



Emerging Methods in Progression Modelling of Alzheimer's Disease

A Comparative Analysis

Jonas Benjamin Rose Aldous
Sandra Helweg Sørensen

Master's Thesis, Mathematics-Economics



AALBORG UNIVERSITY
STUDENT REPORT

Department of Mathematical Sciences

Skjernvej 4A

DK-9220 Aalborg Ø

<http://math.aau.dk>

Title

Emerging Methods in Progression
Modelling of Alzheimer's Disease

Subtitle

A Comparative Analysis

Theme

Alzheimer's Disease

Project Period

September 2023 - June 2024

Project Group

Group 4.111d

Participant(s)

Jonas Benjamin Rose Aldous
Sandra Helweg Sørensen

Supervisor(s)

Rasmus Plenge Waagepetersen

Number of Pages

73

Date of Completion

June 3, 2024

Abstract

This thesis is a study of the progression models for repeated measures (PMRMs) introduced in [Raket, 2022]. Data for the thesis is provided by Novo Nordisk A/S and obtained from the Critical Path for Alzheimer's Disease database. The initial part of the thesis is preliminary theory regarding mixed models as well as a presentation of the PMRMs and the conventionally used constrained longitudinal data analysis (cLDA) model. Furthermore, we present modifications of the PMRMs, including the removal of a correlation structure in the errors, as well as the addition of a random effect. Moreover, a way of implementing the PMRMs in an analysis of heterogeneity of treatment effect between subgroups is explored. We conduct a simulation study, where the performance of the PMRMs is examined in different scenarios, and compared to each other as well as the cLDA model. Here, we find that there are both pros and cons of using the PMRMs, and extensions thereof, compared to the cLDA model. Overall, while the cLDA model offers robust performance and a controlled type I error rate, PMRMs provide better interpretability and higher statistical power in specific scenarios, highlighting their potential applicability in clinical trials. Lastly, we present a way of implementing the PMRMs in health economic modelling, utilising a Markov Model and the assumption of a constant treatment effect over time. Here, a brief overview of how they can be used in a cost-effectiveness analysis is presented. The concluding elements of the thesis discuss the interpretability of the models' estimates and their applicability in health economic modelling and decision-making.

Preface

This master's thesis is written by Jonas Benjamin Rose Aldous and Sandra Helweg Sørensen whilst studying Mathematics-Economics, at the Department of Mathematical Sciences at Aalborg University, Denmark, throughout the last year of the Master's degree programme.


References in the thesis are written in accordance with the American Psychological Associations (APA) reference system, meaning references in the thesis are written [Author(s)'s last name, year, potential pages/chapters]. For website references, the year indicates the year the article was written, unless that information is unavailable in which case the year the website was last visited.

The analysis conducted in this thesis was performed using the programming language R. Therefore, when referring to implementations and packages, these will be respect to R. R is a free software environment for statistical computing and graphics. For additional information about the software see <https://www.r-project.org/>.

The data used within this thesis was provided to us by Novo Nordisk A/S. This data was obtained from the Critical Path for Alzheimer's Disease (CPAD) Database. As such, the investigators within CPAD contributed to the design and implementation of the CPAD database and/or provided data, but did not participate in the analysis of the data or the writing of this thesis. For more information regarding the CPAD database, see <https://c-path.org/programs/cpad/>.

Special thanks to our supervisor(s) for helping us through the thesis, and Novo Nordisk A/S for providing the data set used within this project bringing a practical dimension to the thesis.

Signatures



Jonas Benjamin Rose Aldous
<jaldou19@student.aau.dk>



Sandra Helweg Sørensen
<shsa19@student.aau.dk>

Nomenclature

The nomenclature describes several symbols that will be used within the thesis.

Abbreviations

AD	Alzheimer's Disease
ADAS-Cog	Alzheimer's Disease Assessment Scale – Cognitive Subscale
ADCS-ADL	Alzheimer's Disease Cooperative Study - Activities of Daily Living Inventory
ADCS-iADL	Alzheimer's Disease Cooperative Study - Instrumental Activities of Daily Living
CDR-SB	Clinical Dementia Rating Sum of Boxes
CEA	Cost-Effectiveness Analysis
cLDA	Constrained Longitudinal Data Analysis
CPAD	Critical Path for Alzheimer's Disease
FDA	U.S. Food and Drug Administration
HTA	Health Technology Assessment
iADRS	integrated Alzheimer's Disease Rating Scale
ICER	Incremental Cost-Effectiveness Ratio
LME	Linear Mixed Effects
MCI	Mild Cognitive Impairment
MMSE	Mini-Mental State Examination
NC	No Correlation
PDPMRM	Proportional Reduction in Decline PMRM
PMRM	Progression Model for Repeated Measures
PNLS	Penalised Non-linear Least Squares
PSTPMRM	Proportional Slowing of Disease Progression PMRM
QALY	Quality-Adjusted Life Years
RI	Random Intercept
RS	Random Scaling Factor
TPMRM	Time Based Changes in Disease Progression PMRM

Mathematics

$\det(\cdot)$	Determinant
\sim	Approximately distributed as

$\lambda, \delta, \sigma, \tau$	Variance Parameters
β	Fixed effects
ε	Errors
$\Psi = \tau^2 (\mathbf{\Delta}(\delta)^\top \mathbf{\Delta}(\delta))^{-1}$	Covariance matrix of random effects
$\Sigma = \sigma^2 \mathbf{\Lambda}(\lambda)$	Covariance matrix of errors
M^\top	Matrix transpose
M^{-1}	Matrix inverse
U	Random effects
X	Fixed effects design matrix
Z	Random effects design matrix

Contents

1	Introduction	1
2	Mixed Models	5
2.1	Linear Mixed Models	5
2.1.1	Linear Mixed Model for Repeated Measures	6
2.2	Estimation of Linear Mixed Models	8
2.2.1	Estimation of Fixed Effects	8
2.2.2	Estimation of Covariance Parameters	8
2.2.3	Prediction of Random Effects	10
2.3	Non-Linear Mixed Models	11
2.4	Estimation of Non-linear Mixed Models	12
2.5	Approximation Methods	13
2.5.1	Lindstrom and Bates Algorithm	13
2.5.2	Laplace Approximation	15
2.6	An Extended Non-linear Regression Model	17
2.7	Variance and Correlation Structures of the Errors	18
2.7.1	Variance	18
2.7.2	Correlation	19
3	Modelling the Treatment Effect	21
3.1	Constrained Longitudinal Data Analysis Model	21
3.2	Categorical to Continuous	22
3.3	Progression Models for Repeated Measures	22
3.3.1	Proportional Reduction in Decline	23
3.3.2	Time-based Changes in Disease Progression	24
3.3.3	Proportional Slowing of Disease Progression	25
3.3.4	Subgroup Extension	26
4	Data Analysis	29
4.1	Data Presentation	29
4.1.1	Demographics	29
4.1.2	Treatments	32
4.1.3	Cognitive Assessment Scales	33
4.2	Data Set Selection	33
4.3	Simulation of the Active Arm	38
5	Simulation Study	41
5.1	Performance Analysis	43
5.1.1	Scenarios	43
5.1.2	Correlated Error Terms	46
5.1.3	Random Effects	47
5.2	Calibrated Statistical Power	49
5.3	Time Homogeneity	52

5.4	Subgroup Analysis	53
5.4.1	Estimated Treatment Difference	54
5.4.2	Statistical Power	56
5.5	Summary of Results	57
6	Health Technology Assessment Implementation	59
6.1	Markov Models	60
6.2	Cost-effectiveness Analysis	63
7	Discussion	67
8	Conclusion	71
	Appendices	73
A	Preliminary Theory	75
A.1	Newton-Raphson Algorithm	75
A.2	Natural Cubic Splines	75
B	Additional Evaluations	77
C	Residual Analysis	83
D	Simulation Study Results	87
D.1	Optimisation Algorithm for True Effects	87
D.2	Convergence of Estimates	88
D.3	Convergence of Statistical Power	89
D.4	Parameter Estimates	91
D.5	Statistical Power	100
D.6	Time Homogeneity	101
E	Results from the Subgroup Analysis	103
E.1	Convergence of Estimates	103
E.2	Convergence of Statistical Power	106
E.3	Subgroup Parameter Estimates	109
E.4	Subgroup Statistical Power	130

1 Introduction

Alzheimer's disease, as the leading cause of dementia [Rasmussen and Langerman, 2019], accounting for 60-80% of all cases, has become a topic of great interest in the last couple of years. The National Institute on Aging was, in 2023, supporting 508 active clinical trials on Alzheimer's disease and other dementias, all aiming to significantly improve the treatment for people with Alzheimer's disease [National Institute on Aging, 2023b]. Of these 508 clinical trials, 73 were pharmacological, whereas the remaining consists of non-pharmacological and caregiving trials. It was previously thought that Alzheimer's disease was a part of aging, but this is not the conviction anymore. As of 2019, there were approximately 57 million people living with Alzheimer's disease or other dementias worldwide. The Institute for Health Metrics and Evaluation estimates this number to grow to 153 million people by 2050 [Alzheimer's Disease International, 2022].

Alzheimer's disease and other dementias are very present when looking at annual deaths, as it is the seventh leading cause of death worldwide, where a third of all seniors die with it. Furthermore, in high-income countries, it was the second leading cause of death in 2019, causing 814 thousand deaths worldwide [World Health Organization, 2020]. Naturally, as the number of people with Alzheimer's disease or other dementias is expected to grow, the number of deaths due to Alzheimer's disease will also grow, unless the treatment which they receive also progressively gets better.

Alzheimer's disease is a neurodegenerative disease, which damages and possibly destroys parts of the person's nervous system. It does so progressively, beginning with the prodromal stage, also being known as mild cognitive impairment, which is the stage wherein subjects begin to experience symptoms. At this stage, someone might not be convinced that it is Alzheimer's disease, as the symptoms related hereto often are close to those associated with aging. However, as the disease progresses and possibly leads to dementia, the symptoms get more obvious. A person in the prodromal stage is still able to carry out their daily activities without needing assistance. Whereas when the disease progresses, this individual might, for example, have their driver's license revoked due to the road safety being impacted, not only for the individual but also those around them [National Institute on Aging, 2023a]. This also means that as the disease progresses, the individual will slowly become less able to carry out the same aforementioned daily activities, transferring increasing responsibility onto relatives or the public system. Therefore, the cost of care often increases as the severity of the disease progresses. Two videos which aid to help the public understand this pressure have been created by the Alzheimer's Society and are available on YouTube: <https://www.youtube.com/watch?v=TvKNTIY6uQ8> and <https://www.youtube.com/watch?v=cKSGV-Zuu5M>.

The severity of progressive diseases, such as Alzheimer's disease, can be difficult to measure as many factors might affect them. It has also become a big topic to find biological markers for Alzheimer's disease and other types of dementia, and determine how they can be used to diagnose Alzheimer's disease in the early stages. Examples of hallmarks of Alzheimer's disease are neurofibrillary tangles and amyloid plaques [National Institute on Aging, 2022], however, many more are being investigated to determine their connection to the disease. For example, p-tau217 has not yet been approved by the authorities but has shown to have

performance metrics that rival the gold standard of amyloid plaques [Janelidze et al., 2022]. The neurofibrillary tangles form inside the brain cells due to the hyperphosphorylation of tau protein. Usually, tau protein helps facilitate the neuron's transport system. However, in the case of neurofibrillary tangles, the tau proteins stick together. Once the neurofibrillary tangles become large enough, they block the neuron's transport system, which harms the synaptic communication between neurons. Amyloid plaques are, as the name suggests, plaques of protein fragments in the brain. Normally, these protein fragments are broken down, but in the case of amyloid plaques, the amyloid (beta) protein accumulates to form hard plaques. Two forms of this, beta-amyloid 40 and beta-amyloid 42, have been the main focus of many Alzheimer's disease trials as beta-amyloid 42 is thought to be toxic [National Institute on Aging, 2017]. However, there is no biomarker that can single-handedly conclude at which stage of the disease course the patient is. Therefore, many different cognitive assessment scales based on tests and questionnaires are used to identify which stage of the disease the patient is in. These scales are based upon different content areas such as memory, orientation, judgement and problem solving, community affairs, home and hobbies, and personal care [Mendez, 2021; Basun et al., 2006]. The scale then indicates at which of the following stages the patients are; healthy, questionable dementia, mild dementia, moderate dementia, or severe dementia.

These scales are also used when modelling the progression of the disease in clinical trials and how a treatment affects this progression. The effect of a treatment depends, among other things, on the type of treatment. Since a cure for Alzheimer's disease does not exist, the current practice is to treat the disease as effectively as possible given an individual's severity. The current treatments are either symptomatic- or disease-modifying treatments [Cummings and Fox, 2017]. Symptomatic treatments do not affect the underlying disease but rather the symptoms associated with the disease. As the treatment does not affect the underlying disease, it does not affect the severity of the disease long-term, as seen in Figure 1.1. On the contrary, disease-modifying treatments focus on the underlying disease and affect the severity of the disease long-term, see Figure 1.1. Obviously, Figure 1.1 indicates that early intervention with a disease-modifying treatment (or a combination of symptomatic- and disease-modifying treatment) is the best possible strategy when dealing with Alzheimer's disease, as the individual's cognitive function might be preserved for a longer time period [Rasmussen and Langerman, 2019].

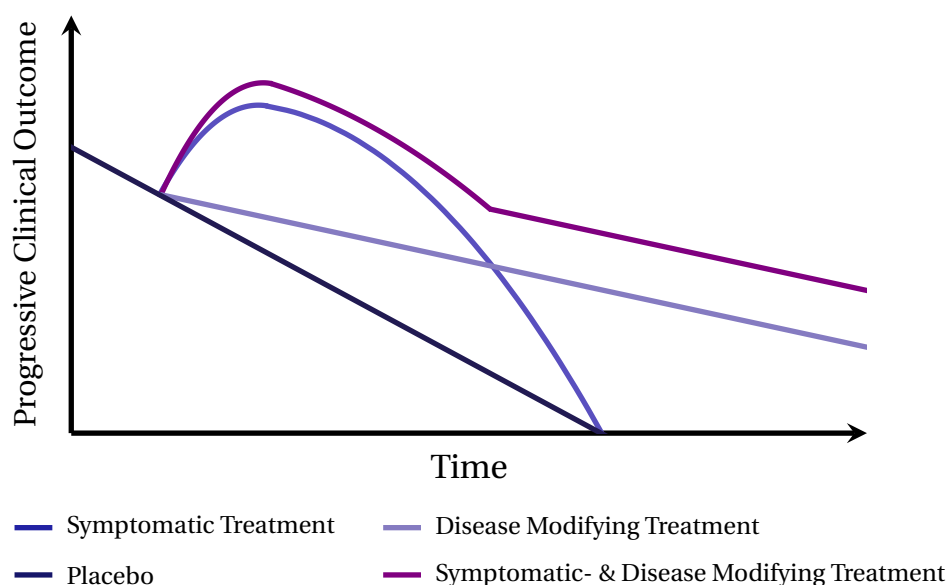


FIGURE 1.1: *Illustration demonstrating how symptomatic treatments, disease-modifying treatments, or a combination of the two affect the progression of Alzheimer's disease.*

So far, multiple treatments for Alzheimer's disease have been developed, and several more are in development. In 2021, the first disease-modifying treatment, aducanemab, was approved by U.S Food and Drug Administration (FDA). However, reimbursement was not granted, leading to a poor uptake, and aducanemab was abandoned in 2024 [Joszt, 2024]. In 2023, lecanemab was approved by FDA and was launched in the US in 2024 [United States Food and Drug Administration, 2023]. Donanemab disclosed promising results from clinical trials in 2023 and anticipates FDA approval after an advisory committee in June 2024. These three treatments are all regarded as disease-modifying treatments that should slow the progression of Alzheimer's disease. Novo Nordisk is conducting two large randomised clinical trials examining if semaglutide can be regarded a disease-modifying treatment [Atri et al., 2022]. The main readout will be in 2025.

As depicted in Figure 1.1, irrespective of the treatment administered, the disease tends to progress over time, leading to a worsening of symptoms. As mentioned, this results in an increased demand for assistance and higher associated costs. In 2019, the global cost of dementia patients was \$1.3 trillion, which is expected to more than double by 2030; \$2.8 trillion [Pillidge and Hanschuh, 2024]. This prediction is based on both the cost of care and that the amount of people living with Alzheimer's disease or other dementias is increasing. Thus, it would not only be beneficial to the people living with Alzheimer's disease but also the different global healthcare sectors and caregivers if an effective treatment was developed. If a new treatment was developed it should be compared to the standard of care to determine if it is more cost-effective.

Health economic modelling provides a framework for structured discussions around the added value of a new treatment and is required in many healthcare systems around the world when initiating discussions on reimbursability of new treatments [Myeloma Patients Europe, 2024]. It seeks to incorporate available evidence about the benefits and costs associated with a new treatment versus comparators of interest. In the setting of clinical drug development, this evidence will often come from clinical trials investigating a new treatment versus comparators. The benefits are often found by modelling the treatment effect of the new treatment, which is

the difference that the new treatment makes compared to placebo. This treatment effect can then be compared to that of the comparators.

The treatment effect can be modelled with models such as linear mixed models, which have especially gained popularity in the medicinal area as they allow subjects to exhibit slight differences [Donohue et al., 2023]. The mixed models can both model the mean trajectory of a group and the subject-specific trajectories, by including random effects such as subject-specific intercepts. This allows more flexibility and is often better at describing the variability in the data of a clinical trial than a standard linear model [Demidenko, 2013, Ch. 1]. However, linear mixed models restrict the behavior of the treatment effect of a treatment to be a linear combination of effects on the outcome scale. Although, this might not be sufficient in all cases. Hence, a new class of models [Raket, 2022] have recently been explored, called progression models. These enable various treatment effect behaviors, and furthermore extends the conventionally used categorical-time parametrisation into a continuous-time framework. These progression models also introduces a relatively new aspect considered in the Alzheimer's disease area by using *time saved* as a treatment effect instead of determining the change outcome scale. This can contribute to a better understanding and clinical meaningfulness of the treatment effect [Raket, 2022; Landhuis, 2024].

However, for the progression models to be used in a clinical trial instead of the standard modelling practice, they should at least experience the same accuracy, both with respect to the estimation- and prediction capabilities. Furthermore, to enable a fair and complete comparison of two treatments, health economic models typically must apply a much longer time horizon than seen in clinical trials. Currently, when conducting a trial for Alzheimer's disease, the trial's duration is often not more than 18 months. However, since Alzheimer's disease is a relatively slow-progressing disease, this has the implication that many of the trial's subjects are not examined long enough to progress to severe Alzheimer's disease. This introduces an extrapolation problem, where statistical modelling assumptions, such as a constant treatment effect over time, should be fulfilled. If this assumption is not met, it is hard to justify which treatment effect should be expected long-term. Hence, the models utilised should not only have great estimation accuracy, but should also be able to be used for extrapolation in some way.

This thesis seeks to examine a new class of models, progression models for repeated measures, across various scenarios, contrasting them with conventional models commonly employed in clinical trials for assessing treatment effects. This should bring insight into the models' shortcomings and strengths, to help assess their applicability in clinical trials and health economic modelling.

2 Mixed Models

This chapter is based upon [Demidenko, 2013], [Jiang and Nguyen, 2021], and [Madsen and Thyregod, 2010], unless stated otherwise.

The progression models will, as mentioned, be compared with one of the models which is commonly used for modelling treatment effect. This model is a linear mixed model, and in this chapter, the theory of mixed models will be presented. Alongside the theory, relevant medical examples are provided to give a greater insight into why these methods are used, and in which cases they are especially useful. This emphasis is due to the central focus of this thesis, which examines clinical trial data for Alzheimer's disease (AD). The chapter begins by describing linear mixed models, and continues by expanding this to non-linear mixed models.

2.1 Linear Mixed Models

Linear regression is one of many statistical methods used to model the relationship between response- and explanatory variables. Let $\mathbf{y} = (y_1, \dots, y_n)^\top$ be an $n \times 1$ vector of response variables. Linear regression then imposes a linear relationship between \mathbf{y} and the explanatory variables, gathered in an $n \times p$ *design matrix* \mathbf{X} . The *standard linear regression model* is given as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is a $p \times 1$ vector of coefficients, also referred to as *fixed effects*. Furthermore, $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ is an $n \times 1$ vector of independent normally distributed errors with mean 0 and covariance $\sigma^2 \mathbf{I}$.

One problem with the standard linear regression model is its inability to handle correlation in the response variables. For example, when conducting a clinical trial, it is common to collect measurements from a subject multiple times to track changes in their health status over time. Naturally, this can create correlation in the response variables. Furthermore, this gives a possible need for a model which can take subject-specific variations into account, since subjects can respond differently to a treatment or other interventions. An extension of the standard linear regression model is thus needed when dealing with such correlated data. This can be done by introducing correlation in the errors and/or *random effects*. *Mixed models*, also known as *mixed effects models*, incorporate random effects, which take correlation between observations into account. Take the example of multiple measurements from different subjects, in which random effects can account for the variability between subjects. This can be done by, for example, including a subject-specific random intercept. Thus, as the name of the mixed model suggests, it uses both the fixed effects as well as the random effects to describe the relationship between the explanatory and response variables.

Similar to the standard linear regression model, the *linear mixed model* imposes a linear relationship between the response- and explanatory variables. Let $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$ be

an $n \times 1$ stochastic vector, with a realisation of the vector denoted as $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$. A *general linear mixed model* is given as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} + \boldsymbol{\varepsilon}, \quad (2.1)$$

where \mathbf{X} and \mathbf{Z} are $n \times p$ and $n \times q$ design matrices, respectively. Furthermore, $\boldsymbol{\beta} \in \mathbb{R}^p$ is a $p \times 1$ vector of fixed effects, and $\mathbf{U} \in \mathbb{R}^q$ is a $q \times 1$ vector of random effects independent of the $n \times 1$ vector of errors $\boldsymbol{\varepsilon} \in \mathbb{R}^n$. In the linear mixed model, the random effects have mean $\mathbf{0}$ and covariance $\boldsymbol{\Psi}(\tau^2)$ noted $\boldsymbol{\Psi}$, whereas the errors have mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma}(\sigma^2)$ noted $\boldsymbol{\Sigma}$. These covariance matrices can be rewritten to simplify notation and can be used to rewrite (2.1) such that the errors are uncorrelated. This could further ease the estimation process. Assuming that $\boldsymbol{\Psi}$ is positive definite, there exists a *precision factor* $\boldsymbol{\Delta}(\delta)$ noted $\boldsymbol{\Delta}$, such that $\boldsymbol{\Psi} = \tau^2 (\boldsymbol{\Delta}^\top \boldsymbol{\Delta})^{-1}$, where $\boldsymbol{\Delta}$ is parameterised by δ . Furthermore, the error's covariance matrix can be expressed as $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{\Lambda}(\lambda)$, where $\boldsymbol{\Lambda}(\lambda)$ noted $\boldsymbol{\Lambda}$ is assumed positive definite and parameterised by λ . The parameter space of the vector $\boldsymbol{\psi} = (\sigma^2, \tau^2, \delta, \lambda)$, is given as

$$\Theta = \{\boldsymbol{\psi} : \sigma^2 > 0, \tau^2 > 0, \delta : \mathbf{z}^\top \boldsymbol{\Sigma} \mathbf{z} > 0, \lambda : \mathbf{x}^\top \boldsymbol{\Psi} \mathbf{x} > 0 \mid \mathbf{0} \neq \mathbf{z} \in \mathbb{C}^q, \mathbf{0} \neq \mathbf{x} \in \mathbb{C}^n\}.$$

In the following, this notation will only be used to ease some of the notation, as the estimation process will be presented for the general case with correlated errors.

One type of a linear mixed model is the *Gaussian linear mixed model*, which gets its name from the assumption of normality for both the random effects and errors. The assumption of normality for the random effects and errors might not hold in practice; hence, such an assumption should be verified before fitting a Gaussian linear mixed model. However, in the remainder of this chapter it will be assumed that the random effects and errors follow a normal distribution, that is

$$\mathbf{U} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}) \quad \text{and} \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}).$$

2.1.1 Linear Mixed Model for Repeated Measures

When describing how observations might be correlated, it was briefly touched upon that subjects could be measured multiple times throughout a clinical trial. Take, for example, a neurodegenerative disorder, such as AD, which, among other things, deteriorates the subject's nervous system over time. To accurately describe the progression of this type of disease, each subject must be measured multiple times throughout the clinical trial. The following theory is a special case of the aforementioned general linear mixed model. This special case allows us to examine the model structure for each subject.

Let $\mathbf{Y}_i = (Y_{i,1}, \dots, Y_{i,n_i})^\top$ be an $n_i \times 1$ stochastic vector for a subject i , where $i = 1, \dots, N$ for $N \in \mathbb{N}$. If $n_i = n$ for all i , it is called a *balanced design*, as every subject has an equal number of observations. Furthermore, let \mathbf{X}_i and \mathbf{Z}_i be $n_i \times p$ and $n_i \times q$ design matrices, respectively, for a subject i . A *general linear mixed model for repeated measures* for a subject i is then given as

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{U}_i + \boldsymbol{\varepsilon}_i, \quad (2.2)$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is a $p \times 1$ vector of fixed effects. Furthermore, $\mathbf{U}_i \in \mathbb{R}^q$ is a $q \times 1$ vector of random effects independent of the $n_i \times 1$ error vector $\boldsymbol{\varepsilon}_i \in \mathbb{R}^{n_i}$. The random effects $\mathbf{U}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}(\tau^2))$

and errors $\varepsilon_i \sim N(0, \Sigma_i(\sigma^2))$ are further assumed to be independent for different i . Notice that the covariance matrix for the errors now depend on the subject i , as indicated by the index.

Example 2.1 demonstrates some of these concepts using orthodontic measurements over time for several subjects.

Example 2.1: Orthodont I.

Consider a linear mixed model for repeated measures and the `Orthodont` data set from the `nlme` package. Figure 2.2 illustrates the subject-specific evolution of the distance from the pituitary gland to the pterygomaxillary fissure for the male subjects in the `Orthodont` data set. Here, we have $N = 16$ subjects, each with $n_i = 4$ observations.

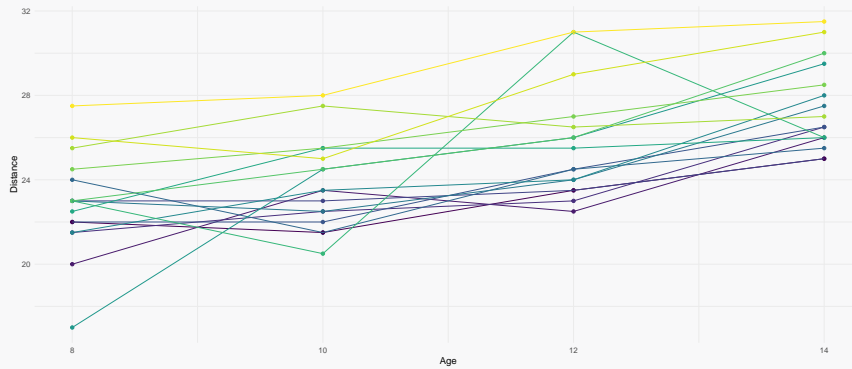


FIGURE 2.2: Orthodontic measurements for 16 male subjects at the ages 8, 10, 12, and 14.

The goal is to model the distance from the pituitary gland to the pterygomaxillary fissure. This model will include fixed effects for the mean intercept and common slope, and a subject-specific random intercept. In R this model would be expressed as `distance ~ age + (1 | Subject)`, where `(1 | Subject)` is the syntax for a subject-specific random intercept.

