UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology

Master of Statistics

Masterthesis

Power calculations for linear mixed effects models

Manjally Ndow

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Biostatistics

SUPERVISOR :

dr. Yannick VANDENDIJCK

dr. Liesbeth BRUCKERS

2017
2018

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

Maastricht University

Faculty of Sciences
School for Information Technology

Master of Statistics

Masterthesis

Power calculations for linear mixed effects models

Manjally Ndow

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Biostatistics

SUPERVISOR :

dr. Yannick VANDENDIJCK

dr. Liesbeth BRUCKERS

Abstract

Background: Power analysis and sample size estimation are very important in designing clinical studies. The two phenomena go hand-in-hand in which the former is often referred to as the process of determining the sample size for a given research. The number of subject participants required in a particular clinical experiment is difficult to determine and is often dependent on the cost at hand to conduct the study. The economic cost is mostly trade-off with meaningful sample size to detect a treatment difference at a certain power. However, the research ethics committees always solicit for justification of the study based on the sample size estimation and statistical power. The main quantities for the direct computation of these two phenomena can be estimated from a pilot study or historical data, otherwise, a reasonable guess is plausible. In this study, we simulate power analysis and sample size estimation for linear mixed-effects models for a continuous response with repeated measures within each subject.

Objective: Investigate power and sample size calculations for longitudinal designs analyzed using linear mixed-effects model with random intercepts, and random intercepts and slopes.

Methods: The analysis involved comparison between the theoretical approach and simulation approach for power and sample size calculations. The theoretical approach include the use of Diggle et al. (2002), Liu and Liang (1997) methods and specific softwares such as longpower. In the simulation approach, for each dataset, we simulate the response, the random effects and the random error and fit both random intercept models and random intercept and slope models to the dataset for estimation of statistical power and sample size. The estimated power is the proportion of the sum of the significant p-values at 0.05 alpha-level per simulation.

Results: Sample size estimates are smaller in random intercept models than in random intercept and slope models. The estimated powers are closer to the nominal power in both models and the unbalanced designs have minimal impact on the estimation of sample size.

Conclusion: Estimating sample size through simulation is greatly important since most specialized software packages have hidden features not explicitly clear to the users on how the methods of computation are performed.

Keywords: Linear mixed-effects, repeated measures, sample size, statistical power, simulation.

Dedication

I dedicate this piece of work to my beloved parents, siblings, spouse and daughter, as well as to all the rest of mankind who contributed directly or indirectly to my success stories in life.

Acknowledgment

I would like to thank everyone for their invaluable support and contribution towards making this thesis possible. I begin by expressing gratitude to the Almighty Allah for granting me the strength and wisdom to accomplish this important work. Secondly, I would like to extend my sincere gratitude and appreciation to my internal supervisor, Dr. Yannick Vandendijck at University Hasselt, Belgium for his constructive criticism and guidance during the course of writing of the thesis. His continuous support and motivation enabled me to accomplish this task.

Moreover, I wouldn't do justice without thanking the University Hasselt administration and academic staff for their commitments and invaluable supports during my study period. The thesis wouldn't have been a smooth task without acquiring the required knowledge and skills from my academic professors. Thanks very much to you all. Finally, special thanks to my beloved and supportive family for their encouragement during my academic struggles. Also, thanking VLIRUOS for granting me the scholarship to pursue this Master program, and my colleagues, friends and loved ones for all the fun and support throughout the study duration.

Manjally Ndow

January 22, 2018

University Hasselt

Diepenbeek, Belgium

Contents

List of Tables

1 Introduction

1.1 Background Information

Power and sample size calculations are important phenomena in the design stage of clinical research or trials. The term power analysis is often referred to as the process of determining the sample size for a given research, and is defined as the probability of detecting a "true or relevant" effect when it exists. However, there exist many derived formulae and/or software programs for determining sample size for particular research situations depending on the primary endpoint. And each of these formulae or programs would require the specification of some factors such as; Type I error rates (α), Type II error rates (β), effect size (δ) and measurement variability (σ^2), which are fundamental in determining statistical power and sample size calculations.

However, statistical power analysis and sample size estimation cannot be overemphasized for numerous ethical research consideration issues. The research ethics committees usually request for justifications of a study based on sample size estimation and statistical power. It is ethically unacceptable to conduct a study that is unable to detect a true effect due to a lack of statistical power and/or recruiting as many subjects as possible when only few subjects are adequate to detect the relevant treatment difference [18]. The trade-off between sample size and cost is important for numerous reasons, as an undersized study can be wasteful for not producing useful results, while an oversized study uses more resources than are necessary [11].

Unfortunately, there is no straightforward answer on how large a sample size should be in order to detect a true effect at a given statistical power, despite larger sample sizes have more statistical power [18]. But the consequences of ignoring sample size and power estimation, especially in an oversized experiment, would lead to unnecessary exposure of participants to a potentially toxic treatment, or denying them a potentially beneficial treatment [11]. However, to avoid under- or over-estimating sample size, it is pivotal to conduct pilot or similar studies in order to obtain estimates of the factors required for power and sample size computations [20]. There is no need for the pilot study or historical data to follow the same design as the planned study [11], but careful consideration is important to obtain estimates for these factors. In the absence of a pilot study or historical data, a plausible reasonable guess to obtain these estimates is necessary [5]. Moreover, it is always essential to conduct pilot studies to estimate the quantities mentioned earlier for sample size calculation. Since the effect size is usually the

'true effect', it is often important to be determined through a scientific knowledge or judgment, otherwise could be estimated from the pilot study. For example, in a hypertensive study, the relevant parameters for measuring diastolic blood pressure (DBP) few years ago might not be the same in recent times or years, therefore, obtaining an estimate for the treatment difference in DPB from such pilot studies may under- or over-estimate the sample size and power analysis.

Furthermore, in this study, we estimate the statistical power and sample size required for linear mixed-effects models by comparing the results obtained from the theoretical approach (such as the longpower function in R, Diggle et al. (2002), Liu and Liang (1997) formulae) [7] with the simulation approach. However, in the simulation approach, we simulate datasets and fit linear mixed-effects models to the simulated response data using the "lme" function in the R software using maximum likelihood (ML) method, and retain the p-values for the covariate of interest per simulation to estimate the statistical power [12]. The estimated power is the proportion of the sum of the significant p-values per simulation in which the null hypothesis would be rejected at a one-tailed significance level (α) of 0.05 (5%).

Finally, for each dataset, we simulate the response, the random effects and the random error and fit both the random intercept model and the random intercept and slope model for the estimation of the statistical power and sample size.

1.2 Objectives

1.2.1 General objective

To investigate power and sample size calculation approaches for longitudinal designs which are analyzed using linear mixed-effects models.

1.2.2 Specific objectives

- Estimate power and sample size for random intercept model and random intercept and slope model using both theoretical and simulation approaches.
- Examine the effect of balanced and unbalanced longitudinal study designs on sample size estimation.
- Evaluate the impact of number of repeated measurements per subject on power analysis through sensitivity analysis by varying the number of repeated measurements.

2 Literature Review

There are several studies that have been conducted to estimate sample size and power analysis in clinical study designs. The main goal is to have an insight of the number of subjects required in a given study trial to achieve a certain power with consideration of various research protocols and ethical principles. Power and sample size calculations mainly depends on several factors or quantities such as Type I and II error probabilities, effect size and measurement variability. However, effect size is one of the key factors in determining the number of subjects required in a given study, as small or large effect sizes would lead to large or small sample sizes respectively.

