1 = \beta_2 = \dots = \beta_{15} = \beta \text{ and } H_1: \beta_i \neq \beta_j \text{ for some } i, j \quad (22)$$

In a balanced design β is identical to the TTE. The CTE model is more parsimonious, which also implies a greater number of degrees of freedom (DF). We need to estimate only $I + 1$ parameters instead twice the number I of subjects. The sum of squares for the error term will be divided by a greater DF value compared to the model described by equation (18), and the MSE will become smaller. It follows that the F-test for the CTE is more likely to reject the Null Hypothesis compared to test for the contrast used to generate the TTE.

These benefits are offset by the fact that the CTE may miss differential treatment effects. In general it may be wise to run both models so that we can apply an HTE test by comparing the two models.

5.1.1 ITEs Estimates

We have carried out the estimation of the ITE using SAS PROC MIXED. This procedure can perform FE linear model analysis by omitting the *Random* option and setting an appropriate estimation method.

In this case we select the MIVQUE0 specification. In this setting SAS PROC MIXED performs a form of Minimum Variance Quadratic Unbiased Estimation (MIVQUE) of the covariance parameters. MIVQUE is a variation of the MINQUE (Minimum Norm Quadratic Unbiased Estimation) method, as reported in McCulloch et al. [20]. The MIVQUE0 form has been developed internally by SAS³ and it is used in a variety of SAS Procedures.

McCulloch et al. [20] report that MINQUE procedure performs an estimation of the variance component that is independent of the FE like REML. In a balanced design data setting this estimation method produces estimates matching those of ANOVA estimators.

The results of the analysis are displayed in Figure 2⁴ and in Table 2, p. 19. As we can see, the S.E. values for each ITE are quite high relative to their estimates. Individually, we can reject the Null Hypothesis only in three subjects (ID = 21719, 57930, and 71386).

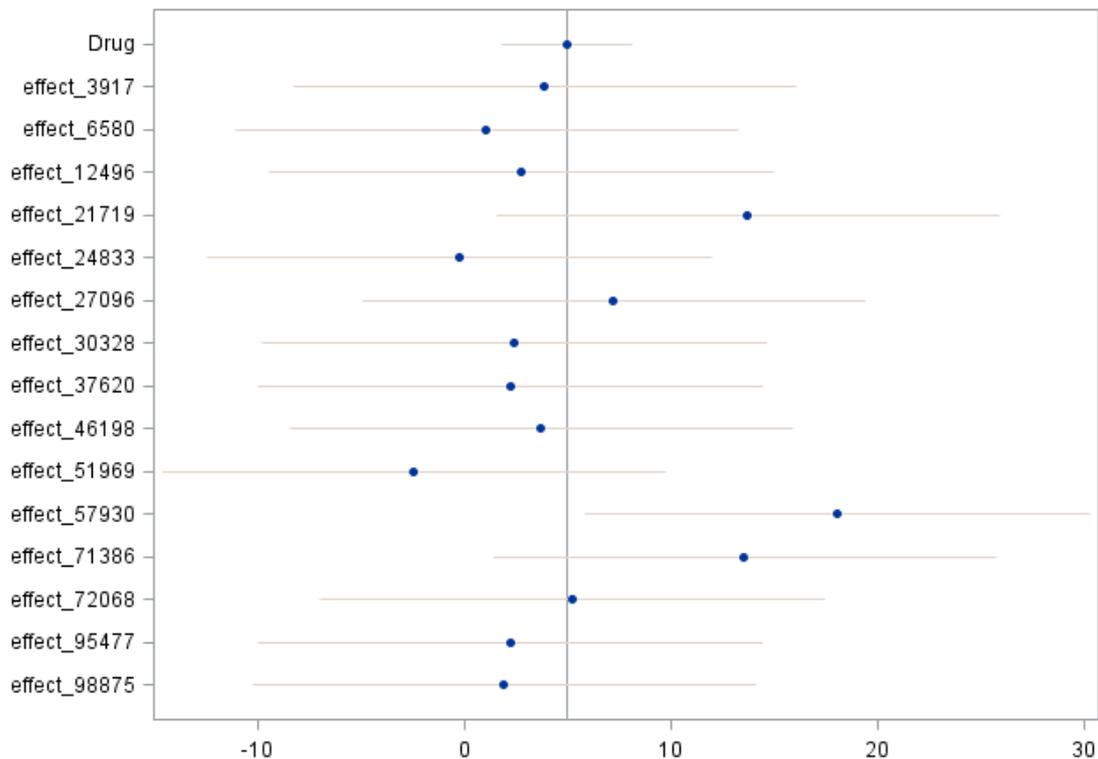


Figure 2: Caterpillar Plot for the FE Linear Model ITEs

We do not require multiple testing restrictions such as Bonferroni in this case. From the individual perspective, the other tests are irrelevant. The Null Hypothesis is concerned with effect of the treatment only in a single subject.

The analysis shows that it would be necessary to increase the individual repeated observations (e.g. additional cycles) to reduce the individual standard errors. However, this approach can become very burdensome for both researchers and patients. In the following sections we show that a RE approach can address this problem.

³As detailed in SAS Technical Report by Goodnight [11].

⁴ This graph is generated with the autocall SAS macro %*CATER*. To use this macro we have used simulated data based on the mean and S.E. estimates of the ITEs.

ID	ITE	S.E.	DF	t Value	Pr > t	Alpha	Lower	Upper
3917	3.8	6.1	60	0.63	0.53	0.05	-8.4	16.1
6580	1.0	6.1	60	0.16	0.87	0.05	-11.2	13.2
12496	2.7	6.1	60	0.44	0.66	0.05	-9.6	14.9
21719	13.7	6.1	60	2.24	0.03	0.05	1.4	25.9
24833	-0.3	6.1	60	-0.05	0.96	0.05	-12.6	11.9
27096	7.2	6.1	60	1.17	0.25	0.05	-5.1	19.4
30328	2.3	6.1	60	0.38	0.70	0.05	-9.9	14.6
37620	2.2	6.1	60	0.35	0.72	0.05	-10.1	14.4
46198	3.7	6.1	60	0.6	0.55	0.05	-8.6	15.9
51969	-2.5	6.1	60	-0.41	0.68	0.05	-14.7	9.7
57930	18.0	6.1	60	2.94	0.005	0.05	5.8	30.2
71386	13.5	6.1	60	2.21	0.03	0.05	1.3	25.7
72068	5.2	6.1	60	0.84	0.40	0.05	-7.1	17.4
95477	2.2	6.1	60	0.35	0.72	0.05	-10.1	14.4
98875	1.8	6.1	60	0.3	0.77	0.05	-10.4	14.1

Table 2: ITE Estimates, FE Linear Model

The analysis indicates that the evidence at individual level is rather weak. That said, the majority of the ITE estimates are positive. This result suggests that the overall treatment effect may well be positive too. Figure 3 shows the scatter plot of the estimated response to the placebo and to the drug for each individual and the scatter plot of the treatment estimates against intercept estimates. These plots suggest that there may be some degree of HTE in this particular sample.

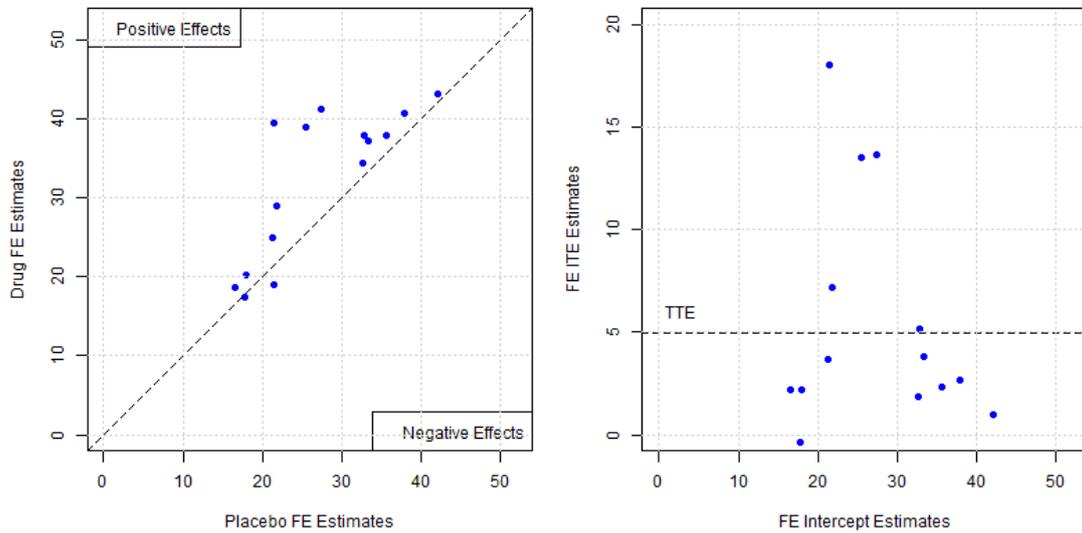


Figure 3: Placebo and Drug Response Relationship

Subjects at the bottom of the scale do not show large improvements (two subjects actually show a decrease in the score). Four subjects with similar placebo results all show a large variation in response. Nevertheless we have seen that the CIs for the ITEs are quite large. HTE may be present but it needs to be formally tested. We will carry out that test in the section 5.1.2.

5.1.2 TTE Estimate

We can estimate the TTE as a contrast. The contrast is set as the average, across individuals, of the difference between the individual intercepts and the individual treatments. The individual intercepts and treatments are represented respectively by the averages, for each individual, of the response to placebo and the response to the drug. The results are displayed in Table 3, with the CTE estimate reported for comparison.

Statistics	TTE as Average ITE	CTE (Constant ITE)
Treatment Effect	4.96	4.96
Standard Error	1.58	1.56
DFs	60	74
T-Value	3.14	3.17
P-Value (2-Tail)	0.003	0.002
CI90% - Upper	8.11	8.07
CI90% - Lower	1.80	1.84

Table 3: N = 15 TTE Estimates - By Modeling Approaches

Table 3 shows that although the TTE estimate has a slightly smaller precision and lower testing power (due to the lower DFs) compared to the CTE estimate, the precision of the TTE estimate is more than adequate. By any standard the evidence for rejecting the Null Hypothesis is very strong and it meets standard clinical guidelines.

It is worth noting that the precision for the estimate of the TTE is the precision for a contrast that includes 15 pairs of estimates (the individual treatment effect for the drug and the placebo effect). The estimate of the CTE is in effect a contrast between only two parameters (the treatment effect for the drug and the placebo effect), controlling for the average response during the trial for each individual.

The t-test shows that within the confines of the trial the selected subjects will on average benefit from a positive treatment effect. Therefore, by pooling the information provided by the N-of-1 trials, it is possible to improve the estimation of the treatment effect across subjects, whilst accounting for variability between subjects. In our case we have strong evidence for an average treatment effect in this sample.

We now test the HTE hypothesis. We do so by considering an F-test between the Residual Sum of Squares (RSS) between the model in equation (18) and the model in equation (21). The model in equation (21) is in fact nested in the model in equation (18), as the latter is based on the interactions between the parameters of the model in equation (21). The Null Hypothesis for the HTE consists of saying that the contribution of the interaction terms on the reduction of the RSS is null. The p-value is 0.577. Therefore there is no evidence for the presence of HTE in this small sample, but we cannot exclude that HTE is present in the general population.

5.2 PTE and the RE Linear Mixed Model Approach

In the previous section we have analyzed the data to estimate both ITEs and the TTE from the internal perspective of the single N-of-1 trials and the RCT as a whole. We now shift the focus on the general population perspectives to estimate the PTE. In this case the ITEs are also estimated in relation to the population perspective.

We estimate the PTE by adding new assumptions. We consider the trial subjects as a random sample of the population of patients in our analysis. To model this assumption we use a Linear Mixed Model (LMM). We estimate PTE that controls for the existence of variance due to random individual intercept and random individual treatment effect. With this mode of analysis we can also estimate more precise (albeit biased) ITEs in comparison to the FE Linear Model.

We propose a LMM with random effects and random parameters, as shown below:

$$y_{i,j} = \alpha + \alpha_i + \beta x_{i,j} + \beta_i x_{i,j} + \varepsilon_{i,j} \quad (23)$$

with:

$i = 1, \dots, 15$ subjects and $j = 1, \dots, 6$ observations for subject

α : Population Mean – Intercept (Fixed Effects)

α_i : Individual Random Intercepts (Random Effects)

β : Population Treatment Effect (Fixed Effects)

β_i : Individual Treatment Random Effects (Random Parameter)

$x_{i,j}$: Dummy Variable, with $x_{i,j} = 0$ for placebo and $x_{i,j} = 1$ for drug

This model assumes that:

α_i : Individual Random Intercepts are i.i.d.r.v. with $\alpha_i \sim N(0, \tau_\alpha^2)$

β_i : Individual Random Treatment term are i.i.d.r.v. with $\beta_i \sim N(0, \tau_\beta^2)$

$\varepsilon_{i,j}$ are i.i.d. r.v., with $\varepsilon_{i,j} \sim N(0, \sigma^2)$

These specifications are fairly conventional. It is worth noting that we may equivalently define $\alpha_i \sim N(\alpha, \tau_\alpha^2)$ and $\beta_i \sim N(\beta, \tau_\beta^2)$. The mean and variance distributional parameters are fixed effects, while the REs are the resulting random variables, instead of being fixed effects.

In this model the PTE is represented by the parameter β . We will show in the following sections that β is a FE population estimate, as it is estimated by taking into account the REs present in the population. The key hypotheses are $H_0: \beta = 0$ and $H_1: \beta \neq 0$. We test the hypothesis that the PTE may not be equal to 0. Subsidiary hypotheses may be set for $H_0: \tau_\beta^2 = 0$ and $H_1: \tau_\beta^2 > 0$, but they will not be discussed here. We do not assume that the random intercept and the random treatment effect are uncorrelated. This point will be discussed in some detail in section 5.2.2. This model specification can model the following factors:

- **Different individual baselines.** Each patient is likely to have different levels of coping with the terminal illness and the side effects of past and current treatments. Different factors such as psychological profile, physical strength and/or weakness, family background may combine in the form of a latent variable that induces a correlation across subject in the general population. This causes different levels of baseline data for different subjects;
- **Individual treatment effects.** We still assume the presence of HTE. It is reasonable to assume, unless proven otherwise, that different patients do not respond to the treatment in the same way. We assume that the new specification is able to detect THE in the form of a treatment RE.

The LMM also allows the explicit estimation of the fixed treatment effect at population level, e.g. a Population Treatment Effect (PTE) and the variance of the random components. The information represented by these two types of estimates is very important from a clinical perspective. The PTE

estimate provides an indication of how large the treatment effect is across the general population on average. The estimate $\hat{\tau}_\beta^2$ provides an indication of the degree of consistency of the treatment effect. The larger the τ_β^2 in the population, the more dispersed the range of responses is to treatment. While the PTE may be positive, the treatment might not be effective or even be potentially harmful for some patients.

We should stress that that this model is more specific compared to the FE Linear Model. We impose a specific distributional structure on the individual effects, that is, there are two individual random effects that follow a normal distribution.

5.2.1 LMM Methodological Basis

We now consider some basic properties of the LMM. We follow the notation and the exposition presented in McCulloch et al. [20]. We define a generic LMM in matrix form as:

$$E(\mathbf{y}|\mathbf{u}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \quad (24)$$

$$\mathbf{u} \sim N(\mathbf{0}, \mathbf{D}) \quad (25)$$

$$\text{var}(\mathbf{y}|\mathbf{u}) = \mathbf{R} \quad (26)$$

where \mathbf{y} is the outcome vector, \mathbf{u} is the random effect vector reflecting our model specification, $\boldsymbol{\beta}$ is the vector of fixed effects, \mathbf{X} and \mathbf{Z} being respectively the design matrix for the fixed and random effects, \mathbf{D} the covariance matrix for the random effects and \mathbf{R} the covariance matrix for the Fixed Effects. We set $E(\mathbf{u}) = \mathbf{0}$ by definition, thus:

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R}) \quad (27)$$

$$\text{var}(\mathbf{y}) = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R} = \mathbf{V} \quad (28)$$

These equations show that: 1) the $\boldsymbol{\beta}$ FEs are represented only in the mean of the distribution of the outcome, and 2) the variance $\mathbf{Z}\mathbf{D}\mathbf{Z}'$ of the random effect is present only in the variance of the outcome. This means that is possible to have mean and variance estimates that are independent of each other. Likewise, the unconditional variance of the outcome \mathbf{y} is composed of two distinct parts: the error term part \mathbf{R} , e.g. the conditional variance, and the random term part $\mathbf{Z}\mathbf{D}\mathbf{Z}'$. Thus it is possible to partition the outcome variance in a fixed component and a random effect component.

This is an important advantage over the FE Linear Model. The partition of the outcome variances in two components indicates that a population characterised by large random effects will show a greater variance compared to a population characterised by relatively small random effects, all other things being equal. In our case, we would expect that if there are no sizable random effects for the treatment, the outcomes for the treatment periods should have a variance equal or smaller than the outcome for the periods under placebo.

The raw data does show in fact a greater variance for observations during the treatment period $\text{var}(FACIT - F) = 119.5$ compared to the placebo period $\text{var}(FACIT - F) = 105.5$. Likewise, the variance of the set of FE ITEs is greater (89.2) than for the FE intercepts (66.4).

A crucial aspect of LMMs is that they define the response mean and variance as being conditional to the realized REs. The consequence of this approach is that the FEs are estimated while controlling for the influence of the random effects on the observed response. Therefore, the variance of the PTE estimates is more realistic (e.g. larger) than the variance of the treatment estimate for a FE Linear Model that does not consider the impact of the REs.

We now approach another important aspect of the LMMs: the estimation (prediction) of \mathbf{u} . As such, the FEs are constants (in the population) that we may wish to estimate. The REs are assumed to be a random selection from an ideal population of effects. To that extent, McCulloch et al. [20] note that it may be preferable to describe the estimation of the REs as a prediction of their expected value conditional to the data available. Such a conditional mean is the best predictor in the sense of minimizing the prediction MSE. Following McCulloch et al. [20], the unconditional variance of the random effect can be expressed as follows:

$$\text{var}(u) = \text{var}(E(u|\mathbf{y})) + E(\text{var}(u|\mathbf{y})) \Rightarrow \text{var}(u) = \text{var}(\tilde{u}) + \text{constant} \quad (29)$$

where $\tilde{u} = E(u|\mathbf{y})$ is the predicted random effect and the constant is the positive value represented by the expected value of $\text{var}(u|\mathbf{y})$. It then follows that the conditional variance of the random effect value will always be less (or at best equal) than the unconditional variance of the random effect itself, that is:

$$\text{var}(\tilde{u}) \leq \text{var}(u) \quad (30)$$

By considering the RE as a random sample, the LMM achieves another important result: to exploit the information provided by all individuals in the sample to predict their realized value. Whereas in the model described by equation (18) the estimate for the ITE of a given i individual depends only on the observations collected for that individual, the precision of the RE predicted values is based on pooling all individual observations to derive an estimate of the variance of the REs.

To obtain the vectors of predicted random effects and their variances we use the following formulas:

$$\tilde{\mathbf{u}} = \mathbf{D}\hat{\mathbf{Z}}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (31)$$

$$\text{var}(\tilde{\mathbf{u}}) = \mathbf{D}\hat{\mathbf{Z}}\mathbf{V}^{-1}\mathbf{Z}\mathbf{D} \quad (32)$$

where $\tilde{\mathbf{u}}$ is the vector of predicted random effect values. These equations show that $\tilde{\mathbf{u}}$ values do not depend on the FEs, but only on the random part of the data: the error term, the random term variances and the REs design matrix. In order to obtain a Best Linear Unbiased Predictors (BLUP) for the $\tilde{\mathbf{u}}$, McCulloch et al. [20] note that it is necessary to replace $\hat{\boldsymbol{\beta}}$ in equation (31) with the ML estimator $\boldsymbol{\beta}^0$.

$$\boldsymbol{\beta}^0 = (\hat{\mathbf{X}}\mathbf{V}^{-1}\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}\mathbf{V}^{-1}\mathbf{y} \quad (33)$$

Thus we write define BLUP $\tilde{\mathbf{u}}^0$ as:

$$\tilde{\mathbf{u}}^0 = \mathbf{D}\hat{\mathbf{Z}}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^0) \quad (34)$$

If the covariance matrix \mathbf{D} , and the unconditional variance matrix \mathbf{V} are unknown, they can be replaced by their estimates $\hat{\mathbf{D}}$, $\hat{\mathbf{V}}$, producing an predicted REs vector $\hat{\mathbf{u}}$ so that:

$$\hat{\mathbf{u}} = \hat{\mathbf{D}}\hat{\mathbf{Z}}\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (35)$$

An important property of the $\tilde{\mathbf{u}}^0$ is that they are orthogonal to the ML FE estimators, so that:

$$\text{cov}(\boldsymbol{\beta}^0, \tilde{\mathbf{u}}^0) = \mathbf{0} \quad (36)$$

This result is important. It means that we can estimate, in principle, the variance of the ITEs estimates simply by summing the estimated variance of the PTE values and the Predicted error variance of the RE values. The estimation of the LMMs FE and RE parameters can be carried out in different ways. The two main methods are the Maximum Likelihood (ML) Estimation and REstricted Maximum Likelihood (REML) Estimation. We restrict our attention to the REML approach because it has an important feature: it estimates the variance component of the model independently of the FEs in a direct manner, unlike, for example, the ML approach, as noted by McCulloch et al. [20]. The implication is that the FEs are modelled separately by ML methods using the results of the estimation of the variance matrix by REML. Furthermore the variance matrix does not vary with the value of $\boldsymbol{\beta}$ by construction. This is an important feature because we will show in the following sections that in the case of small samples the estimation of the covariance matrix for the REs can be difficult. With a REML procedure we can rely on a ‘good’ PTE $\hat{\boldsymbol{\beta}}$ estimate as its derivation will not be influenced by any problem that could arise by the estimation of \mathbf{D} .

According to McCulloch et al. [20], REML maximizes the likelihood of linear combinations of elements of \mathbf{y} . The linear combination is a $\hat{\mathbf{K}}\mathbf{y}$ such that $\hat{\mathbf{K}}\mathbf{y}$ does not include any of the FEs in $\boldsymbol{\beta}$, which means that $\hat{\mathbf{K}}\mathbf{X} = \mathbf{0}$. We need to define a matrix \mathbf{K} composed of the maximum possible number of $\hat{\mathbf{k}}$ vectors orthogonal to \mathbf{X} . This number must be equal to the difference between the number of observations and the rank of the design matrix, $N - \text{rank}(\mathbf{X})$. Thus:

$$\mathbf{K} = [\mathbf{k}_1, \dots, \mathbf{k}_{N-\text{rank}(\mathbf{X})}] \quad (37)$$

The next step is to replace \mathbf{y} in the ML equations with $\hat{\mathbf{K}}\mathbf{y}$. The effect of this substitution is as follows:

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}), \text{ and } \hat{\mathbf{K}}\mathbf{y} \sim N(\mathbf{0}, \hat{\mathbf{K}}\mathbf{V}\mathbf{K}) \quad (38)$$

We consider now the ML equations for the estimation of the variance components:

$$\text{trace}(\mathbf{V}^{-1}\mathbf{Z}_i\hat{\mathbf{Z}}_i) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}\mathbf{Z}_i\hat{\mathbf{Z}}_i\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (39)$$

where $\mathbf{Z}_i\hat{\mathbf{Z}}_i$ is defined as the matrix of partial derivatives of \mathbf{V} with respect to the REs variances. The next step is to replace \mathbf{y} with $\hat{\mathbf{K}}\mathbf{y}$, \mathbf{X} with $\hat{\mathbf{K}}\mathbf{X} = \mathbf{0}$, \mathbf{Z} with $\hat{\mathbf{K}}\mathbf{Z}$, and \mathbf{V} with $\hat{\mathbf{K}}\mathbf{V}\mathbf{K}$ in the equations (39). Then, the estimating equations for the REML variance are:

$$\text{trace}((\hat{\mathbf{K}}\mathbf{V}\mathbf{K})^{-1}\mathbf{Z}_i\hat{\mathbf{Z}}_i) = \hat{\mathbf{K}}\mathbf{y}(\hat{\mathbf{K}}\mathbf{V}\mathbf{K})^{-1}\mathbf{Z}_i\hat{\mathbf{Z}}_i(\hat{\mathbf{K}}\mathbf{V}\mathbf{K})^{-1}\hat{\mathbf{K}}\mathbf{y} \quad (40)$$

As the product $\hat{\mathbf{K}}\mathbf{y}$ does not contain the FE $\boldsymbol{\beta}$, the system of equations (40) is entirely independent of the FEs and can be solved directly for REs variances. Thus the application of orthogonal contrast to the FE component guarantees an independent estimation of the random components and, ultimately, of the REs themselves.

5.2.2 The Analysis

We have analyzed the set of 15 subjects with SAS PROC MIXED, with REML estimation, and *Random statements*: The variance estimation will not be bounded to prevent SAS PROC MIXED to generate an error message and to stop the estimation. We tried two RE Covariance Matrix settings: Diagonal and Unrestricted. The first setting assumes that the error term, random intercept and the random parameters variances are uncorrelated. The second setting allows the random terms (intercept and treatment) to be correlated to each other. As the first setting adds new constraints, we would select an unrestricted matrix. Unfortunately the analysis based on an unrestricted covariance setting returned an estimated RE

covariance matrix with a positive covariance between the random intercept and the random treatment, and, surprisingly, a negative variance for the random treatment.

The SAS PROC MIXED estimation of the random effects has failed, with 0 variances assigned to the predicted random effects and parameters. This problem may be a specific limitation of the SAS PROC MIXED estimation procedure, so that the analysis of small samples with multiple RE implies a risk of degenerate \mathbf{Z} design matrix and consequently of the covariance matrix \mathbf{ZDZ} for the REs.

If simpler models are applied (with either only RE intercepts or only RE treatment effects) the problem does not occur. This is a potential limitation of the analysis of small RCTs based on N-of-1 trials with SAS procedures. A check with the R package *lmer* returns a proper covariance matrix with a positive variance and estimated REs values. The AIC value of the the *lmer* analysis is the same of the SAS model with unrestricted setting. Table 4 compares of the covariance matrix estimates of the three analyses.

	SAS Diagonal		SAS Unrestricted		R lmer Unrestricted	
	Mean	S.E.	Mean	S.E.	Mean	S.E.
$\hat{\tau}_\alpha^2$	46.9	31.6	16.3	44.9	31.6	5.6
$\hat{\tau}_\beta^2$	5.3	9.6	-4.3	14.3	2.1	1.4
$\hat{\sigma}^2$	53.6	9.0	56.1	10.2	54.3	7.4

Table 4: Covariance Matrix Estimates

Having decided to use SAS PROC MIXED for the analysis, we select the diagonal matrix assumption for the purpose of the comparison of results between methods. In addition, the information criteria statistics are more favourable for the diagonal matrix assumption, as shown in Table 5.

Fit Statistics	Diagonal RE Covariance Matrix	Unrestricted RE Covariance Matrix
-2 Res Log Likelihood	637.8	637.3
AIC	643.8	645.3
AICC	644.1	645.8
BIC	645.9	648.1

Table 5: Fit Statistics with Diagonal and Unrestricted RE Covariance Matrix, SAS PROC MIXED

The selected model does show a good fit. The LRT for the Null Model reports a $\chi^2_{df=2} = 35.17$, and a corresponding P-Value lower than 0.0001. We do not perform diagnostic checks on the residuals and misspecification tests as not strictly necessary for the purpose of the thesis.

5.2.3 ITEs

SAS PROC MIXED does not estimate the ITE as the actual total response to treatment for each subject. Therefore this needs to be carried out separately. We define the ITE as the sum of the PTE estimate and the RE treatment estimate. This is a pragmatic choice but is not a straightforward one. The FEs are conditional means relative to the REs, and the REs are conditional means relative to the random components of the data. Thus their linear combination may not necessarily be best represented by a simple sum.

The variance for the ITE estimate is set as the sum of the variances of the PTE and each RE. We do this because the PTE and REs are independent of each other, and their covariance is null. These variances are simply calculated as the sum of the squares of their respective standard error and standard prediction error. Hypothesis testing is not as straightforward. We need to consider what type of distribution the ITE estimates should follow. If we treat the FE and RE estimates as random variables following the Student

T-distribution – which is the standard mode of testing for these values – we would find that the definition of the distribution of their linear combination is a non-trivial task. Berg and Vignat [2] indicate that the resulting distribution is rather complex. The derived ITE estimates and test statistics are shown in Figure 4 and in Table 6, p.30.

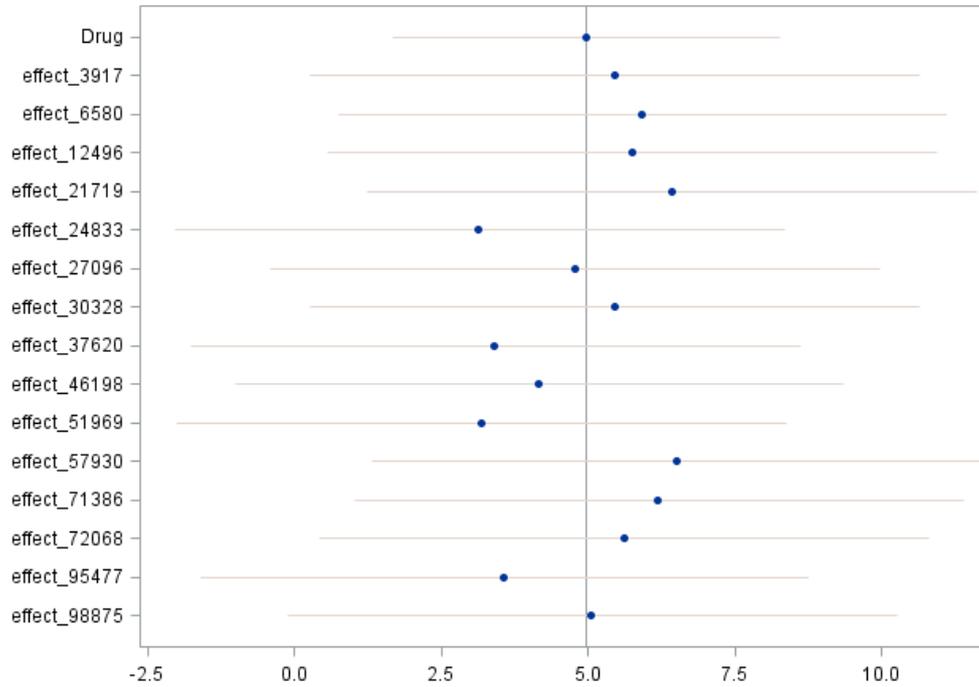


Figure 4: Caterpillar Plot for the Linear Mixed Model ITEs

As a compromise the ITE are tested using a Z-test and a CI based on the Normal Distribution. The resulting tests are optimistic by definition and need to be treated with some caution.

ID	ITE	S.E.	z Value	Pr > z	Alpha	Lower	Upper
3917	5.4	2.6	2.09	0.04	0.05	0.3	10.5
6580	5.9	2.6	2.27	0.02	0.05	0.8	11.0
12496	5.7	2.6	2.21	0.03	0.05	0.6	10.8
21719	6.4	2.6	2.47	0.01	0.05	1.3	11.5
24833	3.1	2.6	1.20	0.23	0.05	-2.0	8.2
27096	4.8	2.6	1.83	0.07	0.05	-0.3	9.9
30328	5.4	2.6	2.09	0.04	0.05	0.3	10.5
37620	3.4	2.6	1.30	0.19	0.05	-1.7	8.5
46198	4.1	2.6	1.59	0.11	0.05	-1.0	9.2
51969	3.2	2.6	1.22	0.22	0.05	-1.9	8.3
57930	6.5	2.6	2.49	0.01	0.05	1.4	11.6
71386	6.2	2.6	2.38	0.02	0.05	1.1	11.3
72068	5.6	2.6	2.15	0.03	0.05	0.5	10.7
95477	3.5	2.6	1.37	0.17	0.05	-1.5	8.6
98875	5.0	2.6	1.94	0.05	0.05	-0.1	10.1

Table 6: ITE estimates and Test Statistics

Granted the limitations of this approach, it can be noted that the ITE estimates of the Mixed Model are reasonably precise, with a 2.6 points standard error.

The Z-test and the CI95% suggest that there is good evidence for stating that the majority of the patients in the trial would benefit from the treatment. For the remainder of the patients the evidence is weaker, with varying degree of confidence in the total effect of the treatment.

However this statement must be qualified by the fact that the ITE estimates have shrunk to some degree towards the PTE estimate. This is a consequence of having imposed a normal distribution on the REs. In the FE model the individual effect are fixed parameters specific to each individual, thus they are free to take any value. In the linear mixed model this is not the case. We have imposed a normal distribution which confines their range. This is a form of shrinkage, whereby we trade bias for precision in the estimation of the ITEs.

5.3.2.1 Feasibility of TTEs based on Mixed Model ITEs

Given the availability of ITEs, it is possible to estimate a TTE value in the same way as for the FE Linear Model. However, there are some problems from an interpretative point of view, as they are composed of a Population Level estimate and individual predicted RE. Further, the linear combination represented by the average of these values cannot offer a greater precision than the actual PTE value itself, which is an estimate at population level. In fact, the variance of this average must be a composite of the PTE estimate variance and the variance of the REs predicted values.

5.2.4 Estimates for the PTE and Variance Components

The Mixed Model PTE and variance estimates are shown in Table 7 and Table 8.