The linear mixed model for repeated measures, for subject i , is then given as in (2.2), where \mathbf{Y}_i is a 4×1 vector, \mathbf{X}_i a 4×2 matrix, $\boldsymbol{\beta}$ a 2×1 vector, \mathbf{Z}_i a 4×1 vector, \mathbf{U}_i a 1×1 vector, and $\boldsymbol{\varepsilon}_i$ a 4×1 vector. Specifically, the design matrices, for subject i , are given as

$$\mathbf{X}_i^\top = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 8 & 10 & 12 & 14 \end{bmatrix} \quad \text{and} \quad \mathbf{Z}_i^\top = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}.$$

Additionally, $\boldsymbol{\Psi} = \tau^2$ and $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}$.

We will return to this example in the following, once we have introduced how the fixed effects, random effects, and the covariance parameters can be determined. The following section will introduce the theory for this.

2.2 Estimation of Linear Mixed Models

This section presents the estimation of the multiple parameters of the linear mixed model. The estimation process differs from that of the linear regression model, as the random effects in the linear mixed model affects the estimates. Additionally, it introduces yet another term which should be estimated. Subsequent sections will cover the estimation of fixed effects β , the covariance parameters ψ , and the prediction of the random effects U . The focus will remain on the Gaussian linear mixed model. For information regarding the estimation of non-Gaussian linear mixed models, see [Jiang and Nguyen, 2021, p. 17-34].

2.2.1 Estimation of Fixed Effects

Under the assumptions of normally distributed random effects and errors, it follows that Y is marginally normally distributed with mean $\mathbb{E}[Y] = X\beta$ and covariance $\text{Var}[Y] = V(\psi) = Z\Psi Z^\top + \Sigma$. Given that Y is marginally normally distributed, its likelihood function is given as

$$L(\beta, \psi; y) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(V(\psi))}} \exp\left(-\frac{1}{2} (y - X\beta)^\top V^{-1}(\psi) (y - X\beta)\right), \quad (2.3)$$

and hence the log-likelihood function is given as

$$\ell(\beta, \psi; y) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \left(\log(\det(V(\psi))) + (y - X\beta)^\top V^{-1}(\psi) (y - X\beta) \right). \quad (2.4)$$

Differentiating this expression with respect to the fixed effects β and equating it to zero yields the maximum likelihood estimate (MLE) of the fixed effect, which is given as

$$\hat{\beta}(\psi) = (X^\top V^{-1}(\psi) X)^{-1} (X^\top V^{-1}(\psi) y), \quad (2.5)$$

conditioned on the existence of the inverse of $X^\top V^{-1}(\psi) X$. For the inverse to exist X must be of full rank ($\text{rank}(X) = p$). If this is not the case, [Jiang and Nguyen, 2021, p. 49-50] proposes using the generalised inverse instead, but this will not be considered here.

If ψ is known, the estimate of the fixed effect in (2.5) is equivalent to the generalised least squares (GLS) estimate. On the contrary, if ψ is not known, it has to be estimated, which is the focus of the following section.

2.2.2 Estimation of Covariance Parameters

If ψ is not known its MLE can be achieved by maximising the profile log-likelihood function

$$\ell(\psi; y) \approx -\frac{1}{2} \log(\det(V(\psi))) - \frac{1}{2} (y - X\hat{\beta}(\psi))^\top V^{-1}(\psi) (y - X\hat{\beta}(\psi)) \quad (2.6)$$

using the estimate $\hat{\beta}$ from (2.5). It is also possible to determine an explicit expression for the estimates of σ^2 , τ^2 , δ , and λ by factorising one of them out of V . For σ^2 this gives

$$V(\psi) = \sigma^2 (\Sigma + \phi Z \Delta \Delta^\top Z^\top) = \sigma^2 \mathcal{V}(\phi, \delta, \lambda), \quad (2.7)$$

where $\phi = \tau^2/\sigma^2$. Reparameterising (2.6) with (2.7), differentiating (2.6) with respect to σ^2 , and equating it to zero yields the MLE for σ^2 , which is given as

$$\hat{\sigma}_{\text{MLE}}^2(\phi, \delta, \lambda) = \frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\beta}(\phi, \delta, \lambda))^\top \mathcal{V}^{-1}(\phi, \delta, \lambda) (\mathbf{y} - \mathbf{X}\hat{\beta}(\phi, \delta, \lambda)). \quad (2.8)$$

This leaves the task of determining the parameters ϕ , δ , and λ , which is done by maximising the profile log-likelihood function using the estimates of β and σ^2 from (2.5) and (2.8), respectively. The profile log-likelihood function for ϕ , δ , and λ is given as

$$\begin{aligned} \ell(\phi, \delta, \lambda; \mathbf{y}) &\approx -\frac{1}{2} \log(\det(\hat{\sigma}^2(\phi, \delta, \lambda) \mathcal{V}(\phi, \delta, \lambda))) - \frac{1}{2} \hat{\sigma}^{-2}(\phi, \delta, \lambda) n \hat{\sigma}^2(\phi, \delta, \lambda) \\ &\equiv \frac{n}{2} \log(\hat{\sigma}^2(\phi, \delta, \lambda)) - \frac{1}{2} \log(\det(\mathbf{\Lambda} + \phi \mathbf{\Delta} \mathbf{\Delta}^\top \mathbf{Z}^\top \mathbf{Z})). \end{aligned}$$

A drawback of the ML method is that MLEs of variance parameters tend to be biased when n is small. Take standard linear regression as an example where $\mathbf{Y} \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2 \mathbf{I})$. Then the MLE of σ^2 , denoted $\hat{\sigma}_{\text{MLE}}^2$, is given as $\frac{1}{n} (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta)$ where

$$\mathbb{E}[\hat{\sigma}_{\text{MLE}}^2] = \mathbb{E}\left[\frac{1}{n} (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta)\right] = \sigma^2 \left(1 - \frac{n-p}{n}\right),$$

and hence σ^2 is in general underestimated. However, it also shows that when n is sufficiently large, the bias becomes negligible.

Restricted Maximum Likelihood Estimation

One method of solving this drawback of the MLE is using what is called the *restricted maximum likelihood* (REML)-method, as it might be less biased for the estimates of the covariance parameters ψ . As n becomes sufficiently large, the REML estimate and MLE become almost identical. Nevertheless, REML can be preferable for both small and large n in some cases when it, for example, eliminates the bias completely. Returning to the example of the standard linear regression, the REML estimate of σ^2 is here not only less biased but unbiased, as

$$\mathbb{E}[\hat{\sigma}_{\text{REML}}^2] = \mathbb{E}\left[\frac{1}{n-p} (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta)\right] = \sigma^2.$$

Hence, using the REML estimate for the linear regression is preferable to the MLE for both small and large n .

The way REML mitigates the bias that appears in the MLE is by making a linear transformation of the data which eliminates the mean. Let \mathcal{A} be an $n \times (n-p)$ matrix for $n > p$ with its columns spanning the orthogonal complement of $\text{span}(\mathbf{X})$ such that $\mathcal{A}^\top \mathbf{X} = \mathbf{0}$. The distribution of $\mathcal{A}^\top \mathbf{Y}$ is therefore given as

$$\mathcal{A}^\top \mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathcal{A}^\top \mathbf{V}(\psi) \mathcal{A}),$$

with a likelihood function given as

$$L(\psi; \mathcal{A}^\top \mathbf{y}) = \frac{1}{(2\pi)^{(n-p)/2} \sqrt{\det(\mathcal{A}^\top \mathbf{V}(\psi) \mathcal{A})}} \exp\left(-\frac{1}{2} (\mathcal{A}^\top \mathbf{y})^\top (\mathcal{A}^\top \mathbf{V}(\psi) \mathcal{A})^{-1} (\mathcal{A}^\top \mathbf{y})\right).$$

Then, using the same computations as before, the log-likelihood function is given as

$$\begin{aligned} \ell(\psi; \mathcal{A}^\top \mathbf{y}) = & -\frac{n-p}{2} \log(2\pi) - \frac{n-p}{2} \log(\sigma^2) - \frac{1}{2} \left(\log(\det(\mathcal{A}^\top \mathcal{V}(\phi, \delta, \lambda) \mathcal{A})) \right. \\ & \left. + \sigma^{-2} (\mathcal{A}^\top \mathbf{y})^\top (\mathcal{A}^\top \mathcal{V}(\phi, \delta, \lambda) \mathcal{A})^{-1} (\mathcal{A}^\top \mathbf{y}) \right). \end{aligned}$$

Differentiating this log-likelihood function with respect to σ^2 and equating to zero yields the REML estimate of σ^2 is given as

$$\begin{aligned} \hat{\sigma}_{\text{REML}}^2(\phi, \delta, \lambda) &= \frac{1}{n-p} (\mathcal{A}^\top \mathbf{y})^\top (\mathcal{A}^\top \mathcal{V}(\phi, \delta, \lambda) \mathcal{A})^{-1} (\mathcal{A}^\top \mathbf{y}) \\ &= \frac{1}{n-p} (\mathbf{y} - \mathbf{X} \hat{\beta}(\phi, \delta, \lambda))^\top \mathcal{V}^{-1}(\phi, \delta, \lambda) (\mathbf{y} - \mathbf{X} \hat{\beta}(\phi, \delta, \lambda)). \end{aligned} \quad (2.9)$$

This can be used in a profile log-likelihood function together with (2.5) to estimate ϕ , δ , and λ in a similar manner as for the MLEs. These are then inserted into (2.5) to find the REML estimate of β .

When comparing the MLE and REML estimate for σ^2 , in (2.8) and (2.9), respectively, the only difference is the denominator. The REML estimate has a denominator of $n - p$ whilst the MLE has a denominator of n , which is what makes the REML estimate less biased compared to the MLE.

Both the ML and REML method will, in general, involve some numerical maximisation of the profile log-likelihood function after eliminating some of the parameters, for example, β and σ^2 . The maximisation of the profile likelihood function of the respective parameters will in practice be performed iteratively, using an algorithm such as Newton-Raphson. A short outline of this algorithm can be found in Section A.1.

2.2.3 Prediction of Random Effects

Given that β and ψ are known, or at the very least estimated, the random effects \mathbf{U} can be predicted. Naturally, the optimal predictor is desired. The optimal predictor of \mathbf{U} is the predictor that minimises the mean square prediction error. Specifically, for a predictor of \mathbf{U} given \mathbf{Y} , $f(\mathbf{Y})$,

$$\mathbb{E}[(\mathbf{U} - \mathbb{E}[\mathbf{U} | \mathbf{Y}]) (\mathbb{E}[\mathbf{U} | \mathbf{Y}] - f(\mathbf{Y}))] = 0.$$

This implies that

$$\mathbb{E}[(\mathbf{U} - f(\mathbf{Y}))^2] = \mathbb{E}[(\mathbf{U} - \mathbb{E}[\mathbf{U} | \mathbf{Y}])^2] + \mathbb{E}[(\mathbb{E}[\mathbf{U} | \mathbf{Y}] - f(\mathbf{Y}))^2] \geq \mathbb{E}[(\mathbf{U} - \mathbb{E}[\mathbf{U} | \mathbf{Y}])^2].$$

It is thus sufficient to look at the conditional expectation $\mathbb{E}[\mathbf{U} | \mathbf{Y}]$ to determine the best prediction of the random effects, where

$$\begin{aligned} \mathbb{E}[\mathbf{U} | \mathbf{Y}] &= (\text{Cov}[\mathbf{Y}, \mathbf{U}])^\top (\text{Cov}[\mathbf{Y}, \mathbf{Y}])^{-1} (\mathbf{Y} - \mathbb{E}[\mathbf{Y}]) \\ &= (\text{Cov}[\mathbf{X}\beta, \mathbf{U}] + \mathbf{Z} \text{Cov}[\mathbf{U}, \mathbf{U}] + \text{Cov}[\boldsymbol{\varepsilon}, \mathbf{U}])^\top (\text{Var}[\mathbf{Y}])^{-1} (\mathbf{Y} - \mathbb{E}[\mathbf{Y}]) \\ &= (\mathbf{Z}\boldsymbol{\Psi})^\top \mathbf{V}^{-1}(\psi) (\mathbf{Y} - \mathbf{X}\beta). \end{aligned}$$

Thus, the prediction of the random effects is given as

$$\hat{\mathbf{U}}(\psi, \beta) = (\mathbf{Z}\boldsymbol{\Psi})^\top \mathbf{V}^{-1}(\psi) (\mathbf{Y} - \mathbf{X}\beta). \quad (2.10)$$

Besides being the best predictor, it can also be seen that it is linear and that $\mathbb{E}[\hat{U}] = \mathbb{E}[U] = 0$, and hence this predictor is the *best linear unbiased predictor*. When β and ψ are not known, their estimates from the previous sections may be used. It should also be noted that depending on whether the MLE or REML estimate for ψ is used, the predictions may vary between the two methods. An example of this can be seen in Example 2.3.

Example 2.3: Orthodont II.

Continuing on Example 2.1 and using the estimation methods described above, the estimates for the fixed effects and covariance parameters are

$$\begin{aligned}\beta_1 &\approx 16.34, & \beta_2 &\approx 0.78, \\ \hat{\tau}_{\text{MLE}}^2 &\approx 2.43, & \hat{\tau}_{\text{REML}}^2 &\approx 2.66, \\ \hat{\sigma}_{\text{MLE}}^2 &\approx 2.76, & \hat{\sigma}_{\text{REML}}^2 &\approx 2.82.\end{aligned}$$

This example also shows that the MLE and REML estimates for the variances do not coincide. The different estimates also provide different predictions of the random effects, as

$$\begin{aligned}\hat{U}_{\text{MLE}} &= \begin{bmatrix} -1.61 & -1.61 & -1.30 & -1.10 & -0.99 & -0.89 & -0.59 & -0.59 \\ -0.59 & -0.08 & 0.13 & 0.74 & 1.15 & 1.35 & 2.27 & 3.69 \end{bmatrix} \\ \hat{U}_{\text{REML}} &= \begin{bmatrix} -1.63 & -1.63 & -1.32 & -1.11 & -1.01 & -0.91 & -0.60 & -0.60 \\ -0.60 & -0.08 & 0.13 & 0.75 & 1.16 & 1.37 & 2.30 & 3.75 \end{bmatrix},\end{aligned}$$

where both are 1×16 vectors representing $(U_1, U_2, \dots, U_{16})$, despite their appearance. The difference, however, is below or equal to 0.06 for all of the predictions, presumably due to the small difference in the estimates of τ^2 and σ^2 .

2.3 Non-Linear Mixed Models

This section is, besides the stated references in the beginning of the chapter, based upon [Pinheiro and Bates, 2000] and [Lindstrom and Bates, 1990].

The linear mixed model described in Section 2.1 is a special case of the broader set of mixed models, called *non-linear mixed models*. Unlike the linear mixed models, the relationship between the explanatory- and response variables is not required to be linear, as the name suggests. This extends the applicability of mixed models to data such as growth data, which often follows a logistic growth curve instead of a linear relation.

Let $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ be an $n \times 1$ stochastic vector. Furthermore, let $\varepsilon \in \mathbb{R}^n$ be an $n \times 1$ vector of errors with mean $\mathbf{0}$ and $n \times n$ covariance matrix $\Sigma(\sigma^2)$ noted Σ . A non-linear mixed model is often presented as a two-stage model, where the first stage is given by

$$\mathbf{Y} = \mathbf{g}(\boldsymbol{\eta}) + \varepsilon, \quad (2.11)$$

where \mathbf{g} is an $n \times 1$ non-linear differentiable vector function. $\mathbf{g}(\boldsymbol{\eta})$ will henceforth be denoted as \mathbf{g} to simplify the following expressions. The second stage, $\boldsymbol{\eta}$, is given as

$$\boldsymbol{\eta} = \mathbf{X}\beta + \mathbf{Z}U. \quad (2.12)$$

Here, \mathbf{X} and \mathbf{Z} are design matrices with dimensions $n \times p$ and $n \times q$, respectively. Moreover, $\boldsymbol{\beta} \in \mathbb{R}^p$ is a $p \times 1$ vector of fixed effects and $\mathbf{U} \in \mathbb{R}^q$ is a $q \times 1$ vector of random effects. Additionally, it is assumed that \mathbf{U} and ε are independent. As in the linear case, \mathbf{U} has mean $\mathbf{0}$ and $q \times q$ covariance matrix $\boldsymbol{\Psi}(\tau^2)$ noted $\boldsymbol{\Psi}$. Furthermore, $\boldsymbol{\Psi}$ is assumed positive definite, such that there exists a precision factor $\boldsymbol{\Delta}(\delta) = \boldsymbol{\Delta}$, meaning $\boldsymbol{\Psi} = \tau^2 (\boldsymbol{\Delta}^\top \boldsymbol{\Delta})^{-1}$, where $\boldsymbol{\Delta}$ is parameterised by δ . Moreover, the error's covariance matrix can be expressed as $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{\Lambda}(\lambda)$, where $\boldsymbol{\Lambda}(\lambda)$ noted $\boldsymbol{\Lambda}$ is assumed positive definite and parameterised by λ . The parameter space of the vector $\boldsymbol{\psi} = (\sigma^2, \tau^2, \delta, \lambda)$, is given as

$$\Theta = \{\boldsymbol{\psi} : \sigma^2 > 0, \tau^2 > 0, \delta : \mathbf{z}^\top \boldsymbol{\Sigma} \mathbf{z} > 0, \lambda : \mathbf{x}^\top \boldsymbol{\Psi} \mathbf{x} > 0 \mid \mathbf{0} \neq \mathbf{z} \in \mathbb{C}^q, \mathbf{0} \neq \mathbf{x} \in \mathbb{C}^n\}.$$

2.4 Estimation of Non-linear Mixed Models

The non-linearity in the model complicates the estimation process. This can be seen by looking at the likelihood function

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\psi}; \mathbf{y}) &= \int_{\mathbb{R}^q} L(\boldsymbol{\beta}, \boldsymbol{\psi}; \mathbf{y}, \mathbf{u}) d\mathbf{u} \\ &= \int_{\mathbb{R}^q} f_n(\mathbf{y} \mid \mathbf{u}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) f_q(\mathbf{u} \mid \boldsymbol{\Psi}) d\mathbf{u}, \end{aligned}$$

for the conditional density function, f_n , of \mathbf{Y} given \mathbf{u} , and the marginal distribution, f_q , of \mathbf{U} . Because of the non-linear relationship between the response variable and the random effects, the random effects cannot be factored out. This is also evident in the specific instance wherein the random effects and errors are assumed Gaussian.

For the non-linear mixed model under the Gaussian assumption, it follows that

$$\mathbf{Y} \mid \mathbf{U} \sim \mathcal{N}(\mathbf{g}, \boldsymbol{\Sigma}).$$

Thus, the likelihood function is given as

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\psi}; \mathbf{y}) &= \int_{\mathbb{R}^q} \frac{1}{(2\pi)^{n/2} \sqrt{\det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{g})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{g})\right) \\ &\quad \frac{1}{(2\pi)^{q/2} \sqrt{\det(\boldsymbol{\Psi})}} \exp\left(-\frac{1}{2}\mathbf{u}^\top \boldsymbol{\Psi}^{-1}\mathbf{u}\right) d\mathbf{u}, \end{aligned} \quad (2.13)$$

and the log-likelihood function as

$$\begin{aligned} \ell(\boldsymbol{\beta}, \boldsymbol{\psi}, \mathbf{y}) &= \log\left(\int_{\mathbb{R}^q} (2\pi)^{-(n+q)/2} \det(\boldsymbol{\Sigma})^{-1/2} \det(\boldsymbol{\Psi})^{-1/2} \right. \\ &\quad \left. \exp\left(-\frac{1}{2}\left((\mathbf{y} - \mathbf{g})^\top (\boldsymbol{\Sigma})^{-1}(\mathbf{y} - \mathbf{g}) + \mathbf{u}^\top (\boldsymbol{\Psi})^{-1}\mathbf{u}\right)\right) d\mathbf{u}\right) \\ &\propto -\frac{1}{2}(\log(\det(\boldsymbol{\Psi})) + \log(\det(\boldsymbol{\Sigma}))) \\ &\quad + \log\left(\int_{\mathbb{R}^q} \exp\left(-\frac{1}{2}\left((\mathbf{y} - \mathbf{g})^\top (\boldsymbol{\Sigma})^{-1}(\mathbf{y} - \mathbf{g}) + \mathbf{u}^\top (\boldsymbol{\Psi})^{-1}\mathbf{u}\right)\right) d\mathbf{u}\right). \end{aligned} \quad (2.14)$$

Hence, even though the non-linear mixed model extends the applicability of mixed models, the inclusion of random effects in a non-linear framework typically results in the absence of

a closed-form expression for the likelihood function. This is also true under the Gaussian assumption, which this thesis addresses, as seen above. Because of this more complex structure, the estimation of the parameters in the non-linear mixed model is more complicated, and therefore, multiple different approximation methods for approximating the likelihood function have been developed. Some of these approximation methods will be described in the following section.

2.5 Approximation Methods

Before delving into a more detailed description of some of these approximation methods, some assumptions are presented and the model is rewritten.

Similar to before, it is assumed that the random effects and errors are normally distributed. With Λ assumed to be positive definite, the square root $\Lambda^{1/2}$ and its inverse exist such that it is possible to express Λ as

$$\Lambda = \Lambda^{\top/2} \Lambda^{1/2} \quad \text{and} \quad \Lambda^{-1} = \Lambda^{-1/2} \Lambda^{-\top/2}.$$

By making a linear transformation with $\Lambda^{-\top/2}$, the model can be rewritten with

$$\mathbf{Y}^* = \mathbf{g}^*(\boldsymbol{\eta}) + \boldsymbol{\varepsilon}^* = \Lambda^{-\top/2} \mathbf{g}(\boldsymbol{\eta}) + \Lambda^{-\top/2} \boldsymbol{\varepsilon} \quad (2.15)$$

as the first stage and (2.12) as the second stage, where the random effects and errors are distributed as

$$\mathbf{U} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Psi}) \quad \text{and} \quad \boldsymbol{\varepsilon}^* \sim \mathbf{N}(\mathbf{0}, \sigma^2 \mathbf{I}).$$

Thereby, there is no heteroscedasticity or correlation in the errors.

Since $d\mathbf{y}^* = \det(\Lambda)^{-1/2} d\mathbf{y}$ the log-likelihood function is given as

$$\begin{aligned} \ell(\boldsymbol{\beta}, \boldsymbol{\psi}; \mathbf{y}) &= \log(f_n(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\psi})) \\ &= \log(f_n(\mathbf{y}^*; \boldsymbol{\beta}, \boldsymbol{\psi})) - \frac{1}{2} \log(\det(\Lambda)) \\ &= \ell(\boldsymbol{\beta}, \boldsymbol{\psi}; \mathbf{y}^*) - \frac{1}{2} \log(\det(\Lambda)). \end{aligned} \quad (2.16)$$

Thus, it is sufficient to approximate $\ell(\boldsymbol{\beta}, \boldsymbol{\psi}; \mathbf{y}^*)$. The following approximation methods are based on (2.13) and (2.14) with $\mathbf{y} = \mathbf{y}^*$, $\mathbf{g} = \mathbf{g}^*$, and $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$.

2.5.1 Lindstrom and Bates Algorithm

One approximation method is called the Lindstrom and Bates algorithm [Lindstrom and Bates, 1990]. This algorithm estimates the likelihood function by alternating between two steps. The first step is called the penalised non-linear least squares (PNLS) step. In this step, the fixed- and the random effects for the non-linear mixed-effects model are estimated and predicted, respectively. The second step, called the linear mixed effects (LME) step, is used to estimate the covariance parameters, $\boldsymbol{\psi}$.

The PNLS Step

In the PNLS step, the fixed- and random effects are estimated for some fixed ψ , thus only terms in the log-likelihood function dependent on either the fixed- or random effects need consideration. That is,

$$r(\beta, u, \psi) = \frac{1}{\sigma^2} (\mathbf{y}^* - \mathbf{g}^*(\beta, u))^\top (\mathbf{y}^* - \mathbf{g}^*(\beta, u)) + \frac{1}{\tau^2} \mathbf{u}^\top \mathbf{\Delta}^\top \mathbf{\Delta} u. \quad (2.17)$$

Here, $\mathbf{g}^*(\beta, u)$ is used instead of $\mathbf{g}^*(\eta)$ to emphasise \mathbf{g}^* 's dependence on the fixed- and random effects in the model. Since σ^2 and τ^2 are fixed and positive, it is only necessary to consider

$$h(\beta, u) = (\mathbf{y}^* - \mathbf{g}^*(\beta, u))^\top (\mathbf{y}^* - \mathbf{g}^*(\beta, u)) + \mathbf{u}^\top \mathbf{\Delta}^\top \mathbf{\Delta} u.$$

By minimising h with respect to β and u , the log-likelihood function, $\ell(\beta, \psi; \mathbf{y}^*)$, will be maximised [Wang, 2015, p. 5358]. Thus, the fixed- and random effects are estimated by minimising the penalised non-linear least squares with respect to β and u . That is,

$$\min_{\beta, u} \left\{ \|\mathbf{y}^* - \mathbf{g}^*(\beta, u)\|^2 + \|\mathbf{\Delta} u\|^2 \right\}. \quad (2.18)$$

For this minimisation problem, the Gauss-Newton algorithm is often employed, which extends the Newton-Raphson algorithm. This involves first rewriting h such that the objective function to be minimised becomes

$$O(\Gamma) = \|\tilde{\mathbf{y}}^* - \tilde{\mathbf{g}}^*(\Gamma)\|^2 \quad (2.19)$$

where

$$\tilde{\mathbf{y}}^* = \begin{pmatrix} \mathbf{y}^* \\ 0 \end{pmatrix}, \quad \tilde{\mathbf{g}}^*(\Gamma) = \begin{pmatrix} \mathbf{g}^*(\Gamma) \\ \mathbf{\Delta} u \end{pmatrix}, \quad \text{and} \quad \Gamma = (\beta, u)$$

The difference between the Newton-Raphson and Gauss-Newton algorithm is that the latter uses an approximation of the Hessian employed in Newton-Raphson algorithm. In the Newton-Raphson algorithm the first and second derivative with respect to Γ , given as

$$\begin{aligned} O'(\Gamma) &= -2\tilde{\mathbf{g}}^{*'}(\Gamma)^\top (\tilde{\mathbf{y}}^* - \tilde{\mathbf{g}}^*(\Gamma)) \\ O''(\Gamma) &= \mathbf{H}(\Gamma) = -2\tilde{\mathbf{g}}^{*''}(\Gamma)^\top (\tilde{\mathbf{y}}^* - \tilde{\mathbf{g}}^*(\Gamma)) + 2\tilde{\mathbf{g}}^{*'}(\Gamma)^\top \tilde{\mathbf{g}}^{*'}(\Gamma), \end{aligned} \quad (2.20)$$

would be used to update the parameters as presented in Section A.1. The Gauss-Newton algorithm instead utilises the expected value of the first term in the second derivative being zero. Hence, making the approximation

$$O''(\Gamma) \approx 2\tilde{\mathbf{g}}^{*'}(\Gamma)^\top \tilde{\mathbf{g}}^{*'}(\Gamma),$$

such that for some current estimate of the random- and fixed effects, $\hat{\Gamma}_{\text{old}}$, the new estimate is given as

$$\hat{\Gamma}_{\text{new}} = \hat{\Gamma}_{\text{old}} - (\tilde{\mathbf{g}}^{*'}(\Gamma)^\top \tilde{\mathbf{g}}^{*'}(\Gamma))^{-1} (-\tilde{\mathbf{g}}^{*'}(\Gamma)^\top (\tilde{\mathbf{y}}^* - \tilde{\mathbf{g}}^*(\Gamma))).$$

Hence, it avoids the computation of the second derivative in order to update its estimates.

Once the fixed- and random effects are estimated, ψ is then estimated in the LME step.