Button et al. [2] studied the relationship between power failure and small sample size, and how the latter undermines the reliability of neuroscience. They showed that the average statistical power of studies in the neurosciences is very low, which could be the consequence of overestimates of effect size and low reproducibility of results. They also argued the importance of appreciating the wastage associated with an underpowered study and claimed even a study achieving only 80% power still presents a 20% possibility that the animals have been sacrificed without the study detecting the underlying true effect. Cohen [3] has determined standardized effect sizes described as "small", "medium" and "large", and these varies for different study designs, depending on the test (e.g. difference between two means or many means, etc).

Sample size estimation and power analysis are paramount but often misunderstood by most researchers. Cunningham and McCrum-Gardner [4] used a simple and freely available statistical software called GPower to address concerns related to sample size and power calculations. This software tool is also user-friendly irrespective of your statistical background and several statistical tests like t-tests, analysis of variance (ANOVA) and chi-square tests can be performed. However, the software program is not applicable to longitudinal data analysis or used for analyzing linear mixed-effects models. Donohue et al. [6] discussed power and sample size estimation for randomized placebo controlled studies for interaction between treatment and time in a linear mixed-effects model. They also demonstrated the relationship between random intercept model and marginal model with exchangeable correlation, and further illustrated on how to derive the correlation and variance-covariance matrix for the random intercept model, and the random intercept and slope model. In another study [7], they introduced the longpower function in R and used the formulae in [5] and [13] to estimate sample size and power.

The importance of conducting a pilot study as discussed by Lenth [11] and Teare et al. [20], emphasized on obtaining an estimate of one or more error variances and specifying the smallest meaningful treatment difference of importance for sample size determination. Moreover, the consequences of undersized or oversized study was also discussed and outlined strategies or methods of overcoming such situations. Leon and Hoe [12] in their paper, calculated sample size to detect various standardized main effects and interaction between two binary fixed effects in a mixed-effects linear regression model with a random intercept using the formula proposed by Diggle et al. (2002) and a simulation approach. In their analysis, the sample size needed to detect an interaction effect is four times that for detecting a main effect of the identical magnitude because the sample size is a linear function of the variance of an effect estimate.

Moreover, since sample size also depends on the number of post-baseline measurements, with smaller sample sizes needed when there are more measurements, Naiji et al. [16] have discussed that longitudinal studies always provide more statistical power than cross-sectional studies, especially when the within-subject measurements are correlated. They further illustrated that the required sample size to detect desired effect size increases as ρ approaches 1 and decreases as ρ approaches 0. This is because repeated measurements are more (as $\rho \sim 1$) or less (as $\rho \sim 0$) similar to each other and provide additional information on the subjects. They also discussed the difference between hypothesis testing and power analysis, of which in the former, it involved testing whether there is evidence against the H_0 based on a specified significance level, while in the latter, both null and alternative hypotheses are fully considered when estimating the power.

Finally, Liu and Liang [13] computed sample size and statistical power for correlated observations through multivariate extension of the work by Self and Mauritsen [19]. They discussed sample size calculation for special case of continuous and binary responses with repeated measurements, and showed that the empirical power estimates and the nominal power were very much similar. Liu and Wu [14] have shown sample size calculation and power analysis for time-averaged difference for unequal sample sizes between two groups for both continuous and binary measures. They also explored the relative importance of number of unique subjects and number of repeated measurements within each subject on statistical power through simulation. Furthermore, they discussed the importance of unbalanced designs for allocating smaller number of subjects to group that is either more expensive, hard to recruit or with limited number of available subjects.

3 Methodology

The analysis involved comparison between the theoretical approach and the simulation approach for power and sample size calculations. The theoretical approach include the use of Diggle et al. (2002), and Liu and Liang (1997) formulae and its implementation in the longpower function in R. In the simulation approach, statistical powers were estimated through testing the significance of the parameter of interest by summing all the significant p-values and divide by simulations.

3.1 The Linear Mixed-Effects Models

Linear mixed-effects models (LMMs) are very useful in modelling datasets with complex, hierarchical structures [10]. The term mixed-effects referred to both the fixed effects and the random effects [21]. The linear mixed-effects model used in this analysis is specified as follows:

$$Y_{ij} = \beta_0 + \beta_1 * treatment_i + \beta_2 * time_{ij} + \beta_3 * treatment_i * time_{ij} + b_{0i} + b_{1i} * time_{ij} + \epsilon_{ij} \quad (1)$$

Hence, Y_{ij} is the continuous response of interest, for the i^{th} subject, measured at time t_{ij}, where (i = 1, ..., m and j = 1, ..., n), β_0 is the intercept, β_1, β_2 and β_3 are the fixed effect parameters, $\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, D\right)$ and D $= \begin{pmatrix} \sigma_b^2 & \rho\sigma_b\sigma_s \\ \rho\sigma_b\sigma_s & \sigma_s^2 \end{pmatrix}$ is the variance-covariance matrix, b_{0i} and b_{1i} are the random effects representing random intercepts and random slopes respectively, whereas $\epsilon_{ij} \sim N(0, \sigma_e^2)$ and denote the random error terms. The random effects and random errors are assumed to be independently and identically normally distributed (iid). Additional covariates such as gender and age would be added to model (1), which is fitted in R using the lme function. Assuming that we have two treatment groups (A = experiment and B = control), therefore, a model for each group can be constructed as follows:

$$Y_{ij} = \begin{cases} \beta_{0A} + \beta_{1A} * treatment_i + \beta_{2A} * time_{ij} + \beta_{3A} * treatment_i * time_{ij} + b_{0i} + \\ b_{1i} * time_{ij} + \epsilon_{ij} \\ \beta_{0B} + \beta_{1B} * treatment_i + \beta_{2B} * time_{ij} + \beta_{3B} * treatment_i * time_{ij} + b_{0i} + \\ b_{1i} * time_{ij} + \epsilon_{ij} \end{cases} \quad (2)$$

Thus, the parameter of interest is β_3 which is interpreted as the change in the response Y_{ij}, compared to subject *i's* own average $\bar{Y}_i$, due to treatment A, and is tested using the following hypotheses: the null hypothesis, $H_0 : \beta_{3A} \leq 0$, against the alternative hypothesis, $H_1 : \beta_{3A} > 0$. The random effects are also interpreted as the additional change in Y_{ij} due to subject i itself, despite of any change due to either treatment effect.

3.2 Power and Sample Size Calculations

The number of subjects (samples) required to achieve a certain power is one of the challenges encountered in most clinical trials. Sample size calculation is based on the nature of the primary endpoint (normal, binary, survival, etc.) and on the planned method of analysis. The trade-off between meaningful sample size and cost is very important to consider, as too large trials are expensive and subjects could be exposed to unknown risk factors, whereas too small trials may miss a biological relevant difference or treatment effect. However, sample size and statistical power calculations go hand-in-hand and the essential factors or quantities that have direct impact on their computations are α, β, δ and σ^2. Additional quantities such as n and ρ are needed for longitudinal studies. Further description of these factors or quantities are as follows:

3.2.1 Type I Error Rate (α)

The probability that the study will reject the null hypothesis (H_0) when it is true. This corresponds to reporting a significant difference in treatment between two groups when in fact there is none. This is typically fixed as significance level and the choice of α often used is 0.05. In general, the control of α, i.e. false-positive, is of primary importance to the agencies. The smaller the alpha, the smaller the probability of rejecting the null hypothesis [17]. Moreover, as α increases the sample size decreases, whereas decreasing α would increase the sample size.