Parameter	Estimate	S.E.	DF	t Value	Pr > t	Alpha	Lower	Upper
Intercept	22.12	3.01	14	7.34	<.0001	0.05	15.7	28.6
Treatment	4.96	1.66	14	2.99	0.01	0.05	1.4	8.5

Table 7: PTE Estimates for Mixed Model

The PTE is equal to the TTE estimate. It is marginally less precise, this being a Population level estimate. The precision of the estimate is however sufficient to reject the Null Hypothesis for the PTE. According to the model, and for this restricted set of data, there would be strong evidence of a positive average treatment effect across the population of patients.

Covariance Parameters	Subject	Estimate	S.E.	Z-Value	Pr Z
Intercept	ID	46.9	31.6	1.48	0.14
Treatment	ID	5.3	9.6	0.56	0.58
Residual		53.6	9.0	5.94	<.0001

Table 8: Variance Components for Mixed Model

The interpretation of the evidence is less straightforward for the Random parameter variance τ_{β}^2 . SAS PROC MIXED treats the variance components estimates as mean estimates, normally distributed, and provides a two-tailed Z-test for the Null Hypothesis of Null Random Effects variance.

This approach has obvious limitations, especially with regards to small variance estimates and small samples. Therefore the results shown in Table 8 should be treated with care. We note that there is a sizeable variance component due to random effects for the intercept. The treatment RE has instead a small estimate and, as we have noted earlier on, estimation in R yields an even smaller estimate. This is

consistent with the graphical inspection carried out in section 4.1, and the assessment of FE estimates in section 5.1.

Most of this random effect variance is attributed to the random intercepts. From the point of view of a clinical researcher, this is a positive result. As the treatment RE variance is comparatively small, it can be assumed that the effect of the treatment across the general population will be reasonably consistent.

Nevertheless the size of the sample cautions against taking these results as definitive. Longford [19] notes that sampling bias may lead to underestimate treatment heterogeneity. For example, as could be in this case, patients completing their treatment cycles in an RCT may be potentially a biased sample of the totality of patients. They may be more homogeneous and generally skewed in their response to treatment in comparison to the general population. Thus we could draw misleading conclusions concerning the consistency of the treatment effect across subjects on the basis of small samples. This point is further discussed in the following chapters.

5.3 The Bayesian Model Approach

In the Bayesian model we add an additional set of assumptions compared to the LMM. Although we still maintain that REs are normally distributed, the means and variances of the random terms are interpreted as r.v. in their own right with a probability distribution. In addition we also state that the error term variances are random variables with their own distributional assumptions.

This set of assumptions is known as the priors and hyperpriors of the model. Although these additional assumptions confer greater flexibility to the estimation of the parameters, the model is now sensitive not only to the information provided by the data, but also the information stated by these distributional assumptions as represented by the priors and hyperpriors. Therefore we need a careful selection of distributions for the parameters in the model, to prevent poor model fit and biased estimation of parameters.

We will not discuss in detail the theoretical properties of Bayesian Regression Analysis for grouped data, which is sometimes defined as Hierarchical Bayes Analysis, as in Gelman et al. [10]. We will point out some key features in comparison with the LMMs.

First, in the Bayesian Model we set the following priors $\alpha_i \sim N(\alpha, \tau_\alpha^2)$, $\beta_i \sim N(\beta, \tau_\beta^2)$, as in the equivalent description given for the LMM in section 5.2. We are specifying for each subject i in this way the equivalent of REs for intercept and treatment. In addition, by stating the α_i , β_i belong to two populations the Bayesian model needs to be able to predict new possible α_i , β_i values. This is exactly the same approach of the Mixed Models. The REs are predicted rather than estimated, as they are random realizations instead of population parameter estimates.

Second, as noted by Gelman et al. [10], the Bayesian model for grouped data produces a posterior distribution for the individual α_i, β_i parameter such that their posterior means, conditional to the priors and the data, are a precision-weighted average of the prior population and the sample mean of the i^{th} group of placebo observations $\bar{y}_{i,placebo}$. For example, the random effect estimate $\hat{\alpha}_i$ is thus defined as:

$$\hat{\alpha}_i = \frac{\frac{\hat{\alpha}}{\hat{\tau}_\alpha^2} + \frac{\bar{y}_{i,placebo}}{\hat{\sigma}^2}}{\frac{1}{\hat{\tau}_\alpha^2} + \frac{1}{\hat{\sigma}^2}} \quad (41)$$

It is evident that if τ_α^2 is very large, then $\hat{\alpha}_i$ converge on the group mean $\bar{y}_{i,placebo}$, whereas if σ^2 is very large then $\hat{\alpha}_i$ will converge on the population parameter α . As generally the $\tau_\alpha^2 < \sigma^2$, then $\hat{\alpha}_i$ will shrink towards the population parameter. The same shrinkage effect applies to $\hat{\beta}_i$. Thus, while we ‘borrow’

information across the subjects in the form of the hyperprior specification, and we improve precision, we pay for this increase in precision with estimates' bias.

Third, the definition of the priors always implies setting the estimation under definite constraints. Priors are conventionally divided between informative (e.g. with high probabilities associated to specific parameter values) and non-informative (with relatively even probabilities for a range of values). Nevertheless many priors that are considered to be non-informative are also making a specific statement such that it influences the estimation of the posterior. Gelman et al. [10] note that uniform priors, conventionally known as non-informative, may create problems of their own in different cases.

Fourth, Gelman et al. [10] note that if we were to use the actual data to estimate the prior parameters, we would perform what is called an Empirical Bayes analysis. However, as noted by McCulloch et al. [20], this method of estimation is what is actually carried out to estimate the REs in a Mixed Model. Therefore, in practice, a Mixed Model analysis relies on a form of Bayesian Analysis.

The actual computation of the model parameters can, as for any other type of Bayesian Modeling, be carried out algebraically by using conjugate priors and hyperpriors or by simulation (Markov Chain Monte Carlo, MCMC). In our case, we will use conjugate priors. In addition we calculate the Intra Class Correlation (ICC) and the ITEs as a combination of the posterior distributions for the RE variance components and the FEs and REs, respectively.

5.3.1 The Analysis

We have analyzed the data using SAS PROC MCMC. We start by outlining the specifications for the model. We define the Bayesian model as follows:

Priors

$$\text{Random Intercept } \alpha_i \sim N(\alpha, \tau_\alpha^2)$$

$$\text{Random Treatment } \beta_i \sim N(\beta, \tau_\beta^2)$$

Hyperpriors

$$\text{Intercept } \alpha \sim N(26,8)$$

$$\text{Treatment } \beta \sim N(0,8)$$

$$\text{Error Term Variance } \sigma^2 \sim U(0,100)$$

$$\text{Random Intercept Variance } \tau_\alpha^2 \sim U(0,100)$$

$$\text{Random Treatment Variance } \tau_\beta^2 \sim U(0,100)$$

Mean Definition

$$\mu = \alpha_i + \beta_i x_{i,j}$$

$x_{i,j}$: Dummy Variable, with $x_{i,j} = 0$ for placebo and $x_{i,j} = 1$ for drug

$$\text{Outcome Distribution: FACIT - F} = y_{i,j} \sim N(\mu, \sigma^2)$$

The Null and Alternative Hypothesis are defined as composite hypothesis, e.g. $H_0: \beta \leq 0$ and $H_1: \beta > 0$.

In SAS PROC MCMC it is possible, for a model, to specify the random term explicitly, and this is an equivalent specification for the hyperprior. Thus the SAS code we will not follow strictly the specification given above. In addition, the computation is faster if the ‘random’ statement is used instead of the hyperprior estimate.

As noted above, we need to model seven parameters: two FE intercept α and treatment β , two REs for the intercept α_i and the treatment β_i , and three variance parameters τ_α^2 , τ_β^2 , σ^2 , that is the variance of the random intercept, the variance of the random treatment and the variance of the error term.

For all these parameters we would like to apply priors that, given the available data and the nature of the model, pose minimal risk to produce unstable or biased estimates. In this case we use conjugate priors to facilitate the estimation of the posteriors. We avoid using the terms non-informative, vague, or weak priors because depending on the circumstances they may still considerably skew the estimation of the data.

According to Gelman et al. [10], it is advisable to gain some degree of information on the plausible values of hyperprior parameters to constrain these values in a finite region. In our case, we would like to consider the extent of variability that can be reasonably be expected for the FACIT-F score (which is bounded) to determine a plausible range of values for the variance component distribution. Thus we will justify the selection of priors in each case in terms of our research context.

FE Intercept Prior. We set as to the intercept prior a normal distribution centred on the middle of the range of the FACIT-F score (which is 26) and a variance that is consistent with the theoretical range of the score. By simple inspection we select a standard deviation of 8 and hence a variance value of 64^5 . We would in fact expect the sample (and the population at large) to have a FE intercept approximately in the middle of the range, with a Standard Deviation (SD) that cannot exceed the set range for the score. By using a normal distribution we can place most of the prior’s mass near the mid-point. This means that we expect the intercept value to be a close value to the mid-point. The hyperprior is thus set as $\alpha \sim N(26,8)$.

FE Treatment Prior. For the FE treatment prior we should adopt a neutral stance concerning the mean. This stance may be described as Clinical Equipose, that is, assigning equal chances to have either positive or negative effects. If we decide to apply a normal distribution, however, we would also implicitly be imposing a sceptic view for the prior, insofar that the prior mass would be placed near the 0 value. We should thus assume low absolute values for the treatment FE rather than higher ones.

An alternative would be a Uniform (flat) prior, with 0 as the mid-point. The prior mass is distributed evenly over the selected range of values, thus making each value equally likely. A less extreme view may be represented by a distribution with moderate to high kurtosis, that is, one with ‘fat’ tails and a prior mass less concentrated near the mode (for example the T-Student distribution). We decide to select a normal prior (e.g. a somewhat sceptic approach) to check if the data has sufficient information to update the prior ‘away’ from the 0. The hyperprior is thus set as $\beta \sim N(0,8)$, with a variance of the same order of the intercept.

Variance Components Priors. Gelman [9] has discussed a variety of approaches to select variance priors for models. Following his suggestion we have selected a Uniform prior because it is less prone to unstable estimation in case there are few grouping units. More common choices, such as the Inverse gamma with near zero scale and shape values suggested by Spiegelhalter et al. [25], are judged by Gelman [9] to be less appropriate for those reasons. We thus select the error variance prior $\sigma^2 \sim U(0,100)$ and variance hyperpriors $\tau_\alpha^2, \tau_\beta^2 \sim U(0,100)$.

⁵ A SD = 8 implies that the chance of exceeding the range [0,52] is less than 0.12% .

The upper limit of the range for the uniform is lower than then variance for the treatment and placebo FACIT-F scores, as shown in Figure 1, p. 19. We also take into account that the sum for the Mixed Model random term variances is approximately 100, as shown in Table 8, p. 31. We set these priors and hyperpriors on the same scale to let them free to assume the most likely value on the basis of the data available. The analysis also requires choosing analytical setting to achieve reliable estimates. These are:

- a long Markov Chain sequence (1M), to ensure a conveniently stable estimate;
- set the burn-in to the maximum value (50,000) to prevent bias due to autocorrelation in the initial chains;
- a high thinning parameter (200), to ensure very low autocorrelation values in the sampling of the posterior distribution and obtain $N = 5,000$ posterior sample, which is well above the minimum for some diagnostic tests on the posterior parameters ;
- a covariance proposal initial estimate based on the quasi Newton-Raphson method, as per advice available in the SAS guide on SAS PROC MCMC.

We also set the SAS PROC MCMC to export the posterior distribution sample ($N = 5,000$) for the model's five parameters and for the thirty REs estimates (Intercept and Treatment for each subject). This makes it possible to estimate the ITEs simply as the sum of two posterior distributions: the posterior for the Treatment FE and, for each ITE, the posterior of each Treatment RE. The precision estimates and test statistics for the ITEs can be derived empirically.

Having set up the analysis on the basis of these choices, we have run SAS PROC MCMC on the data. The estimation took 3.5 minutes on a Intel Dual Core, 64 Bit machine with Clock Speed of 2.5 GHz. The OS is Windows 7.

The analysis included a number of diagnostic tests and graphics to check for the convergence of the estimates. Every parameter was shown to converge on the Geweke test and to be not affected by autocorrelation. The only measure of fit directly available is the Deviance Information Criterion (DIC). The DIC is described by Gelman et al. [10] as a Goodness-of-Fit statistic that approximates the expectation of the predictive deviance. The DIC is defined as:

$$DIC = 2 \left(\frac{1}{L} \sum_{l=1}^L -2 \log p(y|\theta_l) \right) - 2 \log p(y|\hat{\theta}(y)) \quad (42)$$

where the θ_l are posterior simulations, and $\hat{\theta}(y)$ can be a point estimate derived as the mean of the posterior simulations. In particular, the first term is the posterior mean deviance, whereas the second term is the deviance estimate on the $\hat{\theta}$ posterior mean. The value for our analysis is $DIC = 633.5$. We now turn to the results. First, we show the checks on the posterior of key parameters $\beta, \tau_\alpha^2, \tau_\beta^2, \sigma^2$ as shown in Figure 5 p. 36.

We observe that of the four parameters, τ_β^2 , is the one whose convergence plot is more 'jagged'. This is necessarily the case as its estimates are closer to the lower boundary of the distribution. All four parameters show that the autocorrelation between estimates is eliminated relatively quickly. Figure 6, p. 36 shows acceptable results also for the ICC value.

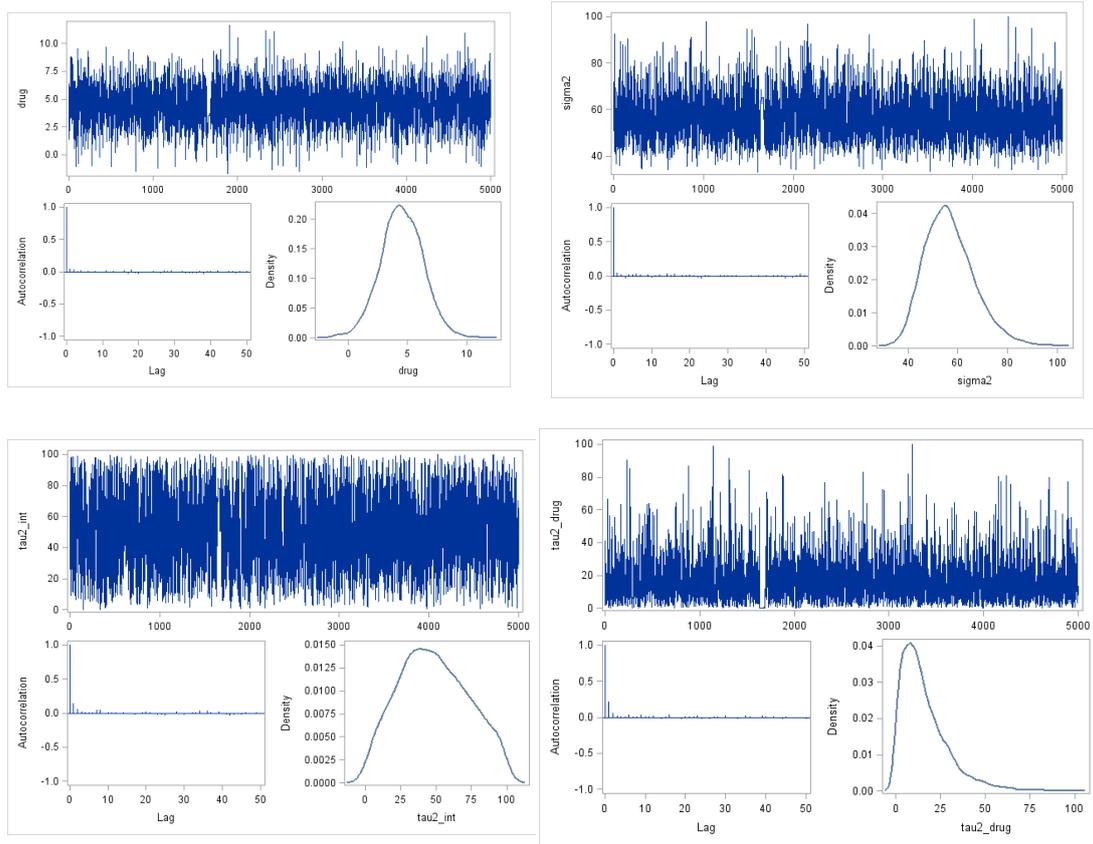


Figure 5: Posterior Diagnostic Graphs for PE and Variance Components (N = 5,000)

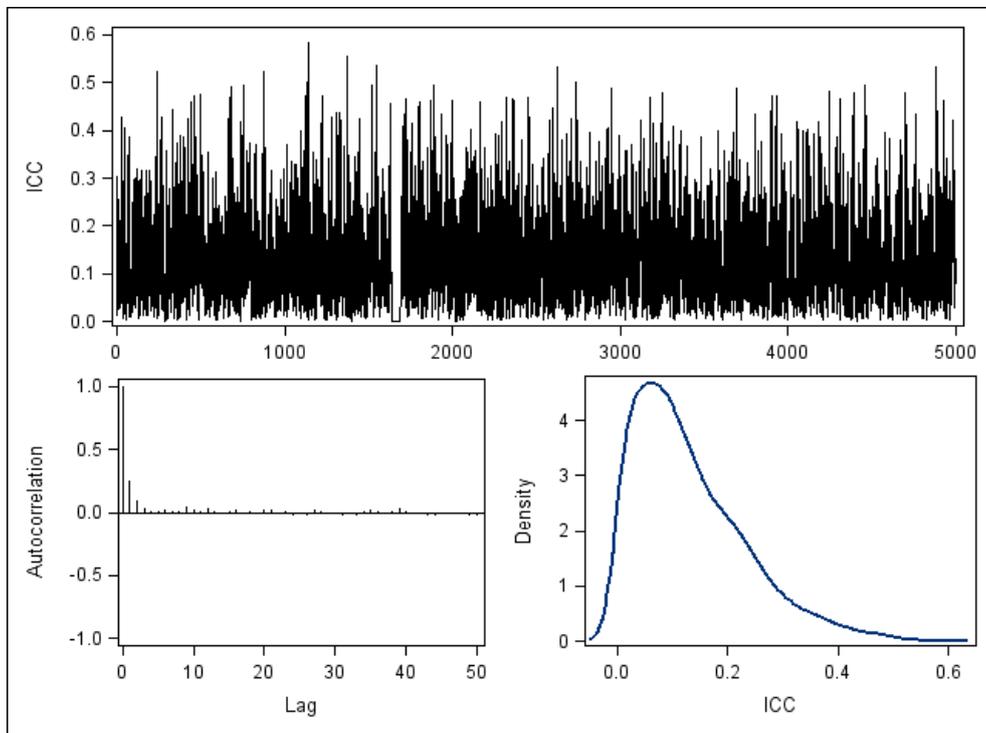


Figure 6: Posterior ICC Diagnostic Graphs for PE and Variance Components (N = 5,000)

We perform a check on the degree of correlation between the parameters in Table 9:

Pearson Correlation Coefficients, N = 5,000					
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\tau}_{\alpha}^2$	$\hat{\tau}_{\beta}^2$
$\hat{\alpha}$	1				
$\hat{\beta}$	-0.622	1			
$\hat{\sigma}^2$	0.014	-0.022	1		
$\hat{\tau}_{\alpha}^2$	0.025	0.025	-0.064	1	
$\hat{\tau}_{\beta}^2$	-0.020	-0.013	0.014	-0.250	1

Table 9: Correlation of Parameters Posterior Distributions

The Bayesian analysis indicates that a moderate negative correlation does exist between the FE intercept and the treatment parameters. The random effect variances are both weakly negative and significantly correlated. This shows that the unrestricted approach for LMM was the correct one. In this sense the Bayesian approach in SAS PROC MCMC has performed better than SAS PROC MIXED, because it has accepted the correlation between the two random term variances without difficulty.

5.3.2 ITEs

The ITEs and the Treatment PTE (labelled ‘drug’) are displayed together in Figure 7⁶ and in Table 10, p. 38. The PTE value is closely aligned to the average ITEs value, as one would expect.

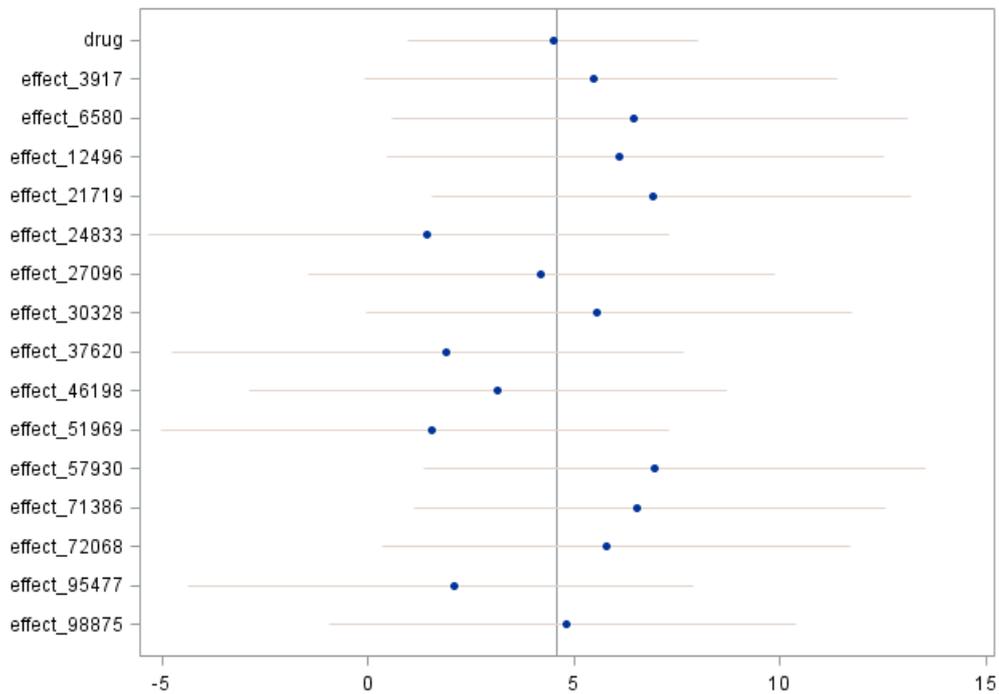


Figure 7: Caterpillar Plot for the Bayes ITEs

Table 10 summarizes the ITE posterior statistics and test statistics. The quantiles figures provide the boundaries for the 90% and 95% credible intervals. The standard errors are calculated as the SD of the posterior distribution of the mean estimate. SAS PROC MCMC can also provide precision estimates (standard errors corrected for autocorrelation based on the MCMC posterior sample).

⁶ This graph is generated with the autocall SAS macro %CATER.

ITE	Mean	S.E.	2.5%	5.0%	50%	95%	97.5%	P(ITE>0)	BF	Log ₁₀ (BF)
effect_3917	5.5	2.9	-0.1	0.8	5.4	10.2	11.4	0.97	0.03	-1.53
effect_6580	6.5	3.2	0.6	1.6	6.3	11.9	13.1	0.98	0.02	-1.68
effect_12496	6.1	3.1	0.5	1.4	5.9	11.5	12.5	0.97	0.03	-1.59
effect_21719	6.9	3.0	1.6	2.2	6.8	12.0	13.2	0.99	0.01	-1.99
effect_24833	1.4	3.2	-5.3	-4.3	1.7	6.4	7.3	0.67	0.49	-0.31
effect_27096	4.2	2.8	-1.4	-0.5	4.2	8.8	9.9	0.93	0.08	-1.12
effect_30328	5.6	3.0	-0.1	0.9	5.4	10.5	11.7	0.97	0.03	-1.51
effect_37620	1.9	3.2	-4.8	-3.6	2.0	6.9	7.7	0.72	0.38	-0.42
effect_46198	3.1	3.0	-2.9	-1.9	3.2	7.8	8.7	0.85	0.17	-0.77
effect_51969	1.6	3.2	-5.0	-4.0	1.8	6.5	7.3	0.69	0.46	-0.34
effect_57930	7.0	3.1	1.4	2.2	6.8	12.2	13.5	0.99	0.01	-1.89
effect_71386	6.5	2.9	1.1	2.0	6.4	11.5	12.6	0.99	0.01	-1.88
effect_72068	5.8	2.9	0.3	1.2	5.7	10.6	11.7	0.98	0.02	-1.65
effect_95477	2.1	3.1	-4.4	-3.4	2.3	7.0	7.9	0.75	0.34	-0.47
effect_98875	4.8	2.9	-1.0	0.1	4.9	9.4	10.4	0.95	0.05	-1.32

Table 10: Bayesian ITEs (N = 5,000 MCMC Sample)

We calculated the BF to validate the Null Hypothesis. Both the BF and the posterior probabilities are estimated parametrically using the means and standard error estimate. In terms of strength of evidence, using Jeffreys' classification of the Log₁₀(BF) statistic, Table 10 shows that the evidence in favour of the Null Hypothesis is very weak, so that Log₁₀(BF) is always less than 0.5. On the other hand 10 subjects show that an ITE value has a more than 90% probability to be greater than 0.

The Bayesian ITEs estimates have the advantage that they are generated in a straightforward manner, as their posterior distributions are simply the sum (e.g. a mixture) of the Treatment PE posterior distributions and the respective Treatment RE posterior distributions. For the same reason it is straightforward to impose greater constraints on the results for Hypothesis Testing. For example, we may require the test statistics to show if the ITEs are greater than a given clinical significance level. The posterior probabilities and the BF statistics can be readily re-calculated and reported as needed.

5.3.3 Estimates for the PTE and Variance Components

In Table 11 we can see that the Bayesian Analysis provides very weak evidence in favour of a negative or null PTE. The selected priors have been updated and they have provided fairly precise estimates.

Parameter	Mean	S.E.	2.50%	5%	50%	95%	97.50%	P(PTE>0)	BF	LOG ₁₀ (BF)
$\hat{\beta}$	4.5	1.8	1.0	1.6	4.5	7.4	8.0	0.99	0.006	-2.2
$\hat{\sigma}^2$	56.8	9.9	40.6	42.4	55.9	74.7	78.9	N/A	N/A	N/A
$\hat{\tau}_\alpha^2$	48.7	24.3	6.0	9.9	47.4	91.3	95.2	N/A	N/A	N/A
$\hat{\tau}_\beta^2$	15.9	13.3	0.7	1.3	12.5	42.1	50.6	N/A	N/A	N/A
$\hat{I}\hat{C}\hat{C}$	0.13	0.10	0.01	0.01	0.11	0.33	0.38	N/A	N/A	N/A

Table 11: PTE estimates (N = 5,000 MCMC Sample).

We note that the $\hat{\tau}_\beta^2$ value is much higher than the LMM estimate, showing the influence of a non-informative prior on the estimate.

However we can consider the ICC estimates as a useful proxy. The ICC measures the correlation between observations within individuals, and may be seen as a measure of how much of the individual response is due to random effects. We have calculated the ICC with respect to τ_β^2 so that:

$$ICC = \frac{\tau_\beta^2}{\tau_\alpha^2 + \tau_\beta^2 + \sigma^2} \quad (43)$$

The \widehat{ICC} value, while not negligible, is reasonably low (0.13). This result suggests that the PTE is reasonably consistent across subjects. The flexibility offered by the availability a posterior distribution implies that we can readily obtain precision estimates⁷ for this value using the posterior distributions of its component parts. Furthermore it would be relatively straightforward to create test statistics for this value using non-parametric methods in order to test the probability $P(\tau_\beta^2 > k)$ for a given k threshold.

5.4 Comparison of Modes of Estimation

We now present the estimates from the different modes of analysis and the strength of the evidence provided by them. We start with the ITEs.

5.4.1 Comparability of the ITEs

We can see in Table 12 that the simple Linear Model results have the disadvantage of a low precision compared to the other two models.

ID	FE Linear Model		RE Mixed Model		RE Bayesian Model	
	ITE	S.E.	ITE	S.E.	ITE	S.E.
3917	3.8	6.1	5.4	2.6	5.5	2.9
6580	1.0	6.1	5.9	2.6	6.5	3.2
12496	2.7	6.1	5.7	2.6	6.1	3.1
21719	13.7	6.1	6.4	2.6	6.9	3.0
24833	-0.3	6.1	3.1	2.6	1.4	3.2
27096	7.2	6.1	4.8	2.6	4.2	2.8
30328	2.3	6.1	5.4	2.6	5.6	3.0
37620	2.2	6.1	3.4	2.6	1.9	3.2
46198	3.7	6.1	4.1	2.6	3.1	3.0
51969	-2.5	6.1	3.2	2.6	1.6	3.2
57930	18.0	6.1	6.5	2.6	7.0	3.1
71386	13.5	6.1	6.2	2.6	6.5	2.9
72068	5.2	6.1	5.6	2.6	5.8	2.9
95477	2.2	6.1	3.5	2.6	2.1	3.1
98875	1.8	6.1	5.0	2.6	4.8	2.9

Table 12: Comparison of FE by Approach

However, these estimates are unbiased and reliable, whereas the ITEs of both the LMM and Bayesian model are by definition biased towards the PTE estimate. Remarkably, as shown in Figure 8, p.40, there is still some degree of correlation between the FE model estimates and the other two approaches. We use Spearman Rho correlations to dispense with normality assumptions. The results show that $\rho = 0.653$ for the FE ITEs with respect to the Bayes ITEs and $\rho = 0.702$ with respect to the LMM ITEs. The bottom two graphs in Figure 8 indicate that the most extreme subjects' estimates did shrink considerably toward the average result.

⁷ This is a far from being a non-trivial exercise with a Mixed Model approach. A cursory check on the distributional properties of the ICC has yielded few, not easily implementable results.

More importantly, the Mixed Model estimates correlate very closely to the Bayesian ones, as shown in the top plot.

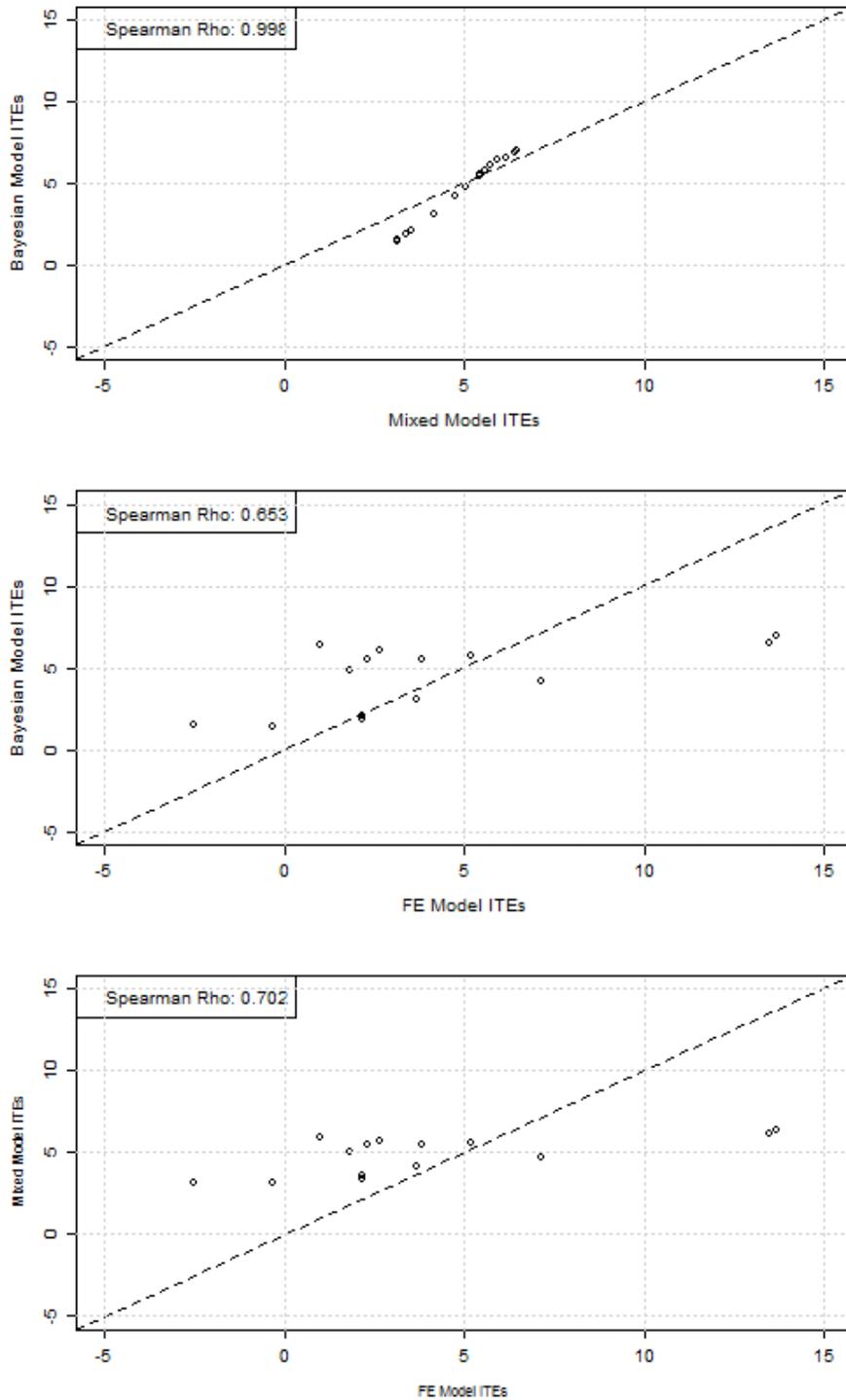


Figure 8: Scatter Plot of ITE Estimates by Modeling Approach

This is to be expected given the similarities between the Mixed Models and the Bayesian Models. As already noted, it is more straightforward to obtain precision estimates and test statistics for the Bayesian estimates for the ITEs. This is not the case with the ITEs estimates derived from the Mixed Model approach. Altogether, the ITEs produced by the FE Linear Model and the Bayesian Model have a greater value for researchers compared to those of the Mixed Model.