The LME Step

The goal of the LME step is to approximate the log-likelihood function for \mathbf{y}^* by approximating the conditional distribution of \mathbf{Y}^* given \mathbf{u} . This involves approximating the residuals of the model around the current estimates of β and \mathbf{U} , denoted as $\hat{\beta}$ and $\hat{\mathbf{u}}$, respectively. The notation

$$\widehat{\mathbf{W}} = \left. \frac{\partial \mathbf{g}^*}{\partial \beta} \right|_{\hat{\beta}, \hat{\mathbf{u}}} \quad \text{and} \quad \widehat{\mathbf{Z}} = \left. \frac{\partial \mathbf{g}^*}{\partial \mathbf{u}} \right|_{\hat{\beta}, \hat{\mathbf{u}}}$$

will in the following be used to ease notation.

By employing a first-order Taylor series expansion of \mathbf{g}^* , the residuals are approximately given as

$$\mathbf{y}^* - \mathbf{g}^*(\beta, \mathbf{u}) \approx \mathbf{y}^* - \left(\mathbf{g}^*(\hat{\beta}, \hat{\mathbf{u}}) + \widehat{\mathbf{W}}\beta - \widehat{\mathbf{W}}\hat{\beta} + \widehat{\mathbf{Z}}\mathbf{u} - \widehat{\mathbf{Z}}\hat{\mathbf{u}} \right).$$

From this approximation it follows that

$$\mathbf{y}^* - \mathbf{g}^*(\hat{\beta}, \hat{\mathbf{u}}) - \widehat{\mathbf{W}}\beta + \widehat{\mathbf{W}}\hat{\beta} - \widehat{\mathbf{Z}}\mathbf{u} + \widehat{\mathbf{Z}}\hat{\mathbf{u}} \mid \mathbf{U} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}),$$

which implies

$$\mathbf{y}^* \mid \mathbf{U} \sim \mathcal{N}\left(\mathbf{g}^*(\hat{\beta}, \hat{\mathbf{u}}) + \widehat{\mathbf{W}}\beta - \widehat{\mathbf{W}}\hat{\beta} + \widehat{\mathbf{Z}}\mathbf{u} - \widehat{\mathbf{Z}}\hat{\mathbf{u}}, \sigma^2 \mathbf{I}\right).$$

Hence, the marginal distribution of \mathbf{y}^* can be approximated, using the above and the distribution of \mathbf{U} , as

$$\mathbf{y}^* \sim \mathcal{N}\left(\mathbf{g}^*(\hat{\beta}, \hat{\mathbf{u}}) + \widehat{\mathbf{W}}\beta - \widehat{\mathbf{W}}\hat{\beta} + \widehat{\mathbf{Z}}\hat{\mathbf{u}}, \sigma^2 \mathbf{I} + \widehat{\mathbf{Z}}\Psi\widehat{\mathbf{Z}}^\top\right).$$

Consequently, the approximation of the log-likelihood function is given as

$$\ell(\beta, \psi; \mathbf{y}^*) \approx -\frac{n}{2} \log(2\pi) - \frac{1}{2} \left(\log(\det(\mathbf{Q})) + (\hat{\mathbf{w}} - \widehat{\mathbf{W}}\beta)^\top \mathbf{Q}^{-1} (\hat{\mathbf{w}} - \widehat{\mathbf{W}}\beta) \right),$$

where

$$\begin{aligned} \hat{\mathbf{w}} &= \mathbf{y}^* - \mathbf{g}^*(\hat{\beta}, \hat{\mathbf{u}}) + \widehat{\mathbf{W}}\hat{\beta} + \widehat{\mathbf{Z}}\hat{\mathbf{u}} \\ \mathbf{Q} &= \sigma^2 \mathbf{I} + \widehat{\mathbf{Z}}\Psi\widehat{\mathbf{Z}}^\top. \end{aligned}$$

Thereby, the problem has been reduced to a linear mixed model with response vector $\hat{\mathbf{w}}$ and design matrices $\widehat{\mathbf{W}}$ and $\widehat{\mathbf{Z}}$. Methods outlined in Section 2.2 can therefore be used to estimate the parameters.

It is worth noting that to maximise $\ell(\beta, \psi; \mathbf{y})$ not only $\ell(\beta, \psi; \mathbf{y}^*)$ but also $-\frac{1}{2} \log(\det(\Lambda))$, seen in (2.16), should be maximised. This will affect the methods outlined in Section 2.2 when estimating δ .

2.5.2 Laplace Approximation

This section is based upon [Wang, 2015].

In Bayesian inference, Laplacian approximations are often utilised to estimate marginal posterior densities. However, they can also be used to approximate integrals such as those

presented in the likelihood function for the non-linear mixed models. The *Laplace approximation* is based on an approximation of \mathbf{r} , defined in (2.17).

Let $\hat{\mathbf{u}}$ be a maximiser of \mathbf{r} , that is

$$\hat{\mathbf{u}} = \underset{\mathbf{u}}{\operatorname{argmax}} \mathbf{r}(\boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\psi}).$$

Then, the second-order Taylor series expansion around $\hat{\mathbf{U}}$ is given as

$$\mathbf{r}(\boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\psi}) \approx \mathbf{r}(\boldsymbol{\beta}, \hat{\mathbf{u}}, \boldsymbol{\psi}) + \frac{1}{2} (\mathbf{u} - \hat{\mathbf{u}})^\top \mathbf{r}''(\boldsymbol{\beta}, \hat{\mathbf{u}}, \boldsymbol{\psi}) (\mathbf{u} - \hat{\mathbf{u}}).$$

This follows from the assumption that $\hat{\mathbf{u}}$ maximises \mathbf{r} , implying that $\mathbf{r}'(\boldsymbol{\beta}, \hat{\mathbf{u}}, \boldsymbol{\psi}) = 0$. Furthermore,

$$\begin{aligned} \frac{\partial \mathbf{r}(\boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\psi})}{\partial \mathbf{u}} &= \mathbf{r}'(\boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\psi}) = -2 \left(\frac{1}{\sigma^2} \mathbf{g}^{*'}(\boldsymbol{\beta}, \mathbf{u})^\top (\mathbf{y}^* - \mathbf{g}^*(\boldsymbol{\beta}, \mathbf{u})) - \frac{1}{\tau^2} \boldsymbol{\Delta}^\top \boldsymbol{\Delta} \mathbf{u} \right), \\ \frac{\partial^2 \mathbf{r}(\boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\psi})}{\partial \mathbf{u} \partial \mathbf{u}^\top} &= \mathbf{r}''(\boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\psi}) = -2 \left(\frac{1}{\sigma^2} \mathbf{g}^{*''}(\boldsymbol{\beta}, \mathbf{u})^\top (\mathbf{y}^* - \mathbf{g}^*(\boldsymbol{\beta}, \mathbf{u})) \right. \\ &\quad \left. - \frac{1}{\sigma^2} \mathbf{g}^{*'}(\boldsymbol{\beta}, \mathbf{u})^\top \mathbf{g}^{*'}(\boldsymbol{\beta}, \mathbf{u}) - \frac{1}{\tau^2} \boldsymbol{\Delta}^\top \boldsymbol{\Delta} \right). \end{aligned}$$

It should be noted that the term involving $\mathbf{g}^{*''}$ in \mathbf{r}'' in general is negligible in comparison to the impact of the term involving $\mathbf{g}^{*'}$ at $\hat{\mathbf{u}}$ [Bates and Watts, 1980]. Therefore,

$$\mathbf{r}''(\boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\psi}) \cong \mathbf{G}(\boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\psi}) = 2 \left(\mathbf{g}^{*'}(\boldsymbol{\beta}, \mathbf{u})^\top (\sigma^2 \mathbf{I})^{-1} \mathbf{g}^{*'}(\boldsymbol{\beta}, \mathbf{u}) + \frac{1}{\tau^2} \boldsymbol{\Delta}^\top \boldsymbol{\Delta} \right).$$

Using the above approximation of \mathbf{r} and the notation $\frac{1}{\sigma^2} \mathbf{I} = \boldsymbol{\Sigma}^{-1}$ and $\frac{1}{\tau^2} \boldsymbol{\Delta}^\top \boldsymbol{\Delta} = \boldsymbol{\Psi}^{-1}$, the log-likelihood function in (2.14) can be rewritten as

$$\begin{aligned} \ell(\boldsymbol{\beta}, \boldsymbol{\psi}, \mathbf{y}) &= -\frac{1}{2} (\log(\det(\boldsymbol{\Psi})) + \log(\det(\boldsymbol{\Sigma}))) + \log \left(\int_{\mathbb{R}^q} \exp \left(-\frac{1}{2} \mathbf{r}(\boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\psi}) \right) d\mathbf{u} \right) \\ &\approx -\frac{1}{2} (\log(\det(\boldsymbol{\Psi})) + \log(\det(\boldsymbol{\Sigma}))) + \log \left(\exp \left(-\frac{1}{2} \mathbf{r}(\boldsymbol{\beta}, \hat{\mathbf{u}}, \boldsymbol{\psi}) \right) \right) \\ &\quad + \log \left(\int_{\mathbb{R}^q} \exp \left(-\frac{1}{4} ((\mathbf{u} - \hat{\mathbf{u}})^\top \mathbf{r}''(\boldsymbol{\beta}, \hat{\mathbf{u}}, \boldsymbol{\psi}) (\mathbf{u} - \hat{\mathbf{u}})) \right) d\mathbf{u} \right) \\ &\propto -\frac{1}{2} \left(\log(\det(\boldsymbol{\Psi})) + \log(\det(\boldsymbol{\Sigma})) - \log \left(\det \left(\frac{1}{2} \mathbf{G}(\boldsymbol{\beta}, \hat{\mathbf{u}}, \boldsymbol{\psi}) \right) \right) \right) \\ &\quad + (\mathbf{y}^* - \mathbf{g}^*(\boldsymbol{\beta}, \hat{\mathbf{u}}))^\top (\boldsymbol{\Sigma})^{-1} (\mathbf{y}^* - \mathbf{g}^*(\boldsymbol{\beta}, \hat{\mathbf{u}})) + \hat{\mathbf{u}}^\top (\boldsymbol{\Psi})^{-1} \hat{\mathbf{u}} \end{aligned}$$

Following the same approach as in Section 2.2, we can estimate some of the parameters by differentiating with respect to this parameter and equating to zero, whilst the remaining parameters can be estimated using the profile likelihood function of the respective parameters.

The Laplace approximation generally provides more accurate estimates than the Lindstrom and Bates algorithm. This is due to the LME step using an expansion around both the estimated fixed- and random effects, whereas the Laplace approximation only uses an expansion around the random effects. However, the Laplace approximation is usually more computationally expensive than the Lindstrom and Bates algorithm [Pinheiro and Bates, 2000, p. 319] and [Wang, 2015, p. 5375].

Another approximation method is the *Adaptive Gaussian Quadrature* approximation, which Laplace approximation is a special case of. The Adaptive Gaussian Quadrature approximation can be made arbitrarily accurate by increasing the number of quadrature points. Quadrature points, are points wherein the integrand is evaluated. Here, one quadrature point corresponds to the Laplace approximation, as we only evaluate it in \hat{u} . Each quadrature point has an associated weight. The associated weight determines how much the integrand evaluated in the quadrature point contributes to the overall approximation. Adaptive Gaussian Quadrature approximation has not yet been implemented in any R-package for non-linear mixed models, and will hence not be elaborated further upon. For more information regarding Adaptive Gaussian Quadrature see [Pinheiro and Bates, 2000, p. 319-322]. Both Laplace approximation and the Lindstrom and Bates algorithm has been implemented in the packages `lme4` and `nLme`, respectively.

2.6 An Extended Non-linear Regression Model

In both the linear- and nonlinear mixed model the correlation in the response variable can be captured by the correlation in the errors and/or by the random effects. However, in some cases the inclusion of one of these is enough to capture the correlation in the response variable. If only including the random effects, the estimation process in Section 2.2 and Section 2.4 still applies with $\Sigma = \sigma^2 I$. Given that we do not include random effects, but only correlation in the errors, the estimation process changes. This will be the focus of this section. This type of model can either be seen as a simple version of the non-linear mixed model or as an extension of a non-linear regression model with uncorrelated errors, and is called an *extended non-linear regression model*. Since there are no random effects, a closed-form expression for the log-likelihood function can be obtained, simplifying the estimation process. The model is defined by the following two stages

$$\begin{aligned} Y &= g(\eta) + \varepsilon, \\ \eta &= X\beta, \end{aligned} \tag{2.21}$$

where g is an $n \times 1$ non-linear differentiable vector function and $\varepsilon \in \mathbb{R}^n$ is an $n \times 1$ error vector with mean 0 and $n \times n$ covariance matrix $\sigma^2 \Lambda(\lambda)$. Here, X is an $n \times p$ design matrix and $\beta \in \mathbb{R}^p$ is a $p \times 1$ vector of fixed effects. Lastly, Λ is assumed positive definite and parameterised by a fixed set of parameters λ . Using the same transformation as in (2.15), it is possible to rewrite (2.21) as

$$\begin{aligned} Y^* &= g^*(\eta) + \varepsilon^*, \\ \eta &= X\beta, \end{aligned}$$

where ε^* is a error vector with mean 0 and covariance matrix $\sigma^2 I$. Assuming normally distributed errors, it follows that

$$Y^* \sim N(g^*(\eta), \sigma^2 I)$$

and hence it follows from (2.16) that the log-likelihood function is given as

$$\ell(\beta, \sigma^2, \lambda; \mathbf{y}) = -\frac{1}{2} \left(n \log(2\pi) + n \log(\sigma^2) + \log(\det(\Lambda)) + \frac{\|\mathbf{y}^* - \mathbf{g}^*(\beta)\|^2}{\sigma^2} \right),$$

where $\mathbf{g}^*(\boldsymbol{\eta})$ is noted $\mathbf{g}^*(\boldsymbol{\beta})$, to allude \mathbf{g}^* 's dependency on $\boldsymbol{\beta}$. By differentiating with respect to σ^2 and equating to zero, the estimate of σ^2 given λ and $\boldsymbol{\beta}$ is

$$\hat{\sigma}^2 = \frac{\|\mathbf{y}^* - \mathbf{g}^*(\boldsymbol{\beta})\|^2}{n}.$$

This gives the following profile log-likelihood function using $\hat{\sigma}^2$

$$\ell(\boldsymbol{\beta}, \lambda; \mathbf{y}) = -\frac{1}{2} \left(n \left(\log \left(\frac{2\pi}{n} \right) + 1 \right) + \log(\det(\Lambda)) + n \log \left(\|\mathbf{y}^* - \mathbf{g}^*(\boldsymbol{\beta})\|^2 \right) \right), \quad (2.22)$$

which should be maximised with respect to $\boldsymbol{\beta}$ and λ to obtain the MLEs of the two parameters. This is achieved by first estimating $\boldsymbol{\beta}$ given an estimate of λ , which from (2.22) can be determined by

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \|\mathbf{y}^* - \mathbf{g}^*(\boldsymbol{\beta})\|^2.$$

This problem can be solved using the Gauss-Newton algorithm described previously. Subsequently, this estimate can then be used to update the estimate of λ . This process iterates until some convergence criterion is satisfied.

This also shows that the estimation process is somewhat less computational expensive compared to that of the non-linear mixed model.

2.7 Variance and Correlation Structures of the Errors

As mentioned, it is possible to include both heteroskedasticity and correlation in the errors in both the mixed models and extended non-linear regression models. Conveniently, it is possible to decompose the covariance matrix of the errors such that the correlation and variance can be modelled separately. The covariance matrix of the errors can be decomposed as

$$\boldsymbol{\Sigma} = \sigma^2 \mathbf{V} \mathbf{C} \mathbf{V},$$

where \mathbf{V} is an $n \times n$ diagonal matrix and \mathbf{C} is an $n \times n$ correlation matrix. Hence,

$$\text{Var}[\varepsilon_j] = \sigma^2 (\mathbf{V}_{j,j})^2 \quad \text{and} \quad \text{Cor}[\varepsilon_j, \varepsilon_k] = \mathbf{C}_{j,k},$$

allowing for the variance and correlation to be modeled separately for $j, k = 1, \dots, n$.

2.7.1 Variance

The variance structure of the errors can be modelled by a variance function as

$$\text{Var}[\varepsilon_j] = \sigma^2 z^2(\mathbf{v}_j, \boldsymbol{\omega}), \quad j = 1, \dots, n.$$

Here, the variance function z is dependent on both the variance covariate vector \mathbf{v}_j and the variance parameters $\boldsymbol{\omega}$. The variance function z is assumed continuous in $\boldsymbol{\omega}$. In R multiple classes of variance functions are defined, each with two main arguments, value and form, for specifying $\boldsymbol{\omega}$ and \mathbf{v}_j , respectively. Two examples of classes of variance functions are `varFixed`

and `varIdent`. `varFixed` only depends on `value`, and can be used when the variance of ε_j is known up to a proportionality constant. For example, the variance could depend on the covariate "age" as

$$\text{Var}[\varepsilon_j] = \sigma^2 \sqrt{\text{age}_j}^2.$$

Thus, the variance function would be the square root of age_j . Contrary, `varIdent` also uses the argument `form`. It is generally used for modelling different variances for some set of stratification variables $s \in \{1, 2, \dots, S\}$ such that

$$\text{Var}[\varepsilon_j] = \sigma^2 \omega_{s_j}^2. \quad (2.23)$$

Here, $\omega_1 = 1$, and ω_l , $l = 2, \dots, S$, represent the ratio between the standard deviation of s_l and s_1 . For example, one can choose `form = ~ 1 | Sex`, indicating that the variance varies depending on the gender in a data set. A more detailed description of these two classes of variance functions and examples of others can be found in [Pinheiro and Bates, 2000, p. 208-214].

2.7.2 Correlation

Correlation structures are often used to describe dependencies between observations in either time-series- or spatial data, indexed by a time variable or spacial location vector. To describe a general correlation structure, assume that ε_j is associated with a position vector \mathbf{p}_j . Assuming that the correlation between ε_j and ε_k only depends on the distance between their corresponding position vectors, $d(\mathbf{p}_j, \mathbf{p}_k)$, the correlation between them is given as

$$\text{Cor}[\varepsilon_j, \varepsilon_k] = c(d(\mathbf{p}_j, \mathbf{p}_k), \boldsymbol{\rho}).$$

Here, $\boldsymbol{\rho}$ represents a correlation parameter vector and c a correlation function. Furthermore, c is defined to take values in the interval $[-1, 1]$, continuous in $\boldsymbol{\rho}$, and $c(0, \boldsymbol{\rho}) = 1$.

Correlation structures can be split into two classes; serial and spatial. This thesis will only include the description of serial correlation structures due to their relevance in this thesis.

Serial correlation structures are used for data observed over time, and in this setup, c is the autocorrelation function. Furthermore, the autocorrelation function only depends on the one-dimensional positions p_j, p_k in this setup through their absolute difference.

As for the variance structures, some of the most commonly used serial- and spatial correlation structures are implemented in R, which have two main arguments, `value` and `form`, for defining the correlation parameters and position vector. Two of the correlation structures that can be implemented are `corCompSymm` and `corSymm`. The complexity of these two structures vary significantly. `corCompSymm` represents a compound symmetry, which means equal correlation between the errors

$$\text{Cor}[\varepsilon_j, \varepsilon_k] = c(i, \boldsymbol{\rho}) = \rho, \quad i = |j - k|, \forall j \neq k, \quad j, k = 1, 2, \dots, n.$$

Contrarily, `corSymm` assumes that all correlations are represented by a different parameter

$$c(i, \boldsymbol{\rho}) = \rho_i, \quad i = |j - k| \quad j, k = 1, 2, \dots, n. \quad (2.24)$$

Whereas `corCompSymm` often assumes a too simple correlation structure, `corSymm` often leads to overparameterisation. Hence, there are advantages and disadvantages to using both. It

is also possible to include autoregressive–moving average correlation structures and, as mentioned, different spatial correlation structures. More information about `corCompSymm`, `corSymm`, and other correlation structures can be found in [Pinheiro and Bates, 2000, p. 227-239].

In summary, this chapter has established a theoretical framework encompassing both linear- and non-linear mixed models. The subsequent chapter will build upon this foundation by introducing the specific models designated for our simulation study, which closely aligns with those presented in this chapter.

3 Modelling the Treatment Effect

This chapter is based upon [Raket, 2022] and [Donohue et al., 2023], unless stated otherwise.

When companies test a new treatment for a specific disease, they are interested in how well the treatment affects the patient, either compared to placebo or another already existing treatment. This effect of a treatment is what is referred to as a *treatment effect*. There are different ways of modelling this treatment effect, such as with a *constrained longitudinal data analysis* (cLDA) model or MMRM models. Here, MMRM refers to a mixed model for repeated measures where specific covariates are included. However, in recent years, another method of measuring the treatment effect has been explored, called *progression models for repeated measures* (PMRMs). This chapter will present the commonly used cLDA model, the PMRMs, and how they differently model the treatment effect. Specifically, it will focus on some of the PMRMs presented in [Raket, 2022]. The structure of the models presented in this chapter resemble extended non-linear regression models for repeated measures with N subjects, as the random effects are not explicitly given. However, as will be explored later, random effects can be added to the models. The models that will be presented are based on the assumption that the subjects are split into two groups: placebo ($i \in \mathcal{I}_{\text{placebo}}$) and active ($i \in \mathcal{I}_{\text{active}}$), where $\mathcal{I}_{\text{placebo}} \cap \mathcal{I}_{\text{active}} = \emptyset$. These models will afterward be extended by assuming that the subjects are split into four groups. Lastly, this thesis focuses on a balanced design, where all subjects have the same amount of visits, $j = 0, 1, \dots, n$. However, the models can be extended to an unbalanced design, see [Donohue et al., 2023]. Here, $j = 0$ refers to baseline, which in a clinical trial is the initial measurement or status of the subject's health or condition just prior to the intervention.

3.1 Constrained Longitudinal Data Analysis Model

The cLDA model and MMRM are quite similar since they are both linear mixed models for repeated measures, and both use visit and the visit-by-treatment as covariates. The MMRM models the change from baseline, and the analysis involves adjusting for baseline values in each treatment group. Contrarily, the cLDA model considers baseline values as a dependent variable, which is equivalent across the treatment groups - the groups are assumed to have same mean baseline. Furthermore, cLDA models are shown to have greater statistical power, compared to the MMRM, given the same sample size, and are also more robust to missing data, and will hence be used [Liu et al., 2009; Lu, 2010].

Considering no explicit random effects, the cLDA model, for subject i , is given as

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \quad (3.1)$$

with design matrix \mathbf{X}_i and fixed effects vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_n, \nu_1, \dots, \nu_n)^\top$. Hence, at post-baseline visit $j = 1, \dots, n$, the model is given as

$$Y_{i,j} = \beta_j \cdot \mathbb{1}_{i \in \mathcal{I}_{\text{placebo}}} + \nu_j \cdot \mathbb{1}_{i \in \mathcal{I}_{\text{active}}} + \varepsilon_{i,j},$$

where $\mathbb{1}_{i \in \mathcal{I}_{\text{active}}}$ equals one if subject i is in the active group and zero otherwise, and similarly $\mathbb{1}_{i \in \mathcal{I}_{\text{placebo}}}$ equals one if subject i is in the placebo group and zero otherwise. Here, the first

term describes the mean of the placebo group while the second term describes the mean of the active group. At baseline $j = 0$

$$Y_{i,0} = \beta_0 + \varepsilon_{i,0},$$

implying a common baseline across treatment groups.

After the model has been fitted, the treatment effect can be determined using several different methods. One method is to determine the difference on the output scale between the mean trajectories of the placebo- and active group at the last visit. Other methods of determining the treatment effect, could include calculating the mean difference area under the curve between the two treatment groups, or use the mean treatment effect over visits.

One of the main differences between the cLDA model and the PMRMs is that the PMRMs use visits as a continuous variable, instead of a categorical variable as the cLDA model, which can increase the statistical power of the model compared to the cLDA model. Therefore, the following section will elaborate on this principle.

3.2 Categorical to Continuous

As mentioned, the cLDA model utilises visits as a categorical variable, and the interim events between visits are typically represented graphically by connecting the data points. The connecting lines are considered a reasonable approximation of interim events. However, visits can also be considered a continuous variable, t , where $t_{i,j}$ represents the time since baseline for subject i 's j th visit, where $t_{i,0} < t_{i,1} < \dots < t_{i,n}$ for $i = 1, \dots, N$. This extends (3.1) to a general non-linear model which for subject i is given as

$$Y_{i,j} = f(t_{i,j}; \mathbf{X}_i \beta) + \varepsilon_{i,j}, \quad (3.2)$$

with design matrix \mathbf{X}_i , fixed effects vector β , and $t_{i,0} = 0$. Here, the shape of f , which maps time to outcome, is determined by $\mathbf{X}_i \beta$. For the cLDA model, $t_{i,j} = t_j$ and the function f is an interpolation of $\mathbf{X}_i \beta$ with t_0, t_1, \dots, t_m as anchor points. When using f in (3.2) for interpolation, the function f describes how subjects behave between visits. Hence, one should have some idea of the subjects' behavior between visits when choosing f , such that it represents this behavior. [Raket, 2022] suggests using a *natural cubic spline* for interpolation in, for example, an Alzheimer's trial, since one would expect it to be continuous and smooth.

As discussed in [Donohue et al., 2023], using visits as a categorical variable has some disadvantages. For example, if a subject does not get their measurement taken at a planned time, it either has to be ignored or in some way carried back. However, using visits as a continuous variable allows subjects to have measurements taken at unscheduled times.

3.3 Progression Models for Repeated Measures

The PMRMs adopt the structure outlined in (3.2) and use visits as a continuous variable. The PMRMs model the mean trajectory of the placebo group by a function, denoted h_0 . This thesis assumes that this function is a natural cubic spline. It is assumed that the natural cubic spline has knots at the scheduled visit times, which is denoted t_j , where $t_{i,j} = t_j$ for all i . The value of the natural cubic spline in each knot is given as the mean of the observations

at the scheduled visits, denoted $\beta = (\beta_0, \beta_{1,\text{placebo}}, \dots, \beta_{n,\text{placebo}})^\top$. Here, β_0 is the mean of all the observations at baseline whereas $\beta_{j,\text{placebo}}$ is the mean of the observations for all the subjects in the placebo group at visit j . An example of this is illustrated in Figure 3.1 with six post-baseline visits.

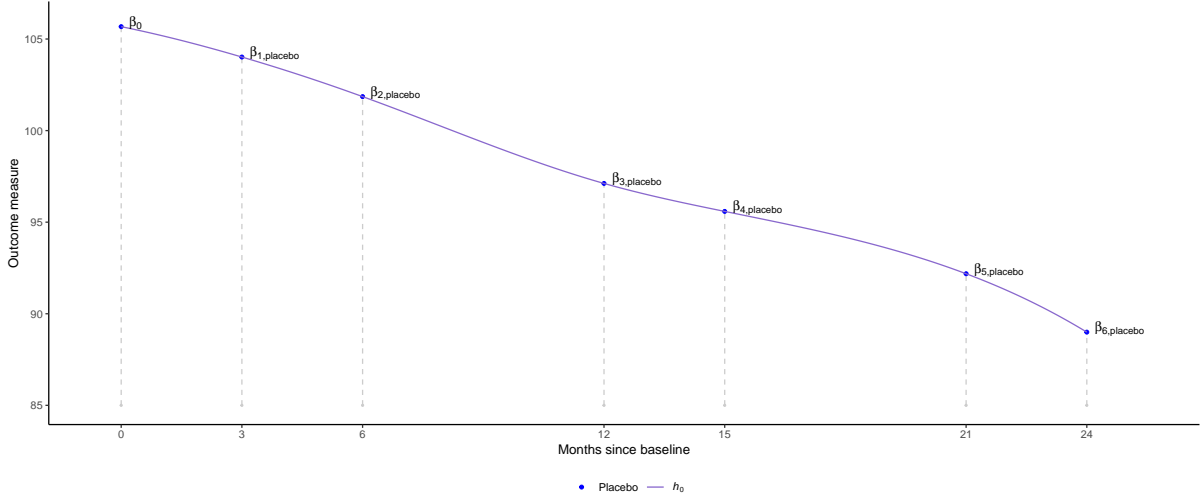


FIGURE 3.1: Mean trajectory of a placebo group visualised using a natural cubic spline, h_0 , with six post-baseline visits used as knot values for the natural cubic spline.