3.2.2 Type II Error Rate (β)

The probability that the study will not reject the null hypothesis when it is false, which corresponds to wrong conclusion of a lack of benefit when in truth one exists. The control of β, i.e. false-negative, is very important for the sponsor. The power of a statistical test (1 - β), is the probability that the test rejects the null hypothesis if the alternative is true. Statistical power is dependent on a number of factors, and is conventionally or often set at 0.80 (80%) to detect the difference in treatment between groups. The smaller the β, the larger the power of the statistical test to detect the true effect, that is, as beta decreases, alpha increases, and power increases, or as beta increases, alpha decreases, and power decreases [17].

3.2.3 Smallest Meaningful Difference (δ)

This is sometimes called effect size or treatment difference which is a measure of the effectiveness of treatment between two comparative groups. However, effect size is always context dependent and should be determined based on scientific knowledge. The null hypothesis is rejected with

high probability when the test is significant [5]. The conventional or standardized measures of effect size δ used in this analysis are categorized as small = 0.2 , medium = 0.5 and large = 0.8 by Cohen (1988) [3]. It is recommended not to set the effect size too high, as this decreases the sample size estimate and so increases the probability of a Type II error [4]. And for a small effect size, the sample size needed to detect this would be larger than that for a moderate or large effect size. The combination of α, β, and a standardized δ completely determines the sample size for any study design. Thus, choosing a small, medium, or large standardized effect size is just a fancy way of seeking for a large, medium, or small sample size, respectively [11].

3.2.4 Measurement Variation (σ^2)

This can be reasonably approximated or derived from pilot studies or previous similar studies, otherwise, a plausible guess would be necessary. The estimates used in this study for random intercept variance σ_b^2 is 55 and random slope variance σ_s^2 is 24, which were used by Donohue, et al. (2016) when computing sample size for the Alzheimer's disease trial [6] and the residual variance σ_e^2 is estimated using each within-subject correlation coefficient. For example, in the random intercept model, the total variance $\text{var}(Y_{ij}) = \text{var}(\epsilon_{ij}) = \sigma^2 = \sigma_b^2 + \sigma_e^2$, and for a given value of the intraclass correlation (ICC), the residual variance $\sigma_e^2 = \sigma_b^2(1 - \text{ICC})/\text{ICC}$, while for the random intercept and slope model, total $\text{var}(Y_{ij}) = \text{var}(\epsilon_{ij}) = \sigma^2 = \sigma_b^2 + t_{ij}^2\sigma_s^2 + 2t_{ij}\text{cov}(b_{0i}, b_{1i}) + \sigma_e^2$ [6], where t_{ij} is the time points (t = 0, 2, 5 and 8) and σ^2 measures the unexplained variability in the response (Y_{ij}). If the variability is small, it will lead to a greater power than if it's large [17], hence, small variability will result in fewer sample sizes to achieve the same power as a large variability.

3.2.5 Number of Repeated Observations per Subject (n)

This may be constrained by practical considerations, or perhaps balanced against the sample size [5]. The investigator could be free to determine the value of n depending on the cost available while considering ethical research issues. Repeated measures data has two dimensions of sample sizes, namely: the number of different subjects m and the number of repeated measurements n from each subject [14], and the relationship between these two entities is worth consideration. However, increasing m by one, means increasing the number of measurements by n, because the new subject gets n repeated measurements as the others [14]. Also, increasing the number of repeated measurements by one, means to increase the number of observations by m, since each subject increases one repeated measurement.

3.2.6 Correlation among Repeated Observations (ρ)

Similar to measurement variation, the pattern of correlation can be estimated from pilot studies or previous studies, otherwise, can reasonably be hypothesized. The correlation values used in this study are 0.2, 0.3, 0.5 and 0.8, as previously used by Diggle, et al. (2002) and Liu and Liang (1997) in their respective studies [5, 13]. The correlation between measurements on the same subject for the random intercept model is the intraclass correlation (ICC) computed using $\sigma_b^2/(\sigma_b^2 + \sigma_e^2)$ [12], while the correlation between random intercept (b_0) and random slope (b_1) for the random intercept and slope model is $\rho = \text{cov}(b_0,b_1)/(\sigma_b * \sigma_s)$ [1, 6].

3.3 Theoretical Approach: Random Intercept Models

A marginal model with an exchangeable correlation structure is equivalent to a random effects model which includes a random intercept for each cluster of correlated observations [6]. However, in the random intercept model, ICC is the correlation between repeated measurements on the same subject, which is equivalent to the marginal model with exchangeable correlation. Consider model (2) without b_{1i} in both treatment groups, assuming an exchangeable correlation, thus, $\text{var}(Y_{ij}) = \text{var}(\epsilon_{ij}) = \sigma^2 = \sigma_b^2 + \sigma_e^2$ and $\text{corr}(Y_{ij},Y_{ik}) = \text{E}[(b_{0i} + e_{ij})(b_{0i} + e_{ij})]/\sigma^2 = \sigma_b^2/\sigma^2$.

3.3.1 Longpower Function in R

This function computes sample size for linear mixed-effects models based on the formula in [5, 13], which are expressed in terms of marginal model or generalized estimating equations (GEE) parameters. The function translate pilot mixed-effect model parameters like random intercept, and/or slope, fixed effects, into marginal model parameters so that either formula is applicable to calculate sample size for two-sample longitudinal designs [7]. Besides the usual quantities needed for sample size calculation, a previously fitted model returned by "lme or lmer" function can also be included in the function when determining the sample size required. Moreover, we can also specify the method to use such as, "diggle" or "liuliang" when estimating the sample size.

3.3.2 Diggle et al. (2002) Method

The method is used to compute the sample size for difference in slopes between two groups for the random intercept model [7]. Assuming an exchangeable correlation structure (equal correlation for repeated measurements on the same subject) [6], $\text{var}(Y_{ij}) = \text{var}(\epsilon_{ij}) = \sigma^2$ and $\text{corr}(Y_{ij}, Y_{ik}) = \text{corr}(\epsilon_{ij}, \epsilon_{ik}) = \rho$ for all subjects where $j \neq k$. The number of subjects required

in each treatment group to obtain a specific nominal power is determined using the following formula:

$$m = \frac{2(z_{1-\frac{\alpha}{2}} + z_{1-\beta})^2(\sigma_b^2 + \sigma_e^2)^2(1 - \sigma_b^2/\sigma^2)}{ns_x^2\delta^2} \tag{3}$$

where α and β are the Type I and Type II error rates respectively, δ is the treatment difference, σ_b^2 is variance of the random intercept, σ_e^2 is variance of the random error, $\sigma^2 = \sigma_b^2 + \sigma_e^2$ is the total residual variance, $s_x^2 = \sum_j (t_j - \bar{x})^2/\text{n}$ is variance of the covariate of interest and n is the number of repeated measurements per subject [5,6]. This method is available in the longpower function in the R software.

3.3.3 Liu and Liang (1997) Method

This method is also used to perform sample size calculation for random intercept model [7]. Considering a special case and an exchangeable correlation structure [6, 13], let π_1 denote the proportion in the experimental group and $\pi_0 = 1$ - π_1, the number of subjects required in each treatment group to obtain a certain nominal power is determined as follows:

$$m = \frac{2(z_{1-\frac{\alpha}{2}} + z_{1-\beta})^2\sigma^2(1 + (n-1)\rho)}{n\pi_0\pi_1\delta^2} \tag{4}$$

with notations similar to formula (3), ρ is the correlation between the repeated measurements. In comparison with the sample size formula for independent observations, the formula in (4) is inflated by a factor of (1 + (n - 1)ρ), which is commonly known as the design effect [13], and this method is also available in the longpower function.

3.4 Theoretical Approach: Random Intercept and Slope Models

3.4.1 Longpower Function in R

The longpower function can also be used to calculate sample size for random intercept and slope model. In the 'lmmpower' function, it is important to specify the appropriate method to make sure that sample size is computed for the random intercept and slope model.