We also like to stress the advantages, compared to the FE linear model, of the Mixed and Bayesian model approaches. We have shown earlier that the FE Linear model provided very weak evidence for the ITEs. The p-values and CIs all suffered from the low precision of the estimates.

In the case of LMM, the application of the REs implies a better control of the variance of outcomes, and, thanks to the ‘BLUP’ properties of the RE predicted values, a much greater level of precision for the ITEs estimates. Because of the relatively small Treatment RE variance we expect the shrinkage to be relatively contained. Therefore 9 out of 15 subjects had adequate evidence to reject the Null Hypothesis at the stipulated test size level. The p-values, in general, exhibited small values. However these test statistics are not very reliable due to the use of Z-tests.

In the case of the Bayesian Measure the interpretation of the evidence was much more straightforward. Ten subjects showed a level of evidence that is considered as strong in a Bayesian perspective, 4 subjects showed adequate levels of evidence and only one subject was not considered to provide adequate evidence for assuming that the ITE is positive. All three test statistics had a straightforward interpretation. Furthermore, the ITEs themselves were obtained as simple sum of posterior distributions, and the relevant precision estimates and test statistics could be derived in a straightforward manner from the same posteriors. This is the strongest advantage that the Bayesian Model has over the Mixed Model approach.

5.4.2 Comparability of the PTE and TTE

Table 13 shows that all methods used to determine an average response (at different levels) produce estimates that are consistent with each other. Each method shows strong evidence in favour of the treatment effect being positive. The test statistics of different types are also consistent with each other.

Parameter	TTE		FE LM		LMM PTE		Bayesian PTE	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
$\hat{\beta}$	4.96	1.58	4.96	1.56	4.96	1.66	4.52	1.78

Table 13: TTE, Linear Model FE and PTE estimates

As noted before, the TTE is an adequate estimator of the treatment effect if restricted to the trial subjects. With regards to the PTE estimates, the estimate based on the Mixed Model has a slight advantage in terms of precision and it is also consistent with the TTE value. It may also have the merit of being ‘objective’ in the sense of not being dependent on the choice of priors, which is a subjective matter. However the Bayesian approach has also an advantage in the sense that it might be possible to consider different priors to reflect different interpretative scenarios, which is not possible with the Mixed Model approach.

5.4.3 Comparability of the Variance Components

Table 14 provides a comparison between variances. The Bayesian approach is more suitable for a small scale RCT provided that we can select a prior that does not cause excessive bias for the $\hat{\tau}_{\beta}^2$ estimate.

Parameter	Mixed Model		Bayesian Model		REs as Standard Deviations	
	Mean	SE	Mean	S.E.	Mixed	Bayesian
$\hat{\sigma}^2$	53.6	9.0	56.8	9.9	7.3	7.5
$\hat{\tau}_{\alpha}^2$	46.9	31.6	48.7	24.3	6.8	7.0
$\hat{\tau}_{\beta}^2$	5.3	9.6	15.9	13.3	2.3	4.0

Table 14: Variance Components

The Bayesian approach however achieves the estimation without any particular difficulty. It also easily provides a realistic range of plausible values in the form of the credible interval. The estimates for the other two variance components are not significantly different. It is worth noting that according to the $\hat{\tau}_\alpha^2$, there is a good deal of heterogeneity in the baseline of the patients belonging to this special population. This has some relevance from a more general research perspective. Finally, we include in Table 14 the standard deviations of the random components to provide a perspective of the degree of variation in the FACIT-F score that may be attributable to each component.

5.4.4 Summary

In summary, we note that each method has specific merits given different purposes. The TTE is an appropriate average estimator for the subjects participating to the trial. The FE ITEs are also unbiased. The Bayesian ITEs seem to us a reasonable alternative as they correlate well with the LMM without relying on unrealistic assumptions. The PTE estimators offered by either the Linear Mixed and Bayesian Model approaches are practically equivalent in terms of the balance between their benefits and drawbacks.

The Bayesian Model approach to the estimation of the variance components is perhaps the best approach if we are constrained to use small scale RCTs based on pooled N-of-1 trials, provided that suitable priors can be provided. A small number of participating subjects do place limits to the ability of SAS PROC MIXED to estimate properly several variance parameters, some of which may be small compared to the others. We will review these conclusions in the next chapter on simulations.

Finally, we note that it is possible to approach the analysis of small sample data in a sequential manner. We start by using the FE Linear Model to test the hypothesis that the TTE is different from 0. The TTE is the most precise estimate of the three estimates provided by the three modes of analysis. If we reject the Null Hypothesis, we continue to test for the presence of a PTE using a Mixed Model or a Bayesian Model. The FE Linear Model can also use the ITEs to test for HTE in the sample, which can inform the modes of analysis for the PTE. The estimate has less precision and the test will have less power, but the PTE implies a generalization of the results of the trial and therefore it is of great importance.

The Bayesian model may be preferred to the LMM for very small samples, if SAS procedures are to be used. If samples of adequate size can be accessed, the LMM is to be preferred because of its lack of dependence on priors selections.

Similar considerations may apply for the estimation of the ITEs. In principle the FE Linear Model is the most reliable – and simple to apply. If suitable priors can be provided, or if we are in a scenario where new subjects need to be assessed in an incremental scenario, then the Bayesian Model approach may be used as a suitable alternative.

6 Simulations

We extend the analysis of the actual data from the RCT data, where a group of patients are recruited in a one-off Trial based on multiple experiments (the N-of-1 trials), to a hypothetical scenario where new patients are continuously included.

This scenario does neither aim to discuss Adaptive Clinical Trial design, nor to explore specific features of open ended trials. The objective is to find the most suitable approach to get:

1) stable PTE estimates for the drug effect and the variance of the random treatment part of the drug effect with the smallest number of participants, and

2) an indication of the minimum sample size that is necessary to identify an individual for which the ITE is greater than a Clinical Significance level with a given level of probability (Power).

Our hypothetical scenario applies a Bayesian Analysis approach where samples of observations can be enlarged through adding new subjects. For the sake of simplicity, we assume that in this scenario the collection and reporting of data is standardised and homogenous, e.g. there is a homogenous measurement error.

Our scenario is based on the simulation of data which reflects the model specification discussed in Chapter 5, e.g., a model with a Random Intercept and a Random Treatment effect. We will vary the size of the variance of the Random Intercept relative to the Random Treatment effect, holding constant the error term variance, to show the implications, following Langford [12] of the magnitude of random effects on the estimation of fixed, e.g. population level, treatment effects.

6.1 Bayesian Analysis by Updating Priors

We first define a procedure in which an initial RCT is conducted and the data is analyzed with a Bayesian Analysis model, using non-informative priors. A posterior distribution is produced for all parameters of interest: PTE, ITE, τ_β . Test statistics are collected: $\text{Log}_{10}(\text{BF})$, DIC, etc. After this first analysis, a new subject is introduced with its own set of observations.

	Initial Sample 1	Initial Sample 2	Initial Sample 3
Very low τ_α, τ_β	$n = 5, \tau_\alpha = 0.1, \tau_\beta = 0.1$	$n = 10, \tau_\alpha = 0.1, \tau_\beta = 0.1$	$n = 15, \tau_\alpha = 0.1, \tau_\beta = 0.1$
Large τ_α , very low τ_β	$n = 5, \tau_\alpha = 7.5, \tau_\beta = 0.1$	$n = 10, \tau_\alpha = 7.5, \tau_\beta = 0.1$	$n = 15, \tau_\alpha = 7.5, \tau_\beta = 0.1$
Large τ_α , small τ_β	$n = 5, \tau_\alpha = 7.5, \tau_\beta = 2.5$	$n = 10, \tau_\alpha = 7.5, \tau_\beta = 2.5$	$n = 15, \tau_\alpha = 7.5, \tau_\beta = 2.5$
Equal τ_α and τ_β	$n = 5, \tau_\alpha = 5.0, \tau_\beta = 5.0$	$n = 10, \tau_\alpha = 5.0, \tau_\beta = 5.0$	$n = 15, \tau_\alpha = 5.0, \tau_\beta = 5.0$
Large τ_β , small τ_α	$n = 5, \tau_\alpha = 2.5, \tau_\beta = 7.5$	$n = 10, \tau_\alpha = 2.5, \tau_\beta = 7.5$	$n = 15, \tau_\alpha = 2.5, \tau_\beta = 7.5$
Large τ_β , very low τ_α	$n = 5, \tau_\alpha = 0.1, \tau_\beta = 7.5$	$n = 10, \tau_\alpha = 0.1, \tau_\beta = 7.5$	$n = 15, \tau_\alpha = 0.1, \tau_\beta = 7.5$

Table 15: Simulation Configurations for Updating Priors

This new subject is then analyzed using the posterior of the previous analysis as the new prior. The process is repeated and tracked until 100 subjects have been analyzed in total. We set out 18 sub-scenarios to control for dependence on initial sample configurations and size, as well as changes in distributional parameters. In these scenarios the intercept $\alpha = 20$, the treatment effect $\beta = 5$ and the error term SD $\sigma = 7.5$ are held constant. The other parameters vary as shown in Table 15 above.

6.1.1 PTE simulation trends

The simulation (Figure 9) shows that the changes in initial size (at least in the range of values used for the simulation) do not affect much the trends towards the achievement of stable estimates for the PTE.

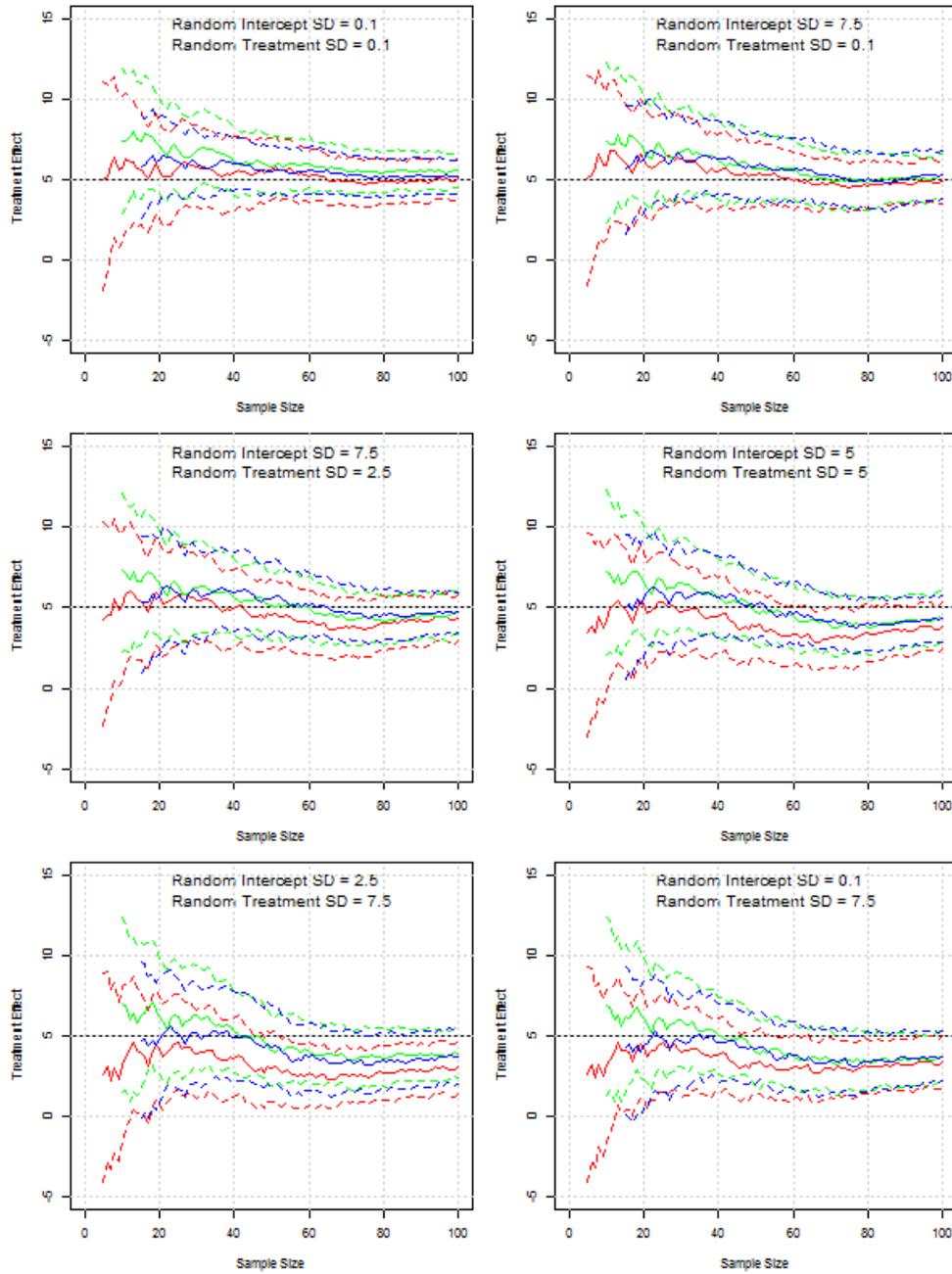


Figure 9: Variation in Drug PTE Estimate by Initial Sample Size and Random Effects Variances. N= 5, Red Line. N = 10, Green Line. N = 15, Blue Line. Dashed Line 95% Credible Interval. Dashed Black Line Population Value.

The random treatment effect variance is a more important factor. Larger τ_{β}^2 values delay the stabilization of results on the PTE..

6.1.2 BF for $H_0: PTE \leq 0$ Simulation Trends

We now consider the behaviour of the Bayesian Factor (in its base 10 logarithmic form) under different conditions (Figure 10). This test statistic is used to check if there a positive PTE.

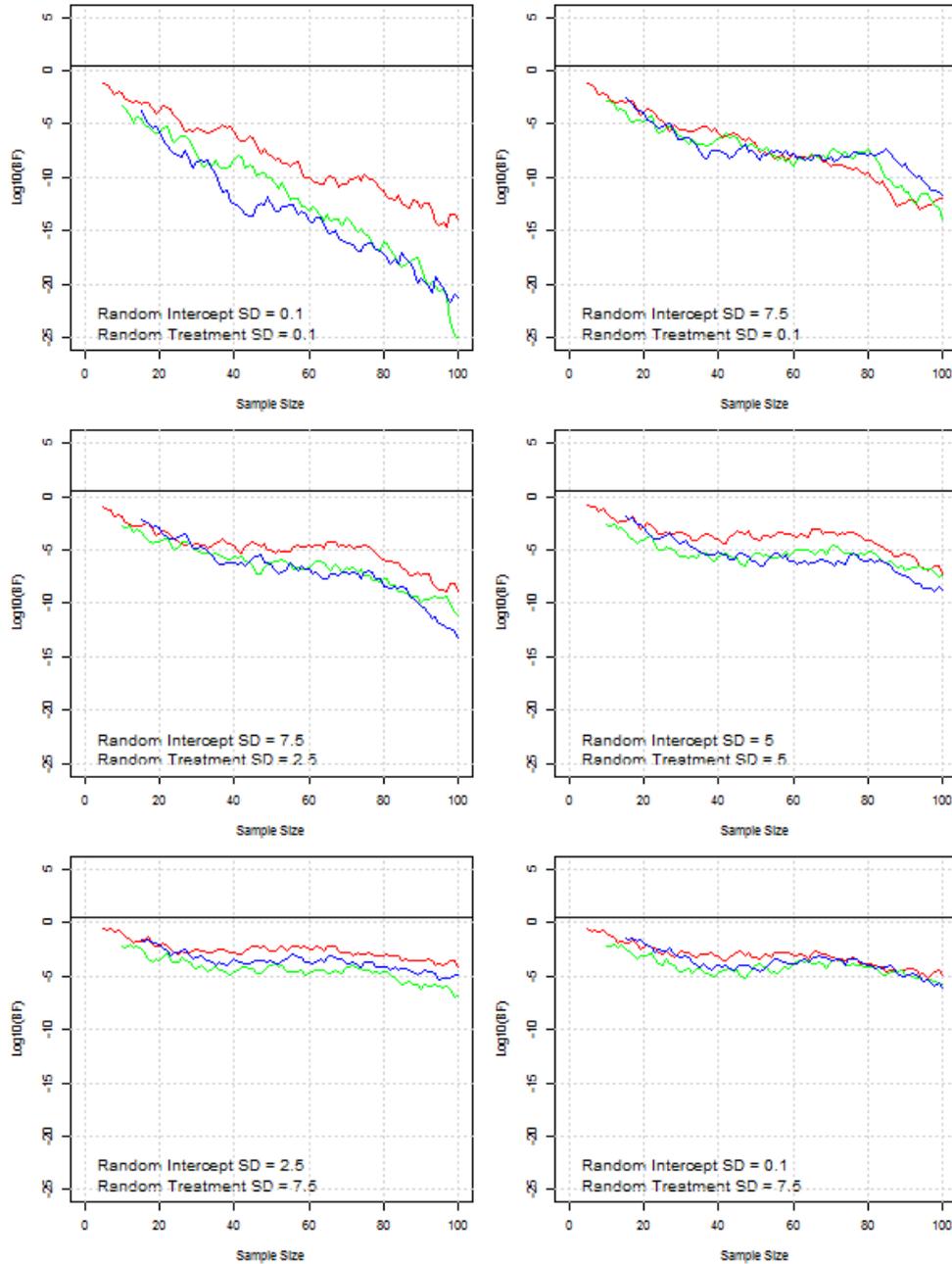


Figure 10: Log₁₀(BF) Simulation Trends by Initial Sample Size and Random Effects Variances. N= 5, Red Line. N = 10, Green Line. N = 15, Blue Line. Black Line as Log₁₀(BF) Reporting Threshold.

The Null Hypothesis is set out as a composite hypothesis, that is: $H_0: PTE \leq 0$. All scenarios reach very quickly strong levels of evidence against the H_0 . The $\text{Log}_{10}(\text{BF})$ statistic is always negative (e.g. for a BF less than 1). However increasing τ_β^2 implies that larger sample size does not alter much the $\text{Log}_{10}(\text{BF})$ values. Remarkably, high τ_β^2 values have also caused the trends for improvement in the test statistic to stall until 40 individuals are added to the initial 40 subjects.

6.1.3 Random Treatment τ_β Simulation Trends

We now consider the impact of the scenario parameters on the estimation of the Random treatment effect standard deviation value τ_β . The trends are shown in Figure 11.

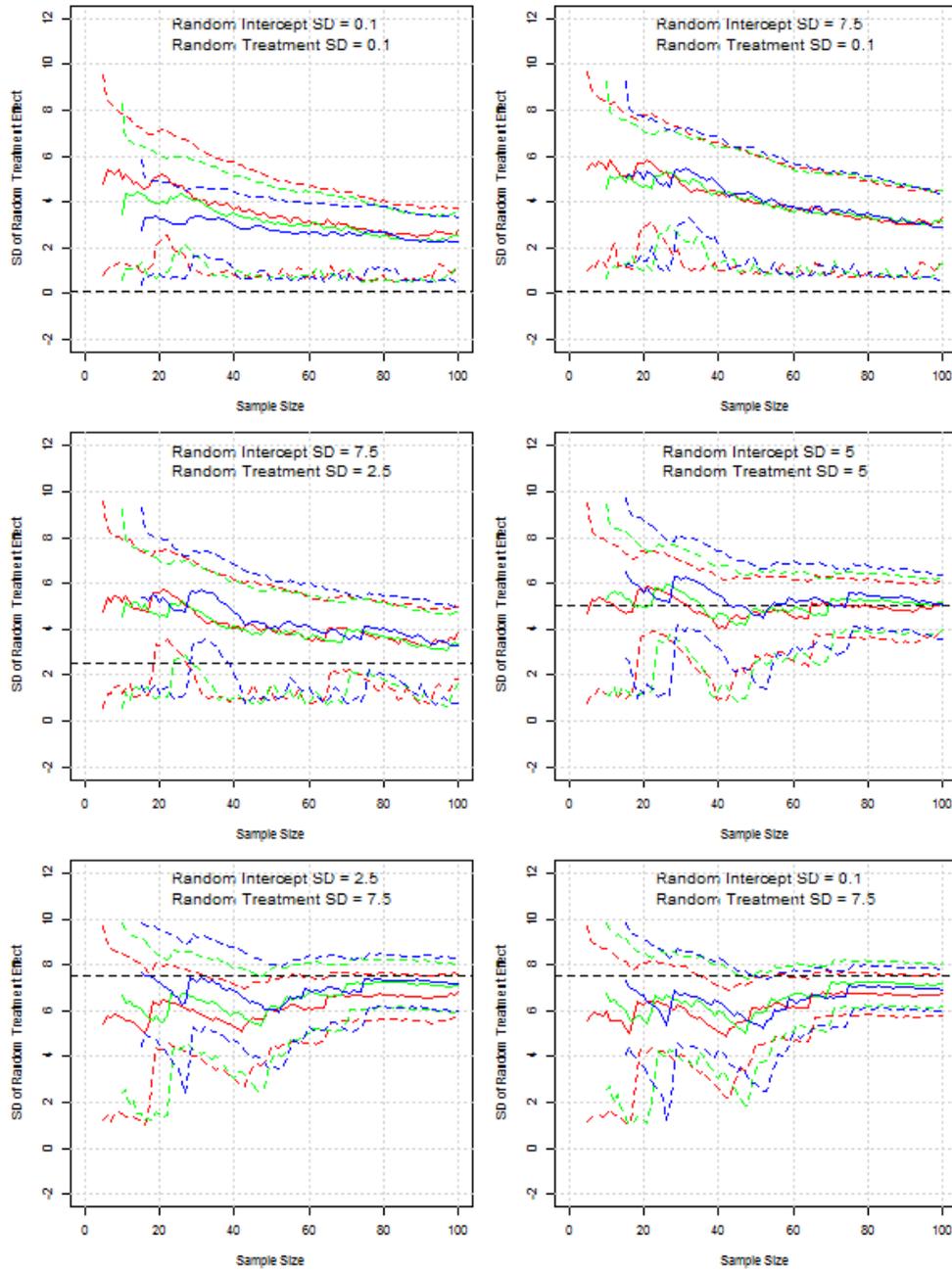


Figure 11: Random Treatment Effect SD by Initial Sample Size and Random Effects Variances. N= 5, Red Line. N = 10, Green Line. N = 15, Blue Line. Dashed Line 95% Credible Interval. Dashed Black Line Population Value.

The value $\hat{\tau}_\beta$ shows a slower convergence to ‘true’ population values compared to the PTE $\hat{\beta}$ value. The trends are also more irregular as the random treatment standard deviation τ_β increases. It must be pointed out that the same non-informative priors have been assigned to the variances of both error terms and random effects.

If there were possible indications of the relative size of these variances, it would also be possible to set more informative priors that may avoid this degree of instability and slow convergence to the true values. It is important to note that only one scenario ($\tau_\alpha = 5.0, \tau_\beta = 5.0$) shows a credible interval that systematically includes the ‘true’ population parameter τ_β .

In all other cases the researchers would have to accept erratic estimates and, unknowingly, biased estimates which may question the validity of the treatment to a large fraction of the subjects. In the next section we extend the simulations to produce multiple starts and assess the actual degree of instability in the assessment of population parameters.

6.1.4 Predicted RE Precision Simulation Trends

The simulation also shows that the precision of the predicted RE values is not dependent on the number of subjects added to the analysis.

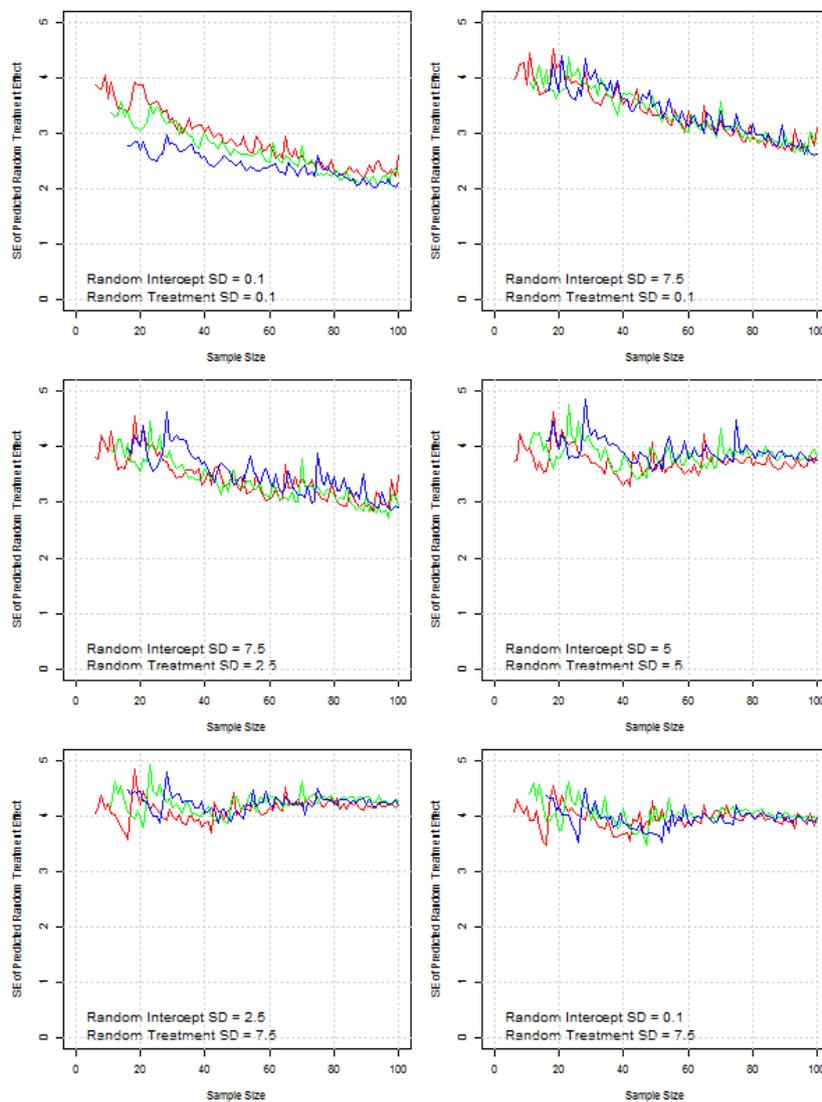


Figure 12: S.E. of Predicted RE estimates by Initial Sample Size and Random Effects Variances. N= 5, Red Line. N = 10, Green Line. N = 15, Blue Line.

In fact there is a downward trend only for the data in which the SD of the RE is very low, that is when any predicted RE value is probably biased. This trend reflects a dominant PTE value.

6.1.5 ITE Precision Simulation Trends

We now show to what extent the different scenarios influence the precision of the ITE estimates.

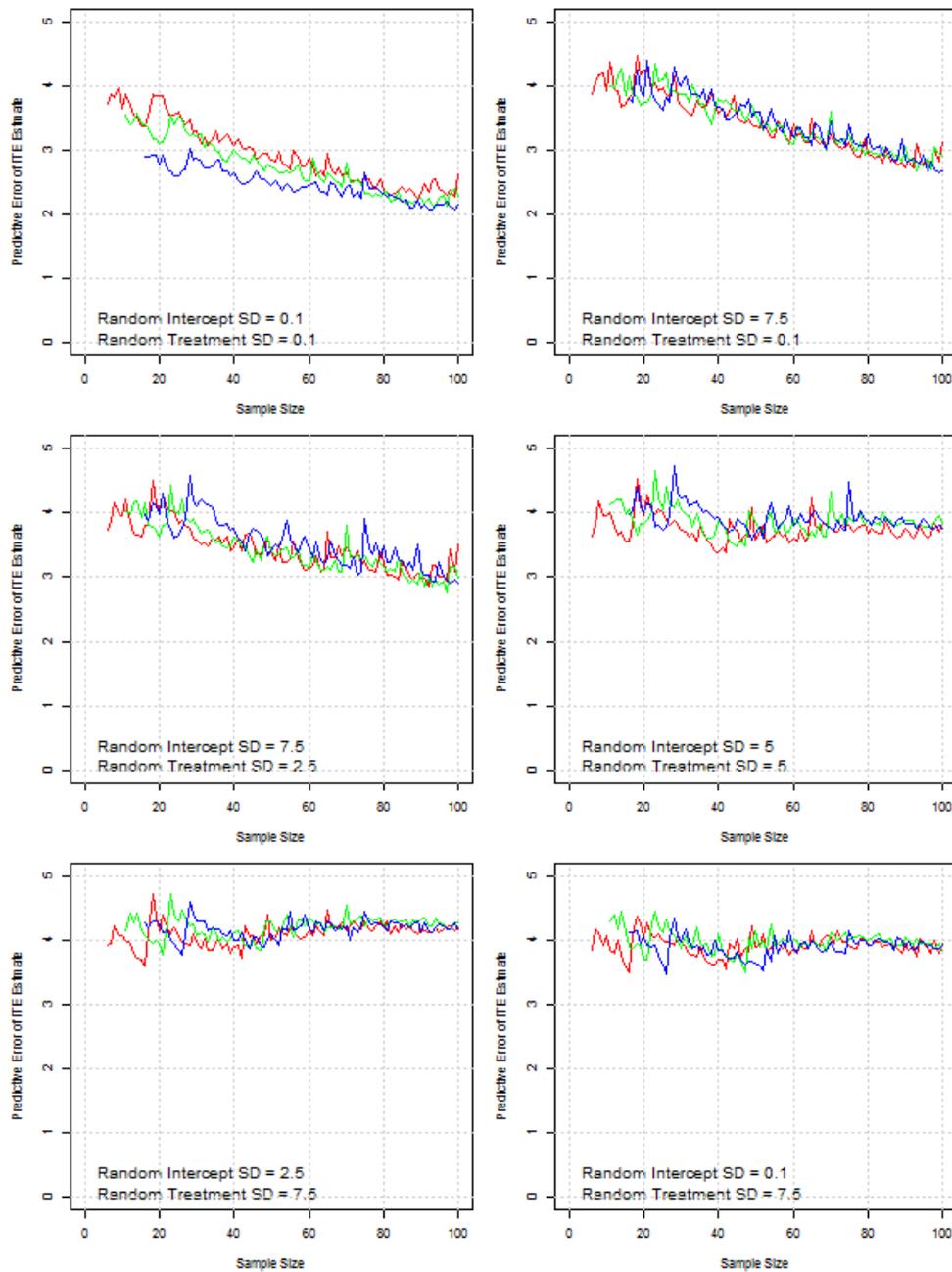


Figure 13: S.E. of ITE estimates by Initial Sample Size and Random Effects Variances. N= 5, Red Line. N = 10, Green Line. N = 15, Blue Line.

The ITE prediction error trend is obtained by tracking the precision estimate for the first new subject added in the analysis as new subjects are added. The ITE prediction closely tracks the trend for the RE predicted error. The trends outlined in the top graphs and the centre left graph show that the estimates are trending towards a nearly null or small value, as one might expect. However this trend manifests itself very slowly and with a fair amount of instability. When nearly all random effect variability is due to the random treatment effect (as in the bottom graphs) the estimated predictive error approaches stability with just about 50 additional subjects.

These trends, although less marked than those in Figure 12, are quite important. An essential function of the RCT scenarios is to evaluate how large is the random effect component of the treatment new subjects. It follows that depending on the relative size of the random effects variance, large numbers of subjects could be required before we acquire consistently precise estimates that can detect subjects that have an ITE greater than the chosen threshold.

For example, the trends are far less satisfactory for the centre left scenario ($\sigma = 7.5$, $\tau_\alpha = 7.5$, $\tau_\beta = 2.5$) which is similar to the scenario suggested in Chapter 5, where $\hat{\sigma} = 7.3$, $\hat{\tau}_\alpha = 6.8$, $\hat{\tau}_\beta = 2.3$. In this case the precision estimates are quite unstable even after 100 subjects are included in the trial. Researchers in this scenario may find it difficult to provide consistent strong evidence for a positive ITE. Any approach that enables a faster converge to stable and consistent prediction error estimates is therefore essential for a formal procedure that assesses a patient on the results of a N-of-1 trial using the accumulated knowledge of updated priors.

6.2 Stability of Bayesian Analysis: Random Sampling Sequences

6.2.1 Non Informative Prior – Small Treatment RE SD

We now consider multiple random starts to check the sensitivity of the Bayesian Analysis to initial configurations. Figure 14 shows the effect of running 30 different random starts for the scenario with $n = 15$, $\tau_\alpha = 7.5$, $\tau_\beta = 2.5$, as it is compatible with the results of the analysis of the set of real data.