The general PMRM proposed by [Raket, 2022], consists of a function, h , used to describe the relation between the mean trajectory of the active group and h_0 . For subject i at visit j , the model is given as

$$Y_{i,j} = h(t_{i,j}; h_0, \beta, W_{i,j}\zeta) + \varepsilon_{i,j}. \quad (3.3)$$

Here, $W_{i,j}$ is a $1 \times (k+1)$ design matrix, ζ a $(k+1) \times 1$ parameter vector, and $\varepsilon_{i,j}$ the error.

We will focus on three of the PMRMs presented in [Raket, 2022], which will be described in the following subsections. These models will be based on the assumption that the modelled outcome measure is decreasing over time. Furthermore, the treatment is assumed to in some way be disease-modifying as described in Chapter 1.

3.3.1 Proportional Reduction in Decline

Assume that the treatment effect over time is characterised by a proportional decrease in decline. That is, the treatment results in a decrease in the rate of deterioration over time, and this decrease is proportional to the change from baseline observed without treatment. Assuming six post-baseline visits, as in Figure 3.1, the cLDA model can fully model this using 13 parameters.

Assuming that the treatment effect over time is characterised by a proportional decrease in decline a more parsimonious model could be used as this assumption suggests that there exists a parameter ζ , such that

$$\beta_{j,\text{active}} - \beta_0 = \zeta(\beta_{j,\text{placebo}} - \beta_0). \quad (3.4)$$

Here, $\beta_{j,\text{active}}$ is the mean of the observations for all the subjects in the active group at visit j . The relation between the mean trajectories can then also be modelled using a PMRM by choosing

$$h(t_{i,j}; \mathbf{h}_0, \beta, \mathbf{W}_{i,j} \boldsymbol{\zeta}) = \mathbf{W}_{i,j} \boldsymbol{\zeta} (\mathbf{h}_0(t_{i,j}; \beta) - \beta_0) + \beta_0, \quad (3.5)$$

where $\boldsymbol{\zeta} = (1, \zeta)^\top$ and

$$\mathbf{W}_{i,j} = \begin{cases} (1, 0) & i \in \mathcal{I}_{\text{placebo}} \\ (0, 1) & i \in \mathcal{I}_{\text{active}}. \end{cases} \quad (3.6)$$

Hence, the treatment effect is described by including the parameter ζ for subjects in the active group. This model will henceforth be referred to as the *proportional reduction in decline PMRM* (PDPMRM).

Both the cLDA model and PDPMRM models the treatment effect based on how the treatment affects the subjects on the outcome scale. However, with a slow-progressing disease such as Alzheimer's disease, the treatment effect measured on the outcome scale over a period of, for example, 24 months can be very small. This can cause the effect of a treatment to be deemed negligible. Nevertheless, a small difference on the outcome scale can correspond to a substantial difference on the time scale, serving to demonstrate the efficacy of the treatment and provide patients and caregivers with an easier interpretation.

The treatment effect in the PMRMs in the following sections is based on how the treatment affects the subjects on the time scale. Heuristically, the treatment effect then describes the additional time a subject gains being in the active group as opposed to the placebo group. Hence, they model the treatment effect as illustrated with the purple dotted lines in Figure 3.2 instead of vertically.

3.3.2 Time-based Changes in Disease Progression

As opposed to the PDPMRM, which assumes that the treatment effect is constant over time, the PMRM considered here, called the *time based changes in disease progression PMRM* (TPMRM), allows the treatment effect to vary over time. Figure 3.2 illustrates the concept behind this model.

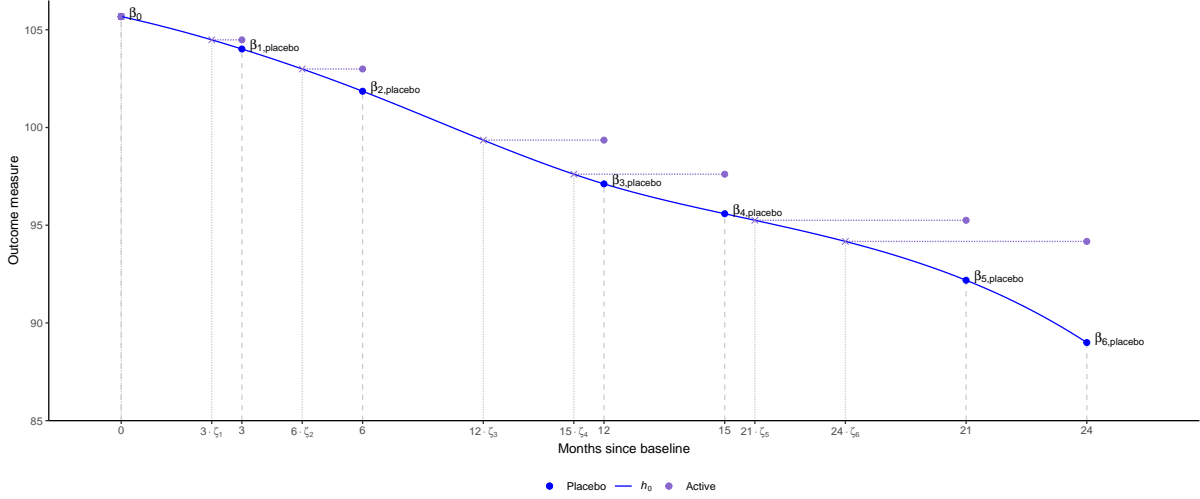


FIGURE 3.2: Mean trajectory of a placebo group visualised using a natural cubic spline, h_0 , and mean trajectory of an active group visualised alongside dotted lines exemplifying the meaning of time saved. The slowing at the six post-baseline visits are $\zeta_1, \zeta_2, \dots, \zeta_6$.

As illustrated in Figure 3.2, the TPMRM assumes that the mean trajectory for the subjects in the active group can be described as the mean trajectory for subjects in the placebo group at another time point. Hence, the model also assumes that the mean trajectory of the active group cannot exhibit improvement or deterioration beyond the bounds observed in the mean trajectory of the placebo group.

The TPMRM is given as

$$h(t_{i,j}; h_0, \beta, W_{i,j} \zeta) = h_0(W_{i,j} \zeta t_{i,j}; \beta), \quad (3.7)$$

where $\zeta = (1, \zeta_1, \dots, \zeta_m)^\top$ and $W_{i,j} = (1, 0, \dots, 0)$ if $i \in \mathcal{I}_{\text{placebo}}$. For $i \in \mathcal{I}_{\text{active}}$, $W_{i,j}$ is a vector of zeroes except for entry $j + 1$, which is one.

Hence, ζ describes the slowing of the disease progression at the different visits for the mean trajectory of the active group compared to that of the placebo group. For example, if $\zeta_6 = 0.75$ as is the case in Figure 3.2, then the active treatment demonstrates a 25% slowing of the disease progression compared to the placebo group. This would indicate a six-month delay in the disease progression at month 24. This can not only demonstrate the efficacy of the treatment, but could also be more tangible for patients than 5 points on an outcome scale.

3.3.3 Proportional Slowing of Disease Progression

As mentioned, the TPMRM assumes that the treatment effect can differ between visits. However, a constant treatment effect between visits can be preferable in some contexts. If a model assumes a constant treatment effect over time, it can simplify the interpretation and analysis of the results. Hence, it could be relevant to test whether the TPMRM is significantly better than a model where the treatment effect is constant over time.

Such a model will be referred to as the *proportional slowing of disease progression PMRM* (PSTPMRM) and is given as the TPMRM, (3.7), where $W_{i,j}$ is given as in (3.6) and $\zeta = (1, \zeta)^\top$. Hence, the PSTPMRM also describes the treatment effect in terms of time saved. However, this treatment effect is considered constant over time.

For progressive diseases, such as Alzheimer's disease, then intuitively $0 \leq \zeta < 1$. This is because $1 < \zeta$ would indicate accelerated disease progression. Additionally, $0 > \zeta$ would indicate subjects in the active group being cured or in some way brought back to a stage possibly before being diagnosed with the disease.

3.3.4 Subgroup Extension

In a clinical trial, it is often analysed how a treatment affects different subgroups in a trial. Here, a subgroup is defined as a subset of the whole population with a specific characteristic. For example, it could be analysed whether there is a difference in the treatment effect between the two subgroups; females and males.

In clinical trials addressing a progressive disease, such as Alzheimer's disease, certain subgroups might progress faster than others. If a treatment effectively slows the disease progression consistently among various subgroups, those characterised by faster progression can derive more significant benefits throughout the duration of the clinical trial. Thus, even if the subgroups have the same treatment effect when measured on the outcome scale, this does not entail that the treatment effect on the time scale is the same due to non-linearity, and vice versa. Generally, the treatment effect in different subgroups can differ both on the outcome- and time scale, and hence models which are able to capture the treatment effect in each subgroup are rather useful.

The three PMRMs presented above can be used to describe the treatment effect between two groups, a placebo- and an active group. They can also be extended such that they are able to accommodate multiple placebo- and active group such that they can determine the treatment effect in each subgroup simultaneously. The simple case of only two subgroups will be considered, and hence the subjects will be split into four groups: two placebo groups ($i \in \mathcal{I}_{\text{placebo1}}$ and $i \in \mathcal{I}_{\text{placebo2}}$) and two active groups ($i \in \mathcal{I}_{\text{active1}}$ and $i \in \mathcal{I}_{\text{active2}}$), where $\mathcal{I}_{\text{placebo1}} \cap \mathcal{I}_{\text{placebo2}} \cap \mathcal{I}_{\text{active1}} \cap \mathcal{I}_{\text{active2}} = \emptyset$. Here, subjects in $\mathcal{I}_{\text{placebo1}}$ or $\mathcal{I}_{\text{active1}}$ are in subgroup 1 and subjects in $\mathcal{I}_{\text{placebo2}}$ or $\mathcal{I}_{\text{active2}}$ are in subgroup 2.

An extension of the PSTPMRM in Subsection 3.3.3, such that it can accommodate two subgroups, can be given as

$$Y_{i,j} = \begin{cases} h_0(t_{i,j}; \beta) + \varepsilon_{i,j} & i \in \mathcal{I}_{\text{placebo1}} \\ h_0(\rho t_{i,j}; \beta) + \varepsilon_{i,j} & i \in \mathcal{I}_{\text{placebo2}} \\ h_0(\zeta t_{i,j}; \beta) + \varepsilon_{i,j} & i \in \mathcal{I}_{\text{active1}} \\ h_0(\rho \gamma t_{i,j}; \beta) + \varepsilon_{i,j} & i \in \mathcal{I}_{\text{active2}} \end{cases} \quad (3.8)$$

Here, we assume that the placebo group in subgroup 2 can be derived from the placebo group in subgroup 1. Specifically, it assumes that the relationship between the two is described as a proportional slowing in disease progression. For the cLDA model to accommodate two subgroups, it would require an additional 12 parameters, resulting in a total of 25 parameters, for six post-baseline visit example. However, compared to the extended PSTPMRM, this model does assume any specific relation between the two placebo groups.

The extended PSTPMRM in (3.8) can be extended further such that it does not assume the specific relation between the two placebo groups, at the cost of additional parameters. However, this thesis will only focus on this simple extension of the PSTPMRM. An extension

of the two other progression models, PDPMRM and TPMRM, can be done in a similar way as for the PSTPMRM, but will not be presented in this thesis.

4 Data Analysis

This chapter will describe and analyse the data used in the remainder of this thesis. As described in the Preface, the data is provided by Novo Nordisk A/S but originates from the Critical Path for Alzheimer's Disease (CPAD). However, as mentioned in the Preface, the CPAD did not participate in the analysis or the writing of this thesis.

The CPAD database is a collection of data from multiple different clinical trials and consists of multiple smaller data sets, each including different classes of data for the subjects in the clinical trials. These data sets include, among others, *ae* (adverse events), *qs* (cognitive assessments), *dm* (demographics), *lb* (laboratory data), *mh* (medical history), and *vs* (vital signs). Some of these data sets will be examined in the following section as we will look at which subjects should be included in the subsequent analysis.

The main focus of this thesis is to compare how well different models describe the treatment effect of an active treatment compared to placebo. Ideally, this analysis should be based on data that mirrors the conditions of a true clinical trial for Alzheimer's disease, which is why the CPAD database is used. The purpose of this chapter is to sort the data such that there are subjects on both active treatment and placebo. From this, it should be possible to model the treatment effect using some of the models previously presented in Chapter 3. Furthermore, we aim to obtain a set of subjects from the same trial. Having subjects from the same trial is preferable due to the significant evolution of care and treatments within the Alzheimer's area. For example, it is only in recent research that it has been shown that Alzheimer's subjects can compensate to some degree for the cognitive decline by receiving more cognitive support, changing the standard of care for Alzheimer's subjects, [Basun et al., 2006]. Therefore, subjects from different trials can evolve differently, not only because of the specific treatment we wish to model the treatment effect of, but also due to variations in the standard of care and treatments available at the time. The CPAD database does, however, not include the identity of individual studies. Thus, it is generally not possible to know if two subjects have been included in the same trial.

4.1 Data Presentation

In this section, there will be a brief presentation of the overall data to give the reader an idea of which populations are generally included in clinical trials for Alzheimer's disease. Additionally, some of the methods used to diagnose Alzheimer's disease and measure disease progression will be presented.

4.1.1 Demographics

This subsection focuses on the *dm* data set, which describes the demographics of the subjects. This includes, among other things, the age, sex, and ethnicity of the subjects. The *dm* data set consists of 14895 subjects, 8769 (59%) of whom are female and 6126 (41%) of whom are male.

Age

Alzheimer's disease predominantly affects the older population, which we expect to be reflected in the CPAD database. It is evident in Figure 4.1, which shows a histogram of the subjects' age, that the majority of the subjects are older than 70 years old. Figure 4.1 further shows that the youngest subject is 45 years old and the oldest is 89 years old.

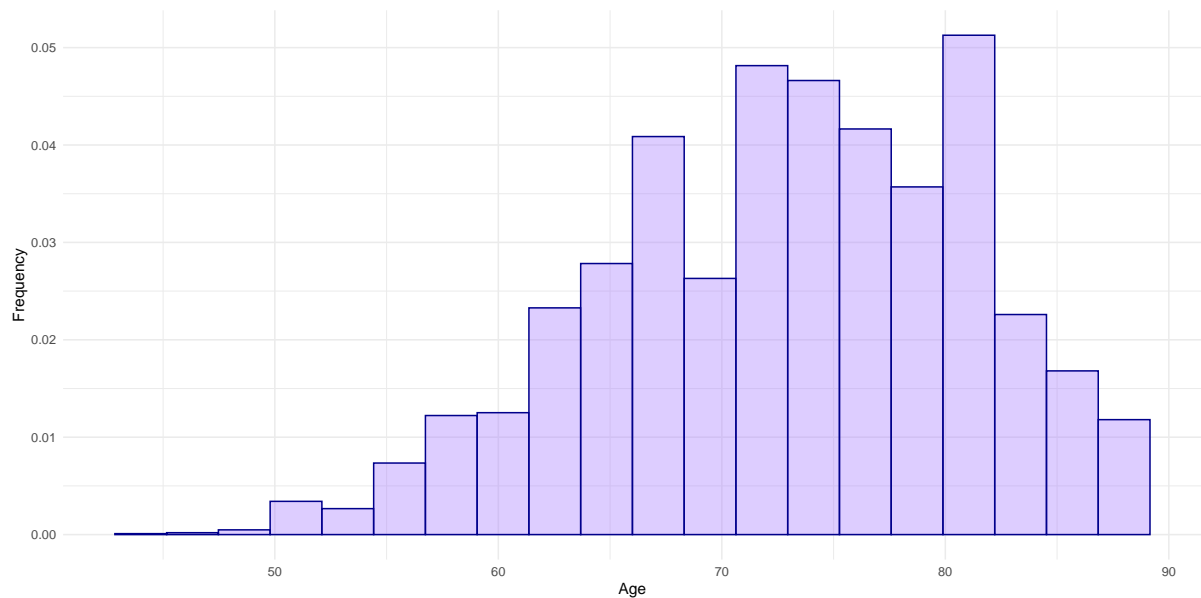


FIGURE 4.1: *Histogram of the ages available in the CPAD database, excluding subjects with ages recorded as 999 or 9999.*

Figure 4.1 excludes some subjects whose ages are recorded as 999 or 9999. These subjects most likely come from countries and/or clinical trials where the age is not recorded or has not been shared with the CPAD database. Although these subjects are excluded from this histogram, they are still included in the subsequent data analysis.

Race, Ethnicity, and Country

For the trials to be representative of the whole population, they must include enough subjects from various racial- and ethnic groups of the population and various countries all around the world. The proportion of subjects within each racial- and ethnic group can be seen in Table 4.2 and Table 4.3, respectively.

Race	%
Not Registered	0.73
American Indian or Alaska Native	0.09
Asian	5.29
Black or African American	1.89
Multiple	0.05
Native Hawaiian or other Pacific Islander	0.05
Other	1.99
Unknown	0.11
White	89.78

TABLE 4.2: *The available racial groups and the percentage of subjects within each racial group.*

As seen in Table 4.2, the majority of the subjects are white, whereas Table 4.3 shows that ethnicity is not registered for the majority of the subjects.

Furthermore, Figure 4.4 illustrates that the majority of subjects are from the USA with 5513 subjects. However, it can be seen that, overall, subjects from most parts of the world are included, though only few are from the African countries.

Ethnicity	%
Not Registered	74.37
Hispanic or Latino	2.45
Not Hispanic or Latino	23.15
Unknown	0.03

TABLE 4.3: *The available ethnic groups and the percentage of subjects within each ethnic group.*

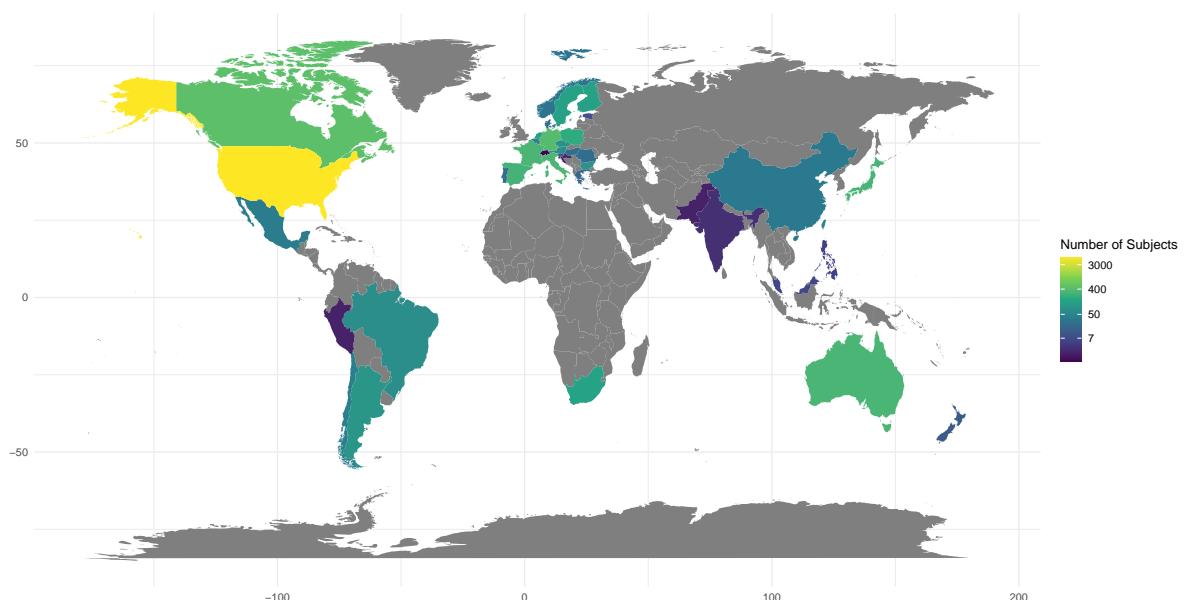


FIGURE 4.4: *World map showing the distribution of subjects. This shows that the majority of countries in North America, Europe, and Oceania are represented, whereas only few countries in South America, Africa and Asia are represented.*

Additionally, it should be noted that 4541 subjects do not have a country registered, why they are not presented in Figure 4.4.

Hence, as expected, the population in the different clinical trials consists mostly of older subjects, equally distributed between men and women, and somewhat represents a broad set of both racial and ethnic groups. As mentioned earlier, the inclusion of many different races, ethnic groups, and countries in a trial is crucial for representing a broad part of the population. This diversity is especially important when seeking approvals for the medicine in a trial, as it can be difficult for a drug to be approved in regions like Asia if no subjects from Asia are included in the clinical trial [National Medical Products Administration, 2020]. However, we are not obligated to have a broad amount of population groups since we are not conducting a formal analysis that requires FDA approval.

4.1.2 Treatments

As mentioned previously, the CPAD database is composed of different clinical trials, meaning that there are multiple different treatments included in the database. When two subjects are on different treatments it is referred to as the subjects being on different treatment arms. The different treatment arms in the CPAD database and the number of subjects within each of these are as follows

- | | | |
|--------------------|----------------------|------------------------|
| • PLACEBO (8543) | • ACTIVE ARM 1 (336) | • DRUG 1 (145) |
| • PLACEBO QD (208) | • ACTIVE ARM 2 (592) | • DRUG 1 DOSE 1 (290) |
| • PBO (963) | • ACTIVE ARM 3 (38) | • DRUG 1 DOSE 2 (597) |
| • CONTROL (566) | • ACTIVE ARM 4 (23) | • DRUG 1 DOSE 3 (2138) |

as well as 455 subjects that have no treatment registered and one subject whose treatment is categorised as BLIND. The first column consists of the placebo arms from the clinical trials, while the other two columns consist of the active arms. As present from the above, the majority of subjects received placebo. Consequently, we have to be attentive not to exclude all subjects receiving an active treatment, when choosing which data set to conduct the data analysis on.

However, we need not only a data set including both subjects in an active and a placebo arm, but we also need to make sure that the subjects have some measure recorded which shows the disease progression. This is needed to analyse the treatment effect. There are two types of measures in the data set used jointly to diagnose Alzheimer's disease: biomarkers and cognitive assessments, [Basun et al., 2006, p. 92]. In the 1b data set, the laboratory test of the different subjects can be seen. Five tests of interest are included: amyloid beta 1-40 ($A\beta$ 1-40), amyloid beta 1-42 ($A\beta$ 1-42), modified amyloid beta 1-40, modified amyloid beta 1-42, and Phosphorylated Tau Protein. These tests are of interest as amyloid plaques and neurofibrillary tangles (see Chapter 1) are believed to be connected to dementia and Alzheimer's disease. Although these biomarkers are of great interest within the Alzheimer's area, these are not yet used to describe disease progression. For this purpose, cognitive assessment scales can be used.

4.1.3 Cognitive Assessment Scales

This subsection focuses on the `qs` data set, which includes patients' scores from different cognitive assessments. Since the CPAD database is composed of different clinical trials, the subjects' disease progression herein are not necessarily evaluated on the same cognitive assessment scale. Below are some of the most commonly used cognitive assessment scales found in the CPAD database, along with the number of subjects for whom these cognitive assessment scales are reported.

- ADAS-Cog (4197)
- ADAS-Cog 14 (2092)
- ADCS-ADL (2092)
- ADAS-Cog 11 (2092)
- CDR-SB (3667)
- iADRS (1068)
- ADAS-Cog 12 (2092)
- MMSE (9247)

An important aspect, which will also become relevant later, is determining a common cognitive assessment scale on which to measure the subjects' cognitive progression. Hence, a brief description of these scales will be presented, based upon [Nationalt Videnscenter for Demens, 2021].

All these scales are based upon tests and questionnaires covering various domains, such as memory, orientation, language, and understanding. All the scores fall within a discrete numerical range from zero and a maximal value which varies between scales. It is important to highlight that the evolution of the measured score in all these cognitive assessment scales for subjects with Alzheimer's disease is minimal within a six-month time frame. Therefore, it is advisable that subjects are measured over an extended period of time, to show significant changes in the disease progression.

The original Alzheimer's Disease Assessment Scale – Cognitive Subscale (ADAS-Cog), represented in the CPAD database as ADAS-Cog or ADAS-Cog 11, has a maximum score of 70. ADAS-Cog 12 and ADAS-Cog 14 are slightly modified versions of ADAS-Cog with different maximum scores. The maximum scores for the Clinical Dementia Rating Sum of Boxes (CDR-SB) scale, Mini-Mental State Examination (MMSE) and Alzheimer's Disease Cooperative Study - Activities of Daily Living Inventory (ADCS-ADL) are 18, 30, and 78, respectively. The integrated Alzheimer's Disease Rating Scale (iADRS), a combination of ADAS-Cog and Alzheimer's Disease Cooperative Study - Instrumental Activities of Daily Living (ADCS-iADL), has a maximum score of 146. Additionally, the interpretation of scores vary, as a higher score does not consistently signify more severe Alzheimer's disease. For the different ADAS-Cog and the CDR-SB scale, a higher score indicates more severe Alzheimer's disease, whereas the opposite applies for the remaining scores.

4.2 Data Set Selection

Now that a short introduction of the subjects included in the CPAD database and some of the cognitive assessment scales herein have been presented, we want to select the data that should be used in the subsequent analysis. As previously mentioned, we need some way to measure the disease progression for the subjects to assess the treatment effect. However, not all of the subjects in the CPAD database are measured on the same cognitive assessment scales (as presented in Subsection 4.1.3).

We aim to include both a placebo arm and an active arm in the data analysis to assess the treatment effect. Hence, we start by examining which treatments are represented when looking at different cognitive assessment scales.

Looking at the subjects for which an ADAS-Cog score has been recorded, both placebo- and active arm subjects are included, as shown in Figure 4.5. However, the longest period of time a subject within an active arm has been recorded is 8 months, with most being recorded for 7 months or less. Hence, due to the short observation period, the disease progression of these subjects is expected to be minimal. Therefore, the only arm that could potentially show any progression measured on the ADAS-Cog seems to be the PLACEBO QD arm.

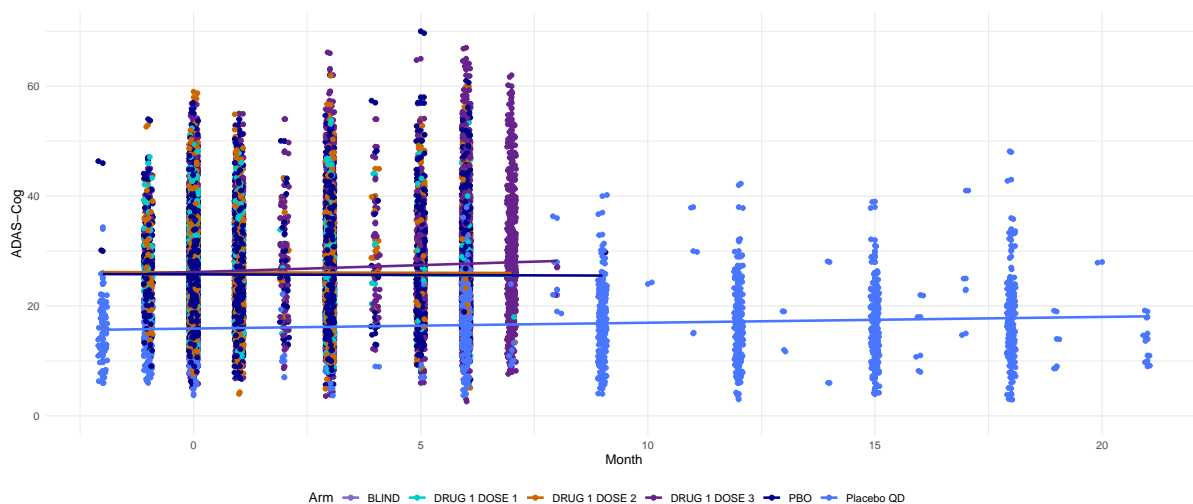


FIGURE 4.5: Data points for subjects recorded with ADAS-Cog scores, alongside a linear mean trajectory of each arm.