3.4.2 Liu and Liang (1997) Method

The method is also used to compute sample size for the random intercept and slope model. For the special case, the formula is similar to the one in (4) but the variance and correlation are expressed as a function of time as discussed previously.

3.5 Simulation Approach

Although there is no real dataset involved in this study, but the dataset used was obtained through simulation, which contains the covariates specified in model (1), namely: treatment (0 = control and 1 = experiment), time = 0, 2, 5, 8 used by Donohue et al. (2016) [6], and additional covariates such as, gender (0 = female and 1 = male) and age. The theoretical approach (formulae) discussed in the previous section may underestimate or overestimate the power or sample size, because in most cases, a closed-form expression for power is rarely available. However, the adequacy of the formulae would be assessed through simulations by determining the closeness between the empirical power and the nominal power [13]. The power and sample size calculations for both the random intercept models, and the random intercept and slope models will be determined using 1000 simulations due to computational involvement.

The mean response (Y_{ij}) in model (1) with additional variables gender and age is calculated for both random intercept models and random intercept and slope models. Initially, values for the fixed effects parameters were arbitrarily selected except for the parameter of interest β_3, which is equal to the value of the effect size. The chosen values for the parameters are $\beta_0 = 5$, $\beta_1 = 2$, $\beta_2 = 9$, $\beta_3 = \delta$, $\beta_4 = 7$ and $\beta_5 = 3$, where $\delta = 0.2$, 0.5, 0.8, and 1.0. Furthermore, the random intercepts were simulated from a normal distribution i.e. rnorm(2*m,mean=0,sd=σ_b) for the random intercept models, while the random effects (intercepts and slopes) were simulated from a multivariate normal distribution i.e. mvrnorm(n=2*m, mu=c(0,0), Sigma=matrix(c(σ_b^2, cov(b_0,b_1), cov(b_0,b_1), σ_s^2),2,2)) for the random intercept and slope models, and the random measurement errors also simulated from a normal distribution i.e. rnorm(2*m*n,mean=0,sd=σ_e) for both models, where m is the sample size per group and n is number of repeated measures.

Finally, all these simulated random values together with the fixed effects were included in the specified model to compute the mean response. For example, in the random intercept and slope models, when $\delta = 0.2$, then $y_{ij} = 5 + 2 * treatment + 9 * time + 0.2 * treatment * time + 7 * gender + 3 * age + b_0 + b_1 * time + e$. Hence, we calculate the mean for each value of the effect size, and fit the model using lme function with ML estimation method to test the significance of β_3 per simulations in order to estimate powers. The estimated powers are calculated by extracting and averaging all the significant p-values of β_3 per 1000 simulations.

4 Results

4.1 Theoretical Approach: Random Intercept Models

The correlation in this model is the intraclass correlation (ICC), which is computed using $\sigma_b^2/(\sigma_b^2 + \sigma_e^2)$. Assuming ICC = 0.2, 0.3, 0.5 and 0.8, and random intercept variance $(\sigma_b^2) = 55$, the estimates for the random error variance (σ_e^2) were calculated using the formula mentioned above, thus, $\sigma_e^2 = 14$, 55, 128, and 220. Total variance $(\sigma^2) = \sigma_b^2 + \sigma_e^2 = 69$, 110, 183, and 275.

4.1.1 Longpower Function in R

Tables (1 - 4) presents the sample size required in each treatment group for detecting the various selected effect sizes δ with a power of 80% at 0.05 significance level. It is observed that, as δ increases, the number of subjects required decreases. Also for each value of σ^2, the sample size decreases as ICC increases. According to Naiji et al. [16], the decrease in sample size could be because repeated measurements are less similar to each other and their respective error variances decreases as well. Conversely, for each value of ICC, sample size increases as σ^2 increases. Thus, smaller/larger variability produces smaller/larger sample size estimates [17]. Moreover, value of σ_b^2 changes with ICC and σ^2, and sample size estimation is driven by σ_e^2.

Table 1: Longpower function for random intercept model with $\delta = 0.2$

ICC	$\sigma^2 = 69$	$\sigma^2 = 110$	$\sigma^2 = 183$	$\sigma^2 = 275$
0.2	465	741	1232	1851
0.3	407	648	1078	1620
0.5	291	463	770	1157
0.8	117	186	308	463

Table 2: Longpower function for random intercept model with $\delta = 0.5$

ICC	$\sigma^2 = 69$	$\sigma^2 = 110$	$\sigma^2 = 183$	$\sigma^2 = 275$
0.2	75	119	198	297
0.3	66	104	173	260
0.5	47	75	124	186
0.8	19	30	50	75

Table 3: Longpower function for random intercept model with $\delta = 0.8$

ICC	$\sigma^2 = 69$	$\sigma^2 = 110$	$\sigma^2 = 183$	$\sigma^2 = 275$
0.2	30	47	77	116
0.3	26	41	68	102
0.5	19	29	49	73
0.8	8	12	20	29

Table 4: Longpower function for random intercept model with $\delta = 1.0$

ICC	$\sigma^2 = 69$	$\sigma^2 = 110$	$\sigma^2 = 183$	$\sigma^2 = 275$
0.2	19	30	50	75
0.3	17	26	44	65
0.5	12	19	31	47
0.8	5	8	13	19

4.1.2 Diggle et al. (2002) Method

Table 5 gives the number of subjects required in each treatment group for some selected values of ICC and total variance σ^2 to detect the various smallest meaningful difference δ with a power of 80% at 0.05 significance level. The sample size estimates are similar to the ones obtained using the Longpower function. Also, the number of subjects required increases with increasing variance and decreasing correlation, while for the smallest variance and highest correlation among repeated measurements, the least sample size is adequate to detect δ.

Table 5: Diggle et al (2002) method for random intercept models

δ	ICC	$\sigma^2 = 69$	$\sigma^2 = 110$	$\sigma^2 = 183$	$\sigma^2 = 275$
0.2	0.2	465	741	1232	1851
	0.3	407	648	1078	1620
	0.5	291	463	770	1157
	0.8	117	186	308	463
0.5	0.2	75	119	198	297
	0.3	66	104	173	260
	0.5	47	75	124	186
	0.8	19	30	50	75
0.8	0.2	30	47	77	116
	0.3	26	41	68	102
	0.5	19	29	49	73
	0.8	8	12	20	29
1.0	0.2	19	30	50	75
	0.3	17	26	44	65
	0.5	12	19	31	47
	0.8	5	8	13	19

4.1.3 Liu and Liang (1997) Method

The results obtained in Table 6 are also similar to those produced using the Longpower function and Diggle et al. (2002) method. Hence, the number of subjects required decreases when the residual variance decreases and the correlation coefficient increases, thus, for the largest residual variance and the smallest correlation among repeated measurements, the largest sample size estimate is much more required for detecting the relevant treatment difference.