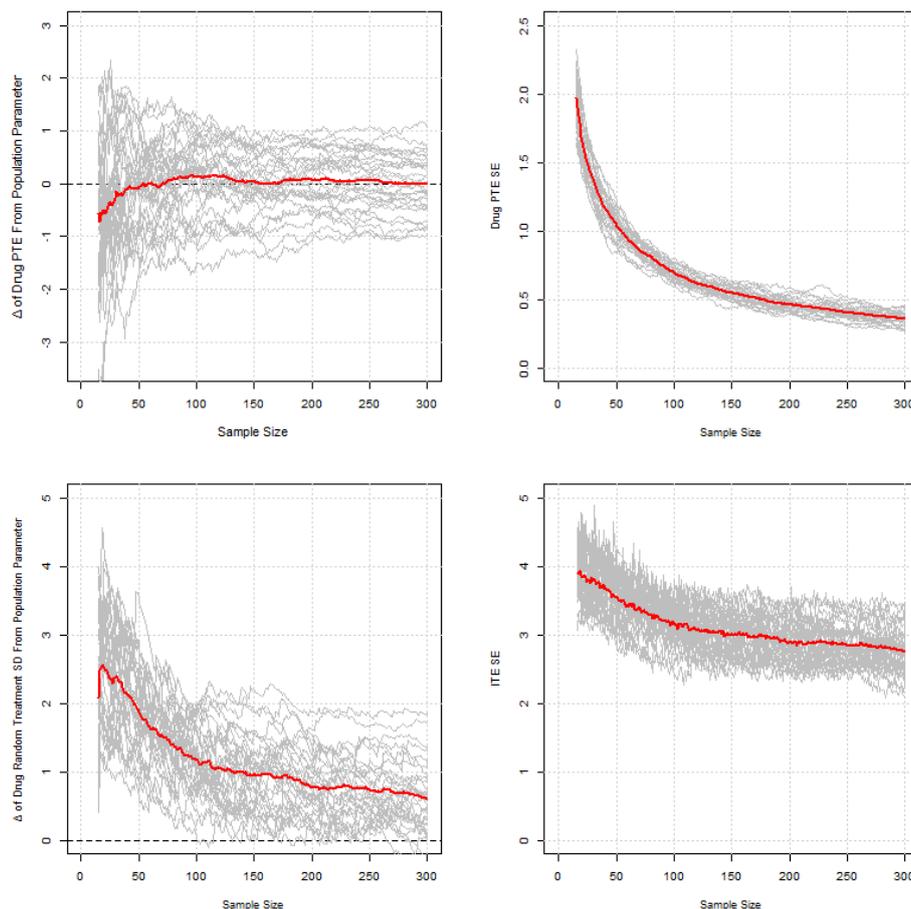


Figure 14: Stability of Model Parameters Estimates With 30 Random Starts

The red lines are smoothed values for the average of the parameters across the randomizations. We simulate a sample of 15 patients to which are added an additional 285 subjects. The analysis of the

simulations tracks four parameters: the drug PTE value, its S.E. estimate, the τ_β SD for the random effect, and the S.E. for a single ITE. The PTE and τ_β values are measured as deviations from their population parameters. The curves are based on Lowess smoothers. A simple visual inspection shows that the average estimated PTE value across simulations (red line) does approach the population parameter very quickly. Some of the trends for the more extreme initial values show that the addition of a single subject at the time does not have the ability to force a quick convergence to the population level. The PTE S.E. follows a predictably asymptotic trend of ever smaller values as the sample size increases.

The average $\hat{\tau}_\beta$ estimate converges far more slowly, being still considerably biased (almost 1 unit in excess of the population parameter 2.5) after 285 subjects have been included. The ITE shows a more irregular and less steep trend because the ITE is the sum of two components: the PTE and the Predicted RE value. The gains in precision are due only to the decreasing S.E. of the PTE estimates. This scenario indicates a potentially serious problem with the Bayesian approach when τ_β is small but not negligible. The non-informative prior we have selected imposes a uniform distribution on τ_β^2 in line with Gelman [9].

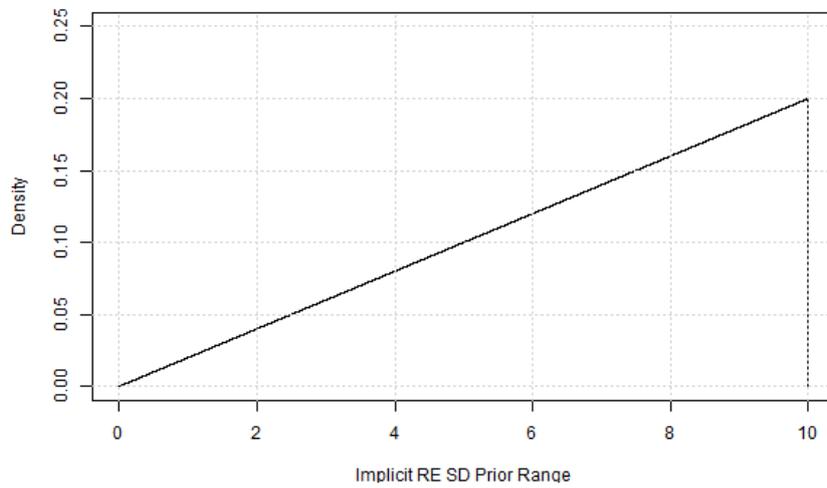


Figure 15: Density of Square Root of U(0,100)

We note that a non-informative uniform prior on τ_β^2 implies an informative prior on τ_β . It can be shown that the probability density function (p.d.f.) for the square root of U(0,100) is $f(x) = 0.02x$. This is a triangular distribution with the mass placed on the high values of τ_β . The prior density for τ_β is therefore actually preventing the posterior from approaching a small value for τ_β . The density increases linearly from 0 to 0.2 on the range [0, 10]. Given the population value $\tau_\beta = 2.5$, the associated probability would be quite small and thus the non-informative Uniform Prior for the τ_β^2 is in practice a sceptical prior for τ_β .

We can set a uniform prior on τ_β , but this means to have definite information on the value τ_β , such that we can assign a prior with a restricted plausible range of values. It implies setting up a user defined prior density of the RE treatment variance to replicate the simulated distribution shown in Figure 16. It can be shown that in this case the pdf of the squared U(0,10) is $f(x) = 0.05x^{-\frac{1}{2}}$.

This is a highly informative prior on the τ_β^2 parameter. Thus an informative prior is required regardless of what we do when small RE Treatment Variance may be present. A similar consideration applies to the PTE estimate. The simulation shows that the addition of a subject at the time may not be able to generate a fast and systematic convergence towards the population parameter. Informative priors are required to make it possible, for any given initial set of data, to converge quickly towards the population parameters. The issue of convergence is important because only once the estimates for the key parameters are reasonably stable and unbiased it is possible to make reliable predictions about new subjects.

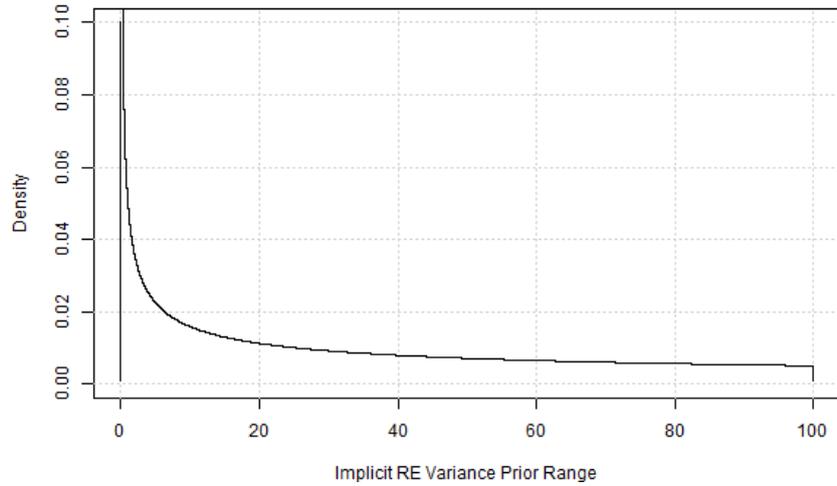


Figure 16: Density of Square of U(0,10)

6.2.2 Non Informative Prior – Increased Treatment RE Variance

We repeat the analysis using the scenario where $n = 15, \tau_\alpha = 5, \tau_\beta = 5$. We have seen that in this scenario the convergence of several parameters is faster and more consistent than in other scenarios.

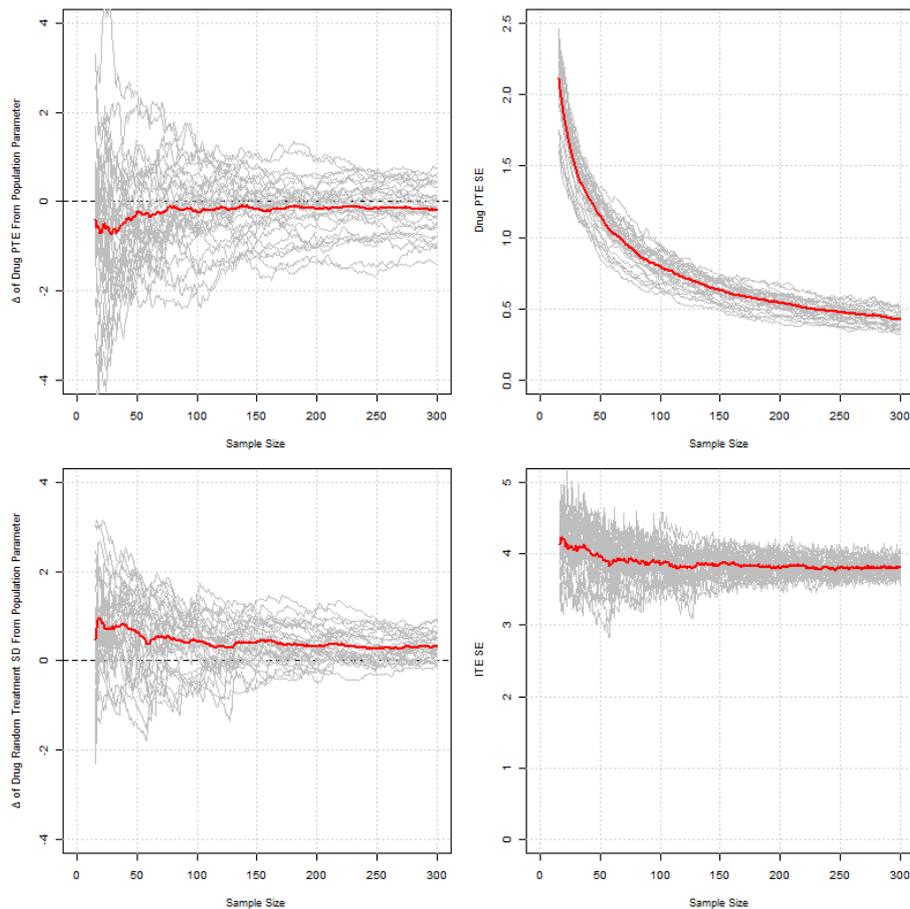


Figure 17: Stability of estimates for higher value of τ_β .

The trends are now much more regular for $\hat{\tau}_\beta$ and the ITE S.E. estimates, although less so for the PTE estimates. The $\hat{\tau}_\beta$ value still shows a degree of upward bias – albeit a reduced one and slowly decreasing

one. PTE estimates also exhibit, on average, some degree of systematic downward bias, and a greater degree of variability at the start of the simulation. It may be noted that setting the Null Hypothesis for ITEs estimates would yield more conservative tests, as the precision of ITE estimates is lower and the range of possible values for the RE (and thus ITE) estimates is wider.

The graphs suggest that with a Bayesian analysis there is a non-favourable trade off between precision and unbiasedness of estimates and their stability. In addition this trade off is influenced by the degree of variability of REs. Another point worth noting is how the ITE precision is influenced by the size of the τ_β . The trends in the precision of PTE estimates are very similar in both scenarios ($\tau_\beta = 2.5$, $\tau_\beta = 5.0$).

The ITE precision in these simulations is shown to be dependent on the magnitude of the τ_β value: the greater the τ_β , the greater the ITE S.E. estimates. If we define a $H_0: ITE \leq 0$, and use conventional decision rules, the estimated ITE must be approximately twice the value of the ITE S.E. value to pass the test decision rule.

This means that in principle greater population τ_β^2 values will necessarily require greater ITE estimates for each subject. All other factors being equal, fewer individuals may be shown to benefit from the chosen treatment if the τ_β is also comparatively large.

6.2.3 Non-Informative Prior for Treatment RE SD

We conclude the assessment with a last simulation. We have noted that the Inverse Gamma distribution with low scale and location parameters has been judged as not being adequate for the modelling of samples with few units of grouped data.

We set aside for the moment this objection and run a simulation where the prior for τ_β^2 is the inverse gamma distribution as $IG(0.175, 0.175)$. These parameters are chosen on purely subjective basis, to provide an example. Figure 18 shows how this distribution behaves as a uniform one, albeit on much wider range of value, for both the τ_β^2 and the τ_β .

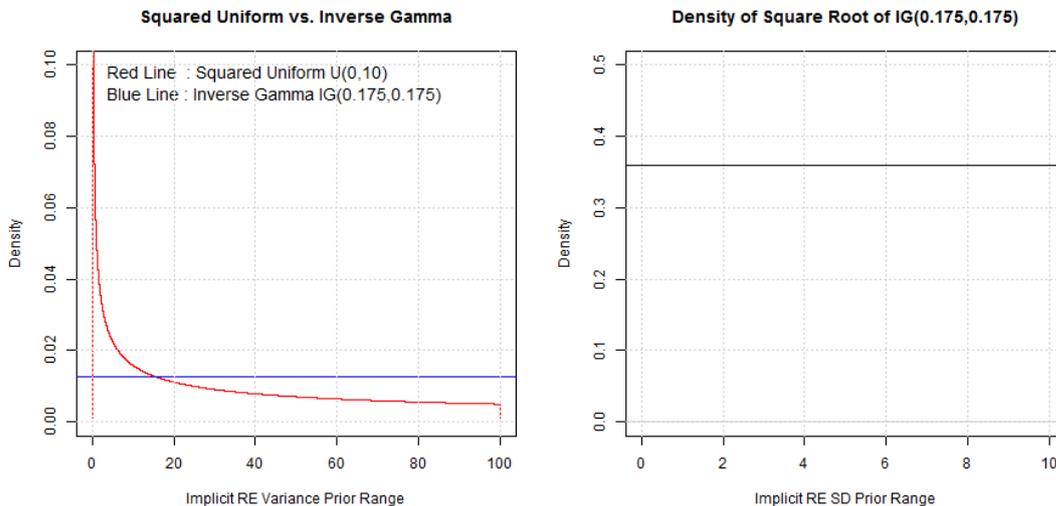


Figure 18: Inverse Gamma for RE Variation and SD

We want to check if the Inverse Gamma may be a suitable substitute as a Non Informative prior for a small RE variance. The simulation is repeated for the scenario where $n = 15$, $\tau_\alpha = 5$, $\tau_\beta = 2.5$.

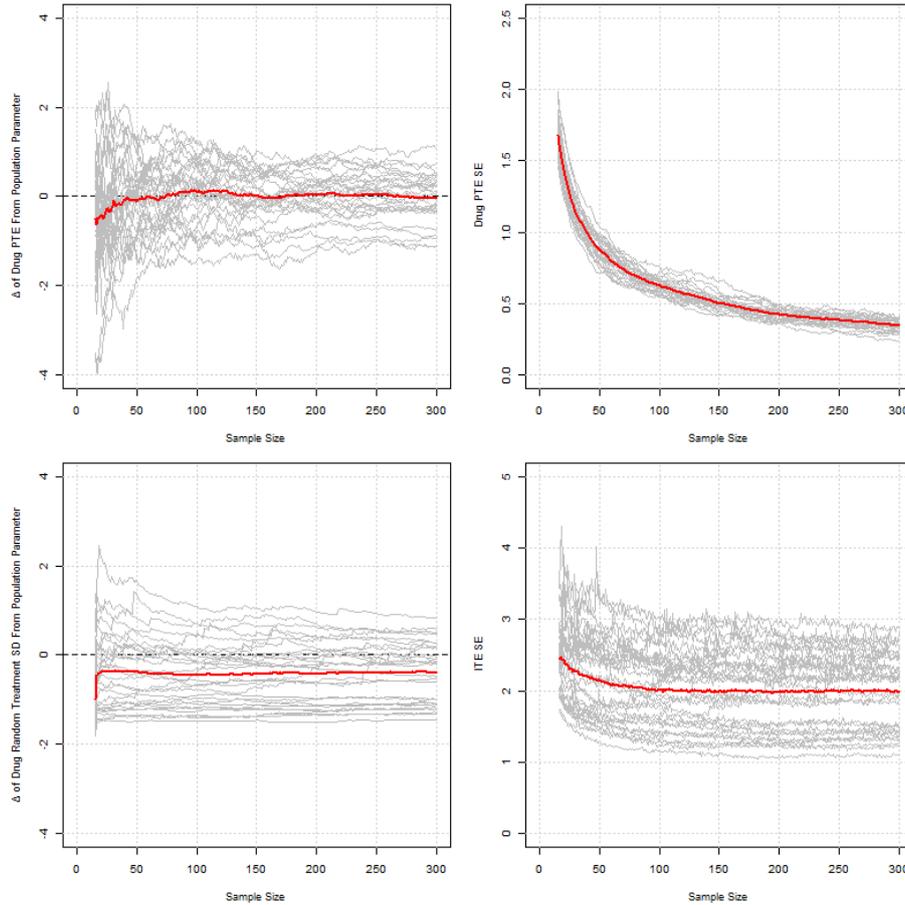


Figure 19: of estimates with a IG (0.175, 0.175) for small τ_β .

Figure 19 shows that the average trend for $\hat{\tau}_\beta$ has improved considerably with this prior. The bias is much smaller and convergence on average is very fast. The same applies indirectly to the estimate for the ITE SE. However we must remember that the Inverse Gamma may not perform as well very small samples. Also, the bottom right graph for the precision of the ITE estimates shows a very wide spread of results. It means that there can be a high degree of inconsistency in the testing of the ITEs.

6.2.4 Implications of Simulations.

The simulations have shown that if $\hat{\tau}_\beta^2$ is not very small the estimates for the model parameters may display instability and bias. This is less of a problem for PTE estimates, but it is a greater concern for the $\hat{\tau}_\beta^2$. The $\hat{\tau}_\beta^2$ value is quite relevant because it determines the degree of consistency of treatment effects across the general population. It follows that to avoid unstable and biased results the Bayesian Analysis requires informative priors, at least as far as the τ_β^2 parameter is concerned. These are very strong limitations in the context of RCTs based on a small set of N-of-1 trials. These trials deal with a situation in which there is a high degree of uncertainty about the effect and consistency of the treatment.

In addition there is the disproportionate impact that sampling biases may cause in the initial analysis. As stated by Longford [19], data characterized by sizeable RE treatment effect variances may generate quite biased estimates of variance components if the sampling procedure does not take into account in some way these sources of variance. The sampling procedure itself requires some degree of knowledge about the structure of the data, which is the same problem defined earlier about informative priors.

7 Conclusions

We suggest that a researcher should consider different analytical perspectives to estimate Individual, Trial or Population effects. We recapitulate the findings of chapter 5 and 6 and outline a range of possible options for the analysis of RCTs based on the pooling of a small set of N-of-1 trials.

7.1 ITEs – FE Linear Model, RE Mixed Model and Bayesian Model

The estimation of ITEs is generally best conducted with a FE Linear model. Although it may yield tests with low power (on account of low precision), the ITEs estimates are unbiased and objective estimates.

There is another advantage from adopting this approach as preliminary step. The ITEs estimates can be combined into a TTE and they can be used to assess the degree of HTE displayed by the data. These estimations are the basis for the extension of the analysis to the estimation of the PTE and the Treatment RE variance. The latter parameter makes it possible to assess the degree of consistency of the treatment effects across the general population.

Bayesian methods may be used to estimate ITEs provided that suitable priors can be used. This approach offers greater precision, the hypothesis testing is simplified and posterior probabilities can be assigned to the posterior estimates. If the variance of the RE treatment effects is large the limited shrinkage effect may be accepted as a reasonable trade-off for the benefits outlined above. In addition, data from new patients can be easily incorporated in the analysis and the overall precision improved – which may not be necessarily the case for a FE Linear model.

The LMM, instead, implies a convolute process for the estimation of the ITEs variance and hypothesis testing of the ITEs would rely on very loose assumptions. The Bayesian model with equivalent specifications produces ITE estimates that are highly correlated to the LMM estimates. Therefore Bayesian ITEs are a better proposition compared to LMM estimates. In general, ITEs should not be estimated by means of LMMs.

7.2 TTE Estimation – FE Linear Model

The TTE estimation is necessarily conducted with a FE Linear Model. Both Mixed and Bayesian Models with RE produce ITEs which include the PTE estimate. Therefore an average of these ITE will have a greater estimation variance compared to the TTE. In addition, it is not obvious how to interpret such averages.

The TTE can be simply estimated and tested by setting a contrast across the ITEs. In our sample of real data the TTE yields a S.E. estimate that is very close to that of a FE Linear Model that excludes ITEs.

While the TTE estimate may be found to reject its Null Hypothesis, it is still important to check for the presence of HTE as suggested in section 7.1. A significant HTE would change considerably the interpretation of the TTE even if limited to the RCT sample.

In effect, significant TTE should be considered as the precondition to proceed with a further stage of analysis for the estimation of a PTE. A significant HTE also provides a rationale for incorporating REs in the analysis aimed to estimate the PTE.

7.3 PTE Estimation – Mixed and Bayesian Model Analysis

Mixed and Bayesian Models are essentially equivalent. The Mixed Model has the advantage of providing unbiased and objective estimates of the PTE. The PTE precision is also better if compared with the Bayesian estimates. With regards to the estimation of Treatment RE variance, the Mixed Model estimation suffers from small sample limitations. The covariance design matrix may become degenerate, at least in SAS PROC MIXED, with highly unstable results. This is less of a problem for reasonably sized samples and / or sizeable τ_{β}^2 values.

If the τ_{β}^2 values are small but not negligible, and the sample is very restricted in size, the Bayesian Model is a better approach. Knowledge of a relatively small τ_{β}^2 makes it possible to design appropriate informative priors for at least this parameter.

The simulations carried out in chapter 6 also show that the Bayesian PTE estimates can converge on average fairly rapidly towards the population estimate. The simulations also suggest that in appropriate conditions the PTE credible intervals will contain the population parameter.

On these bases Bayes Models can be usefully applied to RCTs based on pooled N-of-1 trials for the estimation of PTE. That said, informative priors should be identified for at least the Treatment RE variance and ideally also for the PTE. Such priors speed up convergence to population values.

7.4 Use of SAS PROC MIXED and SAS PROC MCMC

The application of SAS procedures to this type of analysis has highlighted a number of issues concerning SAS PROC MIXED. On the positive side, SAS PROC MIXED is a fairly flexible analytical procedure. It can perform both FE and RE modes of analysis. The covariance specifications will allow for more complex correlation structures in the data. For example, temporal decay and / or longitudinal effects if there is reason to believe that the washout procedures in the N-of-1 trial have not been effective.

On the negative side, small sample estimation of RE is quite unstable and it may require unrealistic covariance specifications, plus specific instructions to ignore lower bound restrictions to the estimation of variance. An alternative method (*R lmer*) provides matching estimates without restrictions. Given these advantages and disadvantages, the selection of SAS PROC MIXED may need to be justified mostly in terms of its general acceptance for the analysis of RCT trials. However, if the drawbacks are revealed as being too great the researchers may want to consider alternative packages such as R or STATA.

SAS PROC MCMC is user-friendly and it is quite flexible. It is possible to specify hyperpriors directly as random effects, which makes it easy for a researcher to replicate SAS PROC MIXED specifications in SAS PROC MCMC. Lessafre and Lawson [17] note that the recent introduction of the ‘random’ option in version 9.3 makes SAS PROC MCMC comparable in performance to WinBugs.

One disadvantage is that dynamic, interactive use of outputs to drive subsequent analysis is not a straightforward process in SAS. SAS procedures’ results are exported in string based files. Extraction and further processing of numerical data from these files require the use of additional SAS PROC routines. In fact vector and matrix treatment of data requires SAS PROC IML, which is in fact a separate programming language within SAS. More complex and challenging simulations should be addressed by a combination of R and WinBugs.

Appendix

A.1 Chapter 5 Codes

A.1.1 Section 5.1 - SAS PROC MIXED for Fixed Effects Linear Model

```
*/ AUTHOR GIULIO FLORE 10/01/2015;

*/ -----;
*/
*/ FE Model - single subject analysis and sample analysis;
*/
*/ -----;

*/ Linear Model approach - period is omitted because of occasional matching with treatment;
*/ MODEL 1;

%let contrast = %sysevalf(1/15);          * contrast weight - required to feed fractional value;

proc mixed data = thesis.long_2 method = MIVQUE0; * LS like estimation method;
class id treatment (ref = first);          * declare categorical variables;
model facit_f = id*treatment / noint s htype = 1; * declare model, require FE estimates, p values and CIs;

*/ require TTE as contrast of interactions;

estimate 'TTE' id*treatment
    &contrast -&contrast &contrast -&contrast &contrast -&contrast &contrast -&contrast &contrast -&contrast
    &contrast -&contrast &contrast -&contrast &contrast -&contrast &contrast -&contrast &contrast -&contrast
    &contrast -&contrast &contrast -&contrast &contrast -&contrast &contrast -&contrast &contrast -&contrast / cl;

lsmeans id*treatment / pdiff cl;          * extract FE ITE, p values and CIs;
ods output diffs = thesis.fixed_ite;      * output results into files;
ods output estimates = thesis.fixed_tte;
run;

*/ reformat output to select only appropriate within subject differences;

data thesis.fixed;
set thesis.fixed;
drop treatment _treatment;
run;

data thesis.fixed_ite;
set thesis.fixed_ite;
if ID ^= _ID then delete;
drop Effect_ID treatment _treatment;
rename Estimate = ITE;
run;

*/ MODEL 2 - Common Treatment Effect Model for comparison purposes;

proc mixed data = thesis.long_2 method = MIVQUE0; * LS like estimation method;
class id treatment (ref = first);
model facit_f = id treatment / s htype = 1;          * sequential type 1 anova test;
lsmeans treatment / pdiff cl;                      * extract FE estimates, p values and CIs;
ods output diffs = thesis.fixed_cte;
run;

*/ Build caterpillar simulation;
*/ Use thesis.fixed_ite and thesis.fixed_tte as basis of simulation;
*/ Extract information to simulate data;

data thesis.fixed_ite;                               * Open file;
set thesis.fixed_ite;

    if ID = 3917 then do;
        call symput('Effect_3917',ite);          * For the matching row of id = 3917;
        call symput('Effect_3917_se',stderr);    * assign to Effect_3917 the value in variable 'ite';
    end;                                           * assign to Effect_3917_se the value in variable 'stderr';
                                                    * Close 'do' instruction;
```

```
if ID = 6580 then do;
    call symput('Effect_6580',ite);
    call symput('Effect_6580_se',stderr);
end;

if ID = 12496 then do;
    call symput('Effect_12496',ite);
    call symput('Effect_12496_se',stderr);
end;

if ID = 21719 then do;
    call symput('Effect_21719',ite);
    call symput('Effect_21719_se',stderr);
end;

if ID = 24833 then do;
    call symput('Effect_24833',ite);
    call symput('Effect_24833_se',stderr);
end;

if ID = 27096 then do;
    call symput('Effect_27096',ite);
    call symput('Effect_27096_se',stderr);
end;

if ID = 30328 then do;
    call symput('Effect_30328',ite);
    call symput('Effect_30328_se',stderr);
end;

if ID = 37620 then do;
    call symput('Effect_37620',ite);
    call symput('Effect_37620_se',stderr);
end;

if ID = 46198 then do;
    call symput('Effect_46198',ite);
    call symput('Effect_46198_se',stderr);
end;

if ID = 51969 then do;
    call symput('Effect_51969',ite);
    call symput('Effect_51969_se',stderr);
end;

if ID = 57930 then do;
    call symput('Effect_57930',ite);
    call symput('Effect_57930_se',stderr);
end;

if ID = 71386 then do;
    call symput('Effect_71386',ite);
    call symput('Effect_71386_se',stderr);
end;

if ID = 72068 then do;
    call symput('Effect_72068',ite);
    call symput('Effect_72068_se',stderr);
end;

if ID = 95477 then do;
    call symput('Effect_95477',ite);
    call symput('Effect_95477_se',stderr);
end;

if ID = 98875 then do;
    call symput('Effect_98875',ite);
    call symput('Effect_98875_se',stderr);
end;
```

```
run;
```

```

data thesis.fixed_tte;
set thesis.fixed_tte;
  if label = "TTE" then do;
    call symput('Drug',estimate);
    call symput('Drug_se',stderr);
  end;
run;

*/ Derived from example: http://blogs.sas.com/content/iml/2014/06/04/simulate-lognormal-data-with-specified-mean-and-variance/;

data thesis.cater_3917;
call streaminit(1);
keep effect_3917;
mean = &Effect_3917; se = &Effect_3917_se;
  do i = 1 to 10000;
    effect_3917 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_6580;
call streaminit(1);
keep effect_6580;
mean = &Effect_6580; se = &Effect_6580_se;
  do i = 1 to 10000;
    effect_6580 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_12496;
call streaminit(1);
keep effect_12496;
mean = &Effect_12496; se = &Effect_12496_se;
  do i = 1 to 10000;
    effect_12496 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_21719;
call streaminit(1);
keep effect_21719;
mean = &Effect_21719; se = &Effect_21719_se;
  do i = 1 to 10000;
    effect_21719 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_24833;
call streaminit(1);
keep effect_24833;
mean = &Effect_24833; se = &Effect_24833_se;
  do i = 1 to 10000;
    effect_24833 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_27096;
call streaminit(1);
keep effect_27096;
mean = &Effect_27096; se = &Effect_27096_se;
  do i = 1 to 10000;
    effect_27096 = rand('Normal', mean, se);
    output;
  end;
run;

```

```

data thesis.cater_30328;
call streaminit(1);
keep effect_30328;
mean = &Effect_30328; se = &Effect_30328_se;
  do i = 1 to 10000;
    effect_30328 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_37620;
call streaminit(1);
keep effect_37620;
mean = &Effect_37620; se = &Effect_37620_se;
  do i = 1 to 10000;
    effect_37620 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_46198;
call streaminit(1);
keep effect_46198;
mean = &Effect_46198; se = &Effect_46198_se;
  do i = 1 to 10000;
    effect_46198 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_51969;
call streaminit(1);
keep effect_51969;
mean = &Effect_51969; se = &Effect_51969_se;
  do i = 1 to 10000;
    effect_51969 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_57930;
call streaminit(1);
keep effect_57930;
mean = &Effect_57930; se = &Effect_57930_se;
  do i = 1 to 10000;
    effect_57930 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_71386;
call streaminit(1);
keep effect_71386;
mean = &Effect_71386; se = &Effect_71386_se;
  do i = 1 to 10000;
    effect_71386 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_72068;
call streaminit(1);
keep effect_72068;
mean = &Effect_72068; se = &Effect_72068_se;
  do i = 1 to 10000;
    effect_72068 = rand('Normal', mean, se);
    output;
  end;
run;

```

```

data thesis.cater_95477;
call streaminit(1);
keep effect_95477;
mean = &Effect_95477; se = &Effect_95477_se;
    do i = 1 to 10000;
        effect_95477 = rand('Normal', mean, se);
        output;
    end;
run;

data thesis.cater_98875;
call streaminit(1);
keep effect_98875;
mean = &Effect_98875; se = &Effect_95477_se;
    do i = 1 to 10000;
        effect_98875 = rand('Normal', mean, se);
        output;
    end;
run;

data thesis.cater_drug;
call streaminit(1);
keep Drug;
mean = &Drug; se = &Drug_se;
    do i = 1 to 10000;
        Drug = rand('Normal', mean, se);
        output;
    end;
run;

*/ Merge files for use in %CATER macro/;

data thesis.fe_effects;
set thesis.cater_3917; set thesis.cater_6580; set thesis.cater_12496; set thesis.cater_21719;
set thesis.cater_24833; set thesis.cater_27096; set thesis.cater_30328; set thesis.cater_37620;
set thesis.cater_46198; set thesis.cater_51969; set thesis.cater_57930; set thesis.cater_71386;
set thesis.cater_72068; set thesis.cater_95477; set thesis.cater_98875; set thesis.cater_drug;
run;

*/ caterpillar plot;

%CATER(data = thesis.fe_effects, var = effect_3917
        effect_6580
        effect_12496
        effect_21719
        effect_24833
        effect_27096
        effect_30328
        effect_37620
        effect_46198
        effect_51969
        effect_57930
        effect_71386
        effect_72068
        effect_95477
        effect_98875
        drug);