ADAS-Cog 11, 12 and 14 only include subjects from the PLACEBO arm over a time period of 40 months (plots for these can be found in Appendix B). The same pattern is observed for ADCS-ADL and iADRS, as illustrated in Figure 4.6 and Figure 4.7, respectively, where the PLACEBO arm, again, is the only represented arm.

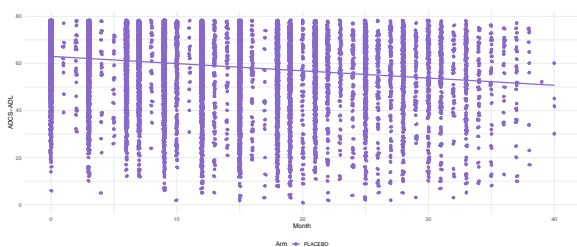


FIGURE 4.6: Data points for subjects recorded with ADCS-ADL scores, alongside a linear mean trajectory.

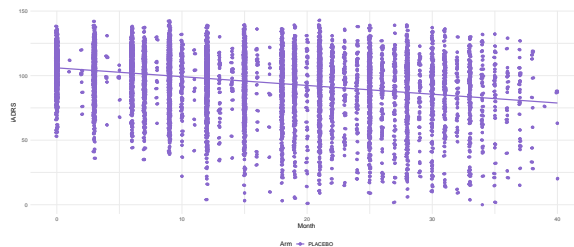


FIGURE 4.7: Data points for subjects with iADRS scores, alongside a linear mean trajectory.

In Figure 4.8 and Figure 4.9, the subjects with recorded CDR-SB and MMSE scores are presented, respectively. Again, both placebo- and active arms subjects are observed. However, as before, the active arms do not seem to be useful, since they are either recorded over a

very short period or do not show any signs of disease progression. Subjects showing no signs of disease progression cannot be used, since this is an unrealistic evolution of a subject with Alzheimer's disease regardless of the treatment, and is hence not representative for the population which we seek to analyse.

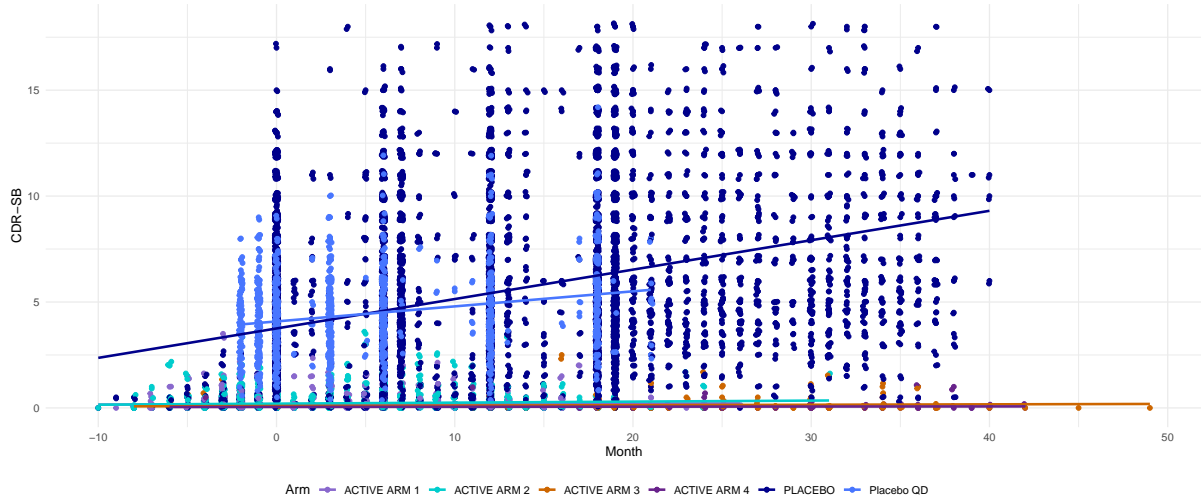


FIGURE 4.8: Data points for subjects recorded with CDR-SB scores, alongside a linear mean trajectory of each arm.

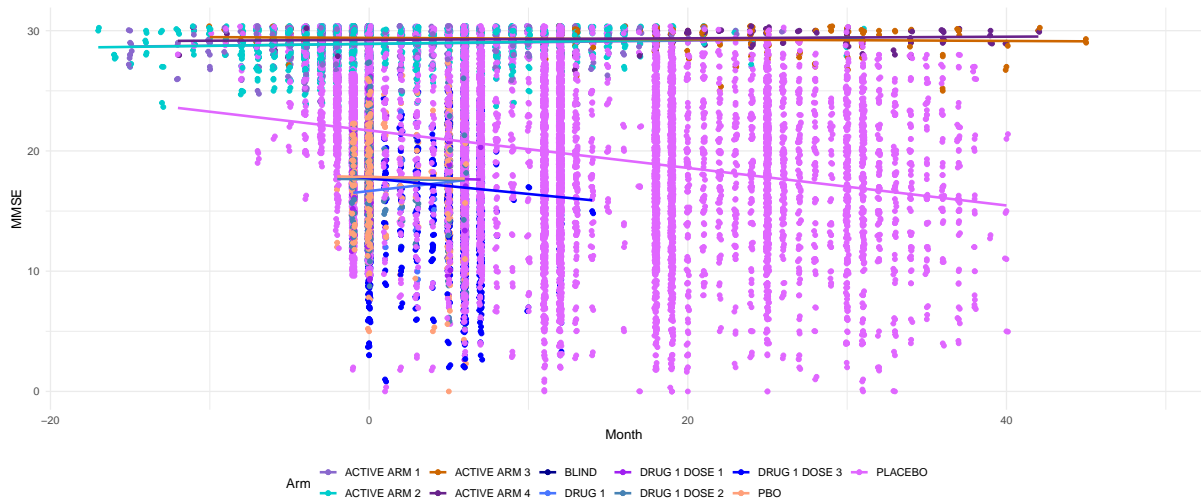


FIGURE 4.9: Data points for subjects recorded with MMSE scores, alongside a linear mean trajectory of each arm

In both Figure 4.8 and Figure 4.9, it is evident that the PLACEBO arm is, as in the plots for ADAS-Cog 11, 12, and 14, clearly the most represented.

From the above analysis, no active arm can be utilised for assessing the treatment effect. This is due to the fact that the active arms are either recorded for a very short period of time or only include subjects that do not show any signs of disease progression. To compare how well different models describe the treatment effect of an active treatment compared to placebo, we need both a placebo- and an active arm. However, as no active arm in the CPAD database

can be utilised, we choose to simulate the active arm instead. This is of course not optimal but necessary. This will be further elaborated upon after choosing the placebo subjects from the CPAD database that we will use.

Now that we do not need to base the selection of data on whether it includes both an active- and a placebo arm, we can consider other important factors. As mentioned in the beginning of this chapter, it would be optimal to only include subjects from the same trial due to changes in standard care, standard treatment, and more for subjects with Alzheimer's disease over time. Furthermore, the models that will be used to model the progression of the disease assumes that the outcome measure follows a normal distribution at each visit. Here, the score of the outcome measure is the score of the chosen cognitive assessment scale. Hence, if the outcome score does not fulfill this assumption, it can affect the models' estimation methods.

Based upon [Honig et al., 2018], we deduce that all subjects with recorded iADRS scores originate from the same trial. This trial involves the treatment named Solanezumab intended for subjects with mild dementia due to Alzheimer's disease. As we derive that this might be the best possibility of only including subject from one specific trial, we choose to only include subjects with recorded iADRS scores. Therefore, our data will include 1068 subjects from the PLACEBO arm, all having recorded scores for the cognitive assessment scales: MMSE, CDR-SB, ADCS-ADL, iADRS, and ADAS-Cog 11, 12, and 14. For the CDR-SB and MMSE scale we assess the number of observations per subject to be inadequate for modelling. For all other scales, there are sufficient observations per subject, with seven to fourteen observations per subject. Before choosing which scale to use for measuring the disease progression we still wish to determine whether the assumption of normality is fulfilled for any of the cognitive assessment scales. Figure 4.10 illustrates the distribution of the iADRS scores at months 3, 6, ..., 21, and 25. These months are chosen as they have the most observations. The histograms illustrating the distribution of the remaining scores can be found in Appendix B.

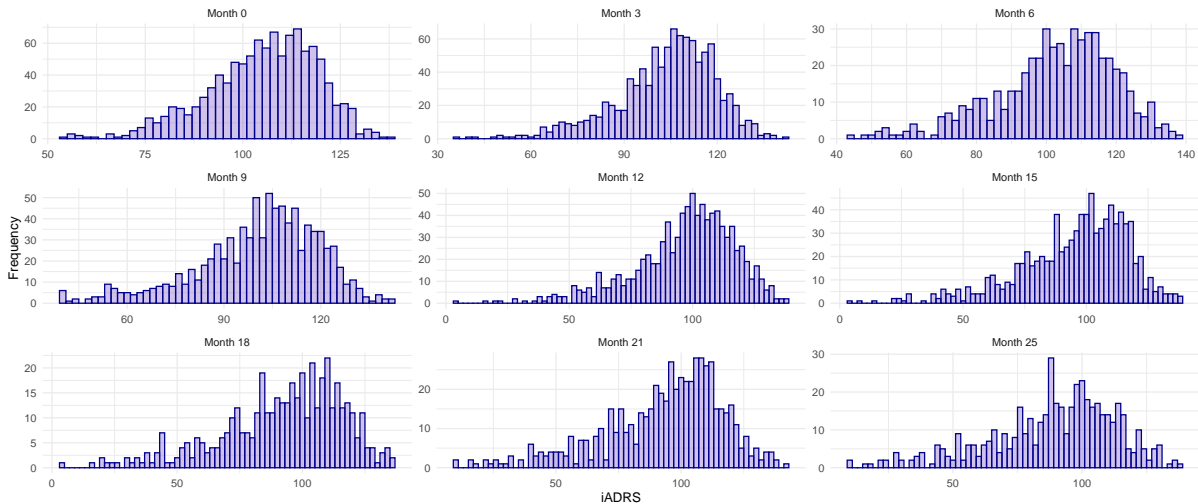


FIGURE 4.10: *Distribution of the iADRS scores at months 3, 6, ..., 21, and 25, as these are the months with the most observations.*

Most scores from the different scales exhibit a distribution that somewhat resembles a normal distribution, but are slightly skewed. However, the ADCS-ADL score is notably left skewed. We opt to utilise the iADRS as our scale for tracking disease progression due to its number of observations, it is being the primary outcome measure in the trial the measurements are

from [Wessels et al., 2021], and it somewhat fulfills the assumption of normality in the first 25 months. Furthermore, iADRS is a combination of two of the other measures, [Honig et al., 2018], more precisely

$$\text{iADRS} = ((-\text{ADAS-Cog13}) + 85) + \text{ADCS-iADL} = ((-\text{ADAS-Cog14}) + 90) + \text{ADCS-iADL},$$

and unlike the others, it has not been used extensively for screening or diagnosing Alzheimer's disease but has been used in recent clinical trials for analysing disease progression.

The subjects within this data are between the age of 55 and 89, with approximately 80% being white, non-Hispanic, or Latino subjects. Furthermore, we observe that approximately half of them are from the USA, and 58% are females. More information regarding the specific trial and the subjects involved is given in Appendix B and [Honig et al., 2018].

The time since baseline and corresponding iADRS score for all the subjects are presented in Figure 4.7, covering a period of 40 months. We observe that the majority of measurements were taken approximately every three months. To streamline our dataset, we have aggregated the measurements to the nearest third month; 0, 3, 6, 9, and so on until month 39. Figure 4.11 presents the iADRS scores when the time points have been aggregated.

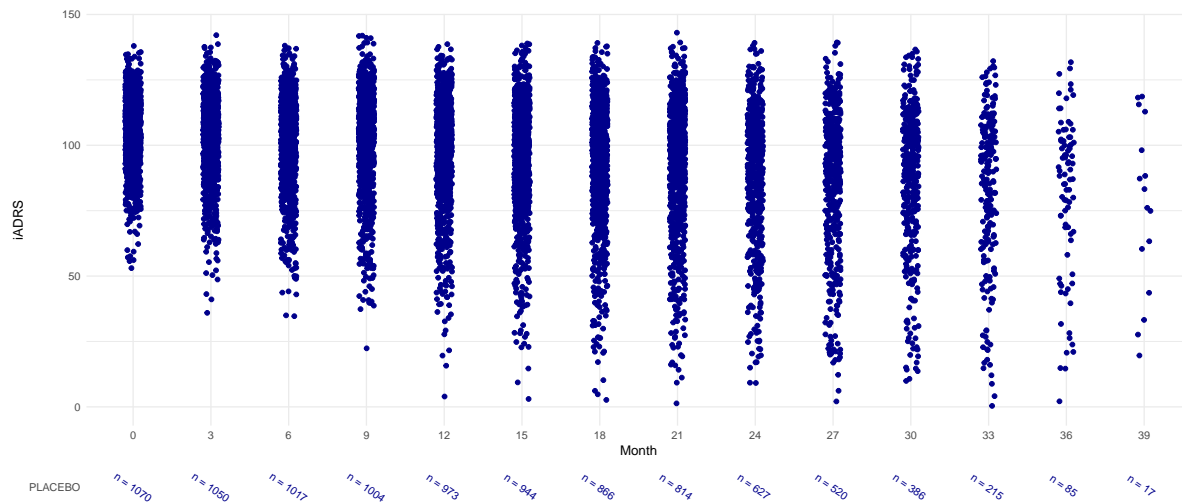


FIGURE 4.11: Measurements of the iADRS score aggregated to the nearest third month (0, 3, 6, 9, and so on until month 39). The data points have been jittered on the time axis to more accurately visualise how the data points are distributed on the iADRS scale. Additionally, labels describing the number of available data points at each third month can be seen below.

By aggregating the data to each third month we will simplify the process of simulating data later on, as well as making the implementation of the PMRMs easier. Every third month can further correspond to a visit which can be used in the cLDA model. Moreover, as can be seen in Figure 4.11, over half of the subjects have dropped out after 24 months. Therefore, we choose to focus the analysis on the data points within the time interval [0, 24].

This concludes our analysis and selection of the data, why we will henceforth be working with the aggregated data presented in Figure 4.11 within the time interval [0, 24], including eight post-baseline visits.

4.3 Simulation of the Active Arm

As concluded in the former section, there are no usable active arms in the CPAD database, why we need to simulate one, as it is necessary for assessing the treatment effect. The drawbacks of simulating an active arm will be discussed in Chapter 7.

This simulation will be highly based on the behavior of the placebo arm presented in Figure 4.11. More precisely, the mean and the covariance structure between visits in the active arm will be based on that of the placebo arm. Hence, we will first present the derived mean and covariance structure of the placebo arm. Both the mean and covariance structure are based upon the aggregated data, in the time interval $[0, 24]$, where each third month is regarded as a visit.

For the nine (eight post-baseline) visits the mean of the placebo group, denoted as $\beta = (\beta_{0,\text{placebo}}, \beta_{1,\text{placebo}}, \dots, \beta_{8,\text{placebo}})^\top$, is given as

$$[105.69 \ 104.03 \ 101.82 \ 100.79 \ 97.06 \ 95.57 \ 93.08 \ 91.99 \ 88.97]. \quad (4.1)$$

The covariance between the visits has been estimated by fitting a linear mixed model to the aggregated placebo data. This model is on the form $y \sim \text{visit} + 0 + (\text{visit} + 0 \mid \text{id})$, where visit is a categorical variable. This means that it has a fixed effect, visit, and no intercept, that is, it is forced to pass through the origin. Additionally, the last term, $(\text{visit} + 0 \mid \text{id})$, specifies random slopes for visit for each subject, with no random intercept. Based on this model, the covariance between the nine visits are given as

	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9
Visit 1	194.18	192.99	207.27	223.24	234.92	247.69	261.44	273.20	280.98
Visit 2	192.99	251.36	251.44	275.54	289.31	310.48	324.80	341.88	350.69
Visit 3	207.27	251.44	312.81	322.68	340.11	367.39	390.46	411.62	422.56
Visit 4	223.24	275.54	322.68	393.40	392.02	424.64	453.31	484.33	492.94
Visit 5	234.92	289.31	340.11	392.02	449.67	466.87	499.06	532.44	550.16
Visit 6	247.69	310.48	367.39	424.64	466.87	543.81	559.26	599.43	624.49
Visit 7	261.44	324.80	390.46	453.31	499.06	559.26	653.34	682.73	710.12
Visit 8	273.20	341.88	411.62	484.33	532.44	599.43	682.73	780.33	792.48
Visit 9	280.98	350.69	422.56	492.94	550.16	624.49	710.12	792.48	879.54

(4.2)

We will, based on this, simulate the active arm in three different ways, corresponding to the three different progression models presented in Section 3.3: proportional reduction in decline PMRM (PDPMRM), time-based changes in disease progression PMRM (TPMRM), and proportional slowing of disease progression PMRM (PSTPMRM).

The mean of the active arm, based on the PDPMRM, at visit $j > 0$, is calculated as

$$\beta_{j,\text{active}} = \zeta (\beta_{j,\text{placebo}} - \beta_{0,\text{placebo}}) + \beta_{0,\text{placebo}}.$$

Here, $\zeta = 0.8$ would for example represent a 20% proportional reduction in decline of the disease progression. The mean of the active arm, based on either the PSTPMRM or the TPMRM, at visit $j > 0$, is calculated as

$$\beta_{j,\text{active}} = \beta_{j \cdot \zeta, \text{placebo}}.$$

Here, $\zeta_j = 0.8$ for $j = 1, 2, \dots, 8$ would represent 20% proportional slowing of the disease progression. Furthermore, $\zeta_j = 0.9 - 0.02857(j - 1)$ for $j = 1, 2, \dots, 8$ would represent a slowing in disease progression changing over visits from 10% at the first visit to 30% at the eighth post-baseline visit. To find $\beta_{j \cdot \zeta_j, \text{placebo}}$, a linear interpolation between $\beta_{j-1, \text{placebo}}$ and $\beta_{j, \text{placebo}}$ is performed using the function `approx`. Furthermore, it is assumed that $\beta_{0, \text{placebo}} = \beta_{0, \text{active}}$, which indicates that the placebo- and active arm have the same mean baseline.

In addition to these three methods, we will also simulate an active arm where the mean of the placebo- and active arm are identical, that is for all j

$$\beta_{j, \text{active}} = \beta_{j, \text{placebo}}.$$

After constructing the mean of the active arms, we use this along with covariance matrix, (4.2), to generate subject-level trajectories from a multivariate normal distribution.

5 Simulation Study

When conducting a clinical trial, one is interested in whether the treatment effect is statistically significant, which relates to the term *statistical power*. Statistical power refers to the probability of correctly rejecting a null hypothesis in a statistical test. In our case, the null hypothesis is that there is no treatment effect. Heuristically, in our case, the statistical power of a statistical test is the ability to detect a treatment effect between the placebo- and active arm when one exists. Hence, it is of interest that the statistical power is as high as possible under a wrong null hypothesis. Different factors can affect the statistical power, such as the *level of significance*. The level of significance is the probability of rejecting a true null hypothesis, and thereby the probability of concluding that there is a treatment effect when non exists. This is referred to as a type I error, and the level of significance, also referred to as the type I error rate, is typically set to 5%. Furthermore, under a true null hypothesis the statistical power should be at the same level as the type I error rate. Furthermore, both the type I error rate and the statistical power relate to the term type II error. A type II error rate is the probability of accepting a wrong null hypothesis. Hence, the statistical power is also given as 1 minus the type II error rate. If the type I error rate increases, the type II error rate will decrease, causing the statistical power to increase. Even though increasing the type I error rate could increase the statistical power and hence decrease the probability of a type II error, it is usually required to have a relatively low type I error rate, as it is of great interest to ensure that a test does not indicate a significant treatment effect when non exists. Another factor that affects the statistical power is the true treatment effect in the data. As the true treatment effect increases, so will the statistical power in general. Moreover, the statistical power is influenced both by the sample size and the standard deviation of the samples. Generally, if either the sample size is increased or the standard deviation of the samples is decreased, statistical power increases [Norton and Strube, 2001].

In practice, statistical power can be determined by simulating multiple data sets; we opt for 1000. A model is then fitted to each data set, and a statistical test, such as an F-test, is then used to determine if the estimated treatment effect is significant. The statistical power is then approximated by the ratio of the number of times the null hypothesis is rejected over the number of simulations. When referring to the statistical power of a model henceforth, it is the statistical power of the statistical test being used. Besides the models' statistical power, their accuracy in estimating the treatment effect is also important. Thus, in this chapter, the comparison of the models from Chapter 3 will focus on statistical power and estimation accuracy, henceforth referred to as the performance of the models.

Data

To compare the performance of the models, we will simulate data using different relations between the placebo- and active arm. The active arm will be simulated using the four methods presented in Section 4.3, providing insight into the comparative performance of the different models and their applicability in these different scenarios. The models will more precisely be analysed in the following scenarios: no effect (NE), 20% proportional slowing (PS), 20% proportional reduction in decline (PD), and 10 to 30% time based changes (TB). An important remark is that, as the scenarios are simulated based on a certain model

specification, we would expect the PDPMRM to estimate the treatment effect well in the 20% PD scenario, and so forth. Ideally, we would test more scenarios to test the behaviour and applicability of the models in these scenarios as well. However, due to the time limitations of this thesis, we opted not to conduct an analysis for additional scenarios.

For each scenario, 1000 data sets will be simulated based on the characteristics of the placebo- and active arms corresponding to that scenario. During testing, we observed that the statistical power and mean estimated treatment effects of the models stabilised within these 1000 simulations; hence, we determined 1000 simulations to be adequate. A figure of the statistical power for all models with respect to the number of simulations, in each scenario, can be seen in Section D.3. Similarly, a figure of the models' mean estimated treatment effect with respect to the number of simulations, in each scenario, can be seen in Section D.4.

We have chosen to simulate the placebo arm based on the placebo arm from the CPAD data, illustrated in Figure 4.11, allowing the placebo arm to behave slightly different in each simulation. The simulation of the placebo arm will be done as described in Section 4.3 using (4.1) and (4.2). Simulating both arms ensures an equal number of subjects in each arm and allows us to determine how the number of subjects in each arm affects the models and the estimation methods. As will be further discussed in Chapter 7, the distribution of the outcome measure in the simulated data differs slightly from that of the CPAD data, which is slightly skewed (see Figure 4.10) compared to the multivariate normal distribution used for our simulations.

The results for the different scenarios will be presented for 500 subjects in each arm, with results for 300 and 1000 subjects in each arm available in Appendix D. As the number of subjects in each arm increases, the mean of the simulated trajectories in each arm should deviate less between simulations. Consequently, the treatment effect in each simulation should not vary as much as it might with fewer subjects in each arm.

The simulated data sets will have a dropout rate of 0, meaning that none of the subjects leave the simulated studies. A dropout rate greater than 0 would require a much more extensive analysis, as many different types of dropout and dropout rates could be considered. Therefore, we have chosen a dropout rate of 0, but the subject of dropout will be briefly discussed in Chapter 7.

The percentages of delaying the disease progression between the placebo- and active arm in the different scenarios closely align with those seen in clinical trials regarding disease-modifying treatments for Alzheimer's disease [Newton, 2023; Lilly, 2023; Wessels et al., 2021]. For example, lecanemab showed a 27% slowing of disease progression after 78 weeks. Similar results have also been observed for donanemab at 32% slowing of disease progression after 76 weeks. In both of these trials the treatment effect was deemed clinically relevant. Furthermore, the covariance matrix used to simulate the data is derived from a placebo arm where we know that all of the subjects are from the same clinical trial. As a result, the simulated data for this study closely resembles data from clinical trials. Though, one thing that makes the simulated data deviate from real data from a clinical trial is that the dropout rate is 0.

Initial Values

Before fitting the PMRMs to the simulated data, initial values must be determined. The initial values required for the PMRMs are the mean values of the placebo arm at each visit. Additionally, an initial value for the treatment effect(s) should be provided and is set to 0 in

all three PMRMs. Initial values for the correlation structure of the errors were also found to be beneficial in ensuring convergence. Hence, determining initial values for the mean trajectory of the placebo arm and correlation structure of the errors is crucial. It is generally a complex problem to determine reasonable initial values for non-linear models. This thesis will utilise a method for determining initial values presented in [Demidenko, 2013, p. xxvi]. This approach involves initiating values based on estimates from a simpler model. We choose to use the cLDA model as the simpler model. The cLDA model provides estimates for the mean trajectory of the placebo arm alongside an estimate of the covariance structure between visits, which we will use as initial values.

This concludes the setup required for us to start our analysis of the models presented in Chapter 3, as we have established both a placebo arm and multiple active arms to model treatment effects upon, as well as initial values for the PMRMs.

5.1 Performance Analysis

The models analysed in the following sections are those presented in Section 3.1 and Section 3.3. Specifically, these include the cLDA model and the three progression models for repeated measure (PMRMs); proportional reduction in decline PMRM (PDPMRM), time-based changes in disease progression PMRM (TPMRM), and proportional slowing of disease progression PMRM (PSTPMRM). The mean trajectory of the placebo arm, h_0 in (3.3), will be modelled by a natural cubic spline for all of the PMRMs considered in this section. For further information about natural cubic splines, see Section A.2. In Appendix C, we test different structures of both the correlation and variance of the errors in one of the PMRMs. Here, we find that, based on the AIC and BIC, an unstructured symmetric correlation structure and the possibility of heterogeneity of variance are the best structures. Hence, we opt for these structures in the correlation and variance of the errors in the PMRMs. The cLDA model that will be used includes an interaction term between treatment and time from baseline, explaining how the response variable is influenced by a treatment over time. Besides the interaction term, the cLDA model includes a random effect term, which allows for a subject-specific variation at each visit.

The performance of the models will be analysed in stages. First, we will compare the performance of the models in different scenarios to see how they are affected by the way the data is simulated. Second, we will analyse how the performance is affected by removing the correlation in the errors. Last, we will examine how the addition of a random effect in the PMRMs affects their performance.

5.1.1 Scenarios

As mentioned, we are interested in how well the models estimate the treatment effect present in the data. In both the TPMRM and the cLDA model, the treatment effect can vary over time and can be analysed using different methods. Although the estimated treatment effects before the last visit contribute to a better understanding of the models, we are only interested in what happens at the last visit. Furthermore, the treatment effect at the last visit is usually used as a primary endpoint in clinical trials, and is hence very important. For the cLDA model, the treatment effect is determined by the difference between the estimated mean trajectories of the placebo- and active arm at the last visit. In the TPMRM, we analyse the parameter in

the model that describes the treatment effect at the last visit. In the remaining models, the treatment effect is given by a single parameter in the model, which can be analysed directly.

To determine if the cLDA model and PMRMs perform well in the different scenarios, we will look at their mean bias, relative bias, and estimated standard error for the 1000 simulations, which are presented in Table 5.1. To determine these values, we have to determine the true effect present in the different data sets. This is done by using an optimisation algorithm to determine the true treatment effect between the mean trajectory of the placebo- and active arm in each simulated data set. Additional details regarding the optimisation procedure can be found in Section D.1.