Table 6: Liu and Liang (1997) method for random intercept models

δ	ICC	$\sigma^2 = 69$	$\sigma^2 = 110$	$\sigma^2 = 183$	$\sigma^2 = 275$
0.2	0.2	465	741	1232	1851
	0.3	407	648	1078	1620
	0.5	291	463	770	1157
	0.8	117	186	308	463
0.5	0.2	75	119	198	297
	0.3	66	104	173	260
	0.5	47	75	124	186
	0.8	19	30	50	75
0.8	0.2	30	47	77	116
	0.3	26	41	68	102
	0.5	19	29	49	73
	0.8	8	12	20	29
1.0	0.2	19	30	50	75
	0.3	17	26	44	65
	0.5	12	19	31	47
	0.8	5	8	13	19

4.2 Theoretical Approach: Random Intercept and Slope Models

The total residual variance (σ^2) for random intercept and slope models is expressed with time using $\text{var}(Y_{ij}) = \text{var}(\epsilon_{ij}) = \sigma^2 = \sigma_b^2 + t_{ij}^2\sigma_s^2 + 2t_{ij}\text{cov}(\text{b}_{0i},\text{b}_{1i}) + \sigma_e^2$ [1, 6], assuming that the correlation between random intercepts and random slopes ρ is 0.8, random intercept variance σ_b^2 is 55, and random slope variance σ_s^2 is 24, the formula $\rho = \text{cov}(\text{b}_0,\text{b}_1)/(\sigma_b * \sigma_s)$ is relevant for computing the correlation and covariance between random intercepts and random slopes. The estimates for the random error variance σ_e^2 are similar to those computed in the random intercept models.

4.2.1 Longpower Function in R

Table 7 provides the sample size required for detecting the various treatment differences δ with a power of 0.80. The required sample size has a positive or negative relationship with the residual variability, that is, increasing/decreasing the residual variance will increase/decrease

the required sample size to achieve the specified nominal power. Hence, more subjects are needed in this model compared to the sample size estimates for the random intercept models, because the random intercept model does not capture variations in the rate of change from subject to subject [6]. As a result, the random slope term can be used to model the rate of improvement or decline within the treatment group, regardless of treatment.

Table 7: Longpower function for random intercept and slope models

δ	$\sigma_e^2 = 14$	$\sigma_e^2 = 55$	$\sigma_e^2 = 128$	$\sigma_e^2 = 220$
0.2	7537	7882	8496	9270
0.5	1206	1262	1360	1484
0.8	472	493	531	580
1.0	302	316	340	371

4.2.2 Liu and Liang (1997) Method

Table 8 results are also similar to those produced using the Longpower function. As illustrated in the previous methods, the relationship between sample size and residual variance is positive, which usually increases with increasing σ^2. Also, it has an inverse relationship with ρ, thereby decreases with increasing correlation of repeated measurements. Intuitively, the sample size increases in the correlation in the first case but decreases in the second case. The reason being the parameter of interest β_3, is the rate of the change in the response variable whose variance is increasing in the correlation in the first case [5]. And in the second case, β_3 is the expected average of the responses for individuals in a group and the variance of the corresponding estimate is decreasing in the correlation.

Table 8: Liu and Liang (1997) method for random intercept and slope models

δ	$\sigma_e^2 = 14$	$\sigma_e^2 = 55$	$\sigma_e^2 = 128$	$\sigma_e^2 = 220$
0.2	7537	7882	8496	9270
0.5	1206	1262	1360	1484
0.8	472	493	531	580
1.0	302	316	340	371

4.3 Unbalanced Designs

In all the previous sections, the results for a balanced study design were presented, that is, equal allocation of number of subjects in both treatment groups. In order to assess the performance of sample size estimation method, we used the method proposed by Liu and Liang (1997) and vary the proportion of treatment allocations. Tables (9 and 10) provides total sample size estimates for unbalanced treatment allocations for random intercept models and random intercept and slope models respectively when $\sigma^2 = 1$. The impact of unbalanced allocations in the estimation of sample size is very minimal [13], but unbalanced design is very important in clinical studies in case a certain treatment group is expensive and/or recruitment of subjects is very difficult or limited number of subjects available for that particular treatment group [14]. Another reason is to put more subjects in the treatment arm than in the control arm because the control is already known [17]. It is further observed that, for each δ, increasing the proportion of allocation would increase the required sample size estimates.

Table 9: Unbalanced treatment allocation for random intercept models

δ	ICC	$\pi_1 = 0.6$	$\pi_1 = 0.4$	$\pi_1 = 0.8$	$\pi_1 = 0.2$
0.2	0.2	258	258	387	387
	0.3	306	306	459	459
	0.5	403	403	604	604
	0.8	548	548	822	822
0.5	0.2	42	42	62	62
	0.3	49	49	74	74
	0.5	65	65	97	97
	0.8	88	88	132	132
0.8	0.2	17	17	25	25
	0.3	20	20	29	29
	0.5	26	26	38	38
	0.8	35	35	52	52
1.0	0.2	11	11	16	16
	0.3	13	13	19	19
	0.5	17	17	25	25
	0.8	22	22	33	33

Table 10: Unbalanced treatment allocation for random intercept and slope models

δ	ρ	$\pi_1 = 0.6$	$\pi_1 = 0.4$	$\pi_1 = 0.8$	$\pi_1 = 0.2$
0.2	0.2	653	653	980	980
	0.3	775	775	1163	1163
	0.5	1020	1020	1531	1531
	0.8	1388	1388	2082	2082
0.5	0.2	104	104	156	156
	0.3	124	124	186	186
	0.5	163	163	245	245
	0.8	222	222	333	333
0.8	0.2	40	40	61	61
	0.3	48	48	72	72
	0.5	63	63	95	95
	0.8	86	86	130	130
1.0	0.2	26	26	39	39
	0.3	31	31	46	46
	0.5	40	40	61	61
	0.8	55	55	83	83

4.4 Simulation Approach

Tables (11 and 12) provides the sample sizes m and estimated powers for random intercept models, and random intercept and slope models for $\sigma^2 = 69$ and $\sigma_e^2 = 14$ respectively through 1000 simulations. The estimated powers are much closer to the nominal power of 0.80 in both models. Furthermore, we vary the number of repeated measurements n per subject to see how the power and sample size are affected. Thus, the sample size decreases as n increases.

Table 11: Power analysis for random intercept models through simulation

		$n = 4$		$n = 6$		$n = 8$	
ICC	δ	m	Power	m	Power	m	Power
0.2	0.2	465	0.807	157	0.786	63	0.808
	0.5	75	0.786	26	0.799	10	0.769
	0.8	30	0.812	10	0.795	4	0.751
	1.0	19	0.787	7	0.810	3	0.863
0.3	0.2	407	0.807	138	0.806	55	0.822
	0.5	66	0.792	22	0.787	9	0.811
	0.8	26	0.800	9	0.790	4	0.802
	1.0	17	0.803	6	0.810	3	0.898
0.5	0.2	291	0.794	98	0.811	39	0.780
	0.5	47	0.794	16	0.779	7	0.817
	0.8	19	0.797	7	0.823	3	0.863
	1.0	12	0.796	4	0.755	2	0.850
0.8	0.2	117	0.801	40	0.807	16	0.789
	0.5	19	0.783	7	0.811	3	0.857
	0.8	8	0.815	3	0.835	2	0.949
	1.0	5	0.770	2	0.833	2	0.989

Table 12: Power analysis for random intercept and slope models through simulation

	$n = 4$		$n = 6$		$n = 8$	
δ	m	Power	m	Power	m	Power
0.2	7537	0.790	7459	0.796	7435	0.830
0.5	1206	0.800	1194	0.791	1190	0.802
0.8	472	0.797	467	0.807	465	0.810
1.0	302	0.790	299	0.786	298	0.805

5 Discussion and Conclusion

The simulation study examined the required sample sizes for detecting the interaction between treatment and time fixed effects in a linear mixed-effects model with both random intercepts and random slopes. The results of this study were obtained through the theoretical and simulation approaches, of which the former involved using the methods or formulae proposed by Diggle et al. (2002) and Liu and Liang (1997), and its implementation in Longpower function in R [5,13]. The simulation approach is specifically carried out in order to determine the adequacy of the theoretical approach in power and sample size estimation.