```

A.1.2 Section 5.2 - SAS PROC MIXED for Random Effects Linear Mixed Model

```
*/ AUTHOR GIULIO FLORE 10/01/2015;

*/ MODEL 1.1 - random intercept and treatment - require diagonal covariance for matrix G ('vc' spec);

proc mixed data = thesis.long_2 covtest method = reml nobound; /* unbounded estimation of variances;
class id; /* declare categorical value;
model facit_f = treatment / s cl htype = 1; /* declare model, require FE estimates, p values and CIs;
random int treatment / sub = id type = vc s g cl; /* declare re on subjects, re covariance matrix, statistics;

ods output SolutionF = thesis.fixed_15; /* output results into files;
ods output SolutionR = thesis.random_15;
ods output CovParms = thesis.varcov_15;

run;

*/ MODEL 1.2 - random intercept and treatment - require unstructures covariance for matrix G ('un' spec);

proc mixed data = thesis.long_2 covtest method = reml nobound;
class id;
model facit_f = treatment / s cl htype = 1;
random int treatment / sub = id type = un s g cl;

ods output SolutionF = thesis.fixed_15_un;
ods output SolutionR = thesis.random_15_un;
ods output CovParms = thesis.varcov_15_un;

run;

*/ MODEL 2 - random treatment only;

proc mixed data = thesis.long_2 covtest method = reml nobound;
class id;
model facit_f = treatment / s cl htype = 1;
random treatment / sub = id type = un s g cl;

ods output SolutionF = thesis.fixed_15_treatment;
ods output SolutionR = thesis.random_15_treatment;
ods output CovParms = thesis.varcov_15_treatment;

run;

*/ MODEL 3 - random intercept only;

proc mixed data = thesis.long_2 covtest method = reml nobound;
class id;
model facit_f = treatment / s cl htype = 1;
random int / sub = id type = vc s g cl;
ods output SolutionF = thesis.fixed_15_int;
ods output SolutionR = thesis.random_15_int;
ods output CovParms = thesis.varcov_15_int;
run;

*/ Add ITE estimates;

*/ select treatment PE estimate;

data thesis.fixed;
set thesis.fixed_15;
if Effect ^= 'treatment' then delete;
run;

*/ extract relevant information;

data thesis.fixed; * extracting PE estimates;
set thesis.fixed;
if Effect = 'treatment' then do;
call symput('slope_pe',Estimate);
call symput('slope_pe_sd',Stderr);
end;
run;
```

```

*/ prepare file with PE data;

proc iml;
pe = j(15,2,&slope_pe);
pe[,2] = j(15,1,&slope_pe_sd);
create thesis.fixed_15_pe from pe;
append from pe;
quit;

data thesis.fixed_15_pe;
set thesis.fixed_15_pe;
rename col1 = PE;
rename col2 = PE_SD;
run;

*/ prepare ite file;

data thesis.estimate;
set thesis.random_15;
if Effect ^= 'treatment' then delete;
run;

*/ Merge data;
*/ Critical - the p-val and CI relies on a Z-test, simply because the linear combination of;
*/ T Distributed random variables is not, itself, T-distributed. As first approximation;
*/ we use a Z-test, this being the least bad assumption, by the fact that the;
*/ distributional assumptions for Beta_re is for normally distributed variables, beta being a constant;
*/ This approach is nevertheless inferior compared to the testing for the FE and Bayesian models;

data thesis.estimate;
set thesis.estimate;
set thesis.fixed_15_pe;

ITE = Estimate + PE;
ITE_SD = (Stderrpred**2+pe_sd**2)**0.5;
ITE_ZVal = quantile('NORMAL',CDF('NORMAL',ITE,0,ITE_SD));
ITE_PVAl = 2*(1-CDF('NORMAL',ITE,0,ITE_SD));
ITE_CI025 = ITE - quantile('NORMAL',0.975)*ITE_SD;
ITE_CI975 = ITE + quantile('NORMAL',0.975)*ITE_SD;

run;

*/ Build caterpillar simulation;
*/ Use thesis.estimate as basis of simulation;

data thesis.estimate;
set thesis.estimate;

if ID = 3917 then do;
call symput('Effect_3917',ite);
call symput('Effect_3917_se',ite_sd);
end;

if ID = 6580 then do;
call symput('Effect_6580',ite);
call symput('Effect_6580_se',ite_sd);
end;

if ID = 12496 then do;
call symput('Effect_12496',ite);
call symput('Effect_12496_se',ite_sd);
end;

if ID = 21719 then do;
call symput('Effect_21719',ite);
call symput('Effect_21719_se',ite_sd);
end;

if ID = 24833 then do;
call symput('Effect_24833',ite);
call symput('Effect_24833_se',ite_sd);
end;

```

```

if ID = 27096 then do;
    call symput('Effect_27096',ite);
    call symput('Effect_27096_se',ite_sd);
end;

if ID = 30328 then do;
    call symput('Effect_30328',ite);
    call symput('Effect_30328_se',ite_sd);
end;

if ID = 37620 then do;
    call symput('Effect_37620',ite);
    call symput('Effect_37620_se',ite_sd);
end;

if ID = 46198 then do;
    call symput('Effect_46198',ite);
    call symput('Effect_46198_se',ite_sd);
end;

if ID = 51969 then do;
    call symput('Effect_51969',ite);
    call symput('Effect_51969_se',ite_sd);
end;

if ID = 57930 then do;
    call symput('Effect_57930',ite);
    call symput('Effect_57930_se',ite_sd);
end;

if ID = 71386 then do;
    call symput('Effect_71386',ite);
    call symput('Effect_71386_se',ite_sd);
end;

if ID = 72068 then do;
    call symput('Effect_72068',ite);
    call symput('Effect_72068_se',ite_sd);
end;

if ID = 95477 then do;
    call symput('Effect_95477',ite);
    call symput('Effect_95477_se',ite_sd);
end;

if ID = 98875 then do;
    call symput('Effect_98875',ite);
    call symput('Effect_98875_se',ite_sd);
end;

run;

*/ Extract info for comparison value;

data thesis.fixed_15;
set thesis.fixed_15;
    if effect = "treatment" then do;
        call symput('Drug',estimate);
        call symput('Drug_se',stderr);
    end;
run;

*/ Derived from example: http://blogs.sas.com/content/iml/2014/06/04/simulate-lognormal-data-with-specified-mean-and-variance/;

data thesis.cater_mixed_3917;
call streaminit(1);
keep effect_3917;
mean = &Effect_3917; se = &Effect_3917_se;
do i = 1 to 10000;
    effect_3917 = rand('Normal', mean, se);
    output;
end;
run;

```

```

data thesis.cater_mixed_6580;
call streaminit(1);
keep effect_6580;
mean = &Effect_6580; se = &Effect_6580_se;
  do i = 1 to 10000;
    effect_6580 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_mixed_12496;
call streaminit(1);
keep effect_12496;
mean = &Effect_12496; se = &Effect_12496_se;
  do i = 1 to 10000;
    effect_12496 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_mixed_21719;
call streaminit(1);
keep effect_21719;
mean = &Effect_21719; se = &Effect_21719_se;
  do i = 1 to 10000;
    effect_21719 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_mixed_24833;
call streaminit(1);
keep effect_24833;
mean = &Effect_24833; se = &Effect_24833_se;
  do i = 1 to 10000;
    effect_24833 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_mixed_27096;
call streaminit(1);
keep effect_27096;
mean = &Effect_27096; se = &Effect_27096_se;
  do i = 1 to 10000;
    effect_27096 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_mixed_30328;
call streaminit(1);
keep effect_30328;
mean = &Effect_30328; se = &Effect_30328_se;
  do i = 1 to 10000;
    effect_30328 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_mixed_37620;
call streaminit(1);
keep effect_37620;
mean = &Effect_37620; se = &Effect_37620_se;
  do i = 1 to 10000;
    effect_37620 = rand('Normal', mean, se);
    output;
  end;
run;

```

```

data thesis.cater_mixed_46198;
call streaminit(1);
keep effect_46198;
mean = &Effect_46198; se = &Effect_46198_se;
  do i = 1 to 10000;
    effect_46198 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_mixed_51969;
call streaminit(1);
keep effect_51969;
mean = &Effect_51969; se = &Effect_51969_se;
  do i = 1 to 10000;
    effect_51969 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_mixed_57930;
call streaminit(1);
keep effect_57930;
mean = &Effect_57930; se = &Effect_57930_se;
  do i = 1 to 10000;
    effect_57930 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_mixed_71386;
call streaminit(1);
keep effect_71386;
mean = &Effect_71386; se = &Effect_71386_se;
  do i = 1 to 10000;
    effect_71386 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_mixed_72068;
call streaminit(1);
keep effect_72068;
mean = &Effect_72068; se = &Effect_72068_se;
  do i = 1 to 10000;
    effect_72068 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_mixed_95477;
call streaminit(1);
keep effect_95477;
mean = &Effect_95477; se = &Effect_95477_se;
  do i = 1 to 10000;
    effect_95477 = rand('Normal', mean, se);
    output;
  end;
run;

data thesis.cater_mixed_98875;
call streaminit(1);
keep effect_98875;
mean = &Effect_98875; se = &Effect_95477_se;
  do i = 1 to 10000;
    effect_98875 = rand('Normal', mean, se);
    output;
  end;
run;

```

```

data thesis.cater_mixed_drug;
call streaminit(1);
keep Drug;
mean = &Drug; se = &Drug_se;
    do i = 1 to 10000;
        Drug = rand('Normal', mean, se);
        output;
    end;
run;

*/ Merge files for use in %CATER macro/;

data thesis.re_effects;
set thesis.cater_mixed_3917; set thesis.cater_mixed_6580; set thesis.cater_mixed_12496;
set thesis.cater_mixed_21719; set thesis.cater_mixed_24833; set thesis.cater_mixed_27096;
set thesis.cater_mixed_30328; set thesis.cater_mixed_37620; set thesis.cater_mixed_46198;
set thesis.cater_mixed_51969; set thesis.cater_mixed_57930; set thesis.cater_mixed_71386;
set thesis.cater_mixed_72068; set thesis.cater_mixed_95477; set thesis.cater_mixed_98875;
set thesis.cater_mixed_drug;
run;

*/ plots;

%CATER(data=thesis.re_effects, var = effect_3917
effect_6580
effect_12496
effect_21719
effect_24833
effect_27096
effect_30328
effect_37620
effect_46198
effect_51969
effect_57930
effect_71386
effect_72068
effect_95477
effect_98875
drug); */ caterpillar plot for random drugs;

```

A.1.3 Section 5.3 - SAS PROC MCMC for Bayesian Model

```
*/ AUTHOR GIULIO FLORE 10/01/2015;

*/ MODEL 1 - Random Intercept and treatment drug model;

*/ In this case we are not making any strong claim concerning the true values;
*/ the PTE prior assumes a 0 value and a large variance;
*/ skepticism expressed by not taking any view on whether the effect is on average
*/ negative, null or positive, and assuming low absolute values as more likely;

ods graphics on;

*/ All posteriors samplings exported in file thesis.long_mcmc;
*/ 1 M chains (nmc)
*/ Burn-in (ntu) set at 50,000 chains;
*/ Sampling off chains 1 every 200 (thin) for N = 5,000 posterior sample;
*/ Quasi-NR for covariance estimation (propcov);

proc mcmc data=thesis.long_2
  outpost=thesis.long_mcmc
  nmc = 1000000
  ntu = 50000
  monitor = (_PARMS_ LOGLIKE LOGPOST LOGPRIOR)
  propcov = quanev
  thin = 200
  diagnostics = all
  plots = all
  seed = 25041945
  DIC
  statistics(percentage=(2.5 5 50 95 97.5)) = all;
  ods select dic Parameters PostSummaries PostIntervals;
  parms intercept 0 drug 0 sigma2 1 tau2_int 1 tau2_drug 1; /* initial value assumptions;

*/ A suitable prior for the intercept can be set bearing in mind the nature of the FACIT-F score;
*/ Hence we set the midpoint as mean for the prior and the var to conditioned by the range of the values;
*/ (0-52). So a SD can be set to have most of the mass of the distribution within the range;
*/ Basic checks shows that a sd = 8 (var = 64) is perfectly adequate;
*/ This is a prior with an initial large variance for the fixed effects;

  prior intercept ~ normal(26, var = 64);

*/ A skeptic prior for the drug is set to give;
*/ equal chances to the effect to be positive or negative. By selecting a normal;
*/ distribution centred on the 0 we also express a skeptic view, e.g. low absolute;
*/ values are judged more likely than higher values. Thus we set the mean at 0;
*/ and the variance at 64, of the same order as the FE intercept;

  prior drug ~ normal( 0, var = 64);

*/ A partially non-informative prior for variance components is set by following the suggestion by Gelman;
*/ (Gelman 2006) whereby a U(0,b) is a suitable solution for a hierarchical model with small numbers
*/ of level units. In the same paper Gelman advises against using a common alternative represented;
*/ represented IG(a,b) with small (near 0) values, because they for few grouping units;
*/ the estimates are unstable and biased. We use values broadly consistent with the sample data variance;

  prior sigma2 ~ uniform(0,100);
  prior tau2_int ~ uniform(0,100);
  prior tau2_drug ~ uniform(0,100);

  random intercept2 ~ normal(0, var = tau2_int) subject = id monitor = (intercept2); /* set the re specification;
  random drug2 ~ normal(0, var = tau2_drug) subject = id monitor = (drug2);

  mu = intercept + intercept2 + drug * treatment + drug2*treatment; /* set a random intercept and drug model for mu;

  model facit_f ~ normal(mu, var = sigma2); /* declare mean model;

  ods output DIC = thesis.DIC; /* output statistics into files;
  ods output PostSummaries = thesis.Posterior_Theta_Long;
  ods output PostIntervals = thesis.Posterior_CI_Long;

run;
```

*/ check on the degree of correlation between the estimated parameters;

```
ods graphics on;  
proc corr data = thesis.long_mcmc plots=matrix(histogram);  
var intercept drug sigma2 tau2_int tau2_drug;  
run;  
ods graphics off;
```

*/ build posteriors of effects, ICC as mixtures of posteriors;

```
data thesis.long_mcmc;  
set thesis.long_mcmc;  
effect_3917 = drug + drug2_3917;  
effect_6580 = drug + drug2_6580;  
effect_12496 = drug + drug2_12496;  
effect_21719 = drug + drug2_21719;  
effect_24833 = drug + drug2_24833;  
effect_27096 = drug + drug2_27096;  
effect_30328 = drug + drug2_30328;  
effect_37620 = drug + drug2_37620;  
effect_46198 = drug + drug2_46198;  
effect_51969 = drug + drug2_51969;  
effect_57930 = drug + drug2_57930;  
effect_71386 = drug + drug2_71386;  
effect_72068 = drug + drug2_72068;  
effect_95477 = drug + drug2_95477;  
effect_98875 = drug + drug2_98875;  
ICC = tau2_drug/(tau2_drug+tau2_int+sigma2);  
run;
```

*/ Format results for effects: Export into file summaries of bayes estimates with empirical quantiles;

```
proc univariate data = thesis.long_mcmc NOPRINT PCTLDEF = 3;  
var effect_3917;  
output out = thesis.bayes_long_summaries1  
mean = Mean std = SE  
pctlpre=P_pctlpts = 2.5 5 50 95 97.5;  
run;
```

```
proc univariate data = thesis.long_mcmc NOPRINT PCTLDEF = 3;  
var effect_6580;  
output out = thesis.bayes_long_summaries2  
mean = Mean std = SE  
pctlpre=P_pctlpts = 2.5 5 50 95 97.5;  
run;
```

```
proc univariate data = thesis.long_mcmc NOPRINT PCTLDEF = 3;  
var effect_12496;  
output out = thesis.bayes_long_summaries3  
mean = Mean std = SE  
pctlpre=P_pctlpts = 2.5 5 50 95 97.5;  
run;
```

```
proc univariate data = thesis.long_mcmc NOPRINT PCTLDEF = 3;  
var effect_21719;  
output out = thesis.bayes_long_summaries4  
mean = Mean std = SE  
pctlpre=P_pctlpts = 2.5 5 50 95 97.5;  
run;
```

```
proc univariate data = thesis.long_mcmc NOPRINT PCTLDEF = 3;  
var effect_24833;  
output out = thesis.bayes_long_summaries5  
mean = Mean std = SE  
pctlpre=P_pctlpts = 2.5 5 50 95 97.5;  
run;
```

```
proc univariate data = thesis.long_mcmc NOPRINT PCTLDEF = 3;  
var effect_27096;  
output out = thesis.bayes_long_summaries6  
mean = Mean std = SE  
pctlpre=P_pctlpts = 2.5 5 50 95 97.5;  
run;
```

```
proc univariate data = thesis.long_mcmc NOPRINT PCTLDEF = 3;
var effect_30328;
output out = thesis.bayes_long_summaries7
    mean = Mean std = SE
    pctlpre=P_ pctlpts = 2.5 5 50 95 97.5;
run;
```

```
proc univariate data = thesis.long_mcmc NOPRINT PCTLDEF = 3;
var effect_37620;
output out = thesis.bayes_long_summaries8
    mean = Mean std = SE
    pctlpre=P_ pctlpts = 2.5 5 50 95 97.5;
run;
```

```
proc univariate data = thesis.long_mcmc NOPRINT PCTLDEF = 3;
var effect_46198;
output out = thesis.bayes_long_summaries9
    mean = Mean std = SE
    pctlpre=P_ pctlpts = 2.5 5 50 95 97.5;
run;
```

```
proc univariate data = thesis.long_mcmc NOPRINT PCTLDEF = 3;
var effect_51969;
output out = thesis.bayes_long_summaries10
    mean = Mean std = SE
    pctlpre=P_ pctlpts = 2.5 5 50 95 97.5;
run;
```

```
proc univariate data = thesis.long_mcmc NOPRINT PCTLDEF = 3;
var effect_57930;
output out = thesis.bayes_long_summaries11
    mean = Mean std = SE
    pctlpre=P_ pctlpts = 2.5 5 50 95 97.5;
run;
```

```
proc univariate data = thesis.long_mcmc NOPRINT PCTLDEF = 3;
var effect_71386;
output out = thesis.bayes_long_summaries12
    mean = Mean std = SE
    pctlpre=P_ pctlpts = 2.5 5 50 95 97.5;
run;
```

```
proc univariate data = thesis.long_mcmc NOPRINT PCTLDEF = 3;
var effect_72068;
output out = thesis.bayes_long_summaries13
    mean = Mean std = SE
    pctlpre=P_ pctlpts = 2.5 5 50 95 97.5;
run;
```

```
proc univariate data = thesis.long_mcmc NOPRINT PCTLDEF = 3;
var effect_95477;
output out = thesis.bayes_long_summaries14
    mean = Mean std = SE
    pctlpre=P_ pctlpts = 2.5 5 50 95 97.5;
run;
```

```
proc univariate data = thesis.long_mcmc NOPRINT PCTLDEF = 3;
var effect_98875;
output out = thesis.bayes_long_summaries15
    mean = Mean std = SE
    pctlpre=P_ pctlpts = 2.5 5 50 95 97.5;
run;
```

```
proc univariate data = thesis.long_mcmc NOPRINT PCTLDEF = 3;
var ICC;
output out = thesis.long_ICC
    mean = Mean std = SE
    pctlpre=P_ pctlpts = 2.5 5 50 95 97.5;
run;
```

```

*/ Add Bayesian test statistics to posterior summaries;

data thesis.posterior_theta_long;
set thesis.posterior_theta_long;
    P = 1-CDF('NORMAL',0,Mean,StdDev);          */ P(PTE>0);
    P_CLINICAL = 1-CDF('NORMAL',3,Mean,StdDev); */ P(PTE>3);
    LOG10_BF = LOG10(BF);                       */ BF Jeffreys version;

run;

*/ merge posterior summaries in single file and add bayesian test statistics;

data thesis.bayes_long_ite_summaries;
set thesis.bayes_long_summaries1 thesis.bayes_long_summaries2 thesis.bayes_long_summaries3
thesis.bayes_long_summaries4 thesis.bayes_long_summaries5 thesis.bayes_long_summaries6
thesis.bayes_long_summaries7 thesis.bayes_long_summaries8 thesis.bayes_long_summaries9
thesis.bayes_long_summaries10 thesis.bayes_long_summaries11 thesis.bayes_long_summaries12
thesis.bayes_long_summaries13 thesis.bayes_long_summaries14 thesis.bayes_long_summaries15;

P_ITE = 1-CDF('NORMAL',0,Mean,SE);          */ P(ITE>0);
BF_ITE = CDF('NORMAL',0,Mean,SE)/(1-CDF('NORMAL',0,Mean,SE)); */ BF against Ho: Treatment < 0;
LOG10_BF_ITE = LOG10(BF_ITE);              */ Jeffreys version;

label Mean = '';                            */ add new labels;
label SE = '';
label P_2_5 = '';
label P_5 = '';
label P_50 = '';
label P_95 = '';
label P_97_5 = '';

input ITE $char12. ;                        */ imput char variable with references;
datalines;
effect_3917
effect_6580
effect_12496
effect_21719
effect_24833
effect_27096
effect_30328
effect_37620
effect_46198
effect_51969
effect_57930
effect_71386
effect_72068
effect_95477
effect_98875
run;

*/ reformat DIC values;

proc transpose data = thesis.DIC
    out = thesis.DIC;
run;

DATA THESIS.DIC;
SET THESIS.DIC;
RENAME COL1 = MeanDev;    */ Mean Posterior Deviance;
RENAME COL2 = DevMean;   */ Deviance for Best (Mean) Parameter Estimate;
RENAME COL3 = Pd;       */ Effective Parameters;
RENAME COL4 = DIC;      */ DIC;
drop _NAME_;
RUN;

```

```

*/ plots - caterpillar plot for treatment effect;

%CATER(data=thesis.long_mcmc, var = effect_3917
      effect_6580
      effect_12496
      effect_21719
      effect_24833
      effect_27096
      effect_30328
      effect_37620
      effect_46198
      effect_51969
      effect_57930
      effect_71386
      effect_72068
      effect_95477
      effect_98875
      drug);

```

```

*/ convergence tests;

```

```

%GEWEKE(data=thesis.long_mcmc, var = intercept
      drug
      sigma2
      tau2_int
      tau2_drug
      drug2_3917
      drug2_6580
      drug2_12496
      drug2_21719
      drug2_24833
      drug2_27096
      drug2_30328
      drug2_37620
      drug2_46198
      drug2_51969
      drug2_57930
      drug2_71386
      drug2_72068
      drug2_95477
      drug2_98875
      ICC
      loglike
      logpost
      logprior);

```

```

*/ convergence plots;

```

```

%tadplot(data=thesis.long_mcmc, var = intercept
      drug
      sigma2
      tau2_int
      tau2_drug
      drug2_3917
      drug2_6580
      drug2_12496
      drug2_21719
      drug2_24833
      drug2_27096
      drug2_30328
      drug2_37620
      drug2_46198
      drug2_51969
      drug2_57930
      drug2_71386
      drug2_72068
      drug2_95477
      drug2_98875
      ICC
      loglike
      logpost
      logprior);

```

A.2 Chapter 6 Codes

A.2.1 Section 6.1 - SAS PROC MCMC for Simulation of Different Populations

```
*/ AUTHOR GIULIO FLORE 10/01/2015;

*/ Cycles: # of treatment cycles;
*/ Periods: # of treatment periods within a cycle;
*/ Patients: # of patients to be randomized;
*/ Randomizer: # randomization seed for repeatability;
*/ beta0: # fixed intercept of re model;
*/ beta1: # fixed drug of re model;
*/ sigma: # error term sd;
*/ tau_int: # random intercept sd;
*/ tau_drug: # random drug sd;

%macro simulator (Cycles,
                 Periods,
                 Patients,
                 Randomizer,
                 beta0,
                 beta1,
                 sigma,
                 tau_int,
                 tau_drug);

*/ macro trial assume a RE structure with random intercept and treatment;

options nonotes; */ exclude output;

data thesis.tmp;
put ID Cycle Period Treatment;
stop;
run;

*/ create file with sequence of random intercept values;

proc iml;
u = j(&Patients,1);
call randseed(&randomizer);
call randgen(u, "Normal",0,&tau_int);
create thesis.randomintercept from u;
append from u;
quit;
*/ Matrix language proc to generate matrix of random values;
*/ allocate vector with # of subjects;
*/ fix randomisation for repeatability;
*/ fill vecotr with random values;
*/ generate new file with randomized values;
*/ stop proc iml - specific command;

*/ add crossreference ID ;

data thesis.randomintercept;
set thesis.randomintercept;
id = _N_;
rename coll = intercept_re;
run;
*/ add index for merger;

*/ create file with sequence of random treatment values;

proc iml;
w = j(&Patients,1);
call randseed(&randomizer+1000);
call randgen(w, "Normal",0,&tau_drug);
create thesis.randomdrug from w;
append from w;
quit;

*/ add crossreference ID ;

data thesis.randomdrug;
set thesis.randomdrug;
id = _N_;
rename coll = drug_re;
run;
```

```

%do i = 1 %to &Patients;                               */ start loop through subjects;

%let random = %sysevalf(&i + &randomizer);

proc plan seed = &random;                               */ seed fix randomisation for repeatability;
  factors Cycle=&Cycles ordered Period=&Periods ordered / noprint; */ gives sequence of periods and cycles for individual;
  treatments Treatment=2 cyclic (1 2);                 */ random assignment to treatment set as dummy;
  output out = thesis.rand
    Treatment random;
run;

data thesis.tmp;                                       */ append new block of trial information;
set thesis.tmp thesis.rand;
if Cycle =. then delete;                               */ clear empty record;
run;

%end;                                                  */ close patient loop;

data thesis.tmp;                                       */ add id;
set thesis.tmp;
id = floor((_N_-1)/(&Cycles*&Periods))+1; */ assign id by blocks of 6;
run;

data thesis.tmp;                                       */ merge random effects;
set thesis.tmp;
merge thesis.tmp thesis.randomintercept thesis.randomdrug;
by id;
run;

data thesis.tmp;                                       */ add outcome value;
call streaminit(&randomizer+2000);                    */ enforce varying random starts;
set thesis.tmp;
treatment = treatment - 1;                             */ set treatment as dummy;
error = rand("Normal",0,&sigma);
facit_f = &beta0 + intercept_re + (treatment*&beta1)+(treatment*drug_re) + error; */ random values for outcome;
facit_f = round(facit_f,0.5);                          */ round value to reflect reported trial values;
run;

options notes;

%mend; */ close macro;

*/ Population: number of patients to be added in simulator;
*/ Cycles: treatment cycles;
*/ Periods: number of treatment periods;
*/ Initial: size of initial group of subjects;
*/ Seeding: randomization parameter for repeatability
*/ beta0, beta1, sigma, tau_int tau_drug simulator parameters;

%macro Bayes_Simulation ( Population,
                          Cycles,
                          Periods,
                          Initial,
                          Seeding,
                          beta0,
                          beta1,
                          sigma,
                          tau_int,
                          tau_drug);

options nonotes; */ suppress output;

*/ Set repository for precision data on ITE values;

data thesis.bayes_sim_ite;
put ITE ITE_SE P2_5 P5 P50 P95 P97_5;
stop;
run;

data thesis.bayes_sim_drug_re;
put Drug_RE Drug_RE_SE P2_5 P5 P50 P95 P97_5;
stop;
run;

```

```

*/ create initial sample;

%simulator(&Cycles,&Periods,&Initial,&Seeding,&beta0,&beta1,&sigma,&tau_int,&tau_drug);

*/ initial estimation with partially informative priors;

*/ initial block marked number of markov chains;
*/ thinning of simulators;
*/ covariance estimation assumptions;
*/ randomization initial value;
*/ outputting of statistics;
*/ outputting of chain estimates for total effect estimation;

*/ Crucial - small sample analysis requires long chains for convergence;
*/ of simulation. Lack of convergence produce biased posteriors;
*/ and distort the analysis;

*/ Apply 1M chains (nmc);
*/ Apply appropriate thinning (thin) to 'defeat' ACF and produce at least 5,000 samplings for convergence statistics;
*/ Apply higher burn-in (ntu): 50000;
*/ Apply proposal covariance (propcov) with quasi-NR algorithm ;

proc mcmc data=thesis.tmp outpost = thesis.mcmc

    nmc = 1000000
    thin = 200
    ntu = 50000
    DIC
    propcov = quanew
    monitor = ( _parms_ LOGLIKE LOGPOST LOGPRIOR)
    seed = &Seeding
    statistics(alpha = 0.05 percentage = (2.5 5 50 95 97.5) ) = all;

parms intercept 0 drug 0 sigma2 1 tau2_int 1 tau2_drug 1; */ speculative initial value assumptions;

*/ priors selected to reflect the nature of the data, with minimal assumptions;

prior intercept ~ normal(26, var = 64); */ this prior assumes the intercept to be in the mid range of facit-f;

prior drug ~ normal(0, var = 64); */ this is a moderately skeptic prior, with equipoise on direction of the treatment effect;

*/ Non informative prior for random effects variance following Gelman (2006);

prior sigma2 ~ uniform(0,100);
prior tau2_int ~ uniform(0,100);
prior tau2_drug ~ uniform(0,100);

*/ set the hierarchical structure;

random intercept2 ~ normal(0, var = tau2_int) subject = id monitor = (intercept2);
random drug2 ~ normal(0, var = tau2_drug) subject = id monitor = (drug2);

mu = intercept + intercept2 + drug * treatment + drug2 * treatment; */ set a random intercept and drug model;

model facit_f ~ normal(mu, var = sigma2);

*/ output file statements;

ods select DIC PostSummaries;
ods output DIC = THESIS.bayes_sim_dic;
ods output PostSummaries = thesis.post;

run;

data thesis.post; */ generate copies conditionally;
set thesis.post;
if _N_ > 5 then delete; */ only distributional parameters;
run;

```

```

*/ Extract sigma, tau intercept and tau treatment estimates;

data thesis.mcmc_random;
set thesis.mcmc;
sigma    = sigma2**0.5;
tau_int  = tau2_int**0.5;
tau_drug = tau2_drug**0.5;
run;

proc univariate data = thesis.mcmc_random NOPRINT PCTLDEF = 3; /* export into files Bayesian summaries;
var sigma;
output out = thesis.bayes_sim_sigma
      mean = Sigma
      std  = Sigma_SE
      pctlpre = P
      pctlpts = 2.5 5 50 95 97.5;
run;

proc univariate data = thesis.mcmc_random NOPRINT PCTLDEF = 3;
var tau_int;
output out = thesis.bayes_sim_tau_int
      mean = Tau_Intercept
      std  = Tau_Intercept_SE
      pctlpre = P
      pctlpts = 2.5 5 50 95 97.5;
run;

proc univariate data = thesis.mcmc_random NOPRINT PCTLDEF = 3;
var tau_drug;
output out = thesis.bayes_sim_tau_drug
      mean = Tau_Drug
      std  = Tau_Drug_SE
      pctlpre = P
      pctlpts = 2.5 5 50 95 97.5;
run;

*/ Transpose DIC results for stacking;

proc transpose data = THESIS.bayes_sim_dic
      out = THESIS.bayes_sim_dic;
run;

*****;
*/
*/ Set initial Parameters for analysis and derivation of statistics outside the loop */
*/
*****;

%let beta1_prior = 0;          /* derived from prior in first estimation;
%let beta1_sd_prior = 8;

*/ also set first priors for use in clinical BF evaluation;

%let beta1_first_prior = 0;
%let beta1_sd_first_prior = 8;

data thesis.post;
set thesis.post;
if Parameter = 'intercept' then do;
      call symput('beta0_prior',mean);
      call symput('beta0_sd_prior',stddev);
end;
else
if Parameter = 'drug' then do;
      call symput('beta1_posterior',mean);          /* get a posterior value for BF estimation;
      call symput('beta1_sd_posterior',stddev);
end;
else
if Parameter = 'sigma2' then do;
      call symput('sigma2_prior',mean);
      call symput('sigma2_sd',stddev);
end;
else

```

```

if Parameter = 'tau2_int' then do;
    call symput('tau2_int_prior',mean);
    call symput('tau2_int_sd',stddev);
end;

else
if Parameter = 'tau2_drug' then do;
    call symput('tau2_drug_prior',mean);
    call symput('tau2_drug_sd',stddev);
end;

run;

*/ The BF needs to be interpreted in relation to the selected prior. If the BF is set against;
*/ the first proposed prior, each new subject inclusion implies a testing of the Null Hypothesis;
*/ as expressed in the first prior - the BF reduces to the posterior odds, as the prior odds are set to 1;

data thesis.post;
set thesis.post;
BF = 0;
if Parameter = 'drug' then
do;
BF = (CDF('NORMAL',0,
    &beta1_posterior,
    &beta1_sd_posterior))/(1-CDF('NORMAL',0,
    &beta1_posterior,
    &beta1_sd_posterior));

end;
run;

*/ derive additional macro variables for execution;
*/ %sysevalf performs floating point calculations;

%let beta1_prior = &beta1_posterior; /* update prior to posterior for feedback;
%let beta1_sd_prior = &beta1_sd_posterior; /* update prior to posterior for feedback;

*/ set variance for prior update;

%let beta0_var_prior = %SYSEVALF(&beta0_sd_prior**2);
%let beta1_var_prior = %SYSEVALF(&beta1_sd_prior**2);

*/ derive uniform parameters for proc mcmc hyperprior;

%let sigma2_prior_a = %SYSEVALF(&sigma2_prior-(&sigma2_sd*(3**0.5)));
%let sigma2_prior_b = %SYSEVALF(&sigma2_prior+(&sigma2_sd*(3**0.5)));

%let tau2_int_prior_a = %SYSEVALF(&tau2_int_prior-(&tau2_int_sd*(3**0.5)));
%let tau2_int_prior_b = %SYSEVALF(&tau2_int_prior+(&tau2_int_sd*(3**0.5)));

%let tau2_drug_prior_a = %SYSEVALF(&tau2_drug_prior-(&tau2_drug_sd*(3**0.5)));
%let tau2_drug_prior_b = %SYSEVALF(&tau2_drug_prior+(&tau2_drug_sd*(3**0.5)));

*/ NEW OBSERVATION INCREASE LOOP;

%do k = 1 %to &Population;

%put Cycle #&k; /* mark loops;

*/ generate a n = 1 sample - randomization parameter linked to loop index for repeatability;

%simulator(&Cycles,&Periods,1,&k,&beta0,&beta1,&sigma,&tau_int,&tau_drug);

*/ analyze new observation using the first analysis posterior as prior;
*/ initial block marked number of Markov chains;
*/ thinning of simulators;
*/ covariance estimation assumptions;
*/ randomization initial value;
*/ outputting of statistics;
*/ randomization linked to loop index for repeatability;

*/ based on analysis of real data ACF curtailed by large burn-in;
*/ for speed set shorter chain after first run with non-informative prior;

*/ Set thinning at lag = 20 to avoid ACF effects on the estimate;