Scenario	Model	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
No effect	cLDA	-0.080	0.003	-3.750	2.935
	PSTPMRM	0.958	0.046	4.802	0.068
	PDPMRM	1.010	-0.004	-0.396	0.144
	TPMRM	0.961	0.043	4.475	0.062
20% Proportional reduction in decline	cLDA	3.289	0.035	1.064	1.310
	PSTPMRM	0.813	-0.051	-6.273	0.140
	PDPMRM	0.810	-0.008	-0.988	0.144
	TPMRM	0.812	0.018	2.217	0.095
20% Proportional slowing	cLDA	3.850	-0.056	-1.455	1.282
	PSTPMRM	0.796	0.004	0.502	0.113
	PDPMRM	0.790	-0.047	-5.949	0.142
	TPMRM	0.781	0.021	2.689	0.091
10 to 30% Time based changes	cLDA	5.078	-0.024	0.414	1.247
	PSTPMRM	0.751	0.020	2.663	0.115
	PDPMRM	0.754	-0.084	-11.141	0.137
	TPMRM	0.700	0.010	1.428	0.087

TABLE 5.1: Mean true treatment effect, mean bias, relative bias, and estimation standard error of the progression models for repeated measures and the constrained longitudinal data analysis model in different scenarios for 500 subjects in each arm. Subset of Table D.9.

The relative bias, calculated as the ratio between the mean bias and mean true treatment effect, in each of the scenarios indicate that the PMRMs generally have worse estimation accuracy than the cLDA model, as their relative bias in most scenarios is higher. Furthermore, Table 5.1 shows what we expected, namely that the PDPMRM performs best (with lowest mean bias and relative bias, and a small estimation standard error) in the 20% PD scenario, the PSTPMRM performs best in the 20% PS scenario, and the TPMRM performs best in the 10 to 30% TB scenario (if we exclude the cLDA model). Moreover, when the data is based on a

model's specification, its relative bias is between 0.5 to 1.5%. When this is not the case, the relative bias generally increases to over 2% in absolute value.

As mentioned, the statistical power of the models can be regarded as a type I error rate when the null hypothesis of no treatment effect is true. Hence, the statistical power in the NE scenario can be regarded as a type I error rate. Unless mentioned otherwise, we will use a type I error rate of 5%.

The statistical power of the models will be derived using an F-test. It is well known that the F-test and the two-tailed t-test are equivalent when testing a single parameter under the assumption that the parameter is symmetrically distributed. Consider the case of a null-hypothesis $b = 0$ for a one-dimensional parameter b . A critical t-test statistic at the 5% level (rejecting for large $|t|$) is then equivalent to a critical F-test statistic at the 5% level (rejecting for large F), since $F = t^2$. For a one-sided t-test with alternative hypothesis $H_1 : b > 0$ we only reject for large t . E.g. if we reject for $t > t_{0.975}$ we obtain a significance level of 2.5%. This is equivalent to rejecting if $F > F_{0.95}$ and $\hat{b} > 0$. This means that the models' statistical power in the NE scenario should be approximately 2.5% (0.025), as we only examine positive treatment effects. The distribution of the bias' can be found in Appendix D. In the other scenarios, we desire the estimates of the treatment effect to be significantly different from 0 and hence a statistical power closer to 1. The statistical power of each model in the various scenarios is presented in Table 5.2.

Scenario	cLDA	PDPMRM	PSTPMRM	TPMRM
No effect	0.019	0.033	0.053	0.001
20% Proportional reduction in decline	0.681	0.799	0.922	0.336
20% Proportional slowing	0.787	0.939	1.000	0.449
10 to 30% Time based changes	0.941	0.992	1.000	0.823

TABLE 5.2: *Statistical power of the progression model for repeated measures and the constrained longitudinal data analysis model in different scenarios for 500 subjects in each arm. Subset of Table D.13.*

Table 5.2 shows that the type I error rate of the PSTPMRM is inflated, whereas the type I error rate of the TPMRM is deflated. The cLDA model and PDPMRM both have a type I error rate of approximately 0.025. An inflated type I error rate indicates that the statistical test is more likely to reject a true null hypothesis. Thus, even though the PSTPMRM achieves great power in the other scenarios, this is probably inflated due to the inflated type I error rate. On the other hand, a deflated type I error rate indicates that the statistical test is less likely to reject a true null hypothesis, which also affects the statistical power of the models in the other scenarios. The inflated type I error rate of the PSTPMRM could be due to the mean bias seen in Table 5.1. Here, we see that on average, the estimates of the treatment effect in the PSTPMRM are too large. Hence, it is reasonable that the null hypothesis will be rejected too often for the PSTPMRM. The mean bias of both the cLDA model and PDPMRM is relatively low, which could result in the type I error rate seen in Table 5.2. Because of different type I error rates, the statistical power of the models in the remaining scenarios are not directly comparable. To make a fair comparison of their statistical power, we require their type I error rates to be similar. Therefore, in a later section, we will attempt to calibrate their individual

type I error rates. As the following sections will focus on modifications of these models, we will wait to compare the statistical power of all of them.

The results from the analysis of the models for 500 subjects in each arm are consistent with those for 300 and 1000 subjects in each arm. When increasing the number of subjects in each arm the relative bias and estimation standard error decreases, and the statistical power slightly increases. An increase in the number of subjects does not seem to significantly increase or decrease the mean bias nor the type I error rate of any of the models.

5.1.2 Correlated Error Terms

A general problem with the PMRMs is that they require the correlation structure in the errors to be estimated, which leads to problems with convergence when fitting the models. Therefore, in this section, we will explore the performance of the PMRMs without correlated errors (NC PMRMs) to determine how they compare to the PMRMs with correlated errors.

Again, we start by examining how well the NC PMRMs estimate the treatment effect in the different scenarios and then compare this to the PMRMs. We only present the 20% PS scenario in Table 5.3, as the trends are the same in the other scenarios.

Scenario	Model	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
20% Proportional slowing	NC PSTPMRM	0.796	0.004	0.503	0.032
	NC PDPMRM	0.790	0.005	0.633	0.034
	NC TPMRM	0.781	0.014	1.793	0.035

TABLE 5.3: Mean true treatment effect, mean bias, relative bias, and estimation standard error of the progression models for repeated measures without correlated errors in the 20% proportional slowing scenario for 500 subjects in each arm. Subset of Table D.9.

Table 5.3 shows that the NC PMRMs have lower or approximately the same mean bias as their respective models with correlated errors. Additionally, they achieve this lower mean bias with a lower estimation standard error. Moreover, the relative bias of the NC PMRMs is seen to be lower, or on par with, that of the PMRMs. Across the different scenarios, we see that the NC TPMRM is the model with the highest relative bias of the NC PMRMs, even in the scenario where the data is based upon this model. Given this relative bias and estimation standard error, it seems that the NC PMRMs perform very well in estimating the treatment effect, indicating that correlation in the errors is not necessary. Given this observation, we are interested in the statistical power of these models, which is presented in Table 5.4.

Scenario	NC PDPMRM	NC PSTPMRM	NC TPMRM
No effect	0.225	0.174	0.007
20% Proportional reduction in decline	0.735	0.699	0.330
20% Proportional slowing	0.787	0.777	0.460
10 to 30% Time based changes	0.862	0.870	0.719

TABLE 5.4: *Statistical power of the progression models for repeated measures without correlated errors in different scenarios for 500 subjects in each arm. Subset of Table D.13.*

Table 5.4 shows that the type I error rate of the NC PDPMRM and the NC PSTPMRM is very inflated, rejecting 22% and 17% of the true null hypotheses, compared to the 2.5% it should be. This is unfortunate as it means these models are very prone to rejecting a true null hypothesis, making them inapplicable even though they achieve great power in the other scenarios. Moreover, despite their type I error rates being more inflated than those of their respective models with correlated errors, the statistical power in the other scenarios is lower. This indicates a generally lower statistical power of the NC PMRMs compared to the PMRMs for these two models. Conversely, the NC TPMRM has a deflated type I error rate, similar to what we observed with the TPMRM. Additionally, the statistical power of the NC TPMRM in the other scenarios is either slightly greater or worse than when including correlation in the errors. This suggests that while the correlation in the errors might not be crucial for estimating the treatment effect, it ensures greater statistical power for the models.

Similar to the results for 300, 500, and 1000 subjects in each arm for the PMRMs with correlated errors, we observe the same tendencies for the NC PMRMs. Specifically, the relative bias and estimation standard error decreases and the statistical power increases as the number of subjects in each arm increases.

We are still interested in determining whether the correlation in the errors in the PMRMs can be avoided because of the experienced convergence problems. Therefore, in the following section, we will analyse how the addition of a random effect can influence the performance of the PMRMs, both with and without correlated errors. The hope is that the random effect, among other things, can compensate for the missing correlation structure, and improve the type I error rate for the models.

5.1.3 Random Effects

We will only implement a random effect in the (NC) PSTPMRM. The random effect can be implemented in the other PMRMs in a similar way. Furthermore, we would expect similar tendencies as we see for the PSTPMRM when implementing random effect in the other PMRMs. The random effect we will implement is either a random intercept (RI) or a random scaling factor (RS). The random intercept and random scaling factor will be implemented as follows, respectively:

$$h(t_{i,j}; h_0, \beta, W_{i,j} \zeta, U_i) = h_0(W_{i,j} \zeta t_{i,j}; \beta) + U_i, \quad (5.1)$$

$$h(t_{i,j}; h_0, \beta, W_{i,j} \zeta, U_i) = \exp(U_i) h_0(W_{i,j} \zeta t_{i,j}; \beta), \quad (5.2)$$

with the notation from Chapter 3 used. Under the assumption that the random effects are normally distributed with mean zero and standard deviation σ_U , it follows that

$$\mathbb{E} [h_0(\mathbf{W}_{i,j}\zeta t_{i,j};\beta) + U_i] = h_0(\mathbf{W}_{i,j}\zeta t_{i,j};\beta)$$

and

$$\mathbb{E} [\exp(U_i) h_0(\mathbf{W}_{i,j}\zeta t_{i,j};\beta)] = \exp\left(\frac{\sigma_U^2}{2}\right) h_0(\mathbf{W}_{i,j}\zeta t_{i,j};\beta),$$

respectively. Hence, the natural cubic spline, which models the mean trajectory of the placebo arm, will, on average, be scaled with some scaling factor depending on the standard deviation of the random effect when including the random effect as in (5.2). As this scaling factor increases, the value of the knots and the amplitude of the natural cubic spline increase. This means that the natural cubic spline describing the mean trajectories will be scaled, which potentially could affect the estimated treatment effect.

Naturally, this change of the mean trajectories is suboptimal, since the mean trajectories should ideally remain unaffected. We attempted to modify the model to keep the mean trajectories unchanged, but these modified models failed to converge. Therefore, we proceeded with our analysis using this model, to evaluate its performance. The potential issues with this implementation of a random effect will be discussed further in Chapter 7.

We again only present the 20% PS scenario in Table 5.5, as the trends are the same in the other scenarios. In Table 5.5, it can be seen that the true effect is the same across all models. This is due to the fact that the models in Table 5.5 have the same type of treatment effect. Hence, the true effect becomes the same when calculating it as described in Section D.1.

Scenario	Model	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
20% Proportional slowing	PSTPMRM	0.796	0.004	0.503	0.113
	NC PSTPMRM	0.796	0.004	0.503	0.032
	RI PSTPMRM	0.796	0.005	0.628	0.113
	NC RI PSTPMRM	0.796	0.000	0.000	0.120
	RS PSTPMRM	0.796	0.005	0.628	0.113
	NC RS PSTPMRM	0.796	-0.010	-1.256	0.122

TABLE 5.5: Mean true treatment effect, mean bias, relative bias, and estimation standard error of the proportional slowing of disease progression progression models for repeated measures with and without correlated errors and a random effect in the 20% proportional slowing scenario for 500 subjects in each arm. Subset of Table D.9.

Table 5.5 reveals that the addition of a random effect in the (NC) PSTPMRM, apart from the NC RS PSTPMRM, leads to estimation accuracy that is better or on par with the PSTPMRM in all scenarios. However, the difference in relative bias between the NC RS PSTPMRM and the other PSTPMRMs is small, indicating that they all perform well. Among the PSTPMRMs

here, the NC RI PSTPMRM and NC PSTPMRM are generally the ones with the best estimation accuracy.

When analysing the random effects of the PMRMs with correlated errors further, we observe that the standard deviation of the random intercept is approximately 7, whereas the standard deviation of the random scaling factor is approximately 0.05, both with mean 0 in all scenarios. This means that the random scaling factor on average affects the natural cubic spline with a scaling factor of approximately 1.001. Thus, similar to the random intercept, the random scaling factor does not change the mean trajectories in any consequential way in practice.

However, when looking at the NC PMRMs, the standard deviations change to approximately 19.5 and 0.2, respectively, for the random intercept and random scaling factor. This means that the random scaling factor on average affects the natural cubic spline with a scaling factor of approximately 1.01. Thus, when the errors are not correlated, the standard deviation of the random scaling factor is larger, which in return affects the mean trajectories slightly more. These results are consistent with the results for 300 and 1000 subjects in each arm.

As we previously saw that the NC PSTPMRM has a very inflated type I error rate, we are also interested in how the addition of a random effect affects the type I error rate of this model. The statistical power of the models, including a random effect, is presented in Table 5.6.

Subjects per arm	Scenario	RI PSTPMRM	NC RI PSTPMRM	RS PSTPMRM	NC RS PSTPMRM
500	No effect	0.052	0.056	0.057	0.057
	20% Proportional reduction in decline	0.937	0.931	0.938	0.930
	20% Proportional slowing	1.000	0.999	1.000	0.998
	10 to 30% Time based changes	1.000	1.000	1.000	1.000

TABLE 5.6: *Statistical power of the progression models for repeated measures with a random effect in different scenarios for 500 subjects in each arm. Subset of Table D.13.*

Table 5.6 shows that the (NC) PSTPMRMs with a random effect have a type I error rate similar to the PSTPMRM. This is consistent with the results for 300 and 1000 subjects in each arm. As these models have a type I error rate similar to the PSTPMRM, we are able to compare their statistical power in the remaining scenarios. Not only are the type I error rate of the models with a random effect similar to that of the PSTPMRM, but the statistical power in the other scenarios is also relatively similar. This indicates that the correlation in the errors can be replaced by a random effect. However, we also wish to determine how all the different PMRMs' statistical power compare to each other, as well as to the cLDA model. This will be the focus in the following section.

5.2 Calibrated Statistical Power

Up until this point we have been unable to compare the statistical power between all of the (NC) PMRMs and the cLDA model due to variations in their type I error rate. However, by calibrating this type I error rate to 0.025, we can compare these models' statistical power to

determine if the previous results still hold. We will calibrate their individual type I error rate by determining the 2.5% quantile value of the p-values collected in the simulations.

Specifically, we are interested in comparing the cLDA model and the PMRMs without random effects and with correlated errors, referred to as the original models. This will allow us see which of the original models, presented in [Raket, 2022], demonstrates the greatest statistical power across the different scenarios. Additionally, we are interested in comparing the original models with the modified models, that is, those without correlated errors and/or a random effect. Hence, we will be able to determine if the correlated errors in the models contribute to a greater statistical power. Moreover, we aim to analyse whether including a random effect contributes to a greater statistical power both with and without correlated errors. This analysis will also provide insight into whether a random effect contributes to a greater statistical power than the correlation in the errors.

In Table 5.7, the calibrated statistical power for all models in the different scenarios is presented for 300, 500, and 1000 subjects in each arm. Table 5.7 shows that when calibrating the type I error rate, the PMRMs have similar or greater power than the cLDA model, with the exception of the (NC) TPMRM, NC PSTPMRM, and NC PDPMRM. These results are consistent for 300, 500, and 1000 subjects in each arm. Herein we also see that across all scenarios, the PSTPMRM is the original model with the greatest statistical power except in the 20% PD scenario for 1000 subjects in each arm.

When we exclude the correlation in the errors from the original PMRMs, we observe that the statistical power decreases significantly for both the PSTPMRM and PDPMRM. This aligns with our observations in Table 5.4. Here, we saw highly inflated type I error rates alongside comparatively lower statistical power across the other scenarios, in contrast to the models with correlated errors. The statistical power of the TPMRM also decreases when removing the correlation in the errors when working with 500 and 1000 subjects in each arm, contrary to the results for 300 subjects in each arm.

In Table 5.7, it is evident that including a random effect in the PSTPMRM, when there is correlation in the errors, only slightly improves the statistical power across all scenarios. Nevertheless, comparing the NC PSTPMRM and its counterpart models including a random effect, the statistical power increases significantly when a random effect is included. In the simulations with 500 and 1000 subjects in each arm the results indicate that the random effect in the models contribute to a greater statistical power than the correlation in the errors does. This is evident as the NC RI/RS PSTPMRM demonstrates equal or greater statistical power compared to the PSTPMRM. However, as the number of subjects in each arm decreases to 300, the PSTPMRM achieves the greatest statistical power among those three models.

Subjects per arm	Scenario	cLDA	PDPMRM	NC PDPMRM	PSTPMRM	NC PSTPMRM	RI PSTPMRM	NC RI PSTPMRM	RS PSTPMRM	NC RS PSTPMRM	TPMRM	NC TPMRM
300	No effect	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025
	20% Proportional reduction in decline	0.439	0.55	0.235	0.784	0.152	0.802	0.495	0.800	0.451	0.144	0.211
	20% Proportional slowing	0.496	0.699	0.219	0.994	0.135	0.994	0.941	0.994	0.928	0.159	0.231
	10 to 30% Time based changes	0.798	0.916	0.347	0.997	0.205	0.997	0.963	0.995	0.956	0.331	0.378
500	No effect	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025
	20% Proportional reduction in decline	0.737	0.789	0.293	0.902	0.267	0.922	0.915	0.922	0.909	0.472	0.406
	20% Proportional slowing	0.84	0.927	0.355	1.000	0.312	1.000	0.999	1.000	0.998	0.569	0.535
	10 to 30% Time based changes	0.96	0.991	0.474	1.000	0.413	1.000	1.000	1.000	0.999	0.861	0.774
1000	No effect	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025
	20% Proportional reduction in decline	0.884	0.959	0.525	0.472	0.541	0.484	0.934	0.488	0.917	0.586	0.328
	20% Proportional slowing	0.959	0.977	0.533	0.997	0.656	1.000	0.998	1.000	1.000	0.656	0.438
	10 to 30% Time based changes	0.996	0.996	0.717	0.999	0.751	1.000	1.000	1.000	1.000	0.957	0.836

cLDA: Constrained Longitudinal Data Analysis, **PDPMRM:** Proportional reduction in decline progression model for repeated measures, **PSTPMRM:** Proportional slowing of disease progression progression model for repeated measures, **TPMRM:** Time based changes in disease progression model for repeated measures, **RI:** Random intercept, **RS:** Random scaling factor, **NC:** Not correlated error terms.

TABLE 5.7: *Calibrated statistical power of the progression models for repeated measures and the constrained longitudinal data analysis model in different scenarios, and for a different number of subjects in each arm. The highlighted numbers in each row (except the scenario: no effect) highlights the model(s) which has(have) the greatest power.*

Even though the calibrated type I error rate makes it possible to compare the models, it does not change the fact that the PMRMs, except the (NC) TPMRM, experience inflated type I error rates, which might deem them unusable for practical applications. This issue is not present in the cLDA model ?mere?. This will be further touched upon in Chapter 7.

5.3 Time Homogeneity

In the former sections we saw that the PSTPMRM performs relatively well. Hence, we are interested in determining whether its assumption of constant treatment effect over time is true. Furthermore, if the treatment can be considered to have a constant effect on the subjects over the time period of a clinical trial, predicting its effect beyond the trial duration becomes more feasible, as will be seen in Chapter 6. We will test this assumption using a likelihood-ratio test between the TPMRM and PSTPMRM. Recall that the TPMRM allows the treatment effect to vary over time, whereas the PSTPMRM assumes that treatment effect can be modelled by a constant. That is, the PSTPMRM is a special case of the TPMRM. Thus, this likelihood-ratio test, tests whether the additional parameters in the TPMRM are necessary. A p-value below 0.05 means that the TPMRM is better than the PSTPMRM and hence we reject the null hypothesis. The p-values for each of the scenarios can be seen in Figure 5.8, with the summarised results being; 95% of the null hypotheses are not rejected for NE, 69% are not rejected for 10 to 30% TB, 14% are not rejected for 20% PD, and 92% are not rejected for 20% PS. Thus, it seems that the assumption of PS is viable in the NE-, 10 to 30% TB-, and 20% PS scenarios, but not in the 20% PD scenario.

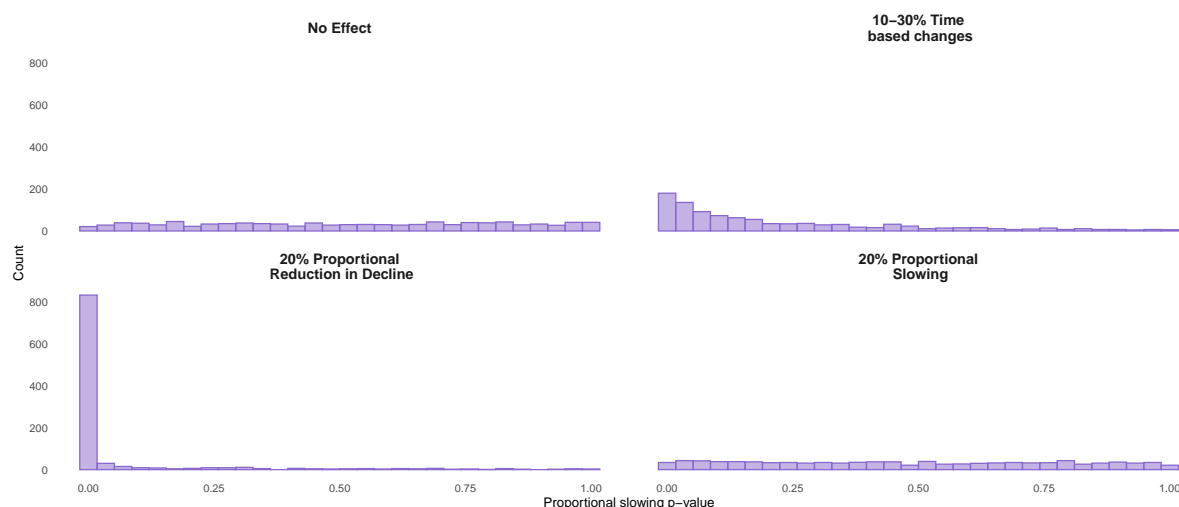


FIGURE 5.8: Time homogeneity p-values testing the assumption of proportional slowing in the scenarios; no effect, 10 to 30% time based changes, 20% proportional reduction in decline, 20% proportional slowing, where there are 500 subjects in each arm.

Hence, these results indicate that in the majority of the scenarios, the assumption of the treatment effect being able to be described by one constant parameter over time holds. In the simulations with 300 and 1000 subjects in each arm we see the same overall tendencies. Note that these results are affected by the used test, and the data, and hence could vary depending on the test and number of subjects in each arm.

5.4 Subgroup Analysis

In the context of Health Technology Assessment (HTA) and in clinical trials the power to detect a difference in treatment effect between subgroups is at least as important as detecting the treatment effect between a placebo- and active arm. This difference in treatment effect between subgroups will henceforth be referred to as the treatment difference. The reason that this is important is, among other things, that a considerable fraction of recommendations from HTA authorities involve restricted recommendations to subgroups [Paget et al., 2011].

Hence, in this section we analyse the power to detect the treatment difference for the cLDA model compared to the PMRMs; PSTPMRM, NC PSTPMRM, NC RI PSTPMRM, and NC RS PSTPMRM. These PMRMs have been chosen, as the PSTPMRM in the former sections seemed to exhibit superior performance compared to the remaining PMRMs. Additionally, an extensive analysis of all 11 models presented in the former sections is not possible due to the time constraints of the thesis.

We will examine the simple case where there are two subgroups, subgroup 1 and subgroup 2, which gives two placebo arms and two active arms to examine the treatment difference. As opposed to using the same PSTPMRM as in the former sections, this section will utilise the subgroup extension of this model, presented in (3.8). Additionally, we will implement a random intercept and a random scaling factor in a similar manner as in Subsection 5.1.3. These extended PMRMs all have one parameter which describes the treatment effect in each group. Hence, we can compare the two parameters describing the treatment effects in the two subgroups, and determine if they are significantly different. The subgroup extension of the cLDA model will include an interaction term between treatment, time from baseline, and subgroup. Besides the interaction term, the cLDA model includes a random effect term, which allows for a subject-specific variation at each visit. To determine whether there is a treatment difference between the two subgroups in the cLDA model, we determine the treatment effect as described in Section 5.1 in each subgroup and determine if these are significantly different.

Data simulation

The placebo arm of subgroup 1 will be simulated with respect to the mean trajectory of the placebo arm from the CPAD database, as the placebo arm in the former sections. The remainder of treatment arms will be simulated on the basis of this placebo arm. The placebo arm in subgroup 2 will be simulated as either a 10% proportional reduction in decline (PD) or 10% proportional slowing of disease progression (PS) of the placebo arm in subgroup 1. In the case where the placebo arm in subgroup 2 is 10% PS of the placebo arm in subgroup 1, both of the active arms will also be simulated as a PS of their respective placebo arms. Similarly, if the placebo arm in subgroup 2 is a 10% PD of the placebo arm in subgroup 1, the active arms will be simulated as a PD of their respective placebo arms. Hence, we will assume that the relation between the arms are of the same type. We have implemented this assumption to minimise the possible combinations of various simulations, primarily due to our uncertainty about which assumptions are accurate. Consequently, we must constrain the variety of scenarios in some fashion.

However, to examine the performance, estimation accuracy, and statistical power of the models we simulate various scenarios. We will in the scenarios fix the treatment effect in subgroup 1 at a 20% PD or PS, and then vary the treatment effect in subgroup 2. We will

simulate a treatment difference of 0%, 5%, 10%, and 15%. For example, in the case where there is a treatment difference of 15% PS, the mean trajectories of the arms, for visit $j > 0$, are given as

$$\beta_{j,\text{active}} = \beta_{j \cdot 0.8, \text{placebo}}, \quad \beta_{j, \text{placebo}_2} = \beta_{j \cdot 0.9, \text{placebo}}, \quad \beta_{j, \text{active}_2} = \beta_{j \cdot 0.65, \text{placebo}_2}.$$

Here, the mean trajectory of the placebo arm in subgroup 1 is given as the vector $\beta = (\beta_{0, \text{placebo}}, \beta_{1, \text{placebo}}, \dots, \beta_{8, \text{placebo}})^\top$. Again, we will assume that all treatment arms have the same mean baseline, that is, $\beta_{0, \text{placebo}} = \beta_{0, \text{active}} = \beta_{0, \text{placebo}_2} = \beta_{0, \text{active}_2}$. The scenario wherein there is no treatment effect in either of the two subgroups the mean trajectories satisfy $\beta_{j, \text{placebo}} - \beta_{j, \text{active}} = \beta_{j, \text{placebo}_2} - \beta_{j, \text{active}_2} = 0$ for all j .

Lastly, we will be conducting the analysis for 1000 simulations as in Section 5.1, as we, again, see that the statistical power and mean estimated treatment difference stabilised within these 1000 simulations. A figure of the statistical power for all models with respect to the number of simulations, in each scenario, can be seen in Section E.2. Similarly, a figure of the models' mean estimated treatment difference with respect to the number of simulations, in each scenario, can be seen in Section E.1.

5.4.1 Estimated Treatment Difference

Instead of analysing the estimated treatment effect in each subgroup as in Section 5.1, we analyse the estimated treatment difference. That is, how well the models estimate the 0%, 5%, 10%, and 15% treatment difference. We will compare the estimated treatment difference from the models with the true treatment difference. Hence, the mean bias presented in Table 5.9 is the mean of the true treatment difference observed in the data subtracted from the estimated treatment difference from the models in each simulation. The true treatment difference is calculated as the difference between the true treatment effects of the two subgroups, where the true treatment effect is calculated as described in Section 5.1, and elaborated upon in Section D.1. Furthermore, the relative bias in Table 5.9 is also determined by comparing the true treatment difference and estimated treatment difference.