In the theoretical approach, the sample size estimates for the linear mixed-effects models with only random intercepts were less than those obtained in the models with both random intercepts and random slopes for the same conventional effect size estimates. This difference could be influenced by the different methods used in estimating the correlation and variability. In the random intercept models, the correlation is the intraclass correlation, while in the random intercept and slope models, correlation and variability are expressed as a function of time. Thus, the estimated sample sizes are similar in all the different methods used for each class of model.

Moreover, in the simulation approach, the estimated powers were at least closer to the nominal power in both models, as a result, both the theoretical and simulation approaches yield similar results. Furthermore, varying the number of repeated measurements per subject have great impact on sample size estimates and power analysis in the random intercept models and minimal impact on power and sample size in the models with both random intercepts and random slopes. Liu and Wu [14], investigated the relative impact of M and n on power when $0 < \rho < 1$. Thus, increasing n by the amount equivalent to $n(1 - \rho + n\rho)/[M - (M + n)\rho]$ is similar to increasing M by 1, and such increment depends on M, n and ρ in estimating or analyzing power. The choice of increasing the number of individual subjects or the number of repeated measurements per subject are driven by several factors depending on the nature of the clinical studies, such as cost and availability of subjects to participate in the study.

Finally, we explored sample size estimation for unbalanced treatment allocations using both models, and the results shows that unbalanced allocations have minimal impact on sample size estimation when $\pi_1 > 0.5$ or less, which is similar to the findings obtained in study [13]. The results suggest that the sample size estimates may not necessarily be overestimated or

underestimated when the proportion of allocation π_1 is greater or less than 0.5. However, the importance of unbalanced designs in clinical trials cannot be overemphasized, especially when recruitment of study participants in a particular treatment group is very difficult or expensive, and also when testing a new drug against a standard drug, whose effect is already known.

In conclusion, power and sample size calculations through simulation is greatly important since most specialized software packages may have some hidden features that are not explicitly clear to the users, especially on the techniques involved in computation and analysis. Simulation is the solution and is possible in any setting, especially useful in complex situations. But it is never exact, only approximation, and can be quite computational intensive. Although, several commercial softwares have programmed procedures to handle complicated or complex cases, but most of them are very expensive to purchase and maintain. There are free software packages available which can be used with great care and attention.

References

[1] Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

[2] Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365-376.

[3] Cohen, J. (1988). Statistical power analysis for the behavioral sciences . Hilsdale. *NJ: Lawrence Earlbaum Associates*, 2.

[4] Cunningham, J. B., & McCrum-Gardner, E. (2007). Power, effect and sample size using GPower: practical issues for researchers and members of research ethics committees. *Evidence-Based Midwifery*, 5(4), 132-137.

[5] Diggle, PJ., Heagerty PJ., Liang K., & Zeger SL. (2002). *Analysis of longitudinal data*. Second Edition. Oxford Statistical Science Series.

[6] Donohue, M. C., Edland, S. D., & Gamst, A. C. (2016). Power for linear models of longitudinal data with applications to Alzheimer's Disease Phase II study design.

[7] Donohue, M. C., Gamst, A. C., & Edland, S. D. (2016). longpower: Power and sample size calculations for longitudinal data, R package version 1.0-16.

[8] Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, 41(4), 1149-1160.

[9] Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175-191.

[10] Gałecki, A., & Burzykowski, T. (2013). Linear mixed-effects model. In *Linear Mixed-Effects Models Using R* (pp. 245-273). Springer New York.

[11] Lenth, R. V. (2001). Some practical guidelines for effective sample size determination. *The American Statistician*, 55(3), 187-193.

[12] Leon, A. C., & Heo, M. (2009). Sample sizes required to detect interactions between two binary fixed-effects in a mixed-effects linear regression model. *Computational statistics & data analysis*, 53(3), 603-608.

[13] Liu, G., & Liang, K. Y. (1997). Sample size calculations for studies with correlated observations. *Biometrics*, 937-947.

[14] Liu, H., & Wu, T. (2005). Sample size calculation and power analysis of time-averaged difference. *Journal of Modern Applied Statistical Methods*, 4(2), 9.

[15] Mayr, S., Erdfelder, E., Buchner, A., & Faul, F. (2007). A short tutorial of GPower. *Tutorials in Quantitative Methods for Psychology*, 3(2), 51-59.

[16] Naiji, L. U., Yu, H. A. N., Tian, C. H. E. N., Douglas, D. G., Yinglin, X. I. A., & Julia, Y. (2013). Power analysis for cross-sectional and longitudinal study designs. *Shanghai archives of psychiatry*, 25(4), 259-262.

[17] Osmena, P. (2010). Statistical power analysis using SAS and R. *Statistics*, 4.

[18] Prajapati, B., Dunne, M., & Armstrong, R. (2010). Sample size estimation and statistical power analyses. *Optometry today*, 16(07), 10-18.

[19] Self, S. G., & Mauritsen, R. H. (1988). Power/sample size calculations for generalized linear models. *Biometrics*, 79-86.

[20] Teare, M. D., Dimairo, M., Shephard, N., Hayman, A., Whitehead, A., & Walters, S. J. (2014). Sample size requirements to estimate key design parameters from external pilot randomised controlled trials: a simulation study. *Trials*, 15(1), 264.

[21] Verbeke, G. & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag, Inc.

Appendix

GPower 3

GPower version three is an excellent free statistical software program that permits high-precision power and sample size analyses [4]. It is capable of computing five (5) different types of power analyses: a priori, post-hoc, compromise, criterion and sensitivity power analysis. An a priori power analysis is the most relevant to sample size estimation, given 1 - β, α and δ [18]. In contrast, a post-hoc analysis is less ideal than a priori analysis and is typically useful for assessing statistical power after the main study while controlling for only α. Its main provision is for a critical evaluation of β, associated with a false decision in favour of the H_0 [8, 9, 15]. Since we have repeated measures within each subject, the repeated measures ANOVA test is ideal to estimate sample size taking into account the within-subject (dependent) factors and between-subject (independent) factors. However, the sample size is computed for all factors and interaction and the largest is selected for the study [18]. But, this may not always be the best option as some studies could have complex interactions uninterested to the investigators, hence, they can specify which factor or interaction is necessary for estimating the sample size.

Table A.1: Repeated measures within factors

		$n=4$		$n=6$		$n=8$	
ρ	δ	m	Power	m	Power	m	Power
0.3	0.2	50	0.8087	40	0.8171	34	0.8225
	0.5	10	0.8442	8	0.8473	8	0.9192
	0.8	6	0.9286	4	0.8044	4	0.9004
	1.0	4	0.8278	4	0.9508	4	0.9871
0.5	0.2	36	0.8073	28	0.8020	24	0.8107
	0.5	8	0.8660	6	0.8307	6	0.9099
	0.8	6	0.9840	4	0.9250	4	0.9760
	1.0	4	0.9328	4	0.9914	4	0.9990
0.8	0.2	16	0.8285	12	0.8029	12	0.8822
	0.5	6	0.9819	4	0.9185	4	0.9728
	0.8	4	0.9924	4	0.9999	4	0.9999
	1.0	4	0.9998	4	0.9999	4	0.9999

Table A.2: Repeated measures between factors

		$n=4$		$n=6$		$n=8$	
ρ	δ	m	Power	m	Power	m	Power
0.3	0.2	96	0.8035	84	0.8013	80	0.8101
	0.5	18	0.8235	16	0.8213	16	0.8476
	0.8	10	0.8936	8	0.8299	8	0.8550
	1.0	8	0.9243	6	0.8065	6	0.8322
0.5	0.2	126	0.8045	118	0.8054	114	0.8059
	0.5	22	0.8055	22	0.8316	20	0.8049
	0.8	12	0.8838	10	0.8258	10	0.8389
	1.0	8	0.8442	8	0.8672	8	0.8786
0.8	0.2	170	0.8030	166	0.8013	164	0.8004
	0.5	30	0.8179	30	0.8253	28	0.8006
	0.8	14	0.8458	14	0.8527	14	0.8561
	1.0	10	0.8506	10	0.8573	8	0.8606