```

```

*/ while obtaining at least 4,000 observations for convergence tests;

proc mcmc data = thesis.tmp outpost = thesis.mcmc

    nmc    = 80000
    thin   = 20
    ntu    = 50000
    propcov = quanev
    monitor = ( _parms_ LOGLIKE LOGPOST LOGPRIOR)
    seed = &k
    DIC
    statistics(alpha=0.05 percentage = (2.5 5 50 95 97.5) ) = all;

*/ parameter declaration and initial values statements based on priors to speed up convergence;

parms intercept &beta0_prior
drug    &beta1_prior
sigma2  &sigma2_prior
tau2_int &tau2_int_prior
tau2_drug &tau2_drug_prior;

*/ these priors are based on the previous cycle posteriors;

prior intercept ~ normal(&beta0_prior, var = &beta0_var_prior);
prior drug     ~ normal(&beta1_prior, var = &beta1_var_prior);
prior sigma2   ~ uniform(&sigma2_prior_a, &sigma2_prior_b);
prior tau2_int ~ uniform(&tau2_int_prior_a, &tau2_int_prior_b);
prior tau2_drug ~ uniform(&tau2_drug_prior_a, &tau2_drug_prior_b);

*/ random part statement - subject as grouping unit;

random intercept2 ~ normal(0, var = tau2_int) subject = id monitor = (intercept2);
random drug2     ~ normal(0, var = tau2_drug) subject = id monitor = (drug2);

*/ model mean statement;

mu = intercept + intercept2 + (drug * treatment) + (drug2 * treatment);
model facit_f ~ normal(mu, var = sigma2);

*/ output file statements;

ods select DIC PostSummaries;
ods output DIC = thesis.tmp_DIC;
ods output PostSummaries = thesis.tmp_posterior_theta;

run;

*/ extract variance posteriors and redefine them as SD and SD s.e.;

data thesis.mcmc_random;
set thesis.mcmc;
sigma = sigma2**0.5;
tau_int = tau2_int**0.5;
tau_drug = tau2_drug**0.5;
run;

proc univariate data = thesis.mcmc_random NOPRINT PCTLDEF = 3;
var sigma;
output out = thesis.sigma_tmp
    mean = Sigma
    std = Sigma_SE
    pctlpre = P
    pctlpts = 2.5 5 50 95 97.5;
run;

proc univariate data = thesis.mcmc_random NOPRINT PCTLDEF = 3;
var tau_int;
output out = thesis.tau_int_tmp
    mean = Tau_Intercept
    std = Tau_Intercept_SE
    pctlpre = P
    pctlpts = 2.5 5 50 95 97.5;
run;

```

```

proc univariate data = thesis.mcmc_random NOPRINT PCTLDEF = 3;
var tau_drug;
output out = thesis.tau_drug_tmp
  mean = Tau_Drug
  std = Tau_Drug_SE
  pctlpre = P
  pctlpts = 2.5 5 50 95 97.5;
run;

*/ combine random variance as sd estimates;

data thesis.bayes_sim_sigma;
set thesis.bayes_sim_sigma thesis.sigma_tmp;
run;

*/ Stack the data;

data thesis.bayes_sim_tau_int;
set thesis.bayes_sim_tau_int thesis.tau_int_tmp;
run;

data thesis.bayes_sim_tau_drug;
set thesis.bayes_sim_tau_drug thesis.tau_drug_tmp;
run;

*/ extract ite distribution as mixture;

data thesis.mcmc_ite;
set thesis.mcmc;
ITE = DRUG + DRUG2_1;
KEEP DRUG2_1 ITE;
run;

proc univariate data = thesis.mcmc_ite NOPRINT PCTLDEF = 3;
var ite;
output out = thesis.ite_tmp
  mean = ITE
  std = ITE_SE
  pctlpre = P
  pctlpts = 2.5 5 50 95 97.5;
run;

proc univariate data = thesis.mcmc_ite NOPRINT PCTLDEF = 3;
var drug2_1;
output out = thesis.drug_re_tmp
  mean = Drug_RE
  std = Drug_RE_SE
  pctlpre = P
  pctlpts = 2.5 5 50 95 97.5;
run;

*/ combine re estimate and ite estimate;

data thesis.bayes_sim_ite;
set thesis.bayes_sim_ite thesis.ite_tmp;
run;

*/ Stack the data;

data thesis.bayes_sim_drug_re;
set thesis.bayes_sim_drug_re thesis.drug_re_tmp;
run;

*/ Update DIC output;

proc transpose data = thesis.tmp_DIC
  out = thesis.tmp_DIC;
run;

data THESIS.bayes_sim_dic;
set THESIS.bayes_sim_dic thesis.tmp_DIC;
run;

```

```

*/ Update data repository of posteriors ;

data thesis.tmp_post;          */ generate copies conditionally;
set thesis.tmp_posterior_theta;
if _N_ > 5 then delete;      */ only distributional parameters;
run;

*****;
*/                               *;
*/ Update all relevant posterior parameters *;
*/                               *;
*****;

data thesis.tmp_post;
set thesis.tmp_post;

if Parameter = 'intercept' then do;
    call symput('beta0_prior',mean);
    call symput('beta0_sd_prior',stddev);
end;

else

if Parameter = 'drug' then do;
    call symput('beta1_posterior',mean);
    call symput('beta1_sd_posterior',stddev);
end;
*/ get a posterior value for BF estimation;

else

if Parameter = 'sigma2' then do;
    call symput('sigma2_prior',mean);
    call symput('sigma2_sd',stddev);
end;

else

if Parameter = 'tau2_int' then do;
    call symput('tau2_int_prior',mean);
    call symput('tau2_int_sd',stddev);
end;

else

if Parameter = 'tau2_drug' then do;
    call symput('tau2_drug_prior',mean);
    call symput('tau2_drug_sd',stddev);
end;

run;

*/ The BF needs to be interpreted in relation to the selected prior. If the BF is set against;
*/ the first proposed prior, each new subject inclusion implies a testing of the Null Hypothesis;
*/ as expressed in the first prior;
*/ the prior odd calculations are still expressed to made the formula explicit ;

data thesis.tmp_post;
set thesis.tmp_post;
BF = 0;
if Parameter = 'drug' then
do;
BF = (CDF('NORMAL',0,
    &beta1_posterior,
    &beta1_sd_posterior))/(1-CDF('NORMAL',0,
    &beta1_posterior,
    &beta1_sd_posterior)) ;

end;
run;

*/ derive additional macro variables for execution;
*/ %sysevalf performs floating point calculations;

%let beta1_prior = &beta1_posterior;          */ update prior to posterior for feedback;
%let beta1_sd_prior = &beta1_sd_posterior;    */ update prior to posterior for feedback;

%let beta0_var_prior = %SYSEVALF(&beta0_sd_prior**2);  */ set variance update for prior;
%let beta1_var_prior = %SYSEVALF(&beta1_sd_prior**2);

```

```

%let sigma2_prior_a = %SYSEVALF(&sigma2_prior-(&sigma2_sd*(3**0.5))); /* derive a parameter for proc mcmc hyperprior;
%let sigma2_prior_b = %SYSEVALF(&sigma2_prior+(&sigma2_sd*(3**0.5))); /* derive b parameter for proc mcmc hyperprior;

%let tau2_int_prior_a = %SYSEVALF(&tau2_int_prior-(&tau2_int_sd*(3**0.5)));
%let tau2_int_prior_b = %SYSEVALF(&tau2_int_prior+(&tau2_int_sd*(3**0.5)));

%let tau2_drug_prior_a = %SYSEVALF(&tau2_drug_prior-(&tau2_drug_sd*(3**0.5)));
%let tau2_drug_prior_b = %SYSEVALF(&tau2_drug_prior+(&tau2_drug_sd*(3**0.5)));

*/ CLOSE PARAMETER UPDATING SECTION;

*/ FORMAT OUTPUT FILES;

data thesis.post; /* stack posterior estimates;
set thesis.post thesis.tmp_post;
run;

*/ clear temporary files;

proc datasets library = thesis NOLIST;
delete      mcmc tmp tmp_post tmp_posterior_theta
            drug_re_tmp ite_tmp mcmc_ite mcmc_random
            sigma_tmp tau_int_tmp tau_drug_tmp;
run;

%end; /* close loop;

options notes; /* restore output;

*/ remove labels from proc univariate outputs and add sample size;

data thesis.bayes_sim_sigma;
set thesis.bayes_sim_sigma;
label Sigma = '';
label Sigma_SE = '';
label P2_5 = '';
label P5 = '';
label P50 = '';
label P95 = '';
label P97_5 = '';
Sample = _N_+&Initial-1;
run;

data thesis.bayes_sim_tau_int;
set thesis.bayes_sim_tau_int;
label Tau_Intercept = '';
label Tau_Intercept_SE = '';
label P2_5 = '';
label P5 = '';
label P50 = '';
label P95 = '';
label P97_5 = '';
Sample = _N_+&Initial-1;
run;

data thesis.bayes_sim_tau_drug;
set thesis.bayes_sim_tau_drug;
label Tau_Drug = '';
label Tau_Drug_SE = '';
label P2_5 = '';
label P5 = '';
label P50 = '';
label P95 = '';
label P97_5 = '';
Sample = _N_+&Initial-1;
run;

```

```

*/ Add sample informations to ite info;

data thesis.bayes_sim_ite;
set thesis.bayes_sim_ite;
Sample = _N_+&Initial; */ add sample size data - no need to subtract initial value as not calculated at first analysis;
run;

data thesis.bayes_sim_drug_re;
set thesis.bayes_sim_drug_re;
Sample = _N_+&Initial;
run;

*/ Format DIC table;

DATA THESIS.bayes_sim_dic;
SET THESIS.bayes_sim_dic;
RENAME COL1 = MeanDev;      */ Mean Posterior Deviance;
RENAME COL2 = DevMean;     */ Deviance for Best (Mean) Parameter Estimate;
RENAME COL3 = Pd;          */ Effective Parameters;
RENAME COL4 = DIC;         */ DIC;
drop _NAME_;
Sample = _N_+&Initial-1;    */ add sample size data;
RUN;

*/ split data for separate analysis of parameters;

data thesis.bayes_sim_intercept
      thesis.bayes_sim_drug
      thesis.bayes_sim_sigma2
      thesis.bayes_sim_tau2_int
      thesis.bayes_sim_tau2_drug;
set thesis.post;
if parameter = 'intercept' then output thesis.bayes_sim_intercept;
if parameter = 'drug'      then output thesis.bayes_sim_drug;
if parameter = 'sigma2'   then output thesis.bayes_sim_sigma2;
if parameter = 'tau2_int' then output thesis.bayes_sim_tau2_int;
if parameter = 'tau2_drug' then output thesis.bayes_sim_tau2_drug;
run;

*/ add sample size;

data thesis.bayes_sim_intercept;
set thesis.bayes_sim_intercept;
Sample = _N_+&Initial-1;
drop BF BF_CLINICAL;
run;

*/ add sample size and Bayesian test statistics;

data thesis.bayes_sim_drug;
set thesis.bayes_sim_drug;
Sample = _N_+&Initial-1;
P      = 1-CDF('NORMAL',0,Mean,StdDev);      */ P(PTE>0);
P_CLINICAL = 1-CDF('NORMAL',3,Mean,StdDev); */ P(PTE>3);
LOG10_BF = LOG10(BF);                        */ Jeffreys version;
LOG10_BF_CLINICAL = LOG10(BF_CLINICAL);
run;

data thesis.bayes_sim_sigma2;
set thesis.bayes_sim_sigma2;
Sample = _N_+&Initial-1;
drop BF BF_CLINICAL;
run;

data thesis.bayes_sim_tau2_int;
set thesis.bayes_sim_tau2_int;
Sample = _N_+&Initial-1;
drop BF BF_CLINICAL;
run;

```

```

data thesis.bayes_sim_tau2_drug;
set thesis.bayes_sim_tau2_drug;
Sample = _N_+&Initial-1;
drop BF BF_CLINICAL;
run;

%mend; */ close macro;

***** DEPENDENCE ON INITIAL SET *****;

**** NEW POPULATION SETTING - HIGH RANDOM INTERCEPT VARIANCE, VERY LOW RANDOM TREATMENT
VARIANCE;

* intercept = 20
* drug      = 5
* sigma     = 7.5
* tau_int   = 7.5
* tau_drug  = 0.1;

* initial set N = 5;

%Bayes_Simulation(95,3,2, 5,25051945,20,5,7.5,7.5,0.1);

*/ Rename output files;

data thesis.bayes_sim_intercept_vlow_n5;
set thesis.bayes_sim_intercept;
run;

data thesis.bayes_sim_drug_vlow_n5;
set thesis.bayes_sim_drug;
run;

data thesis.bayes_sim_sigma2_vlow_n5;
set thesis.bayes_sim_sigma2;
run;

data thesis.bayes_sim_tau2_int_vlow_n5;
set thesis.bayes_sim_tau2_int;
run;

data thesis.bayes_sim_tau2_drug_vlow_n5;
set thesis.bayes_sim_tau2_drug;
run;

data thesis.bayes_sim_sigma_vlow_n5;
set thesis.bayes_sim_sigma;
run;

data thesis.bayes_sim_tau_int_vlow_n5;
set thesis.bayes_sim_tau_int;
run;

data thesis.bayes_sim_tau_drug_vlow_n5;
set thesis.bayes_sim_tau_drug;
run;
data thesis.bayes_sim_ite_vlow_n5;
set thesis.bayes_sim_ite;
run;

data thesis.bayes_sim_drug_re_vlow_n5;
set thesis.bayes_sim_drug_re;
run;

data thesis.bayes_sim_dic_vlow_n5;
set thesis.bayes_sim_dic;
run;

* initial set N = 10;

%Bayes_Simulation(90,3,2,10,25051945,20,5,7.5,7.5,0.1);

*/ Rename output files;

```

```

data thesis.bayes_sim_intercept_vlow_n10;
set thesis.bayes_sim_intercept;
run;

data thesis.bayes_sim_drug_vlow_n10;
set thesis.bayes_sim_drug;
run;

data thesis.bayes_sim_sigma2_vlow_n10;
set thesis.bayes_sim_sigma2;
run;

data thesis.bayes_sim_tau2_int_vlow_n10;
set thesis.bayes_sim_tau2_int;
run;

data thesis.bayes_sim_tau2_drug_vlow_n10;
set thesis.bayes_sim_tau2_drug;
run;

data thesis.bayes_sim_sigma_vlow_n10;
set thesis.bayes_sim_sigma;
run;

data thesis.bayes_sim_tau_int_vlow_n10;
set thesis.bayes_sim_tau_int;
run;

data thesis.bayes_sim_tau_drug_vlow_n10;
set thesis.bayes_sim_tau_drug;
run;

data thesis.bayes_sim_ite_vlow_n10;
set thesis.bayes_sim_ite;
run;

data thesis.bayes_sim_drug_re_vlow_n10;
set thesis.bayes_sim_drug_re;
run;

data thesis.bayes_sim_dic_vlow_n10;
set thesis.bayes_sim_dic;
run;

* initial set N = 15;

%Bayes_Simulation(85,3,2,15,25051945,20,5,7.5,7.5,0.1);

*/ Rename output files;

data thesis.bayes_sim_intercept_vlow_n15;
set thesis.bayes_sim_intercept;
run;

data thesis.bayes_sim_drug_vlow_n15;
set thesis.bayes_sim_drug;
run;

data thesis.bayes_sim_sigma2_vlow_n15;
set thesis.bayes_sim_sigma2;
run;

data thesis.bayes_sim_tau2_int_vlow_n15;
set thesis.bayes_sim_tau2_int;
run;

data thesis.bayes_sim_tau2_drug_vlow_n15;
set thesis.bayes_sim_tau2_drug;
run;

```

```

data thesis.bayes_sim_sigma_vlow_n15;
set thesis.bayes_sim_sigma;
run;

data thesis.bayes_sim_tau_int_vlow_n15;
set thesis.bayes_sim_tau_int;
run;

data thesis.bayes_sim_tau_drug_vlow_n15;
set thesis.bayes_sim_tau_drug;
run;

data thesis.bayes_sim_ite_vlow_n15;
set thesis.bayes_sim_ite;
run;

data thesis.bayes_sim_drug_re_vlow_n15;
set thesis.bayes_sim_drug_re;
run;

data thesis.bayes_sim_dic_vlow_n15;
set thesis.bayes_sim_dic;
run;

**** NEW POPULATION SETTING - HIGH RANDOM INTERCEPT VARIANCE, LOW RANDOM TREATMENT
VARIANCE;

* initial set N = 5;

%Bayes_Simulation(95,3,2, 5,25051945,20,5,7.5,7.5,2.5);

*/ Rename output files;

data thesis.bayes_sim_intercept_low_n5;
set thesis.bayes_sim_intercept;
run;

data thesis.bayes_sim_drug_low_n5;
set thesis.bayes_sim_drug;
run;

data thesis.bayes_sim_sigma2_low_n5;
set thesis.bayes_sim_sigma2;
run;

data thesis.bayes_sim_tau2_int_low_n5;
set thesis.bayes_sim_tau2_int;
run;

data thesis.bayes_sim_tau2_drug_low_n5;
set thesis.bayes_sim_tau2_drug;
run;

data thesis.bayes_sim_sigma_low_n5;
set thesis.bayes_sim_sigma;
run;

data thesis.bayes_sim_tau_int_low_n5;
set thesis.bayes_sim_tau_int;
run;

data thesis.bayes_sim_tau_drug_low_n5;
set thesis.bayes_sim_tau_drug;
run;
data thesis.bayes_sim_ite_low_n5;
set thesis.bayes_sim_ite;
run;

data thesis.bayes_sim_drug_re_low_n5;
set thesis.bayes_sim_drug_re;
run;

```

```

data thesis.bayes_sim_dic_low_n5;
set thesis.bayes_sim_dic;
run;

* initial set N = 10;

%Bayes_Simulation(90,3,2,10,25051945,20,5,7.5,7.5,2.5);

*/ Rename output files;

data thesis.bayes_sim_intercept_low_n10;
set thesis.bayes_sim_intercept;
run;

data thesis.bayes_sim_drug_low_n10;
set thesis.bayes_sim_drug;
run;

data thesis.bayes_sim_sigma2_low_n10;
set thesis.bayes_sim_sigma2;
run;

data thesis.bayes_sim_tau2_int_low_n10;
set thesis.bayes_sim_tau2_int;
run;

data thesis.bayes_sim_tau2_drug_low_n10;
set thesis.bayes_sim_tau2_drug;
run;

data thesis.bayes_sim_sigma_low_n10;
set thesis.bayes_sim_sigma;
run;

data thesis.bayes_sim_tau_int_low_n10;
set thesis.bayes_sim_tau_int;
run;

data thesis.bayes_sim_tau_drug_low_n10;
set thesis.bayes_sim_tau_drug;
run;

data thesis.bayes_sim_ite_low_n10;
set thesis.bayes_sim_ite;
run;

data thesis.bayes_sim_drug_re_low_n10;
set thesis.bayes_sim_drug_re;
run;

data thesis.bayes_sim_dic_low_n10;
set thesis.bayes_sim_dic;
run;

* initial set N = 15;

%Bayes_Simulation(85,3,2,15,25051945,20,5,7.5,7.5,2.5);

*/ Rename output files;

data thesis.bayes_sim_intercept_low_n15;
set thesis.bayes_sim_intercept;
run;

data thesis.bayes_sim_drug_low_n15;
set thesis.bayes_sim_drug;
run;

data thesis.bayes_sim_sigma2_low_n15;
set thesis.bayes_sim_sigma2;
run;

```

```
data thesis.bayes_sim_tau2_int_low_n15;  
set thesis.bayes_sim_tau2_int;  
run;
```

```
data thesis.bayes_sim_tau2_drug_low_n15;  
set thesis.bayes_sim_tau2_drug;  
run;
```

```
data thesis.bayes_sim_sigma_low_n15;  
set thesis.bayes_sim_sigma;  
run;
```

```
data thesis.bayes_sim_tau_int_low_n15;  
set thesis.bayes_sim_tau_int;  
run;
```

```
data thesis.bayes_sim_tau_drug_low_n15;  
set thesis.bayes_sim_tau_drug;  
run;
```

```
data thesis.bayes_sim_ite_low_n15;  
set thesis.bayes_sim_ite;  
run;
```

```
data thesis.bayes_sim_drug_re_low_n15;  
set thesis.bayes_sim_drug_re;  
run;
```

```
data thesis.bayes_sim_dic_low_n15;  
set thesis.bayes_sim_dic;  
run;
```

```
**** NEW POPULATION SETTING - EQUAL RANDOM INTERCEPT AND RANDOM TREATMENT VARIANCE;
```

```
* initial set N = 5;
```

```
%Bayes_Simulation(95,3,2, 5,25051945,20,5,7.5,5,5);
```

```
*/ Rename output files;
```

```
data thesis.bayes_sim_intercept_medium_n5;  
set thesis.bayes_sim_intercept;  
run;
```

```
data thesis.bayes_sim_drug_medium_n5;  
set thesis.bayes_sim_drug;  
run;
```

```
data thesis.bayes_sim_sigma2_medium_n5;  
set thesis.bayes_sim_sigma2;  
run;
```

```
data thesis.bayes_sim_tau2_int_medium_n5;  
set thesis.bayes_sim_tau2_int;  
run;
```

```
data thesis.bayes_sim_tau2_drug_medium_n5;  
set thesis.bayes_sim_tau2_drug;  
run;
```

```
data thesis.bayes_sim_ite_medium_n5;  
set thesis.bayes_sim_ite;  
run;
```

```
data thesis.bayes_sim_sigma_medium_n5;  
set thesis.bayes_sim_sigma;  
run;
```

```
data thesis.bayes_sim_tau_int_medium_n5;  
set thesis.bayes_sim_tau_int;  
run;
```

```

data thesis.bayes_sim_tau_drug_medium_n5;
set thesis.bayes_sim_tau_drug;
run;

data thesis.bayes_sim_drug_re_medium_n5;
set thesis.bayes_sim_drug_re;
run;

data thesis.bayes_sim_dic_medium_n5;
set thesis.bayes_sim_dic;
run;

* initial set N = 10;

%Bayes_Simulation(90,3,2,10,25051945,20,5,7.5,5,5);

*/ Rename output files;

data thesis.bayes_sim_intercept_medium_n10;
set thesis.bayes_sim_intercept;
run;

data thesis.bayes_sim_drug_medium_n10;
set thesis.bayes_sim_drug;
run;

data thesis.bayes_sim_sigma2_medium_n10;
set thesis.bayes_sim_sigma2;
run;

data thesis.bayes_sim_tau2_int_medium_n10;
set thesis.bayes_sim_tau2_int;
run;

data thesis.bayes_sim_tau2_drug_medium_n10;
set thesis.bayes_sim_tau2_drug;
run;

data thesis.bayes_sim_sigma_medium_n10;
set thesis.bayes_sim_sigma;
run;

data thesis.bayes_sim_tau_int_medium_n10;
set thesis.bayes_sim_tau_int;
run;

data thesis.bayes_sim_tau_drug_medium_n10;
set thesis.bayes_sim_tau_drug;
run;

data thesis.bayes_sim_ite_medium_n10;
set thesis.bayes_sim_ite;
run;

data thesis.bayes_sim_drug_re_medium_n10;
set thesis.bayes_sim_drug_re;
run;

data thesis.bayes_sim_dic_medium_n10;
set thesis.bayes_sim_dic;
run;

* initial set N = 15;

%Bayes_Simulation(85,3,2,15,25051945,20,5,7.5,5,5);

*/ Rename output files;

data thesis.bayes_sim_intercept_medium_n15;
set thesis.bayes_sim_intercept;
run;

```

```
data thesis.bayes_sim_drug_medium_n15;  
set thesis.bayes_sim_drug;  
run;
```

```
data thesis.bayes_sim_sigma2_medium_n15;  
set thesis.bayes_sim_sigma2;  
run;
```

```
data thesis.bayes_sim_tau2_int_medium_n15;  
set thesis.bayes_sim_tau2_int;  
run;
```

```
data thesis.bayes_sim_tau2_drug_medium_n15;  
set thesis.bayes_sim_tau2_drug;  
run;
```

```
data thesis.bayes_sim_sigma_medium_n15;  
set thesis.bayes_sim_sigma;  
run;
```

```
data thesis.bayes_sim_tau_int_medium_n15;  
set thesis.bayes_sim_tau_int;  
run;
```

```
data thesis.bayes_sim_tau_drug_medium_n15;  
set thesis.bayes_sim_tau_drug;  
run;
```

```
data thesis.bayes_sim_ite_medium_n15;  
set thesis.bayes_sim_ite;  
run;
```

```
data thesis.bayes_sim_drug_re_medium_n15;  
set thesis.bayes_sim_drug_re;  
run;
```

```
data thesis.bayes_sim_dic_medium_n15;  
set thesis.bayes_sim_dic;  
run;
```

```
**** NEW POPULATION SETTING - LOW RANDOM INTERCEPT VARIANCE, HIGH RANDOM TREATMENT  
VARIANCE;
```

```
* initial set N = 5;
```

```
%Bayes_Simulation(95,3,2, 5,25051945,20,5,7.5,2.5,7.5);
```

```
*/ Rename output files;
```

```
data thesis.bayes_sim_intercept_high_n5;  
set thesis.bayes_sim_intercept;  
run;
```

```
data thesis.bayes_sim_drug_high_n5;  
set thesis.bayes_sim_drug;  
run;
```

```
data thesis.bayes_sim_sigma2_high_n5;  
set thesis.bayes_sim_sigma2;  
run;
```

```
data thesis.bayes_sim_tau2_int_high_n5;  
set thesis.bayes_sim_tau2_int;  
run;
```

```
data thesis.bayes_sim_tau2_drug_high_n5;  
set thesis.bayes_sim_tau2_drug;  
run;
```

```
data thesis.bayes_sim_sigma_high_n5;  
set thesis.bayes_sim_sigma;  
run;
```

```

data thesis.bayes_sim_tau_int_high_n5;
set thesis.bayes_sim_tau_int;
run;

data thesis.bayes_sim_tau_drug_high_n5;
set thesis.bayes_sim_tau_drug;
run;

data thesis.bayes_sim_ite_high_n5;
set thesis.bayes_sim_ite;
run;

data thesis.bayes_sim_drug_re_high_n5;
set thesis.bayes_sim_drug_re;
run;

data thesis.bayes_sim_dic_high_n5;
set thesis.bayes_sim_dic;
run;

* initial set N = 10;

%Bayes_Simulation(90,3,2,10,25051945,20,5,7.5,2.5,7.5);

*/ Rename output files;

data thesis.bayes_sim_intercept_high_n10;
set thesis.bayes_sim_intercept;
run;

data thesis.bayes_sim_drug_high_n10;
set thesis.bayes_sim_drug;
run;

data thesis.bayes_sim_sigma2_high_n10;
set thesis.bayes_sim_sigma2;
run;

data thesis.bayes_sim_tau2_int_high_n10;
set thesis.bayes_sim_tau2_int;
run;

data thesis.bayes_sim_tau2_drug_high_n10;
set thesis.bayes_sim_tau2_drug;
run;

data thesis.bayes_sim_sigma_high_n10;
set thesis.bayes_sim_sigma;
run;

data thesis.bayes_sim_tau_int_high_n10;
set thesis.bayes_sim_tau_int;
run;

data thesis.bayes_sim_tau_drug_high_n10;
set thesis.bayes_sim_tau_drug;
run;

data thesis.bayes_sim_ite_high_n10;
set thesis.bayes_sim_ite;
run;

data thesis.bayes_sim_drug_re_high_n10;
set thesis.bayes_sim_drug_re;
run;

data thesis.bayes_sim_dic_high_n10;
set thesis.bayes_sim_dic;
run;

* initial set N = 15;

%Bayes_Simulation(85,3,2,15,25051945,20,5,7.5,2.5,7.5);

```

*/ Rename output files;

```
data thesis.bayes_sim_intercept_high_n15;  
set thesis.bayes_sim_intercept;  
run;
```

```
data thesis.bayes_sim_drug_high_n15;  
set thesis.bayes_sim_drug;  
run;
```

```
data thesis.bayes_sim_sigma2_high_n15;  
set thesis.bayes_sim_sigma2;  
run;
```

```
data thesis.bayes_sim_tau2_int_high_n15;  
set thesis.bayes_sim_tau2_int;  
run;
```

```
data thesis.bayes_sim_tau2_drug_high_n15;  
set thesis.bayes_sim_tau2_drug;  
run;
```

```
data thesis.bayes_sim_sigma_high_n15;  
set thesis.bayes_sim_sigma;  
run;
```

```
data thesis.bayes_sim_tau_int_high_n15;  
set thesis.bayes_sim_tau_int;  
run;
```

```
data thesis.bayes_sim_tau_drug_high_n15;  
set thesis.bayes_sim_tau_drug;  
run;
```

```
data thesis.bayes_sim_ite_high_n15;  
set thesis.bayes_sim_ite;  
run;
```

```
data thesis.bayes_sim_drug_re_high_n15;  
set thesis.bayes_sim_drug_re;  
run;
```

```
data thesis.bayes_sim_dic_high_n15;  
set thesis.bayes_sim_dic;  
run;
```

**** NEW POPULATION SETTING - VERY LOW RANDOM INTERCEPT VARIANCE, HIGH RANDOM TREATMENT VARIANCE;

* initial set N = 5;

```
%Bayes_Simulation(95,3,2, 5,25051945,20,5,7.5,0.1,7.5);
```

*/ Rename output files;

```
data thesis.bayes_sim_intercept_vhigh_n5;  
set thesis.bayes_sim_intercept;  
run;
```

```
data thesis.bayes_sim_drug_vhigh_n5;  
set thesis.bayes_sim_drug;  
run;
```

```
data thesis.bayes_sim_sigma2_vhigh_n5;  
set thesis.bayes_sim_sigma2;  
run;
```

```
data thesis.bayes_sim_tau2_int_vhigh_n5;  
set thesis.bayes_sim_tau2_int;  
run;
```

```

data thesis.bayes_sim_tau2_drug_vhigh_n5;
set thesis.bayes_sim_tau2_drug;
run;

data thesis.bayes_sim_sigma_vhigh_n5;
set thesis.bayes_sim_sigma;
run;

data thesis.bayes_sim_tau_int_vhigh_n5;
set thesis.bayes_sim_tau_int;
run;

data thesis.bayes_sim_tau_drug_vhigh_n5;
set thesis.bayes_sim_tau_drug;
run;

data thesis.bayes_sim_ite_vhigh_n5;
set thesis.bayes_sim_ite;
run;

data thesis.bayes_sim_drug_re_vhigh_n5;
set thesis.bayes_sim_drug_re;
run;

data thesis.bayes_sim_dic_vhigh_n5;
set thesis.bayes_sim_dic;
run;

* initial set N = 10;

%Bayes_Simulation(90,3,2,10,25051945,20,5,7.5,0.1,7.5);

*/ Rename output files;

data thesis.bayes_sim_intercept_vhigh_n10;
set thesis.bayes_sim_intercept;
run;

data thesis.bayes_sim_drug_vhigh_n10;
set thesis.bayes_sim_drug;
run;

data thesis.bayes_sim_sigma2_vhigh_n10;
set thesis.bayes_sim_sigma2;
run;

data thesis.bayes_sim_tau2_int_vhigh_n10;
set thesis.bayes_sim_tau2_int;
run;

data thesis.bayes_sim_tau2_drug_vhigh_n10;
set thesis.bayes_sim_tau2_drug;
run;

data thesis.bayes_sim_sigma_vhigh_n10;
set thesis.bayes_sim_sigma;
run;

data thesis.bayes_sim_tau_int_vhigh_n10;
set thesis.bayes_sim_tau_int;
run;

data thesis.bayes_sim_tau_drug_vhigh_n10;
set thesis.bayes_sim_tau_drug;
run;

data thesis.bayes_sim_ite_vhigh_n10;
set thesis.bayes_sim_ite;
run;

data thesis.bayes_sim_drug_re_vhigh_n10;
set thesis.bayes_sim_drug_re;
run;

```

```

data thesis.bayes_sim_dic_vhigh_n10;
set thesis.bayes_sim_dic;
run;

* initial set N = 15;

%Bayes_Simulation(85,3,2,15,25051945,20,5,7.5,0.1,7.5);

*/ Rename output files;

data thesis.bayes_sim_intercept_vhigh_n15;
set thesis.bayes_sim_intercept;
run;

data thesis.bayes_sim_drug_vhigh_n15;
set thesis.bayes_sim_drug;
run;

data thesis.bayes_sim_sigma2_vhigh_n15;
set thesis.bayes_sim_sigma2;
run;

data thesis.bayes_sim_tau2_int_vhigh_n15;
set thesis.bayes_sim_tau2_int;
run;

data thesis.bayes_sim_tau2_drug_vhigh_n15;
set thesis.bayes_sim_tau2_drug;
run;

data thesis.bayes_sim_sigma_vhigh_n15;
set thesis.bayes_sim_sigma;
run;

data thesis.bayes_sim_tau_int_vhigh_n15;
set thesis.bayes_sim_tau_int;
run;

data thesis.bayes_sim_tau_drug_vhigh_n15;
set thesis.bayes_sim_tau_drug;
run;

data thesis.bayes_sim_ite_vhigh_n15;
set thesis.bayes_sim_ite;
run;

data thesis.bayes_sim_drug_re_vhigh_n15;
set thesis.bayes_sim_drug_re;
run;

data thesis.bayes_sim_dic_vhigh_n15;
set thesis.bayes_sim_dic;