As an example, if the true treatment effect in subgroup 1 is a 20% PS and it in subgroup 2 is a 35% PS, the true treatment difference is a 15% PS. Furthermore, if the estimated treatment effect in subgroup 1 is a 19% PS and it in subgroup 2 is a 35% PS, the estimated treatment difference is a 16% PS. In this example the bias is then $0.15 - 0.16 = -0.01$.

Table 5.9 displays the mean bias, the relative bias, and the estimated standard error over the 1000 simulations, together with the mean of the true treatment difference. Table 5.9 only presents the results for the simulations with 500 subjects in each arm in the scenario with a 15% PS or PD treatment difference. The results for 300 and 1000 subjects in each arm and the remaining scenarios (0%, 5%, and 10% treatment difference) for 500 subjects in each arm can be found in Appendix E. The scenarios with a 15% treatment difference is the only scenario presented, since similar tendencies are seen for the remaining scenarios.

Scenario	Model	Mean True Treatment Difference	Mean Bias	Relative Bias (%)	Estimation Standard Error
15% difference in PS	cLDA	0.761	-0.020	-2.628	1.840
	PSTPMRM	0.146	0.036	24.658	0.239
	NC PSTPMRM	0.146	0.002	1.370	0.038
	NC RI PSTPMRM	0.146	0.009	6.164	0.155
	NC RS PSTPMRM	0.146	0.065	44.521	0.236
15% difference in PD	cLDA	1.920	0.037	1.927	1.800
	PSTPMRM	0.141	0.144	102.128	0.308
	NC PSTPMRM	0.141	-0.002	-1.418	0.155
	NC RI PSTPMRM	0.141	0.085	60.284	0.146
	NC RS PSTPMRM	0.141	0.121	85.816	0.390

TABLE 5.9: Mean true treatment difference, mean bias, relative bias, and estimation standard error of the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in the 15% difference in PD/PS scenarios for 500 subjects in each arm. Subset of Table E.11.

The relative bias here is generally larger than in Section 5.1, which could be due to the true treatment difference being closer to zero than the true treatment effects in Section 5.1. Thus, low biases can result in a large relative bias. This is because the denominator of the formula for the relative bias being the mean true treatment differences, which is close to zero, especially in the 0% difference scenarios. However, the relative bias of the extended PMRMs indicate that the estimation accuracy of the NC PSTPMRM is the best, similar to what we saw in Table 5.5.

Table 5.9 shows that there is a general tendency for the estimated treatment difference to be larger than the true treatment difference, since the mean bias is positive. This pattern persists when considering the simulations with 300 and 1000 subjects in each arm for the PSTPMRM and the models including a random effect, but not for the NC PSTPMRM and the cLDA model. Even though there is an indication of a bias in the estimated treatment differences for some of the extended PMRMs, this is not the case in the estimates of each treatment effect in the two subgroups. The results for each of the estimated treatment effects can be seen in Appendix E. This applies to all of the models across all scenarios. The tendencies of the bias, relative bias and estimated standard error of each of the treatment effects in the individual subgroups are very similar to those seen in Section 5.1. Hence, we will not comment further on these results. As expected, the mean bias and estimated standard error of the treatment difference is larger than when analysing the treatment effects in the two subgroups individually, since it includes the uncertainty of two independent parameter estimates compared to one.

As in Section 5.1 the mean bias and relative bias of the estimated treatment difference decreases as the number of subjects increase. Similarly, the estimated standard error decreases when the number of subjects increase. Furthermore, the results indicate that the extended PMRMs estimate the treatment difference more accurately in the PS scenarios as opposed

to the PD scenarios. It does, however, not seem like the cLDA model favors one of the two scenarios. In Table E.11 and Table E.15, it can be seen that for 500 and 1000 subjects in each arm the mean bias and estimated standard error of the extended PMRMs without correlated errors is smaller or on par with those of the PSTPMRM. This again indicates that the models without correlated errors provide more accurate parameter estimates than those with correlated errors. However, when there are 300 subjects in each arm, the estimated standard error in the NC PSTPMRM and NC RS/RI PSTPMRM varies significantly across the different scenarios and is sometimes rather large.

5.4.2 Statistical Power

Besides analysing the estimation accuracy of the models we will analyse their power to detect a treatment difference. Again, statistical power relates to an F-test, with the null hypothesis being that the treatment difference is zero. Hence, for a type I error rate of 9%, the statistical power in the 0% treatment difference scenarios, where the null hypothesis is true, should be approximately 0.045, using the same reasoning as presented above Table 5.2. The type I error rate is chosen higher than in Section 5.1 based on the arguments seen in [Paget et al., 2011]. The type I error rate of each model is presented in Table 5.10 for both a 10% PD and a 10% PS relation between the two placebo arms.

Scenario	cLDA	PSTPMRM	NC PSTPMRM	NC RI PSTPMRM	NC RS PSTPMRM
0% difference in PS	0.052	0.167	0.068	0.034	0.028
0% difference in PD	0.056	0.171	0.062	0.012	0.014

TABLE 5.10: *Type I error rate of the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model for 500 subjects in each arm. Subset of Table E.19.*

Table 5.10 shows that the type I error rate of the cLDA model is slightly inflated, while the type I error rate of the PSTPMRM is highly inflated. The type I error rate of the NC RI/RS PSTPMRM is, however, quite deflated in both scenarios. The type I error rate of the NC PSTPMRM is only slightly inflated, meaning that the removal of the correlation structure in PSTPMRM decreases the type I error rate. Thus, contrary to Section 5.1, it seems that the correlated errors do not seem to result in greater statistical power. Thus, as opposed to before, it would seem that perhaps the estimation of the correlation structure in the extended PMRMs is negatively affecting the results.

To compare the statistical power of the models in the remaining scenarios we will again calibrate the type I error rate of the models, using the same method as in Section 5.2. However, the calibration will be done separately for the PD and PS scenarios, as these have two different 0% difference scenarios. The calibrated statistical power is presented in Table 5.11.

Scenario	cLDA	PSTPMRM	NC	NC RI	NC RS
			PSTPMRM	PSTPMRM	PSTPMRM
0% difference in PS	0.045	0.045	0.045	0.045	0.045
0% difference in PD	0.045	0.045	0.045	0.045	0.045
5% difference in PS	0.945	0.118	0.323	0.478	0.525
5% difference in PD	0.965	0.099	0.313	0.687	0.687
10% difference in PS	0.975	0.154	0.372	0.703	0.743
10% difference in PD	0.976	0.130	0.363	0.818	0.796
15% difference in PS	0.994	0.198	0.487	0.851	0.866
15% difference in PD	0.986	0.155	0.458	0.889	0.89

TABLE 5.11: *Calibrated statistical power of the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model for 500 subjects in each arm. The highlighted numbers in each row (except the scenarios with 0% difference) highlights the model(s) which has(have) the greatest statistical power. Subset of Table E.20.*

Comparing the calibrated statistical power of the extended PMRMs we see that the PSTPMRM and NC PSTPMRM have significantly lower statistical power than the models including a random effect. Furthermore, there is an indication that the model with a random intercept generally achieves greater statistical power than the model with a random scaling factor. Furthermore, an advantage of the extended PMRMs with a random effect is that their uncalibrated type I error rates are not inflated (see Table 5.10). Hence, this indicates that the inclusion of a random effect in the extended PMRMs can replace the correlation in the errors. Additionally, in this analysis it seems to significantly improve the extended PMRMs' performance. Moreover, we observe that the calibrated statistical power is higher for the cLDA model compared to the extended PMRMs. Furthermore, we also observe that the statistical power increases as the true treatment difference and/or the number of subjects increases. This is consistent with the notion that the power to detect each treatment effect also increases with the number of subjects.

In Table 5.11, we further see that the (NC) PSTPMRM achieves the greatest power in the PS scenarios compared to the PD scenarios, whereas the extended PMRMs with a random effect seem to do the opposite. However, this varies slightly when looking at different number of subjects.

5.5 Summary of Results

Overall, we see that the estimation accuracy of all the models in Section 5.1 becomes better when increasing the number of subjects in each arm. Additionally, we observe that most PMRMs exhibit inflated type I error rates, especially the NC PDPMRM and NC PSTPMRM. The very inflated type I error rates for the NC PDPMRM and NC PSTPMRM also result in these models having the poorest calibrated statistical power of all the models. When comparing the calibrated statistical power for all of the models, most of the PMRMs had greater statistical

power than the cLDA model. Specifically, the PSTPMRM and its modifications with a random effect achieved the highest statistical power among all of the models, in most scenarios. Generally, we see that the addition of a random effect can compensate for the correlation in the errors, especially when the number of subjects in each arm is sufficiently high. When decreasing the number of subjects in each arm to 300, we see that the models with a random effect should advantageously include correlation in the errors.

For the subgroup models, much of the same can be seen. The estimation accuracy improves when the number of subjects in each arm increases. Furthermore, the estimation accuracy of the extended PMRMs is better in the proportional slowing of disease progression scenarios compared to the proportional reduction in decline scenario. However, contrary to the statistical power in Table 5.7, the subgroup model extension of the cLDA model has higher statistical power than either of the extended PMRMs (see Table E.19). Thus, even though the PMRMs previously seemed to outperform the cLDA model, this is not the case here. However, it was seen that including a random effect significantly increases the statistical power of the extended PMRMs. In the simulation with 1000 subjects in each arm, the statistical power of these extended models perform almost on par with the cLDA model in most scenarios. This again suggests that the random effects can replace the complicated correlation structure in the errors in the extended PMRMs and even improve their performance. However, again when decreasing the number of subjects in each arm to 300, we see that the random effect models should advantageously include correlation in the errors.

Hence, from these analyses the NC RI/RS PSTPMRM and their subgroup extensions perform very similar to the conventionally used cLDA model in both the first analysis between one placebo- and active arm, as well as in the subgroup analysis. In the analysis between a placebo- and active arm the type I error rate is somewhat controllable. Moreover, their type I error rates are deflated in the subgroup analysis, as opposed to the cLDA model which experiences an inflated type I error rate. Hence, using any of these models in clinical trials, the inflated type I error rate should be kept in mind. Raket suggests recalibrating the type I error rate based on resampling techniques. The remaining models experience either very inflated type I error rates and/or lower statistical power than these models in either the first analysis between one placebo- and active arm, or in the subgroup analysis. Furthermore, the choice of models should be chosen based on prior knowledge about how treatment effects behave, proportional reduction in decline, proportional slowing of disease progression, etc..

6 Health Technology Assessment Implementation

This chapter is based upon [Ito et al., 2023], [Whitehead and Ali, 2010], and [MIT Critical Data, 2016, Ch. 24].

Besides being interested in the performance of the PMRMs in terms of estimation accuracy and statistical power, we are interested in the further use of these models in a Health Technology Assessment (HTA) context.

When assessing new treatments, HTA bodies must evaluate the treatment's efficacy and cost in comparison to existing alternatives. The evaluation of a new treatment by HTA bodies is intended to assist decision-makers in evidence-based policy-making to maximise the benefits that can be achieved with the available resources. Typically, HTA bodies employ health economic models to effectively capture the impact of a treatment. Clinical studies, such as those regarding Alzheimer's disease (AD), typically do not last longer than 18 months, during which subjects do not necessarily reach the severe stages of the disease. Hence, HTA bodies use extrapolation to determine the long-term effects of a treatment and meaningfully capture how it affects the total cost and disease progression over time.

A comparison of treatments can be done by performing a *cost-effectiveness analysis* (CEA), which compares the cost and effectiveness of different treatments. It is beneficial to measure the cost-effectiveness on a common scale across diseases and treatments, as it makes the comparison of cost-effectiveness easier across diseases and healthcare areas, and is also recommended by the National Institute for Health and Clinical Excellence (NICE), which is a HTA authority. An often used scale is the *incremental cost-effectiveness ratio* (ICER).

The ICER is calculated by taking the difference in the total costs between two treatments and dividing it by the difference in effects, typically measured in *quality-adjusted life years* (QALYs). Specifically, the ICER is given as

$$\text{ICER} = \frac{\Delta \text{Cost}}{\Delta \text{QALY}}. \quad (6.1)$$

Here, the total cost includes both the cost of the treatment, the cost if a subject has to receive institutional care, the caregivers' lost income, and much more. QALY is a way of valuing a life year and is calculated by multiplying the duration of time spent in a health state by the *utility score* of this health state. Here, the utility score, typically ranging from 0 to 1, is a way of valuing health state of a disease to reflect the desirability of being in a particular health state. Take, for example, AD, which can be split into multiple stages including, among others, mild cognitive impairment (MCI) due to AD, mild AD dementia, and moderate to severe AD dementia. The utility score, and hence QALY, will for a subject with AD decrease as the severity of the disease increases. Conversely, the cost will increase because of the increasing amount of care required. Thus, the CEA of a treatment in the context of AD, relies heavily on the time a subject spends in each stage, and hence it is rather important how a treatment extends or reduces the time in the different stages of the disease.

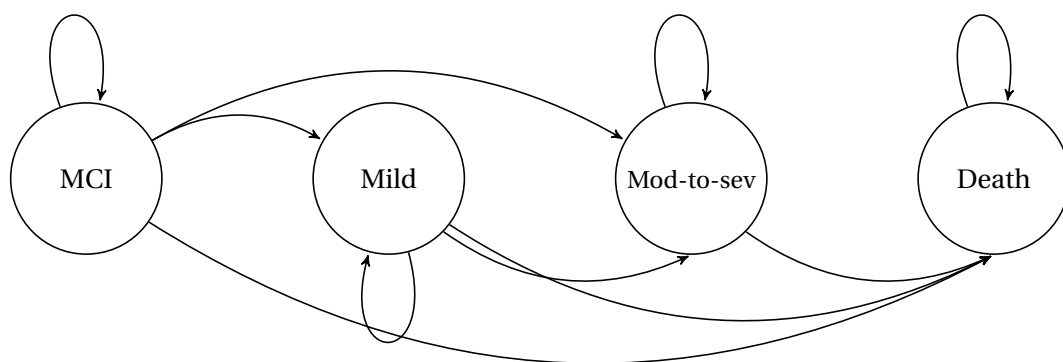
In this chapter, we will demonstrate how the PMRMs can be used in the context of a CEA. Specifically, we will introduce one possible way of using the PMRMs in modelling how a treatment affects the time spent in each health state long-term. For this purpose, we will be using *Markov models*, since they are one of the most commonly used tools in a CEA. This implementation of the PMRM closely corresponds to that presented in [Jönsson et al., 2024].

6.1 Markov Models

A Markov model consists of a state space, usually consisting of different discrete health states, and *transition probabilities* between these states. Furthermore, it is a stochastic process that undergoes transitions from one state to another. Additionally, Markov models satisfy the first-order Markov property, meaning that the probability of moving to another state only depends on the current state and not the former states.

In the following, we will be using a simple Markov model which has constant transition probabilities over time. The model can be extended in multiple different ways, depending on the assumptions, such as transition probabilities that vary over time, not satisfying the first-order Markov property, etc..

In Alzheimer's disease, the health states could be the three aforementioned stages of the disease as well as death. Then the subjects transition from one state to another with some probability over some period of time. This period of time is called a *cycle*, which can be a day, month, year, and so on. The possible health states and transitions in the Alzheimer's disease example are illustrated using a state transition diagram, seen in Figure 6.1.



MCI: Mild cognitive impairment due to Alzheimer's disease, **Mild:** Mild Alzheimer's disease dementia, **Mod-to-sev:** Moderate to severe Alzheimer's disease.

FIGURE 6.1: State transition diagram for cognitive decline in Alzheimer's Disease.

Here, the nodes represent the different health states and the arrows represent the possible transitions. In Figure 6.1, some of the assumptions are presented. First, it is possible to transition to death from all other health states, and death is an *absorbing state*, meaning a subject can not leave the health state after transitioning to it. Second, it is not possible to transition to a less severe health state than the one a subject is already in. Last, it is possible for a subject to stay in the same health state for multiple cycles.

Each arrow has a corresponding probability of transitioning in each cycle - their transition probability. All of these transition probabilities can be presented in a transition matrix, which for Alzheimer's disease could resemble the one in Table 6.2.

	MCI	Mild	Mod-to-sev.	Death
MCI	0.978	0.015	0.005	0.002
Mild	0.000	0.983	0.010	0.007
Mod-to-sev.	0.000	0.000	0.988	0.012
Death	0.000	0.000	0.000	1.000

MCI: Mild cognitive impairment due to Alzheimer's disease, **Mild:** Mild Alzheimer's disease dementia, **Mod-to-sev:** Moderate to severe Alzheimer's disease.

TABLE 6.2: *Transition matrix (should be read from left to right)*

These are external transition probabilities based on other clinical trials [Jönsson et al., 2024; Lin et al., 2023], and they represent the transition probabilities for a group of subjects in a placebo arm. These transition probabilities are based on monthly cycles, allowing subjects to transition to another health state each month.

The transition matrix in Table 6.2 should be read from left to right, meaning that the value in the first row and second column describes the probability of transitioning from the MCI due to AD state to the mild AD dementia state in one month. The transition probabilities in each row should further add up to one. As mentioned, the subjects can not transition to a less severe health state, which is reflected in the transition matrix, as the values below the diagonal are all 0.

Based on the transition matrix, it is possible to calculate the probability of a subject being in the different health states after each cycle. However, Table 6.2 only describes the transition probabilities for subjects in a placebo arm, and hence we need to use the PMRMs' estimated treatment effect to analyse how the subjects in the active arm transition over time.

Assuming we are working with the proportional slowing of disease progression PMRM (PSTPMRM), we can model the relation between the mean trajectory of the placebo- and active arm. The PSTPMRM contains one parameter describing the percentage in time the progression of the disease is slowed by, by being on an active treatment. Hence, for example, assuming the model estimates a 20% slowing of disease progression, the mean trajectory of the active arm will at month 20 have the value that the mean trajectory of the placebo arm had at month 16. We will incorporate this in the Markov model by using the same transition matrix as that of the placebo arm but changing the length of the cycles, as proposed by [Jönsson et al., 2024]. Hence, if we continue assuming that the treatment slows the progression of the disease by 20% and each cycle is one month for the subjects in the placebo arm, the cycle for the subjects in the active arm will then be 1.25 months.

We will look at a time period of 15 years (180 months), during which the placebo group will go through 180 cycles of one month each, whereas the active group will go through 144 cycles of length 1.25 months each. We have chosen 15 years based on [Jönsson et al., 2024] and the fact that the life expectancy of a subject with Alzheimer's disease typically does not exceed 10 years [Zanetti et al., 2009]. Assuming that all subjects will start in the MCI due to AD state, the

probability of being in the different health states, after n cycles, is calculated as the $(1, 0, 0, 0)$ vector multiplied by the transition matrix to the power of n . The probability of being in each health state after one cycle is given in Table 6.3.

MCI	Mild	Mod-to-sev.	Death
0.978	0.015	0.005	0.002

MCI: Mild cognitive impairment due to Alzheimer's disease, **Mild:** Mild Alzheimer's disease dementia, **Mod-to-sev:** Moderate to severe Alzheimer's disease.

TABLE 6.3: Probabilities of being in each health state after one cycle.

The probability of being in each health state over time is presented in Figure 6.4.

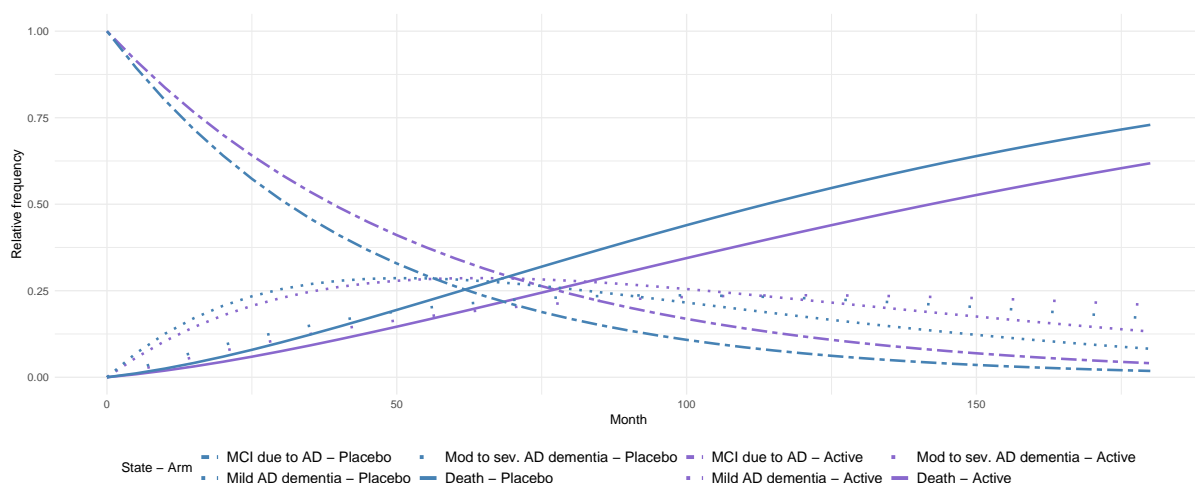


FIGURE 6.4: Estimated proportion of which health states the subjects are in over time, using the transition matrix in Table 6.2.

It can be seen that the proportion of subjects in the MCI due to AD state is larger in the active arm than in the placebo arm. Hence, this implies that the subjects are generally in the mildest state of Alzheimer's disease for a longer period of time. Furthermore, Figure 6.4 shows that the proportion of subjects who are dead is smaller in the active arm than in the placebo arm, meaning the treatment extends the life expectancy of the subjects.

Assuming 1000 subjects receiving placebo in the MCI due to AD state, it is possible, based on Figure 6.4, to determine how many subjects there theoretically should be in each health state over time. For example, after one cycle 978 subjects should be in MCI due to AD, 15 in mild AD dementia, 5 in moderate to severe AD dementia and 2 dead. Meaning Figure 6.4 can be used to analyse the allocation of a group of Alzheimer's disease subjects over time, and thereby simulate the disease progression in the two arms.

Figure 6.4 is, however, not as intuitive if you are only interested in how the treatment, on average, extends or reduces the time spent in each of the different states. This is illustrated in Figure 6.5, where the first two bars are the sum of the three other bars of the same color. From the two first bars, we can again deduce that the subjects on average live for a longer period of

time when on an active treatment. More precisely, in this example, they live, on average, for 24 months longer than the subjects in the placebo arm, considering the 15 year time period.

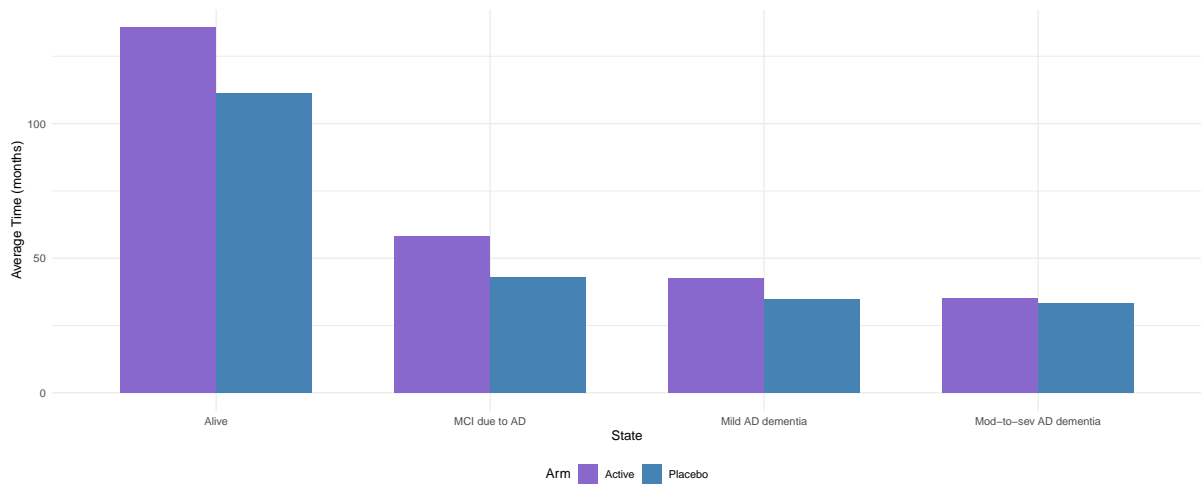


FIGURE 6.5: Average time spent in the different health states over the course of 15 years (180 months) using the transition matrix in Table 6.2.

Here, it can also be seen that the subjects in the active arm, over 15 years, generally spend longer time in all health states. Specifically, in Figure 6.5, we observe that the active group spends 14 months longer in the MCI due to AD state, 7 months longer in the mild AD dementia state, and 3 months longer in the moderate to severe AD dementia state.

Contrary to the placebo arm, the subjects in the active arm receive a, possibly costly, treatment. This treatment extends their life expectancy and delays their disease progression, but also comes at an additional cost. Hence, a CEA should be conducted to help determine if the treatment is cost-effective. In this example, the CEA compares an active treatment to placebo, however, placebo could be replaced by another treatment. Then, the comparison could help determine which treatment is more cost-effective, which could help decision-makers determine which treatment is the better choice as the standard of care.

Furthermore, as described in Section 5.4, some of the extended PMRMs can be used to determine the treatment effect in two subgroups simultaneously. Additionally, it can then be tested whether there is heterogeneity in the treatment effect between subgroups. The treatment effects observed in each subgroup could be regarded as two different treatments, wherein an ICER would determine how they compare with respect to their cost-effectiveness. A substantial ICER could lead to the treatment only being offered for one of the subgroups.

6.2 Cost-effectiveness Analysis

The CEA is conducted by associating a cost and utility to each health state and calculating the expected cost and utility score in each arm over time. From [Ito et al., 2023], the utility score and cost of the subjects in each health state are

	MCI	Mild	Mod-to-sev.	Death
Utility	0.85	0.84	0.82	0
Cost (\$/month)	3011	4536	7577.5	0

MCI: Mild cognitive impairment due to Alzheimer's disease, **Mild:** Mild Alzheimer's disease dementia, **Mod-to-sev:** Moderate to severe Alzheimer's disease.

TABLE 6.6: *The utility and cost of the different states associated with Alzheimer's disease.*

As previously mentioned, the utility score decreases and cost increases as the disease becomes more severe. Apart from the costs presented in Table 6.6, the subjects on the active treatment will further have a cost of the treatment added in each health state, which [Ito et al., 2023] estimates to be 1333 \$/month. Assuming there are 1000 subjects in each arm starting in the MCI due to AD state, it is possible to determine the total QALY and total cost for each arm over time. This is illustrated in Figure 6.7 and Figure 6.8, respectively. However, for simplicity, note that this example is presented without discounting, that is, the cost and QALY is not brought into the present.

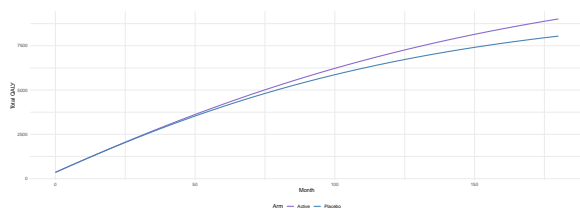


FIGURE 6.7: *Total QALY for each arm over time for 1000 subjects in each arm.*

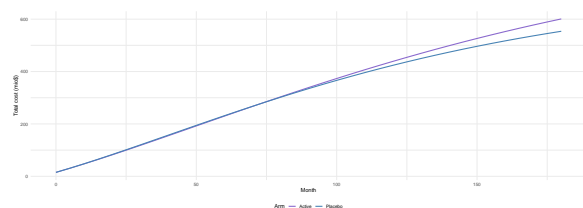


FIGURE 6.8: *Total cost for each arm over time for 1000 subjects in each arm.*

Figure 6.7 and Figure 6.8 illustrate how the total QALY and total cost evolve over time, respectively. To determine this, we determine the number of subjects in each health state after each cycle and multiply with the health state's respective utility score or cost, and then add them together. For example, the total cost of the first two months is

$$1000 \cdot (3011 + 1333) + 978 \cdot (3011 + 1333) + 15 \cdot (4536 + 1333) + 5 \cdot (7577.5 + 1333) + 2 \cdot 0 = 8725019.5\$.$$

In Figure 6.7 and Figure 6.8, we see that the total QALY for the subjects in the active arm always exceeds that of the subjects in the placebo arm. Furthermore, the total cost of the subjects in the placebo arm exceeds that of the subjects in the active arm for approximately the first six years. However, it decreases below the total cost of the subjects in the active arm beyond this point. Calculating the ICER as in (6.1) at the last time point, after 15 years, we get

$$\text{ICER} = \frac{600860075\$ - 553712120\$}{9014.1649 - 8045.2522} = 48660.69\$/\text{QALY}$$

which decision-makers then should decide whether or not is a cost-effective treatment. Using this method, it is also possible to analyse which treatment effect the PSTPMRM should estimate to get a specific ICER. In Table 6.9, we see how the ICER changes relative to the

PSTPMRM's estimated treatment effect, where it is clear that the higher the treatment effect is, the less it costs per additional QALY.