Table A.3: Repeated measures within-between interaction

		$n = 4$		$n = 6$		$n = 8$	
ρ	δ	m	Power	m	Power	m	Power
0.3	0.2	50	0.8087	40	0.8171	34	0.8225
	0.5	10	0.8442	8	0.8473	8	0.9192
	0.8	6	0.9286	4	0.8044	4	0.9004
	1.0	4	0.8278	4	0.9508	4	0.9871
0.5	0.2	36	0.8073	28	0.8020	24	0.8107
	0.5	8	0.8660	6	0.8307	6	0.9099
	0.8	6	0.9840	4	0.9250	4	0.9760
	1.0	4	0.9328	4	0.9914	4	0.9990
0.8	0.2	16	0.8285	12	0.8029	12	0.8822
	0.5	6	0.9819	4	0.9185	4	0.9728
	0.8	4	0.9924	4	0.9999	4	0.9999
	1.0	4	0.9998	4	0.9999	4	0.9999

Tables (A.1 - A.3) presents the total sample size and the actual power estimates for both groups using ANOVA repeated measures within-factors, between-factors and within-between factor interactions respectively. In all the tables, the sample size estimates decreases with increasing effect size for each specific correlation coefficient. The actual power estimates are also observed to be very similar to the nominal power or at least slightly larger. The sample size estimates in Table A.2 increases with increase in correlation coefficient and are much larger compared to other two tables. Hence, the investigator can make decision on which interesting factors or interactions are useful for estimating sample size and whether to increase the number of individual subjects or increase the number of repeated measures within each subject to achieve a certain power, depending on the resources or time available for conducting the study.

Table A.4 gives the sample size estimates for unbalanced designs using the formula proposed by Liu and Wu [14]. The proportion of sample size in the treatment group is 0.6 and in the control group is 0.4, and as a result, the treatment group has more subjects than the control group. This method is ideal in situations we are interested in testing the effect of a new drug against a standard drug having already known the effect of the standard drug. Because sample size and statistical power are positively related, therefore, more subjects are required in the treatment arm than the control arm to achieve the desired statistical power.

Table A.4: Sample size estimates for unbalanced treatment allocations

δ	ρ	$\pi_1 = 0.6$	$\pi_2 = 0.4$
0.2	0.2	1176	784
	0.3	1396	931
	0.5	1387	1225
	0.8	2499	1666
0.5	0.2	188	125
	0.3	223	148
	0.5	294	196
	0.8	399	267
0.8	0.2	73	49
	0.3	87	58
	0.5	114	77
	0.8	156	104
1.0	0.2	47	31
	0.3	55	38
	0.5	73	49
	0.8	99	67