```

A.2.2 Sections 6.2 and 6.3 - SAS PROC MCMC for Multiple Simulations

```
*/ AUTHOR GIULIO FLORE 10/01/2015;

*/ Cycles: # of treatment cycles;
*/ Periods: # of treatment periods within a cycle;
*/ Patients: # of patients to be randomized;
*/ Randomizer: # randomization seed for repeatability;
*/ beta0: # fixed intercept of re model;
*/ beta1: # fixed drug of re model;
*/ sigma: # error term sd;
*/ tau_int: # random intercept sd;
*/ tau_drug: # random drug sd;

%macro simulator (Cycles,
                 Periods,
                 Patients,
                 Randomizer,
                 beta0,
                 beta1,
                 sigma,
                 tau_int,
                 tau_drug);

*/ macro trial assume a RE structure with random intercept and drug;

options nosource nonotes; */ exclude output;

data thesis.tmp;
put ID Cycle Period Treatment;
stop;
run;

*/ create file with sequence of random intercept values;

proc iml;
u = j(&Patients,1);
call randseed(&randomizer);
call randgen(u, "Normal",0,&tau_int);
create thesis.randomintercept from u;
append from u;
quit;
*/ Matrix language proc to generate matrix of random values;
*/ allocate vector with # of subjects;
*/ fix randomization for repeatability;
*/ fill it */;
*/ generate new file with randomized values;
*/ stop proc iml - specific command;

*/ add crossreference ID ;

data thesis.randomintercept;
set thesis.randomintercept;
id = _N_;
rename col1 = intercept_re;
run;
*/ add index for merger;

*/ create file with sequence of random drug values;

proc iml;
w = j(&Patients,1);
call randseed(&randomizer+1000);
call randgen(w, "Normal",0,&tau_drug);
create thesis.randomdrug from w;
append from w;
quit;

*/ add crossreference ID ;

data thesis.randomdrug;
set thesis.randomdrug;
id = _N_;
rename col1 = drug_re;
run;
```

```

%do i = 1 %to &Patients;

%let random = %sysevalf(&i + &randomizer);

proc plan seed = &random;                                /* seed fix randomisation for repeatability;
  factors Cycle=&Cycles ordered Period=&Periods ordered / noprint; /* gives sequence of periods and cycles for individual;
  treatments Treatment=2 cyclic (1 2);                  /* random assignment to treatment set as dummy;
  output out = thesis.rand
  Treatment random;
run;

data thesis.tmp;                                        /* append new block of trial information;
set thesis.tmp thesis.rand;
if Cycle = . then delete;                               /* clear empty record;
run;

%end; /* close patient loop;

data thesis.tmp;                                        /* add id, merge random drugs, create outcome value;
set thesis.tmp;
id = floor((_N_-1)/(&Cycles*&Periods))+1; /* assign id by blocks of 6;
run;

data thesis.tmp;                                        /* add id, merge random drugs, create outcome value;
set thesis.tmp;
merge thesis.tmp thesis.randomintercept thesis.randomdrug;
by id;
run;

data thesis.tmp;                                        /* add id, merge random drugs, create outcome value;
call streaminit(&randomizer+2000); /* enforce repeatability of error random values;
set thesis.tmp;
treatment = treatment - 1;                               /* set treatment as dummy;
error = rand("Normal",0,&sigma);
facit_f = &beta0 + intercept_re + (treatment*&beta1)+(treatment*drug_re) + error; /* random values for outcome;
facit_f = round(facit_f,0.5);                             /* round value to reflect reported trial values;
run;

%mend; /* close macro;

/* Population: number of patients to be added in simulator;
/* Cycles: treatment cycles;
/* Periods: number of treatment periods;
/* Initial: size of initial group of subjects;
/* Seeding: randomization parameter for repeatability
/* beta0, beta1, sigma, tau_int tau_drug simulator parameters;

%macro Bayes_Simulation ( Population,
                          Cycles,
                          Periods,
                          Initial,
                          Seeding,
                          beta0,
                          beta1,
                          sigma,
                          tau_int,
                          tau_drug);

options nosource nonotes; /* suppress output;

/* Set repository for precision data on ITE values;

data thesis.bayes_sim_ite;
put ITE ITE_SE P2_5 P5 P50 P95 P97_5;
stop;
run;

data thesis.bayes_sim_drug_re;
put Drug_RE Drug_RE_SE P2_5 P5 P50 P95 P97_5;
stop;
run;

/* create initial sample;

```

```

%simulator(&Cycles,&Periods,&Initial,&Seeding,&beta0,&beta1,&sigma,&tau_int,&tau_drug);

*/ initial block marked number of markov chains;
*/ thinning of simulators;
*/ covariance estimation assumptions;
*/ randomization initial value;
*/ outputting of statistics;
*/ outputting of chain estimates for total effect estimation;

*/ Crucial - single subject analysis requires long chains;
*/ of simulation. Lack of convergence produce biased posteriors;
*/ and distort the analysis - Also ;

*/ Apply 1M chains (nmc);
*/ Apply appropriate thinning (thin) to 'defeat' ACF and produce at least 4,000 sampling for convergence statistics;
*/ Apply higher burn-in (ntu): 50000;
*/ Apply proposal covariance (propcov) with quasi-NR algorithm ;

proc mcmc data=thesis.tmp outpost = thesis.mcmc

    nmc = 1000000
    thin = 250
    ntu = 50000
    DIC
    propcov = quanev
    monitor = ( _parms_ LOGLIKE LOGPOST LOGPRIOR)
    seed = &Seeding
    statistics(alpha = 0.05 percentage = (2.5 5 50 95 97.5) ) = all;

parms intercept 26 drug 0 sigma2 50 tau2_int 50 tau2_drug 50; */ speculative initial value assumptions;

*/ priors selected to reflect the nature of the data, with minimal assumptions;

prior intercept ~ normal(26, var = 64);      */ this prior assumes the intercept to be in the mid range of facit-f;
prior drug ~ normal( 0, var = 64);      */ this is a moderately skeptic prior, with equipoise on direction of the treatment effect;

*/ Non informative prior for random effects variance following Gelman (2006);

prior sigma2 ~ uniform(0,100);
prior tau2_int ~ uniform(0,100);
prior tau2_drug ~ uniform(0,100);

*/ set the hierarchical structure;

random intercept2 ~ normal(0, var = tau2_int) subject = id monitor = (intercept2);
random drug2 ~ normal(0, var = tau2_drug) subject = id monitor = (drug2);

mu = intercept + intercept2 + drug * treatment + drug2 * treatment; */ set a random intercept and drug mean model;
model facit_f ~ normal(mu, var = sigma2);

*/ output file statements;

ods select DIC PostSummaries;
ods output DIC = THESIS.bayes_sim_dic;
ods output PostSummaries = thesis.post;

run;

data thesis.post;          */ generate copies conditionally;
set thesis.post;
if _N_ > 5 then delete;  */ only distributional parameters;
run;

*/ Extract sigma, tau intercept and tau treatment estimates;

data thesis.mcmc_random;
set thesis.mcmc;
sigma = sigma2**0.5;
tau_int = tau2_int**0.5;
tau_drug = tau2_drug**0.5;
run;

```

```

proc univariate data = thesis.mcmc_random NOPRINT PCTLDEF = 3; /* output Bayesian estimates into file with quantiles;
var sigma;
output out = thesis.bayes_sim_sigma
    mean = Sigma
    std = Sigma_SE
    pctlpre = P
    pctlpts = 2.5 5 50 95 97.5;
run;

proc univariate data = thesis.mcmc_random NOPRINT PCTLDEF = 3;
var tau_int;
output out = thesis.bayes_sim_tau_int
    mean = Tau_Intercept
    std = Tau_Intercept_SE
    pctlpre = P
    pctlpts = 2.5 5 50 95 97.5;
run;

proc univariate data = thesis.mcmc_random NOPRINT PCTLDEF = 3;
var tau_drug;
output out = thesis.bayes_sim_tau_drug
    mean = Tau_Drug
    std = Tau_Drug_SE
    pctlpre = P
    pctlpts = 2.5 5 50 95 97.5;
run;

/* generate total effect estimates in each sampled markov chain;
/* rough solution - no dynamic extension to pre-specified number of subjects;
/* must invest time on function that counts vars with initial string drug2_ ;

/* Transpose DIC results for stacking;

proc transpose data = THESIS.bayes_sim_dic
    out = THESIS.bayes_sim_dic;
run;

*****
*/
/* Set initial Parameters for analysis and derivation of statistics outside the loop */
/*
*****

%let beta1_prior = 0; /* derived from vague prior in first estimation;
%let beta1_sd_prior = 8;

/* also set first priors for use in clinical BF evaluation;

%let beta1_first_prior = 0;
%let beta1_sd_first_prior = 8;

data thesis.post;
set thesis.post;

if Parameter = 'intercept' then do;
    call symput('beta0_prior',mean);
    call symput('beta0_sd_prior',stddev);
end;
else

if Parameter = 'drug' then do;
    call symput('beta1_posterior',mean); /* get a posterior value for BF estimation;
    call symput('beta1_sd_posterior',stddev);
end;
else

if Parameter = 'sigma2' then do;
    call symput('sigma2_prior',mean);
    call symput('sigma2_sd',stddev);
end;
else

```

```

if Parameter = 'tau2_int' then do;
    call symput('tau2_int_prior',mean);
    call symput('tau2_int_sd',stddev);
end;
else
if Parameter = 'tau2_drug' then do;
    call symput('tau2_drug_prior',mean);
    call symput('tau2_drug_sd',stddev);
end;
run;

*/ The BF needs to be interpreted in relation to the selected prior. If the BF is set against;
*/ the first proposed prior, each new subject inclusion implies a testing of the Null Hypothesis;
*/ as expressed in the first prior - the BF reduces to the posterior odds, as the prior odds are set to 1;
*/ a clinical significance Ho is also set;

data thesis.post;
set thesis.post;
BF = 0;
if Parameter = 'drug' then
do;
BF = (CDF('NORMAL',0,
    &beta1_posterior,
    &beta1_sd_posterior))/(1-CDF('NORMAL',0,
    &beta1_posterior,
    &beta1_sd_posterior));
end;
run;

*/ derive additional macro variables for execution;
*/ %sysevalf performs floating point calculations;

%let beta1_prior = &beta1_posterior; */ update prior to posterior for feedback;
%let beta1_sd_prior = &beta1_sd_posterior; */ update prior to posterior for feedback;

*/ set variance for prior update;

%let beta0_var_prior = %SYSEVALF(&beta0_sd_prior**2);
%let beta1_var_prior = %SYSEVALF(&beta1_sd_prior**2);

*/ derive uniform parameters for proc mcmc hyperprior;

%let sigma2_prior_a = %SYSEVALF(&sigma2_prior-(&sigma2_sd*(3**0.5)));
%let sigma2_prior_b = %SYSEVALF(&sigma2_prior+(&sigma2_sd*(3**0.5)));

%let tau2_int_prior_a = %SYSEVALF(&tau2_int_prior-(&tau2_int_sd*(3**0.5)));
%let tau2_int_prior_b = %SYSEVALF(&tau2_int_prior+(&tau2_int_sd*(3**0.5)));

%let tau2_drug_prior_a = %SYSEVALF(&tau2_drug_prior-(&tau2_drug_sd*(3**0.5)));
%let tau2_drug_prior_b = %SYSEVALF(&tau2_drug_prior+(&tau2_drug_sd*(3**0.5)));

*/ NEW OBSERVATION INCREASE LOOP;

%do k = 1 %to &Population;

%put Cycle #&k; */ mark loops;

*/ generate a n = 1 sample;

%let starter = %sysevalf(&k + &seeding);

*/ randomization parameter linked to loop index for repeatability;

%simulator(&Cycles,&Periods,1,&starter,&beta0,&beta1,&sigma,&tau_int,&tau_drug);

*/ analyse new observation using the first analysis posterior as prior;
*/ initial block marked number of markov chains;
*/ thinning of simulators;
*/ covariance estimation assumptions;
*/ randomization initial value;
*/ outputting of statistics;

```

```

*/ randomization linked to loop index for repeatability;

*/ based on analysis of real data ACF curtailed by large burn-in;
*/ for speed set shorter chain after first run with non-informative prior;

*/ Set thinning at lag = 20 to avoid ACF effects on the estimate;
*/ while obtaining at least 4,000 observations for convergence tests;

proc mcmc data = thesis.tmp outpost = thesis.mcmc

    nmc    = 80000
    thin   = 20
    ntu    = 50000
    propcov = quanew
    monitor = ( _parms_ LOGLIKE LOGPOST LOGPRIOR)
    seed = &k
    DIC
    statistics(alpha=0.05 percentage = (2.5 5 50 95 97.5) ) = all;

*/ parameter declaration and initial values statements based on priors to speed up convergence;

Parms    intercept  &beta0_prior
         drug        &beta1_prior
         sigma2      &sigma2_prior
         tau2_int    &tau2_int_prior
         tau2_drug   &tau2_drug_prior;

*/ these priors are based on the previous cycle posteriors;

prior intercept ~ normal(&beta0_prior, var = &beta0_var_prior);
prior drug      ~ normal(&beta1_prior, var = &beta1_var_prior);
prior sigma2    ~ uniform(&sigma2_prior_a, &sigma2_prior_b);
prior tau2_int  ~ uniform(&tau2_int_prior_a, &tau2_int_prior_b);
prior tau2_drug ~ uniform(&tau2_drug_prior_a, &tau2_drug_prior_b);

*/ random part statement - subject as grouping unit;

random intercept2 ~ normal(0, var = tau2_int) subject = id monitor = (intercept2); */ set the hierarchical aspect;
random drug2      ~ normal(0, var = tau2_drug) subject = id monitor = (drug2);

*/ mean model statement;

mu = intercept + intercept2 + (drug * treatment) + (drug2 * treatment);
model facit_f ~ normal(mu, var = sigma2);

*/ output file statements;

ods select DIC PostSummaries;
ods output DIC = thesis.tmp_DIC;
ods output PostSummaries = thesis.tmp_posterior_theta;

run;

*/ extract variance distributions to track them as SD and SD s.e;

data thesis.mcmc_random;
set thesis.mcmc;
sigma = sigma2**0.5;
tau_int = tau2_int**0.5;
tau_drug = tau2_drug**0.5;
run;

proc univariate data = thesis.mcmc_random NOPRINT PCTLDEF = 3; */ Output Bayesian summaries into files with quantiles;
var sigma;
output out = thesis.sigma_tmp
    mean = Sigma
    std  = Sigma_SE
    pctlpre = P
    pctlpts = 2.5 5 50 95 97.5;
run;

```

```

proc univariate data = thesis.mcmc_random NOPRINT PCTLDEF = 3;
var tau_int;
output out = thesis.tau_int_tmp
  mean = Tau_Intercept
  std = Tau_Intercept_SE
  pctlpre = P
  pctlpts = 2.5 5 50 95 97.5;
run;

proc univariate data = thesis.mcmc_random NOPRINT PCTLDEF = 3;
var tau_drug;
output out = thesis.tau_drug_tmp
  mean = Tau_Drug
  std = Tau_Drug_SE
  pctlpre = P
  pctlpts = 2.5 5 50 95 97.5;
run;

*/ combine random variance as sd estimates;

data thesis.bayes_sim_sigma;                                */ Stack the data;
set thesis.bayes_sim_sigma thesis.sigma_tmp;
run;

data thesis.bayes_sim_tau_int;
set thesis.bayes_sim_tau_int thesis.tau_int_tmp;
run;

data thesis.bayes_sim_tau_drug;
set thesis.bayes_sim_tau_drug thesis.tau_drug_tmp;
run;

*/ extract ite distribution as mixture;

data thesis.mcmc_ite;
set thesis.mcmc;
ITE = DRUG + DRUG2_1;
KEEP DRUG2_1 ITE;
run;

proc univariate data = thesis.mcmc_ite NOPRINT PCTLDEF = 3; */ output Bayesian summaries into files with quantiles;
var ite;
output out = thesis.ite_tmp
  mean = ITE
  std = ITE_SE
  pctlpre = P
  pctlpts = 2.5 5 50 95 97.5;
run;

proc univariate data = thesis.mcmc_ite NOPRINT PCTLDEF = 3;
var drug2_1;
output out = thesis.drug_re_tmp
  mean = Drug_RE
  std = Drug_RE_SE
  pctlpre = P
  pctlpts = 2.5 5 50 95 97.5;
run;

*/ combine re estimate and ite estimate;

data thesis.bayes_sim_ite;                                */ Stack the data;
set thesis.bayes_sim_ite thesis.ite_tmp;
run;

data thesis.bayes_sim_drug_re;
set thesis.bayes_sim_drug_re thesis.drug_re_tmp;
run;

*/ Update DIC output;

proc transpose data = thesis.tmp_DIC
  out = thesis.tmp_DIC;
run;

```

```

data THESIS.bayes_sim_dic;
set THESIS.bayes_sim_dic thesis.tmp_DIC;
run;

*/ Update data repository of posteriors ;

data thesis.tmp_post;          */ generate copies conditionally;
set thesis.tmp_posterior_theta;
if _N_ > 5 then delete;      */ only distributional parameters;
run;

/******;
*/                               *;
*/ Update all relevant posterior parameters  *;
*/                               *;
/******;

data thesis.tmp_post;
set thesis.tmp_post;
if Parameter = 'intercept' then do;
    call symput('beta0_prior',mean);
    call symput('beta0_sd_prior',stddev);
end;

else

if Parameter = 'drug' then do;
    call symput('beta1_posterior',mean); */ get a posterior value for BF estimation;
    call symput('beta1_sd_posterior',stddev);
end;

else

if Parameter = 'sigma2' then do;
    call symput('sigma2_prior',mean);
    call symput('sigma2_sd',stddev);
end;

else

if Parameter = 'tau2_int' then do;
    call symput('tau2_int_prior',mean);
    call symput('tau2_int_sd',stddev);
end;

else

if Parameter = 'tau2_drug' then do;
    call symput('tau2_drug_prior',mean);
    call symput('tau2_drug_sd',stddev);
end;

run;

*/ The BF needs to be interpreted in relation to the selected prior. If the BF is set against;
*/ the first proposed prior, each new subject inclusion implies a testing of the Null Hypothesis;
*/ as expressed in the first prior;

*/ the prior odd calculations are still expressed to made the formula explicit ;

data thesis.tmp_post;
set thesis.tmp_post;
BF = 0;
if Parameter = 'drug' then
do;
BF = (CDF('NORMAL',0,
    &beta1_posterior,
    &beta1_sd_posterior))/(1-CDF('NORMAL',0,
    &beta1_posterior,
    &beta1_sd_posterior)) ;

end;
run;

*/ derive additional macro variables for execution;
*/ %sysevalf performs floating point calculations;

```

```

%let beta1_prior = &beta1_posterior;          /* update prior to posterior for feedback;
%let beta1_sd_prior = &beta1_sd_posterior;    /* update prior to posterior for feedback;

%let beta0_var_prior = %SYSEVALF(&beta0_sd_prior**2); /* set variance update for prior;
%let beta1_var_prior = %SYSEVALF(&beta1_sd_prior**2);

%let sigma2_prior_a = %SYSEVALF(&sigma2_prior-(&sigma2_sd*(3**0.5)));
%let sigma2_prior_b = %SYSEVALF(&sigma2_prior+(&sigma2_sd*(3**0.5)));

%let tau2_int_prior_a = %SYSEVALF(&tau2_int_prior-(&tau2_int_sd*(3**0.5)));
%let tau2_int_prior_b = %SYSEVALF(&tau2_int_prior+(&tau2_int_sd*(3**0.5)));

%let tau2_drug_prior_a = %SYSEVALF(&tau2_drug_prior-(&tau2_drug_sd*(3**0.5)));
%let tau2_drug_prior_b = %SYSEVALF(&tau2_drug_prior+(&tau2_drug_sd*(3**0.5)));

*/ CLOSE PARAMETER UPDATING SECTION;

*/ FORMAT OUTPUT FILES;

data thesis.post; /* stack posterior estimates;
set thesis.post thesis.tmp_post;
run;

*/ clear temporary files;

proc datasets library = thesis NOLIST;
delete mcmc tmp tmp_post
      tmp_posterior_theta drug_re_tmp ite_tmp mcmc_ite mcmc_random
      sigma_tmp tau_int_tmp tau_drug_tmp;
run;

%end; /* close loop;

*/ remove labels from proc univariate outputs and add sample size;

data thesis.bayes_sim_sigma;
set thesis.bayes_sim_sigma;
label Sigma = '';
label Sigma_SE = '';
label P2_5 = '';
label P5 = '';
label P50 = '';
label P95 = '';
label P97_5 = '';
Sample = _N_+&Initial-1;
run;

data thesis.bayes_sim_tau_int;
set thesis.bayes_sim_tau_int;
label Tau_Intercept = '';
label Tau_Intercept_SE = '';
label P2_5 = '';
label P5 = '';
label P50 = '';
label P95 = '';
label P97_5 = '';
Sample = _N_+&Initial-1;
run;

data thesis.bayes_sim_tau_drug;
set thesis.bayes_sim_tau_drug;
label Tau_Drug = '';
label Tau_Drug_SE = '';
label P2_5 = '';
label P5 = '';
label P50 = '';
label P95 = '';
label P97_5 = '';
Sample = _N_+&Initial-1;
run;

*/ Add sample information to ite info;

```

```

data thesis.bayes_sim_ite;
set thesis.bayes_sim_ite;
Sample = _N_; /* add sample size data;
run;

data thesis.bayes_sim_drug_re;
set thesis.bayes_sim_drug_re;
Sample = _N_+&Initial; /* add sample size data - no need to subtract initial value as not calculated at first analysis;
run;

/* Format DIC table;

DATA THESIS.bayes_sim_dic;
SET THESIS.bayes_sim_dic;
RENAME COL1 = MeanDev; /* Mean Posterior Deviance;
RENAME COL2 = DevMean; /* Deviance for Best (Mean) Parameter Estimate;
RENAME COL3 = Pd; /* Effective Parameters;
RENAME COL4 = DIC; /* DIC;
drop _NAME_;
Sample = _N_+&Initial-1; /* add sample size data;
RUN;

/* split data for separate analysis of parameters;

data thesis.bayes_sim_intercept
thesis.bayes_sim_drug
thesis.bayes_sim_sigma2
thesis.bayes_sim_tau2_int
thesis.bayes_sim_tau2_drug;
set thesis.post;
if parameter = 'intercept' then output thesis.bayes_sim_intercept;
if parameter = 'drug' then output thesis.bayes_sim_drug;
if parameter = 'sigma2' then output thesis.bayes_sim_sigma2;
if parameter = 'tau2_int' then output thesis.bayes_sim_tau2_int;
if parameter = 'tau2_drug' then output thesis.bayes_sim_tau2_drug;
run;

/* add sample size;

data thesis.bayes_sim_intercept;
set thesis.bayes_sim_intercept;
Sample = _N_+&Initial-1;
drop BF;
run;

data thesis.bayes_sim_drug;
set thesis.bayes_sim_drug;
Sample = _N_+&Initial-1;
P = 1-CDF('NORMAL',0,Mean,StdDev); /* P(PTE>0);
P_CLINICAL = 1-CDF('NORMAL',3,Mean,StdDev); /* P(PTE>3);
LOG10_BF = LOG10(BF); /* Jeffreys version;
drop Parameter;
run;

data thesis.bayes_sim_sigma2;
set thesis.bayes_sim_sigma2;
Sample = _N_+&Initial-1;
drop BF;
run;

data thesis.bayes_sim_tau2_int;
set thesis.bayes_sim_tau2_int;
Sample = _N_+&Initial-1;
drop BF;
run;

data thesis.bayes_sim_tau2_drug;
set thesis.bayes_sim_tau2_drug;
Sample = _N_+&Initial-1;
drop Parameter BF;
run;

%mend; /* close macro;

```

```

***** CONVERGENCE CHECK *****;

%macro powercycle(Simulations_p,
                 Population_p,
                 Cycles_p,
                 Periods_p,
                 Initial_p,
                 Seeding_p,
                 beta0_p,
                 beta1_p,
                 sigma_p,
                 tau_int_p,
                 tau_drug_p);

data thesis.bayes_drug_convergence;
put N Mean StdDev P2_5 P5 P50 P95 P97_5 Sample
   P P_Clinical Log10_BF Log10_BF_Clinical;
stop;
run;

data thesis.bayes_tau2_drug_convergence;
put N Mean StdDev P2_5 P5 P50 P95 P97_5 Sample;
stop;
run;

data thesis.bayes_ite_convergence;
put ITE ITE_SE P2_5 P5 P50 P95 P97_5 Sample;
stop;
run;

data thesis.bayes_tau_drug_convergence;
put Tau_Drug Tau_Drug_se P2_5 P5 P50 P95 P97_5 Sample;
stop;
run;

%do counter = 1 %to &Simulations_p;                */ start randomization loop;

%let seeded = %sysevalf(&Seeding_p+(&counter*1000)); */ change randomization at each cycle;
%put Hypercycle #&counter;                          */ mark loops;

%Bayes_Simulation( &Population_p,&Cycles_p,&Periods_p,&Initial_p,&seeded,
                  &beta0_p,&beta1_p,&sigma_p,&tau_int_p,&tau_drug_p);

*/ stack output files;

data thesis.bayes_drug_convergence;
set thesis.bayes_drug_convergence thesis.bayes_sim_drug;
run;

data thesis.bayes_tau2_drug_convergence;
set thesis.bayes_tau2_drug_convergence thesis.bayes_sim_tau2_drug;
run;

data thesis.bayes_tau_drug_convergence;
set thesis.bayes_tau_drug_convergence thesis.bayes_sim_tau_drug;
run;

data thesis.bayes_ite_convergence;
set thesis.bayes_ite_convergence thesis.bayes_sim_ite;
run;

%end; */ end do loop;

data thesis.bayes_drug_convergence;
set thesis.bayes_drug_convergence;
Simulation = ceil(_N_/(&Population_p+1));
run;

data thesis.bayes_tau2_drug_convergence;
set thesis.bayes_tau2_drug_convergence;
Simulation = ceil(_N_/(&Population_p+1));
run;

```

```
data thesis.bayes_tau_drug_convergence;
set thesis.bayes_tau_drug_convergence;
Simulation = ceil(_N_/(&Population_p+1));
run;
```

```
*/ ite estimates are taken only for the first initial subject onwards;
```

```
data thesis.bayes_ite_convergence;
set thesis.bayes_ite_convergence;
Simulation = ceil(_N_/(&Population_p));
run;
```

```
%mend; */ close macro;
```

```
**** NEW POPULATION SETTING - HIGH RANDOM INTERCEPT VARIANCE AND LOW RANDOM TREATMENT VARIANCE;
```

```
* initial set N = 15, TOTAL 300 SUBJECTS, 30 random start samples sample;
```

```
options nonotes;
```

```
%powercycle(30,285,3,2,15,300|2015,20,5,7.5,7.5,2.5);
```

```
data thesis.bayes_drug_convergence_1;
set thesis.bayes_drug_convergence;
run;
```

```
data thesis.bayes_tau2_drug_convergence_1;
set thesis.bayes_tau2_drug_convergence;
run;
```

```
data thesis.bayes_tau_drug_convergence_1;
set thesis.bayes_tau_drug_convergence;
run;
```

```
data thesis.bayes_ite_convergence_1;
set thesis.bayes_ite_convergence;
run;
```

```
%powercycle(30,285,3,2,15,300|2015,20,5,7.5,5,5);
```

```
data thesis.bayes_drug_convergence_2;
set thesis.bayes_drug_convergence;
run;
```

```
data thesis.bayes_tau2_drug_convergence_2;
set thesis.bayes_tau2_drug_convergence;
run;
```

```
data thesis.bayes_tau_drug_convergence_2;
set thesis.bayes_tau_drug_convergence;
run;
```

```
data thesis.bayes_ite_convergence_2;
set thesis.bayes_ite_convergence;
run;
```

A.2.3 Section 6.4 - SAS PROC MCMC for Simulation with Different Prior

```
*/ AUTHOR GIULIO FLORE 10/01/2015;

*/ Cycles:          # of treatment cycles;
*/ Periods:         # of treatment periods within a cycle;
*/ Patients:        # of patients to be randomized;
*/ Randomizer:      # randomization seed for repeatability;
*/ beta0:           # fixed intercept of re model;
*/ beta1:           # fixed drug of re model;
*/ sigma:          # error term sd;
*/ tau_int:         # random intercept sd;
*/ tau_drug:        # random drug sd;

%macro simulator ( Cycles,
                  Periods,
                  Patients,
                  Randomizer,
                  beta0,
                  beta1,
                  sigma,
                  tau_int,
                  tau_drug);

*/ macro trial assume a RE structure with random intercept and drug;

options nosource nonotes; */ exclude output;

data thesis.tmp;
put ID Cycle Period Treatment;
stop;
run;

*/ create file with sequence of random intercept values;

proc iml;                                */ Matrix language proc to generate matrix of random values;
u = j(&Patients,1);                       */ allocate vector with # of subjects;
call randseed(&randomizer);               */ fix randomisation for repeatability;
call randgen(u, "Normal",0,&tau_int);      */ fill it ;
create thesis.randomintercept from u;      */ generate new file with randomized values;
append from u;
quit;                                       */ stop proc iml - specific command;

*/ add crossreference ID;

data thesis.randomintercept;
set thesis.randomintercept;
id = _N_; */ add index for merger;
rename col1 = intercept_re;
run;

*/ create file with sequence of random drug values;

proc iml;
w = j(&Patients,1);
call randseed(&randomizer+1000);
call randgen(w, "Normal",0,&tau_drug);
create thesis.randomdrug from w;
append from w;
quit;

*/ add cross reference ID ;

data thesis.randomdrug;
set thesis.randomdrug;
id = _N_; */ add index for merger;
rename col1 = drug_re;
run;
```

```

%do i = 1 %to &Patients;

%let random = %sysevalf(&i + &randomizer);

proc plan seed = &random;                                */ seed fix randomisation for repeatability;
factors Cycle=&Cycles ordered Period=&Periods ordered / noprint; */ gives sequence of periods and cycles for individual;
treatments Treatment=2 cyclic (1 2);                    */ random assignment to treatment set as dummy;
output out = thesis.rand
Treatment random;
run;

data thesis.tmp;                                        */ append new block of trial information;
set thesis.tmp thesis.rand;
if Cycle = . then delete;                               */ clear empty record;
run;

%end; */ close patient loop;

data thesis.tmp;                                       */ add id, merge random drugs, create outcome value;
set thesis.tmp;
id = floor((_N_-1)/(&Cycles*&Periods))+1; */ assign id by blocks of 6;
run;

data thesis.tmp;                                       */ add id, merge random drugs, create outcome value;
set thesis.tmp;
merge thesis.tmp thesis.randomintercept thesis.randomdrug;
by id;
run;

data thesis.tmp;                                       */ add id, merge random drugs, create outcome value;
call streaminit(&randomizer+2000);                    */ enforce repeatability of error random values;
set thesis.tmp;
treatment = treatment - 1;                             */ set treatment as dummy;
error = rand("Normal",0,&sigma);
facit_f = &beta0 + intercept_re + (treatment*&beta1)+(treatment*drug_re) + error; */ random values for outcome;
facit_f = round(facit_f,0.5);                          */ round value to reflect reported trial values;
run;

%mend; */ close macro;

*/ Population: number of patients to be added in simulator;
*/ Cycles: treatment cycles;
*/ Periods: number of treatment periods;
*/ Initial: size of initial group of subjects;
*/ Seeding: randomization parameter for repeatability
*/ beta0, beta1, sigma, tau_int tau_drug simulator parameters;

%macro Bayes_Simulation_IG (Population,
                           Cycles,
                           Periods,
                           Initial,
                           Seeding,
                           beta0,
                           beta1,
                           sigma,
                           tau_int,
                           tau_drug);

options nosource nonotes; */ suppress output;

*/ Set repository for precision data on ITE values;

data thesis.bayes_sim_ite;
put ITE ITE_SE P2_5 P5 P50 P95 P97_5;
stop;
run;

data thesis.bayes_sim_drug_re;
put Drug_RE Drug_RE_SE P2_5 P5 P50 P95 P97_5;
stop;
run;

```

```

*/ create initial sample;

%simulator(&Cycles,&Periods,&Initial,&Seeding,&beta0,&beta1,&sigma,&tau_int,&tau_drug);

*/ initial estimation with vague priors;

*/ initial block marked number of markov chains;
*/ thinning of simulators;
*/ covariance estimation assumptions;
*/ randomization initial value;
*/ outputting of statistics;
*/ outputting of chain estimates for total effect estimation;

*/ Crucial - single subject analysis requires long chains;
*/ of simulation. Lack of convergence produce biased posteriors;
*/ and distort the analysis - Also ;

*/ Apply 1M chains (nmc);
*/ Apply appropriate thinning (thin) to 'defeat' ACF and produce at least 5,000 sampling for convergence statistics;
*/ Apply higher burn-in (ntu): 50000;
*/ Apply proposal covariance (propcov) with quasi-NR algorithm ;

proc mcmc data=thesis.tmp outpost = thesis.mcmc

    nmc    = 1000000
    thin   = 250
    ntu    = 50000
    DIC
    propcov = quanew
    monitor = ( _parms_ LOGLIKE LOGPOST LOGPRIOR)
    seed   = &Seeding
    statistics(alpha = 0.05 percentage = (2.5 5 50 95 97.5) ) = all;

parms intercept 26 drug 0 sigma2 50 tau2_int 50 tau2_drug 50; */ speculative initial value assumptions;

*/ priors selected to reflect the nature of the data, with minimal assumptions;

prior intercept ~ normal(26, var = 64); */ this prior assumes the intercept to be in the mid range of facit-f;

prior drug ~ normal( 0, var = 64); */ this is a moderately skeptic prior with equipoise on the direction of the treatment effect;

*/ Non informative prior for random effects variance following Gelman (2006);

prior sigma2 ~ uniform(0,100);
prior tau2_int ~ uniform(0,100);

*/ Informative prior for small random effects variance following Spiegelhalter;

prior tau2_drug ~ igamma(0.001 , scale = 0.001);

*/ set the hierarchical structure;

random intercept2 ~ normal(0, var = tau2_int) subject = id monitor = (intercept2);
random drug2 ~ normal(0, var = tau2_drug) subject = id monitor = (drug2);

*/ set a random intercept and drug model for the mean;

mu = intercept + intercept2 + drug * treatment + drug2 * treatment;
model facit_f ~ normal(mu, var = sigma2);

*/ output file statements;

ods select DIC PostSummaries;
ods output DIC = THESIS.bayes_sim_dic;
ods output PostSummaries = thesis.post;

run;

data thesis.post;          */ generate copies conditionally;
set thesis.post;
if _N_ > 5 then delete;  */ only distributional parameters;
run;

```

```

*/ Extract sigma, tau intercept and tau treatment estimates;

data thesis.mcmc_random;
set thesis.mcmc;
sigma = sigma2**0.5;
tau_int = tau2_int**0.5;
tau_drug = tau2_drug**0.5;
run;

proc univariate data = thesis.mcmc_random NOPRINT PCTLDEF = 3; /* Output Bayesian summaries into files with quantiles;
var sigma;
output out = thesis.bayes_sim_sigma
      mean = Sigma
      std = Sigma_SE
      pctlpre = P
      pctlpts = 2.5 5 50 95 97.5;
run;

proc univariate data = thesis.mcmc_random NOPRINT PCTLDEF = 3;
var tau_int;
output out = thesis.bayes_sim_tau_int
      mean = Tau_Intercept
      std = Tau_Intercept_SE
      pctlpre = P
      pctlpts = 2.5 5 50 95 97.5;
run;

proc univariate data = thesis.mcmc_random NOPRINT PCTLDEF = 3;
var tau_drug;
output out = thesis.bayes_sim_tau_drug
      mean = Tau_Drug
      std = Tau_Drug_SE
      pctlpre = P
      pctlpts = 2.5 5 50 95 97.5;
run;

*/ generate total effect estimates in each sampled markov chain;
*/ rough solution - no dynamic extension to pre-specified number of subjects;
*/ must invest time on function that counts vars with initial string drug2_ ;

*/ Transpose DIC results for stacking;

proc transpose      data = THESIS.bayes_sim_dic
                   out = THESIS.bayes_sim_dic;
run;

*****;
*/
*/ Set initial Parameters for analysis and derivation of statistics outside the loop *;
*/
*****;

%let beta1_prior = 0;          /* derived from vague prior in first estimation;
%let beta1_sd_prior = 8;

*/ also set first priors for use in clinical BF evaluation;

%let beta1_first_prior = 0;   /* derived from vague prior in first estimation;
%let beta1_sd_first_prior = 8;

data thesis.post;
set thesis.post;
if Parameter = 'intercept' then do;
      call symput('beta0_prior',mean);
      call symput('beta0_sd_prior',stddev);
end;
else
if Parameter = 'drug' then do;
      call symput('beta1_posterior',mean); /* get a posterior value for BF estimation;
      call symput('beta1_sd_posterior',stddev);
end;
else

```

```

if Parameter = 'sigma2' then do;
    call symput('sigma2_prior',mean);
    call symput('sigma2_sd',stddev);
end;
else
if Parameter = 'tau2_int' then do;
    call symput('tau2_int_prior',mean);
    call symput('tau2_int_sd',stddev);
end;
else
if Parameter = 'tau2_drug' then do;
    call symput('tau2_drug_prior',mean);
    call symput('tau2_drug_sd',stddev);
end;
run;

*/ The BF needs to be interpreted in relation to the selected prior. If the BF is set against;
*/ the first proposed prior, each new subject inclusion implies a testing of the Null Hypothesis;
*/ as expressed in the first prior - the BF reduces to the posterior odds, as the prior odds are set to 1;
*/ a clinical significance Ho is also set;

data thesis.post;
set thesis.post;
BF = 0;
if Parameter = 'drug' then
do;
BF = (CDF('NORMAL',0,
    &beta1_posterior,
    &beta1_sd_posterior))/(1-CDF('NORMAL',0,
    &beta1_posterior,
    &beta1_sd_posterior));

end;
run;

*/ derive additional macro variables for execution;
*/ %sysevalf performs floating point calculations;

%let beta1_prior = &beta1_posterior; */ update prior to posterior for feedback;
%let beta1_sd_prior = &beta1_sd_posterior; */ update prior to posterior for feedback;

*/ set variance for prior update;

%let beta0_var_prior = %SYSEVALF(&beta0_sd_prior**2);
%let beta1_var_prior = %SYSEVALF(&beta1_sd_prior**2);

*/ derive uniform parameters for proc mcmc hyperprior;

%let sigma2_prior_a = %SYSEVALF(&sigma2_prior-(&sigma2_sd*(3**0.5)));
%let sigma2_prior_b = %SYSEVALF(&sigma2_prior+(&sigma2_sd*(3**0.5)));

%let tau2_int_prior_a = %SYSEVALF(&tau2_int_prior-(&tau2_int_sd*(3**0.5)));
%let tau2_int_prior_b = %SYSEVALF(&tau2_int_prior+(&tau2_int_sd*(3**0.5)));

*/ derive IG parameters for proc mcmc hyperprior;

%let tau2_drug_prior_location = %SYSEVALF(((&tau2_drug_prior/&tau2_drug_sd)**2) + 2);
%let tau2_drug_prior_scale = %SYSEVALF(&tau2_drug_prior*(((&tau2_drug_prior/&tau2_drug_sd)**2) + 1));

*/ NEW OBSERVATION INCREASE LOOP;

%do k = 1 %to &Population;

%put Cycle #&k; */ mark loops;

*/ generate a n = 1 sample;

%let starter = %sysevalf(&k + &seeding);

*/ randomization parameter linked to loop index for repeatability;

```

```

%simulator(&Cycles,&Periods,1,&starter,&beta0,&beta1,&sigma,&tau_int,&tau_drug);

*/ analyse new observation using the first analysis posterior as prior;
*/ initial block marked number of Markov chains;
*/ thinning of simulators;
*/ covariance estimation assumptions;
*/ randomization initial value;
*/ outputting of statistics;
*/ randomization linked to loop index for repeatability;

*/ based on analysis of real data ACF curtailed by large burn-in;
*/ for speed set shorter chain after first run with non-informative prior;

*/ Set thning at lag = 20 to avoid ACF effects on the estimate;
*/ while obtaining at least 4,000 observations for convergence tests;

proc mcmc data = thesis.tmp outpost = thesis.mcmc

  nmc = 80000
  thin = 20
  ntu = 50000
  propcov = quanev
  monitor = ( _parms_ LOGLIKE LOGPOST LOGPRIOR)
  seed = &k
  DIC
  statistics(alpha=0.05 percentage = (2.5 5 50 95 97.5) ) = all;

  */ parameter declaration and initial values statements based on priors to speed up convergence;

  parms intercept &beta0_prior
    drug &beta1_prior
    sigma2 &sigma2_prior
    tau2_int &tau2_int_prior
    tau2_drug &tau2_drug_prior;

  */ these priors are based on the previous cycle posteriors;

  prior intercept ~ normal(&beta0_prior, var = &beta0_var_prior);
  prior drug ~ normal(&beta1_prior, var = &beta1_var_prior);

  prior sigma2 ~ uniform(&sigma2_prior_a, &sigma2_prior_b);
  prior tau2_int ~ uniform(&tau2_int_prior_a, &tau2_int_prior_b);

  prior tau2_drug ~ igamma(&tau2_drug_prior_location, scale = &tau2_drug_prior_scale);

  */ random part statement - subject as hierarchical unit;

  random intercept2 ~ normal(0, var = tau2_int) subject = id monitor = (intercept2); */ set the grouping;
  random drug2 ~ normal(0, var = tau2_drug) subject = id monitor = (drug2);

  */ mean model statement;

  mu = intercept + intercept2 + (drug * treatment) + (drug2 * treatment); */ set a random drug model;
  model facit_f ~ normal(mu, var = sigma2);

  */ output file statements;

  ods select DIC PostSummaries;
  ods output DIC = thesis.tmp_DIC;
  ods output PostSummaries = thesis.tmp_posterior_theta;
run;

*/ extract variance distributions to track them as SD and SD s.e;

data thesis.mcmc_random;
set thesis.mcmc;
sigma = sigma2**0.5;
tau_int = tau2_int**0.5;
tau_drug = tau2_drug**0.5;
run;

```

```

proc univariate data = thesis.mcmc_random NOPRINT PCTLDEF = 3; /* Output Bayesian summaries into files with quantiles;
var sigma;
output out = thesis.sigma_tmp
      mean = Sigma
      std = Sigma_SE
      pctlpre = P
      pctlpts = 2.5 5 50 95 97.5;
run;

proc univariate data = thesis.mcmc_random NOPRINT PCTLDEF = 3;
var tau_int;
output out = thesis.tau_int_tmp
      mean = Tau_Intercept
      std = Tau_Intercept_SE
      pctlpre = P
      pctlpts = 2.5 5 50 95 97.5;
run;

proc univariate data = thesis.mcmc_random NOPRINT PCTLDEF = 3;
var tau_drug;
output out = thesis.tau_drug_tmp
      mean = Tau_Drug
      std = Tau_Drug_SE
      pctlpre = P
      pctlpts = 2.5 5 50 95 97.5;
run;

/* combine random variance as sd estimates;

data thesis.bayes_sim_sigma; /* Stack the data;
set thesis.bayes_sim_sigma thesis.sigma_tmp;
run;

data thesis.bayes_sim_tau_int;
set thesis.bayes_sim_tau_int thesis.tau_int_tmp;
run;

data thesis.bayes_sim_tau_drug;
set thesis.bayes_sim_tau_drug thesis.tau_drug_tmp;
run;

/* extract ite distribution to track RE pre err and ITE s.e;

data thesis.mcmc_ite;
set thesis.mcmc;
ITE = DRUG + DRUG2_1;
KEEP DRUG2_1 ITE;
run;

proc univariate data = thesis.mcmc_ite NOPRINT PCTLDEF = 3; /* Output Bayesian summaries into files with quantiles;
var ite;
output out = thesis.ite_tmp
      mean = ITE
      std = ITE_SE
      pctlpre = P
      pctlpts = 2.5 5 50 95 97.5;
run;

proc univariate data = thesis.mcmc_ite NOPRINT PCTLDEF = 3;
var drug2_1;
output out = thesis.drug_re_tmp
      mean = Drug_RE
      std = Drug_RE_SE
      pctlpre = P
      pctlpts = 2.5 5 50 95 97.5;
run;

/* combine re estimate and ite estimate;

data thesis.bayes_sim_ite; /* Stack the data;
set thesis.bayes_sim_ite thesis.ite_tmp;
run;

```

```

data thesis.bayes_sim_drug_re; /* Stack the data;
set thesis.bayes_sim_drug_re thesis.drug_re_tmp;
run;

*/ Update DIC output;

proc transpose data = thesis.tmp_DIC
    out = thesis.tmp_DIC;
run;

data THESIS.bayes_sim_dic; /* Stack the data;
set THESIS.bayes_sim_dic thesis.tmp_DIC;
run;

*/ Update data repository of posteriors ;

data thesis.tmp_post; /* generate copies conditionally;
set thesis.tmp_posterior_theta;
if _N_ > 5 then delete; /* only distributional parameters;
run;

*/*****;
*/
*/ Update all relevant posterior parameters */;
*/
*/*****;

data thesis.tmp_post;
set thesis.tmp_post;
if Parameter = 'intercept' then do;
    call symput('beta0_prior',mean);
    call symput('beta0_sd_prior',stddev);
end;

else
if Parameter = 'drug' then do;
    call symput('beta1_posterior',mean); /* get a posterior value for BF estimation;
    call symput('beta1_sd_posterior',stddev);
end;

else
if Parameter = 'sigma2' then do;
    call symput('sigma2_prior',mean);
    call symput('sigma2_sd',stddev);
end;

else
if Parameter = 'tau2_int' then do;
    call symput('tau2_int_prior',mean);
    call symput('tau2_int_sd',stddev);
end;

else
if Parameter = 'tau2_drug' then do;
    call symput('tau2_drug_prior',mean);
    call symput('tau2_drug_sd',stddev);
end;

run;

*/ The BF needs to be interpreted in relation to the selected prior. If the BF is set against;
*/ the first proposed prior, each new subject inclusion implies a testing of the Null Hypothesis;
*/ as expressed in the first prior - the prior odd calculations are still expressed to made the formula explicit ;

data thesis.tmp_post;
set thesis.tmp_post;
BF = 0;
if Parameter = 'drug' then
do;
BF = (CDF('NORMAL',0,
    &beta1_posterior,
    &beta1_sd_posterior))/(1-CDF('NORMAL',0,
    &beta1_posterior,
    &beta1_sd_posterior)) ;

end;
run;

```

```

*/ derive additional macro variables for execution;
*/ %sysevalf performs floating point calculations;

%let beta1_prior = &beta1_posterior;          */ update prior to posterior for feedback;
%let beta1_sd_prior = &beta1_sd_posterior;    */ update prior to posterior for feedback;

%let beta0_var_prior = %SYSEVALF(&beta0_sd_prior**2); */ set variance update for prior;
%let beta1_var_prior = %SYSEVALF(&beta1_sd_prior**2);

%let sigma2_prior_a = %SYSEVALF(&sigma2_prior-(&sigma2_sd*(3**0.5)));
%let sigma2_prior_b = %SYSEVALF(&sigma2_prior+(&sigma2_sd*(3**0.5)));

%let tau2_int_prior_a = %SYSEVALF(&tau2_int_prior-(&tau2_int_sd*(3**0.5)));
%let tau2_int_prior_b = %SYSEVALF(&tau2_int_prior+(&tau2_int_sd*(3**0.5)));

*/derive location parameter for proc mcmc hyperprior;

%let tau2_drug_prior_location = %SYSEVALF(((&tau2_drug_prior/&tau2_drug_sd)**2) + 2);
%let tau2_drug_prior_scale = %SYSEVALF(&tau2_drug_prior*(((&tau2_drug_prior/&tau2_drug_sd)**2) + 1));

*/ CLOSE PARAMETER UPDATING SECTION;

data thesis.post; */ stack posterior estimates;
set thesis.post thesis.tmp_post;
run;

*/ clear temporary files;

proc datasets library = thesis NOLIST;
delete      mcmc tmp tmp_post
            tmp_posterior_theta drug_re_tmp ite_tmp mcmc_ite mcmc_random
            sigma_tmp tau_int_tmp tau_drug_tmp;
run;

%end; */ close loop;

*/ remove labels from proc univariate outputs and add sample size;

data thesis.bayes_sim_sigma;
set thesis.bayes_sim_sigma;
label Sigma = '';
label Sigma_SE = '';
label P2_5 = '';
label P5 = '';
label P50 = '';
label P95 = '';
label P97_5 = '';
Sample = _N_+&Initial-1;
run;

data thesis.bayes_sim_tau_int;
set thesis.bayes_sim_tau_int;
label Tau_Intercept = '';
label Tau_Intercept_SE = '';
label P2_5 = '';
label P5 = '';
label P50 = '';
label P95 = '';
label P97_5 = '';
Sample = _N_+&Initial-1;
run;

data thesis.bayes_sim_tau_drug;
set thesis.bayes_sim_tau_drug;
label Tau_Drug = '';
label Tau_Drug_SE = '';
label P2_5 = '';
label P5 = '';
label P50 = '';
label P95 = '';
label P97_5 = '';
Sample = _N_+&Initial-1;
run;

```

```

*/ Add sample informations to ite info;

data thesis.bayes_sim_ite;
set thesis.bayes_sim_ite;
Sample = _N_; */ add sample size data;
run;

data thesis.bayes_sim_drug_re;
set thesis.bayes_sim_drug_re;
Sample = _N_+&Initial; */ add sample size data - no need to subtract initial value as not calculated at first analysis;
run;

*/ Format DIC table;

DATA THESIS.bayes_sim_dic;
SET THESIS.bayes_sim_dic;
RENAME COL1 = MeanDev;    */ Mean Posterior Deviance;
RENAME COL2 = DevMean;    */ Deviance for Best (Mean) Parameter Estimate;
RENAME COL3 = Pd;         */ Effective Parameters;
RENAME COL4 = DIC;        */ DIC;
drop _NAME_;
Sample = _N_+&Initial-1;  */ add sample size data;
RUN;

*/ split data for separate analysis of parameters;

data thesis.bayes_sim_intercept
thesis.bayes_sim_drug
thesis.bayes_sim_sigma2
thesis.bayes_sim_tau2_int
thesis.bayes_sim_tau2_drug;
set thesis.post;
if parameter = 'intercept' then output thesis.bayes_sim_intercept;
if parameter = 'drug' then output thesis.bayes_sim_drug;
if parameter = 'sigma2' then output thesis.bayes_sim_sigma2;
if parameter = 'tau2_int' then output thesis.bayes_sim_tau2_int;
if parameter = 'tau2_drug' then output thesis.bayes_sim_tau2_drug;
run;

*/ add sample size;

data thesis.bayes_sim_intercept;
set thesis.bayes_sim_intercept;
Sample = _N_+&Initial-1;
drop BF;
run;

data thesis.bayes_sim_drug;
set thesis.bayes_sim_drug;
Sample = _N_+&Initial-1;
P = 1-CDF('NORMAL',0,Mean,StdDev); */ P(PTE>0);
P_CLINICAL = 1-CDF('NORMAL',3,Mean,StdDev); */ P(PTE>3);
LOG10_BF = LOG10(BF); */ Jeffreys version;
drop Parameter;
run;

data thesis.bayes_sim_sigma2;
set thesis.bayes_sim_sigma2;
Sample = _N_+&Initial-1;
drop BF;
run;

data thesis.bayes_sim_tau2_int;
set thesis.bayes_sim_tau2_int;
Sample = _N_+&Initial-1;
drop BF;
run;

data thesis.bayes_sim_tau2_drug;
set thesis.bayes_sim_tau2_drug;
Sample = _N_+&Initial-1;
drop Parameter BF;

```

```

run;

%mend; */ close macro;

***** CONVERGENCE CHECK *****;

%macro powercycle_IG(Simulations_p,
    Population_p,
    Cycles_p,
    Periods_p,
    Initial_p,
                                Seeding_p,
    beta0_p,
    beta1_p,
    sigma_p,
    tau_int_p,
    tau_drug_p);

data thesis.bayes_drug_convergence;
put N Mean StdDev P2_5 P5 P50 P95 P97_5 Sample
    P P_Clinical Log10_BF Log10_BF_Clinical;
stop;
run;

data thesis.bayes_sigma2_convergence;
put N Mean StdDev P2_5 P5 P50 P95 P97_5 Sample;
stop;
run;

data thesis.bayes_tau2_int_convergence;
put N Mean StdDev P2_5 P5 P50 P95 P97_5 Sample;
stop;
run;

data thesis.bayes_tau2_drug_convergence;
put N Mean StdDev P2_5 P5 P50 P95 P97_5 Sample;
stop;
run;

data thesis.bayes_ite_convergence;
put ITE ITE_SE P2_5 P5 P50 P95 P97_5 Sample;
stop;
run;

data thesis.bayes_tau_drug_convergence;
put Tau_Drug Tau_Drug_se P2_5 P5 P50 P95 P97_5 Sample;
stop;
run;

%do counter = 1 %to &Simulations_p;

%let seeded = %ysevalf(&Seeding_p+(&counter*1000));

%put Hypercycle #&counter; */ mark loops;

%Bayes_Simulation_IG(&Population_p,&Cycles_p,&Periods_p,&Initial_p,&seeded,
    &beta0_p,&beta1_p,&sigma_p,&tau_int_p,&tau_drug_p);

*/ stack output files;

data thesis.bayes_drug_convergence;
set thesis.bayes_drug_convergence thesis.bayes_sim_drug;
run;

data thesis.bayes_sigma2_convergence;
set thesis.bayes_sigma2_convergence thesis.bayes_sim_sigma2;
run;

data thesis.bayes_tau2_int_convergence;
set thesis.bayes_tau2_int_convergence thesis.bayes_sim_tau2_int;
run;

```

```

data thesis.bayes_tau2_drug_convergence;
set thesis.bayes_tau2_drug_convergence thesis.bayes_sim_tau2_drug;
run;

data thesis.bayes_tau_drug_convergence;
set thesis.bayes_tau_drug_convergence thesis.bayes_sim_tau_drug;
run;

data thesis.bayes_ite_convergence;
set thesis.bayes_ite_convergence thesis.bayes_sim_ite;
run;

%end; */ close loop;

data thesis.bayes_drug_convergence;
set thesis.bayes_drug_convergence;
Simulation = ceil(_N_/(&Population_p+1));
run;

data thesis.bayes_sigma2_convergence;
set thesis.bayes_sigma2_convergence;
Simulation = ceil(_N_/(&Population_p+1));
run;

data thesis.bayes_tau2_int_convergence;
set thesis.bayes_tau2_int_convergence;
Simulation = ceil(_N_/(&Population_p+1));
run;

data thesis.bayes_tau2_drug_convergence;
set thesis.bayes_tau2_drug_convergence;
Simulation = ceil(_N_/(&Population_p+1));
run;

data thesis.bayes_tau_drug_convergence;
set thesis.bayes_tau_drug_convergence;
Simulation = ceil(_N_/(&Population_p+1));
run;

*/ ite estimates are taken only for the first initial subject onwards;

data thesis.bayes_ite_convergence;
set thesis.bayes_ite_convergence;
Simulation = ceil(_N_/(&Population_p));
run;

%mend;

**** NEW POPULATION SETTING - VERY HIGH RANDOM INTERCEPT, LOW RANDOM TREATMENT;

* initial set N = 15, TOTAL 300 SUBJECTS, 30 random start samples sample;

options nonotes;

%powercycle_IG(30,285,3,2,15,300|2015,20,5,7.5,7.5,2.5);

data thesis.bayes_drug_convergence_IG;
set thesis.bayes_drug_convergence;
run;

data thesis.bayes_sigma2_convergence_IG;
set thesis.bayes_sigma2_convergence;
run;

data thesis.bayes_tau2_int_convergence_IG;
set thesis.bayes_tau2_int_convergence;
run;

data thesis.bayes_tau2_drug_convergence_IG;
set thesis.bayes_tau2_drug_convergence;
run;

```

```
data thesis.bayes_tau_drug_convergence_IG;  
set thesis.bayes_tau_drug_convergence;  
run;
```

```
data thesis.bayes_ite_convergence_IG;  
set thesis.bayes_ite_convergence;  
run;
```

References

- [1] AVINS, ANDREW L., STEPHEN BENT, AND JOHN M. NEUHAUS. "Use of an Embedded N-of-1 Trial to Improve Adherence and Increase INFORMATION FROM A CLINICAL STUDY." *CONTEMPORARY CLINICAL TRIALS* 26, NO. 3 (JUNE 2005): 397–401. DOI:10.1016/J.CCT.2005.02.004.
- [2] BERG, C., AND C. VIGNAT. "ON THE DENSITY OF THE SUM OF TWO INDEPENDENT STUDENT - RANDOM VECTORS." *STATISTICS & PROBABILITY LETTERS* 80, NO. 13–14 (JULY 15, 2010): 1043–55. DOI:10.1016/J.SPL.2010.02.019.
- [3] CASELLA, GEORGE, AND ROGER L. BERGER. *STATISTICAL INFERENCE*. DUXBURY THOMSON LEARNING, 2002.
- [4] CHEN X, CHEN P (2014) A COMPARISON OF FOUR METHODS FOR THE ANALYSIS OF N-OF-1 TRIALS. *PLOS ONE* 9(2): E87752. DOI:10.1371/JOURNAL.PONE.0087752.
- [5] CUSHING, CHRISTOPHER C, RYAN W WALTERS, AND LESA HOFFMAN. "AGGREGATED N-OF-1 RANDOMIZED CONTROLLED TRIALS: MODERN DATA ANALYTICS APPLIED TO A CLINICALLY VALID METHOD OF INTERVENTION EFFECTIVENESS." *JOURNAL OF PEDIATRIC PSYCHOLOGY* 39, NO. 2 (MARCH 2014): 138–50. DOI:10.1093/JPEPSY/JST083.
- [6] DUAN, NAIHUA, RICHARD L KRAVITZ, AND CHRISTOPHER H SCHMID. "SINGLE-PATIENT (N-OF-1) TRIALS: A PRAGMATIC CLINICAL DECISION METHODOLOGY FOR PATIENT-CENTERED COMPARATIVE EFFECTIVENESS RESEARCH." *JOURNAL OF CLINICAL EPIDEMIOLOGY* 66, NO. 8 SUPPL (AUGUST 2013): S21–28. DOI:10.1016/J.JCLINEPI.2013.04.006.
- [7] FACEY, KAREN, ALICIA GRANADOS, GORDON GUYATT, ALASTAIR KENT, NILAY SHAH, GERT JAN VAN DER WILT, AND DURHANE WONG-RIEGER. "GENERATING HEALTH TECHNOLOGY ASSESSMENT EVIDENCE FOR RARE DISEASES." *International Journal of Technology Assessment in Health Care*, November 19, 2014, 1–7. doi:10.1017/S0266462314000464.
- [8] GABLER, NICOLE B, NAIHUA DUAN, SUNITA VOHRA, AND RICHARD L KRAVITZ. "N-OF-1 TRIALS IN THE MEDICAL LITERATURE: A SYSTEMATIC REVIEW." *MEDICAL CARE* 49, NO. 8 (AUGUST 2011): 761–68. DOI:10.1097/MLR.0B013E318215D90D.
- [9] GELMAN, ANDREW. "PRIOR DISTRIBUTIONS FOR VARIANCE PARAMETERS IN MODELS (COMMENT ON ARTICLE BY BROWNE AND DRAPER)." *BAYESIAN ANALYSIS* 1, NO. 3 (SEPTEMBER 2006): 515–34. DOI:10.1214/06-BA117A.
- [10] GELMAN, ANDREW, JOHN B. CARLIN, HAL S. STERN, AND DONALD B. RUBIN. *BAYESIAN DATA ANALYSIS, SECOND EDITION*. CRC PRESS, 2003.
- [11] GOODNIGHT, J. "COMPUTING MIVQUE0 ESTIMATES OF VARIANCE COMPONENTS." SAS INSTITUTE INC, TECHNICAL REPORT, 1978.
- [12] GUYATT, GORDON H., JANA L. KELLER, ROMAN JAESCHKE, DAVID ROSENBLOOM, JONATHAN D. ADACHI, AND MICHAEL T. NEWHOUSE. "THE N-OF-1 RANDOMIZED CONTROLLED TRIAL: CLINICAL USEFULNESS. OUR THREE-YEAR EXPERIENCE." *ANNALS OF INTERNAL MEDICINE* 112, NO. 4 (FEBRUARY 15, 1990): 293–99. DOI:10.7326/0003-4819-112-4-293.
- [13] JANOSKY, J E. "USE OF THE SINGLE SUBJECT DESIGN FOR PRACTICE BASED PRIMARY CARE RESEARCH." *POSTGRADUATE MEDICAL JOURNAL* 81, NO. 959 (SEPTEMBER 2005): 549–51. DOI:10.1136/PGMJ.2004.031005.
- [14] KASS, ROBERT E., AND ADRIAN E. RAFTERY. "BAYES FACTORS." *JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION* 90, NO. 430 (JUNE 1, 1995): 773–95. DOI:10.1080/01621459.1995.10476572.

- [15] KRAVITZ, RICHARD L, NAIHUA DUAN, EDMUND J NIEDZINSKI, M CAMERON HAY, SASKIA K SUBRAMANIAN, AND THOMAS S WEISNER. "WHAT EVER HAPPENED TO N-OF-1 TRIALS? INSIDERS' PERSPECTIVES AND A LOOK TO THE FUTURE." THE MILBANK QUARTERLY 86, NO. 4 (DECEMBER 2008): 533–55. DOI:10.1111/J.1468-0009.2008.00533.X.
- [16] LESAFFRE, EMMANUEL, AND ANDREW B. LAWSON. BAYESIAN BIOSTATISTICS. JOHN WILEY & SONS, 2012.
- [17] LILFORD, R J, J G THORNTON, AND D BRAUNHOLTZ. "CLINICAL TRIALS AND RARE DISEASES: A WAY OUT OF A CONUNDRUM." BMJ (CLINICAL RESEARCH ED.) 311, NO. 7020 (DECEMBER 16, 1995): 1621–25.
- [18] LILLIE, ELIZABETH O, BRADLEY PATAY, JOEL DIAMANT, BRIAN ISSELL, ERIC J TOPOL, AND NICHOLAS J SCHORK. "THE N-OF-1 CLINICAL TRIAL: THE ULTIMATE STRATEGY FOR INDIVIDUALIZING MEDICINE?" PERSONALIZED MEDICINE 8, NO. 2 (MARCH 2011): 161–73. DOI:10.2217/PME.11.7.
- [19] LONGFORD, NICHOLAS T. "SELECTION BIAS AND TREATMENT HETEROGENEITY IN CLINICAL TRIALS." STATISTICS IN MEDICINE 18, NO. 12 (JUNE 30, 1999): 1467–74. DOI:10.1002/(SICI)1097-0258(19990630)18:12<1467::AID-SIM149>3.0.CO;2-H.
- [20] MCCULLOCH, CHARLES E., SHAYLE R. SEARLE, AND JOHN M. NEUHAUS. GENERALIZED, LINEAR, AND MIXED MODELS. JOHN WILEY & SONS, 2011.
- [21] RICE, JOHN. MATHEMATICAL STATISTICS AND DATA ANALYSIS. CENGAGE LEARNING, 2006.
- [22] ROCHON, J. "A STATISTICAL MODEL FOR THE 'N-OF-1' STUDY." JOURNAL OF CLINICAL EPIDEMIOLOGY 43, NO. 5 (1990): 499–508.
- [23] SCHLUTER, P J, AND R S WARE. "SINGLE PATIENT (N-OF-1) TRIALS WITH BINARY TREATMENT PREFERENCE." STATISTICS IN MEDICINE 24, NO. 17 (SEPTEMBER 15, 2005): 2625–36. DOI:10.1002/SIM.2132.
- [24] SCHMID CH, DUAN N, THE DECIDE METHODS CENTER N-OF-1 GUIDANCE PANEL. STATISTICAL DESIGN AND ANALYTIC CONSIDERATIONS FOR N-OF-1 TRIALS. IN: KRAVITZ RL, DUAN N, EDS, AND THE DECIDE METHODS CENTER N-OF-1 GUIDANCE PANEL (DUAN N, ESLICK I, GABLER NB, KAPLAN HC, KRAVITZ RL, LARSON EB, PACE WD, SCHMID CH, SIM I, VOHRA S). DESIGN AND IMPLEMENTATION OF N-OF-1 TRIALS: A USER'S GUIDE. AHRQ PUBLICATION NO. 13(14)-EHC122-EF. ROCKVILLE, MD: AGENCY FOR HEALTHCARE RESEARCH AND QUALITY; FEBRUARY 2014: CHAPTER 4, PP. 33-53.
- [25] SPIEGELHALTER, DAVID J., KEITH R. ABRAMS, AND JONATHAN P. MILES. BAYESIAN APPROACHES TO CLINICAL TRIALS AND HEALTH-CARE EVALUATION. VOL. 3. JOHN WILEY & SONS, INC., 2004. DOI/10.1002/PST.130.
- [26] TSAPAS, A, AND D R MATTHEWS. "N OF 1 TRIALS IN DIABETES: MAKING INDIVIDUAL THERAPEUTIC DECISIONS." DIABETOLOGIA 51, NO. 6 (JUNE 2008): 921–25. DOI:10.1007/S00125-008-0983-2.
- [27] ZUCKER, D R, C H SCHMID, M W MCINTOSH, R B D'AGOSTINO, H P SELKER, AND J LAU. "COMBINING SINGLE PATIENT (N-OF-1) TRIALS TO ESTIMATE POPULATION TREATMENT EFFECTS AND TO EVALUATE INDIVIDUAL PATIENT RESPONSES TO TREATMENT." JOURNAL OF CLINICAL EPIDEMIOLOGY 50, NO. 4 (APRIL 1997): 401–10.
- [28] ZUCKER, DEBORAH R, ROBIN RUTHAZER, AND CHRISTOPHER H SCHMID. "INDIVIDUAL (N-OF-1) TRIALS CAN BE COMBINED TO GIVE POPULATION COMPARATIVE TREATMENT EFFECT ESTIMATES: METHODOLOGIC CONSIDERATIONS." JOURNAL OF CLINICAL EPIDEMIOLOGY 63, NO. 12 (DECEMBER 2010): 1312–23. DOI:10.1016/J.JCLINEPI.2010.04.020.