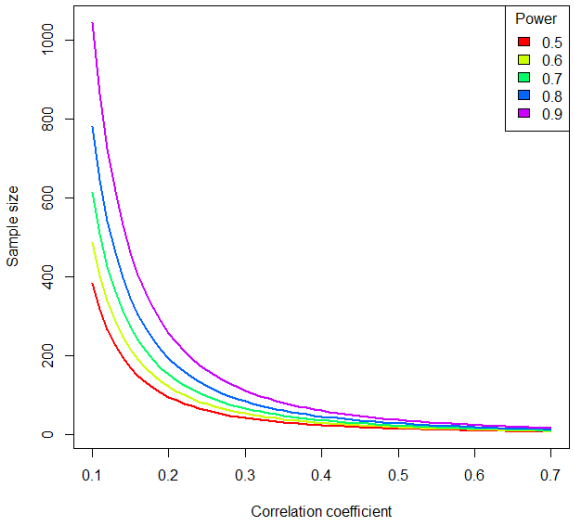

Figure A.1: Sample size and correlation

R Program Codes

```
## Required packages
library(longpower)
library(lme4)
library(Matrix)
library(nlme)
library(MASS)

#-----------------------------------

## Time points
t = c(0, 2, 5, 8)
n = length(t)
## Treatment difference
effectsize = c(0.2, 0.5, 0.8, 1.0)

#-----------------------------------

## THEORETICAL APPROACH
## RANDOM INTERCEPT MODELS:

## Random intercept variance
sigma2.b = c(55, 55, 55, 55)
## Random slope variance
sigma2.s = c(0, 0, 0, 0)
## Random error variance
sigma2.e = c(14, 55, 128, 220)
## Total residual variance
sigma2 = sigma2.b + sigma2.e
## Within-subject Correlation
rho = sigma2.b/(sigma2.b + sigma2.e)
rho = c(0.2, 0.3, 0.5, 0.8)
```

```
#-------------------------------------

## LONGPOWER
## Random intercept models
# NB: The value of var.b changes with icc and var
# Thus, sample size estimation depends on var.e
icc = 0.2 # (0.2,0.3,0.5,0.8)
var = 69 # (69,110,183,275)
var.b = icc*var
var.e = var - var.b

lmmpower(delta = 0.2, t = t, sig2.e = var.e, sig2.i = var.b, sig2.s = 0,
cov.s.i = icc*sqrt(0)*sqrt(55), alternative = "one.sided", power = 0.80)

## DIGGLE ET AL. (2002)
## Random intercept models
diggle.int <- outer(rho, sigma2,
                    Vectorize(function(rho, sigma2){
                      ceiling(diggle.linear.power(
                        delta = 0.2,
                        t = t,
                        sigma2 = sigma2,
                        R = rho, sig.level = 0.05,
                        alternative = "one.sided",
                        power = 0.80)$n)}))
colnames(diggle.int) <- paste("sigma2 =", sigma2)
rownames(diggle.int) <- paste("rho =", rho)

## LIU AND LANG (1997)
## Random intercept models
u = list(u1 = t, u2 = rep(0,n))
v = list(v1 = cbind(1,1,t),
         v2 = cbind(1,0,t))
liuliang.int = outer(rho, sigma2,
```

```
                    Vectorize(function(rho, sigma2){
                      ceiling(liu.liang.linear.power(
                        delta = 0.2, u = u, v = v,
                        sigma2 = sigma2, sig.level = 0.05,
                        R = rho, alternative = "one.sided",
                        power = 0.80)$N/2)}))
colnames(liuliang.int) = paste("sigma2 =", sigma2)
rownames(liuliang.int) = paste("rho =", rho)

#------------------------------------------------------------
#------------------------------------------------------------

## RANDOM INTERCEPT AND SLOPE MODELS:

## Random intercept variance
var.intercept = 55
## Random slope variance
var.slope = 24
## Random error variance
var.error = 14 #c(14, 55, 128, 220)
## Treatment difference
effect.size = 0.2
## Correlation of random effects
rho.slope.intercept = 0.8
## Covariance of intercept and slope
cova.slope.intercept = rho.slope.intercept*sqrt(var.intercept)*sqrt(var.slope)

#------------------------------------------------------------

## LONGPOWER
## Random intercept and slope models
lmmpower(delta = effect.size, t = t, sig2.e = var.error, sig2.i = var.intercept,
sig2.s=var.slope, cov.s.i=cova.slope.intercept, alternative = "one.sided",power=0.80)
```

```
## LIU AND LANG (1997)
## Random intercept and slope models
cov.s.i <- rho.slope.intercept*sqrt(var.intercept)*sqrt(var.slope)
cov.t <- function(t1, t2, var.intercept, var.slope, cov.s.i){
  var.intercept + t1*t2*var.slope + (t1+t2)*cov.s.i
}

R = outer(t, t, function(x,y){cov.t(x,y, var.intercept, var.slope, cov.s.i)})
R = R + diag(var.error, n, n)
u = list(u1 = t, u2 = rep(0,n))
v = list(v1 = cbind(1,1,t), v2 = cbind(1,0,t))
liu.liang.linear.power(d = effect.size, u=u, v=v, R=R, sig.level = 0.05,
power = 0.80, alternative = "one.sided")

#-----------------------------------------------------------------
#-----------------------------------------------------------------

## SAMPLE SIZE FOR UNBALANCED DESIGNS:
## Liu and Liang (1997) method
## delta = c(0.2, 0.5, 0.8, 1.0)
## rho = c(0.2, 0.3, 0.5, 0.8)
unbalDesign <- data.frame(cbind(
  rho = rep(c(0.2,0.3,0.5,0.8), 2),
  pi1 = c(rep(0.6, 4), rep(0.4, 4))))
m <- c()
for(i in 1:nrow(unbalDesign)){
  R <- matrix(unbalDesign$rho[i], nrow = 4, ncol = 4)
  diag(R) <- 1
  m <- c(m, ceiling(liu.liang.linear.power(
    delta = 0.2,
    u = list(u1 = rep(1, 4), # treatment
             u2 = rep(0, 4)), # control
    v = list(v1 = rep(1, 4), v2 = rep(1, 4)), # intercept
```

```
    sigma2 = 1,
    Pi = c(unbalDesign$pi1[i], 1-unbalDesign$pi1[i]),
    R = R, alternative = "one.sided",
    power = 0.80)$N))
}
cbind(unbalDesign, m)

#---------------------------------------

## Unbalanced treatment allocations
## Honghu Liu and Tongtong Wu (2005) formula
n = 4
var = 1
pi = 0.6
d = 0.2
rho = 0.2
numerator = (1.96 + 0.84)^2*var*(1 + (n - 1)*rho)
denominator = n*(1 - pi - pi^2)*d^2
M = numerator/denominator
# sample size in treatment group
m1 = pi*M
# sample size in control group
m2 = (1 - pi)*M

#---------------------------------------------------
#---------------------------------------------------

## SIMULATION APPROACH (1):
## Random intercept models

## number of subjects per group
m = 465
## time points (n=4, n=6 and n=8)
tp = c(0, 2, 5, 8)
```

```
tp = c(0, 2, 5, 8, 10, 12)
tp = c(0, 2, 5, 8, 10, 12, 15, 18)
## number of repeated measures per subject
n = length(tp)
## treatment difference
effectSize = 0.2
## correlation
rho = 0.2
## total variance
sigma2 = 69
## random intercept variance
sigma2.int = sigma2*rho
## random error variance
sigma2.esp = sigma2 - sigma2.int
## number of simulations
Nsimulations = 1000
## Predictor variables:
subject = rep(1:(2*m), each = n)
treatment = rep(c(0,1), each = n)
time = rep(c(0, 2, 5, 8))
trt_time = treatment*time
gender = rep(c(0,1), each = m*n)
set.seed(1234)
age = sample(seq(15,50, by = 1), size = m*n*2, replace = TRUE)
data.frame(subject = subject, treatment = treatment, time = time, trt_time = trt_time,
                      gender = gender, age = age)
## Response variable:
beta0 = 5; beta1 = 2; beta2 = 9; beta3 = effectSize; beta4 = 7; beta5 = 3
meanResponse = (beta0 + beta1*treatment + beta2*time + beta3*trt_time +
                              beta4*gender + beta5*age)
## View structure of the simulated data
mysimdata = data.frame(subject = subject, treatment = treatment, time = time,
      trt_time = trt_time, gender = gender, age = age, meanResponse = meanResponse)
```

```
set.seed(1234)
pval_ri_model = numeric(Nsimulations)
for(i in 1:Nsimulations){
  random.intercepts = rnorm(2*m, mean=0, sd=sqrt(sigma2.int))
  response1 = meanResponse + random.intercepts[mysimdata$subject] +
                       rnorm(nrow(mysimdata), mean=0, sd=sqrt(sigma2.esp))

  simulatedDataset = data.frame(subject = subject, treatment = treatment, time = time,
             trt_time = trt_time, gender = gender, age = age, response1 = response1)

  myfit_rim <- lme(response1 ~ treatment + time + trt_time + gender + age,
                random = list(~ 1|subject), method =  "ML", data = simulatedDataset)

  pval_ri_model[i] <- summary(myfit_rim)$tTable["trt_time",5]
  print(i)
}

## Estimate power
sum(pval_ri_model < 0.10)/Nsimulations

#-------------------------------------------------------------

## SIMULATION APPROACH (2):
## Random intercept and slope models

m = 7537                            #number of subjects per group
Nsim = 1000                         #number of simulations
effectSize = 0.2                    #treatment difference
rho.slope.intercept = 0.8           #correlation of random intercept and slope
tp = c(0, 2, 5, 8)                  # 4 time points
tp = c(0, 2, 5, 8, 10, 12)          # 6 time points
tp = c(0, 2, 5, 8, 10, 12, 15, 18) # 8 time points
n = length(tp)                      #number of repeated measurements per subject
sigma2.b0 = 55                      #random intercept variance
```

```
sigma2.b1 = 24                      #random slope variance
sigma2.ei = 14                      #random error variance
#covariance of random intercept and slope
cova.b0b1 = rho.slope.intercept*sqrt(sigma2.b0)*sqrt(sigma2.b1)

set.seed(2017)
pvalues = vector()
for(s in 1:Nsim){
  subject = rep(1:(2*m), each = n)
  treatment = rep(c(0,1), each = n)
  time = rep(c(0, 2, 5, 8))
  trt_time = treatment*time
  gender = rep(c(0,1), each = m*n)
  age = sample(seq(15,50, by = 1), size = m*n*2, replace = TRUE)

  beta0 = 5; beta1 = 2; beta2 = 9; beta3 = effectSize; beta4 = 7; beta5 = 3
  meanResponse = (beta0 + beta1*treatment + beta2*time + beta3*trt_time +
                              beta4*gender + beta5*age)

  simdat = data.frame(subject = subject, treatment = treatment, time = time,
     trt_time = trt_time, gender = gender, age = age, meanResponse = meanResponse)

  random.effects = mvrnorm(n=2*m, mu=c(0,0),
                Sigma=matrix(c(sigma2.b0,cova.b0b1,cova.b0b1,sigma2.b1),2,2))
  random.intercepts = rnorm(2*m, mean=0, sd=sqrt(sigma2.b0))
  random.slopes = rnorm(2*m, mean=0, sd=sqrt(sigma2.b1))

  response2 = meanResponse + random.effects[simdat$subject,1] +
                         random.effects[simdat$subject,2]*simdat$time +
                             rnorm(nrow(simdat), mean=0, sd=sqrt(sigma2.ei))

  simDataset = data.frame(subject = subject, treatment = treatment, time = time,
         trt_time = trt_time, gender = gender, age = age, response2 = response2)
```

```
  fit_rism <- lme(response2 ~ treatment + time + trt_time + gender + age,
                random = list(~ time|subject), method = "ML", data = simDataset,
                        control=lmeControl(returnObject=TRUE, opt='optim'))

  pvalues[s] <- coef(summary(fit_rism))[4,5]
  print(s)

}

## Estimate power
sum(pvalues < 0.10)/Nsim
```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Power calculations for linear mixed effects models

Richting: **Master of Statistics-Biostatistics**
Jaar: **2018**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Ndow, Manjally

Datum: **22/01/2018**