Treatment effect (%)	20	40	60	80
ICER (\$/QALY)	48660.69	42496.23	33900.98	21171.35

TABLE 6.9: *ICER for different treatment effects for the PSTPMRM - estimated proportional slowing of the disease when on treatment.*

This is just one way of implementing the PMRMs in a HTA context. Another way could be extrapolating the trajectories of the subjects using the PMRMs. Each value on the outcome score will then be associated with a health state, and thereby a cost and utility score, allowing us to conduct a CEA based upon this. This will, however, not be elaborated further upon.

7 Discussion

In this thesis, we examined a new class of models, progression models for repeated measures (PMRMs), some of which estimated the time saved in a progressive disease, such as Alzheimer’s disease. We found that this new class of models had relative good estimation accuracy (small relative bias and estimation standard error), as well as high power to detect a treatment effect, in various scenarios, compared to the conventionally used constrained longitudinal data analysis (cLDA) model. Furthermore, we examined some subgroup extensions of the PMRMs, where we found that the cLDA model generally outperformed them. Additionally, we provided an example of how the estimate of the time saved can be used in a cost-effectiveness analysis and to examine how a treatment extends the time in the different stages of Alzheimer’s disease, which led to the average life expectancy being extended. Throughout this thesis, we have discovered some of the shortcomings of the progression models and the assumptions made in Chapter 4. These will be the focus of this chapter.

Data Simulation

In the simulation study, we opted to simulate both the placebo- and active arm, where some simplifying assumptions were made, which could have caused the data to deviate from real world data.

First, as we determined that neither of the active arms in the CPAD database were sufficient for an analysis, we opted to simulate both arms based on the mean trajectory of the chosen placebo arm. Due to this, we assumed that the covariance structure of the placebo arm and the active arm were identical. Due to insufficient information about the behavior of an active arm in the CPAD database, we cannot determine if this assumption is realistic. However, in [Raket, 2022], the same covariance matrix is used to simulate the placebo- and active arm, which suggests that the assumption is acceptable.

Second, the simulated data used in Chapter 5 was simulated using a multivariate normal distribution. All models that have been analysed throughout the simulation study, Chapter 5, have been implemented using R functions that assume normality. Hence, the data were simulated according to the model assumptions. However, we saw in Figure 4.10, that the assumption of normality was not fulfilled for the iADRS data. This could indicate that real-world data does not necessarily fulfill the assumption of normality as the simulated data we have used does, which could affect the results. Moreover, we conducted a simulation study for non-linear mixed models, that showed that the distribution slightly affects the estimation accuracy. However, this study only analysed models on the form of the NC RI/RS PSTPMRMs, and it could hence also be advantageous to test how the other models perform when the distribution of the data deviates from a normal distribution. This simulation study can be found on <https://github.com/loprrtq/Master-s-Thesis-Results>, and is not included in the thesis since it did not fit into the scope. In short, it concentrated on comparing the `nlme` function’s estimation- and extrapolation accuracy to that of the `nlmer` function. The `nlme` function is the function used to fit the (NC) RI/RS PSTPMRMs. The simulation study thus compared the Lindstrom and Bates algorithm, utilised by the `nlme` function, to the Laplace approximation, utilised by the `nlmer` function. The simulation study showed that the `nlme`

function was more robust to deviations from normality than the `nlmer` function.

Third, we assumed that data points only occur every three months, aggregating any small deviations to the nearest third month. This removes small deviations in when subjects were measured, even though the progression models, in theory, should be able to include these small deviations. This was done to make the simulation of data and the implementation of the progression models easier. Furthermore, every third month could then correspond to a visit, which was used in the cLDA model. This aggregation of the data points could lead to some bias in the data. However, as Alzheimer's disease is a rather slow-progressing disease, aggregating these data points to every third month should not change the general tendencies of the data, as could be the case for a faster-progressing disease, where one month might significantly affect the outcome measure. The possibility of including the small deviation should improve the performance of the progression models compared to the cLDA model because of the natural cubic spline interpolation.

Last, even though the aggregation of the data might not have changed the mean trajectories and general results, something that could have affected these is the implementation of dropout and discontinuation, which were not considered in the thesis. This was chosen as it could result in a substantial analysis, given their complexity and potential impact. Furthermore, we would not have anything to base an assumption of a specific dropout rate on for the active arm, but only for the placebo arm. For the placebo arm, we saw that in the iADRS data there were approximately a 5% dropout rate every third month. Even though we did not implement dropout, it is very important to conduct an analysis of how the performance of the models could be affected by this, since no Alzheimer's trial is without it. However, an analysis of the models under the assumption of no dropout and discontinuation should still provide a good indication of the models' general performance.

Some of the assumptions made, make the simulated data deviate from real data. Though, it still align somewhat with real data as the covariance of the arms and the mean trajectory of the placebo arm are based on real data. Furthermore, the chosen treatment effect closely align with that seen in real clinical trials.

Models

The cLDA model and the PDPMRM both provide estimates of the treatment effect with respect to an outcome measure, which can be very difficult to understand as a patient or a caregiver. The PSTPMRM and TPMRM, on the other hand, provide estimates with respect to time saved. This estimate is easier to interpret for patients and caregivers, since they, for example, will know how many additional months they will be able to pick up their grandkids or go grocery shopping themselves. Thus, even though a treatment effect on the time scale is not deemed clinically meaningful, it could be meaningful for the subjects living with the disease. Hence, these model introduce a new possible way of measuring the meaningfulness of a treatment. Furthermore, in a clinical trial, multiple outcome measures are usually modelled to analyse the treatment effect instead of just one. These outcome measures are usually on different scales (minimum and maximum values) and can hence be difficult to compare. Besides the better interpretability of the estimates from the PSTPMRM and TPMRM, it makes it easier to compare treatment effects among multiple outcome measures since they are on the same scale. However, a downside of measuring the treatment effect on the time scale is that this measure is not yet well used among different clinical Alzheimer's trials. It can thus be difficult to compare to the treatment effect observed in other clinical trials.

As described in Subsection 5.1.3, we included a random intercept and a random scaling factor in the PSTPMRM model. Naturally, the inclusion of a random effect changes the model's estimation method. As already mentioned, the mean trajectories of the PSTPMRM with a random intercept do not change with respect to the mean trajectories of the PSTPMRM, as opposed to the PSTPMRM with a random scaling factor. The mean trajectories of the PSTPMRM with a random scaling factor change with respect to the variance of the random scaling factor. As we saw in Subsection 5.1.3, the random scaling factor's variance was approximately 0.2^2 in the RS PSTPMRM and 0.4^2 in the NC RS PSTPMRM. Thus, the amplitude and knot values of the natural cubic splines were only slightly increased. In practice, changing the mean trajectories in any way might not be desirable. We therefore tried to implement a similar model, wherein the mean trajectories were not affected. This model failed to converge when attempting to fit it to the simulated data. Hence, we deemed that it was not possible to implement within the analysis' framework. However, as seen in Subsection 5.1.3, the inclusion of a random scaling factor achieved good estimation accuracy and great statistical power. Thus, even though it changed the mean trajectories, it performed well. The inclusion of a random effect could also increase the interpretability of the model, as random effects allow us to determine the subject-specific variations.

Subgroups

In the subgroup extension of the PSTPMRM, it was assumed that not only the relation between the placebo- and active arms, but also the relation between one of the placebo arms in the two subgroups is a proportional slowing of the other. In Section 5.4, we found that given a sufficient number of subjects in each arm, the extended PSTPMRMs performed relatively well, especially when including a random effect. However, in Section 5.4, the data were simulated to exhibit some relation between the two placebo arms. If the assumption that the relation between the two placebo arms cannot be described by a single parameter, these models would be expected to perform worse than seen in Section 5.4 compared to the cLDA model. Alternatively, it could be assumed that the two subgroups exhibit no relation, such that each placebo arm is described using its own natural cubic spline. Let these natural cubic splines be denoted h_{01} and h_{02} for subgroup 1 and subgroup 2, respectively, then

$$Y_{i,j} = \begin{cases} h_{01}(t_{i,j}; \beta) + \varepsilon_{i,j} & i \in \mathcal{I}_{\text{placebo1}} \\ h_{02}(t_{i,j}; \beta) + \varepsilon_{i,j} & i \in \mathcal{I}_{\text{placebo2}} \\ h_{01}(\zeta t_{i,j}; \beta) + \varepsilon_{i,j} & i \in \mathcal{I}_{\text{active1}} \\ h_{02}(\gamma t_{i,j}; \beta) + \varepsilon_{i,j} & i \in \mathcal{I}_{\text{active2}} \end{cases} \quad (7.1)$$

Even though this increases the number of parameters in the model, it would still be more parsimonious than the cLDA model.

Type I error rate

As we saw in Section 5.1, the type I error rate for most PMRMs was inflated, which is a concern if these were to be used in practice. We calibrated the type I error rate of all models to be 0.025, which helped compare the models' statistical power. In practice, calibrating the type I error rate would be possible by simulating data sets without a treatment effect. Nevertheless, calibrating the type I error rate in practice could be an issue. This is because healthcare authorities are concerned about an inflated type I error rate of the models [U.S. Food and

Drug Administration, 2022], as this suggest a higher tendency of falsely deeming a treatment effective. As most of the PMRMs experience inflated type I error rates, then even though Raket in [Raket, 2022] suggests that these are reasonably controlled, authorities might not feel the same way. However, in the subgroup analysis, we saw that the extended PMRMs with a random effect were the only ones of the analysed models for subgroups which did not experience an inflated type I error rate. Hence, the PMRMs could be preferred over the cLDA model in some cases with regard to the type I error rate.

Numerical Approximations

To determine the bias of the estimated treatment effect we had to determine the true treatment effect in the data, as described in Section D.1. Hence, the true treatment effect is determined numerically instead of being known, which creates the risk that the bias does not only represent the bias of the treatment effect estimated by the models but is also affected by the numerical computation error of the true effect. Therefore, the presented bias and estimation standard error do not necessarily correctly represent the PMRMs' estimation accuracy. Upon further testing of the optimisation algorithm, it became evident that there is typically only error present in the third decimal place. This testing included comparing the results to another optimisation method, which did not only utilise the mean trajectories but the PMRMs themselves. Furthermore, we also conducted some tests of the used optimisation methods where the true treatment effect were know, which showed similar results. This indicates that the bias should not be significantly influenced by the estimation of the true effect.

Health Technology Assessment

In Chapter 6, a very simple way of implementing the estimated treatment effect from a PSTPMRM in a Markov model, was explored. Naturally, some of the assumptions made in Chapter 6 may have been be too simple, such as constant transition probabilities and no discounting. Furthermore, it was assumed that the treatment effect was constant over time and over health states. These assumptions imply that the time taken to reach the state of death is impacted by a similar degree of slowing as the other health states. All of these assumptions will need to be investigated and will probably be deemed too simple in practice. For example, the probability of transitioning to death usually becomes larger as the subject ages. Nonetheless, this way of implementing the PSTPMRMs in a Markov model still offers great insight into how PMRMs can be employed within a health technology assessment context.

8 Conclusion

Throughout this thesis, theory regarding linear- and non-linear mixed models has been introduced. Moreover, the constrained longitudinal data analysis (cLDA) model, which is a linear mixed model, has been introduced, as it is widely used to model the treatment effect between the placebo- and active arm in a clinical trial. Furthermore, theory regarding progression models for repeated measures (PMRMs) has been introduced. These models have in recent years also been introduced to model the treatment effect. There are several differences between the cLDA model and PMRMs, particularly in their estimated treatment effects. The cLDA model estimates the treatment effect as the difference between the placebo- and active arm with respect to some outcome measure. In contrast, the PMRMs provide percentage estimates of the change from baseline or differences between arms on the time scale, making these estimates easier to interpret for caregivers and patients. Additionally, the PMRMs were extended by implementing a random effect, specifically a random intercept and a random scaling factor. Besides extending the PMRMs to include a random effect, we also extended these, as well as the cLDA model, to determine if they were able to detect heterogeneity in the treatment effect between subgroups, which is important in the context of health technology assessment.

In the analysis of the performance of the models when having to model one placebo- and active arm, we observed that some of the PMRMs performed on par or better than the conventionally used cLDA model. Specifically, they achieved the same or higher statistical power, as well as better estimation accuracy. This was especially true for the PSTPMRMs. Furthermore, a random effect seemed to improve the performance of the PSTPMRM and could, in some cases, replace the correlation structure in the errors. This could be useful when using these models in practice, as the correlation structure caused convergence issues in some cases. Despite some PMRMs showing higher statistical power than the cLDA model, they often had inflated type I error rates, which affects their practical utility in a clinical trial. In the subgroup analysis, we observed that the cLDA model performed better than the extended PMRMs, however, it too experienced inflated type I error rates. However, provided enough subjects in each arm, the extended PMRMs with a random effect performed almost on par with the cLDA model, without having an inflated type I error rate. An advantage of the PMRMs is that it provides an easily interpretable treatment effect on the time scale. Measuring the treatment effect on the time scale could be important for subjects living with Alzheimer's disease, as it could provide them with how much longer they can expect to live and how it will affect their health status in this time. Thus, this measure of the treatment effect could open up for a new interpretation of what a meaningful treatment effect is for subjects with Alzheimer's disease.

Additionally, we proposed a method for incorporating the PSTPMRM's treatment effect estimate into a Markov model for health economic modelling. As we saw in Section 5.3 that the assumption of a constant treatment effect over time was reasonable in most scenarios, we assumed the treatment effect as constant over time. This approach simplifies the integration of PSTPMRM into the Markov model, aiding in the projection of long-term treatment outcomes and cost-effectiveness.

Hence, there are both pros and cons of using the PMRMs, and extensions thereof, compared to the conventionally used cLDA model. Overall, while the cLDA model offers robust performance and a controlled type I error rate, PMRMs provide better interpretability and higher statistical power in specific scenarios, highlighting their potential applicability in clinical trials and health economic modelling. Furthermore, all of these models can be extended using multiple additional variables which could affect their performance. This could be an interesting topic and another method to extend both model types.

Appendices

A Preliminary Theory

A.1 Newton-Raphson Algorithm

The Newton-Raphson algorithm is a widely used algorithm to determine some optimal parameters θ that maximise a log-likelihood function $\ell(\theta; \mathbf{y})$. It does so iteratively by updating the parameters to θ_{new} using a second order Taylor series expansion around the current parameters θ_{old} . That is,

$$\theta_{\text{new}} = \theta_{\text{old}} - \mathbf{H}^{-1}(\theta_{\text{old}}) \nabla \ell(\theta_{\text{old}}; \mathbf{y}), \quad (\text{A.1})$$

where θ_{new} becomes θ_{old} in the next step and so forth until it converges to a local maximum. Here, \mathbf{H} is the negative Hessian matrix and $\nabla \ell(\theta; \mathbf{y})$ the gradient of the log-likelihood function. In practice, one might use the profile log-likelihood function for the different parameters and update them simultaneously by iterations until convergence is achieved.

One drawback of the Newton-Raphson algorithm is that if its initial parameters are far from maximum, it might fail. Hence, a justified initial guess could make a significant difference. For more information on the subject, see [Kornerup and Muller, 2006].

A.2 Natural Cubic Splines

Natural cubic splines, as opposed to linear splines, are more flexible due to the additional degrees of freedom, which means that they are able to describe more complex data structures. Specifically, as the name implies, natural cubic splines use cubic polynomials to describe what happens before, between and after the *knots* of the spline. Given there are m knots, then there are also m cubic splines. Thus, a natural cubic spline can also be written as the piecewise function

$$P(t) = \begin{cases} P_0(t) & t_0 < t < t_1 \\ \vdots & \\ P_{m-1}(t) & t_{m-1} < t < t_m \end{cases}$$

To ensure that the natural cubic spline is sufficiently smooth it has to fulfill the following four requirements

$$\begin{aligned} P_{j-1}(t_j) &= P_j(t_j), & P'_{j-1}(t_j) &= P'_j(t_j), \\ P''_{j-1}(t_j) &= P''_j(t_j), & P''_0(t_0) &= P''_{m+1}(t_m) = 0, \end{aligned}$$

for $j = 1, \dots, m-1$, where the difference between a cubic spline and a natural cubic spline is the last requirement. To shorten the notation in the following, we will use the following:

$$\begin{aligned} h_j &= (t_j - t_{j-1}), & v_j &= 2(h_{j-1} + h_j), \\ b_j &= \frac{1}{h_j}(\beta_{j+1, \text{placebo}} - \beta_{j, \text{placebo}}), & u_j &= 6(b_j - b_{j-1}), \end{aligned}$$

as well as $z_j = P''(t_j)$ and $\beta_{j,\text{placebo}} = P(t_j)$ for $j = 0, \dots, m$. We can, however, make this into a linear spline problem, as the second derivative of the cubic spline is a linear spline. This can be utilised to give us the following expressions

$$\begin{aligned} P_j''(t) &= z_j \frac{t_{j+1} - t}{h_j} + z_{j+1} \frac{t - t_j}{h_j}, \\ P_j'(t) &= \frac{z_{j+1}}{2h_j} (t - t_j)^2 - \frac{z_j}{2h_j} (t_{j+1} - t)^2 + b_j - \frac{h_j}{6} (z_{j+1} - z_j) \\ P_j(t) &= \frac{z_{j+1}}{6h_j} (t - t_j)^3 + \frac{z_j}{6h_j} (t_{j+1} - t)^3 + \left(\frac{\beta_{j+1,\text{placebo}}}{h_j} - \frac{z_{j+1}}{6} h_j \right) (t - t_j) \\ &\quad + \left(\frac{\beta_{j,\text{placebo}}}{h_j} - \frac{z_j}{6} h_j \right) (t_{j+1} - t) \end{aligned}$$

which can be solved using a tridiagonal system of linear equations, given as:

$$\begin{bmatrix} v_1 & h_1 & & & \\ h_1 & v_2 & h_2 & & \\ & h_2 & \ddots & \ddots & \\ & & \ddots & v_{m-2} & h_{m-2} \\ & & & h_{m-2} & v_{m-1} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{m-2} \\ z_{m-1} \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{m-2} \\ u_{m-1} \end{bmatrix}$$

This tridiagonal system of linear equations can be solved to give an exact solution by using Gaussian elimination. The simplified version of Gaussian elimination is isolating z_1 in the first equation, which gives $z_1 = \frac{u_1}{v_1} - \frac{h_1}{v_1} z_2$. Substituting this into the second equation, isolating for z_2 and substituting it into the third equation and so forth will give the solution to the system.

B Additional Evaluations

This chapter includes additional tables and figures not included in Chapter 4.

ADAS-Cog Scores

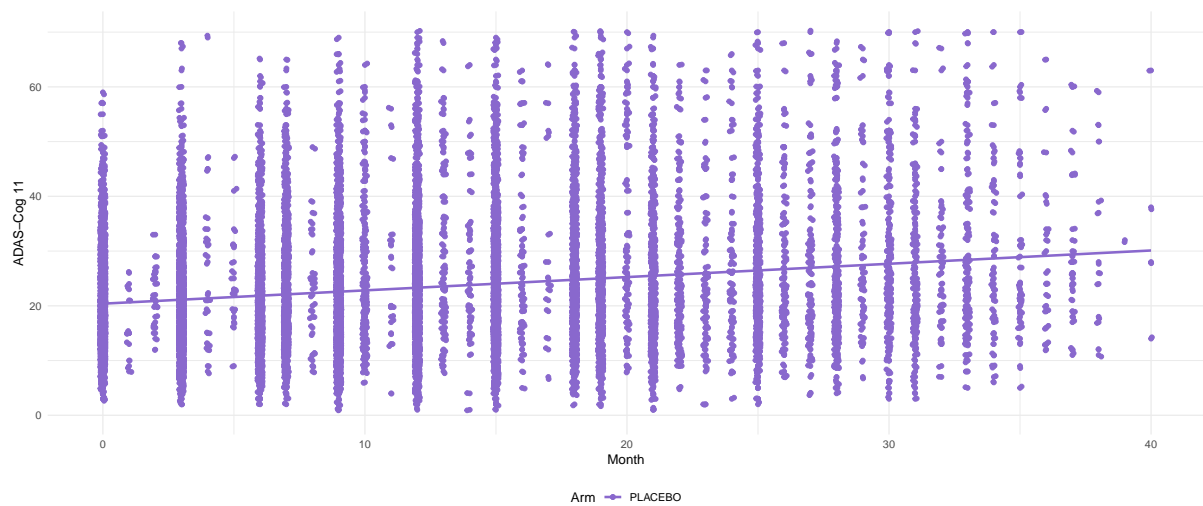


FIGURE B.1: Available data points for subjects with an ADAS-Cog 11 score, alongside a linear mean trajectory.

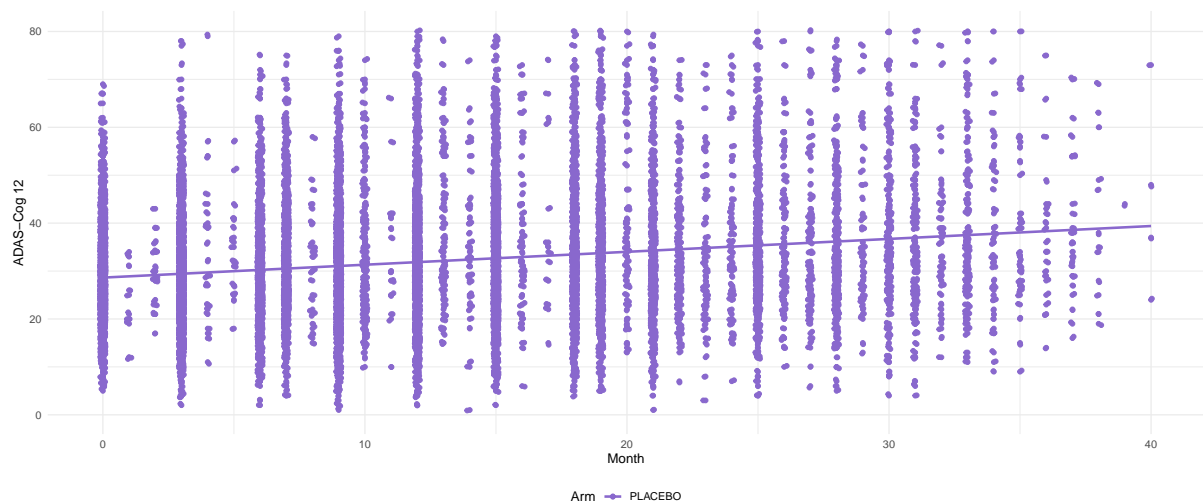


FIGURE B.2: Available data points for subjects with an ADAS-Cog 12 score, alongside a linear mean trajectory.

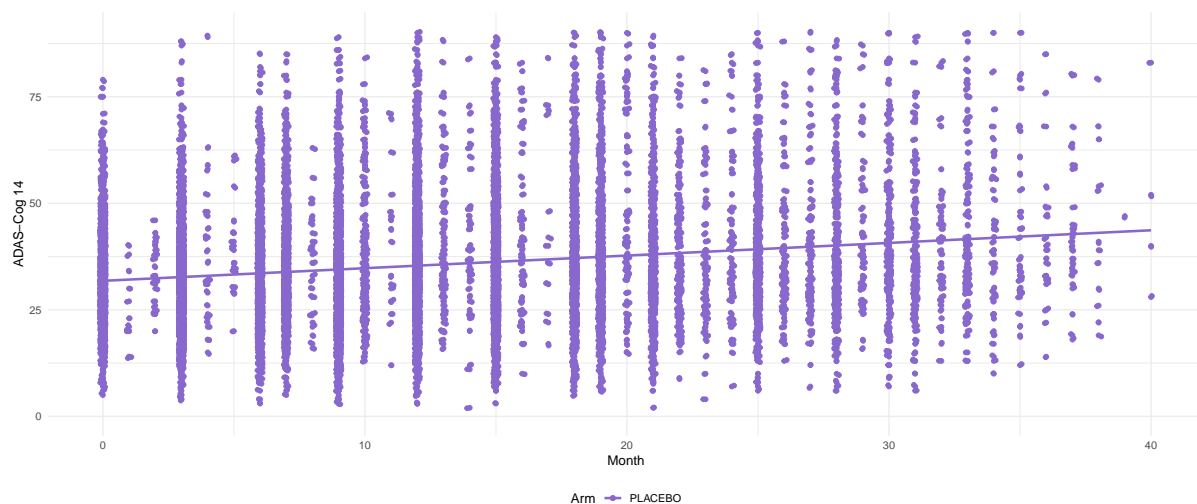


FIGURE B.3: Available data points for subjects with an ADAS-Cog 14 score, alongside a linear mean trajectory.

Distribution of Cognitive Assessment Scales

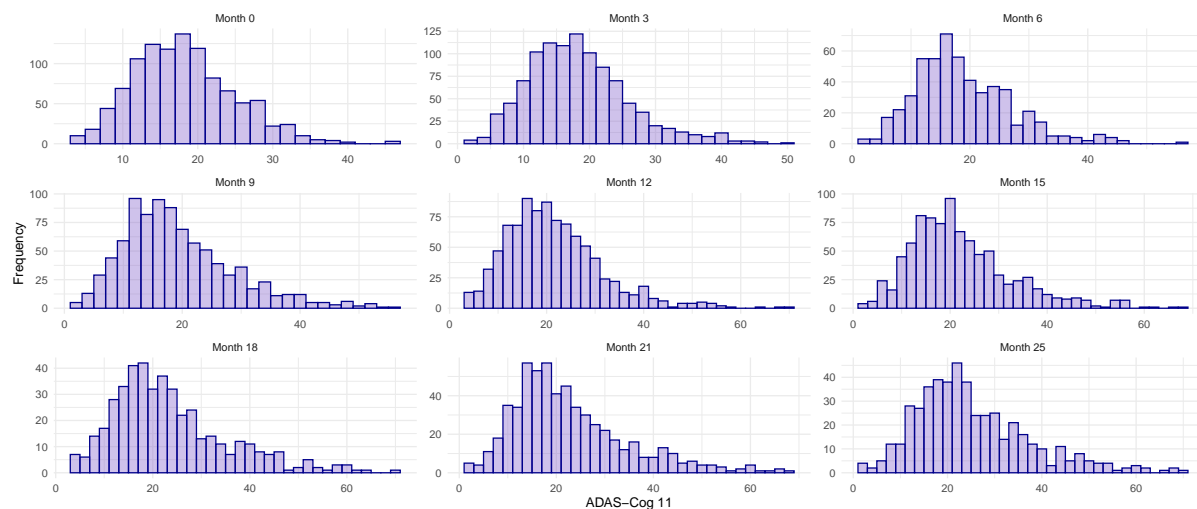


FIGURE B.4: Distribution of the ADAS-Cog 11 scores at months 3, 6, ..., 21, and 25, as these are the months with the most observations.

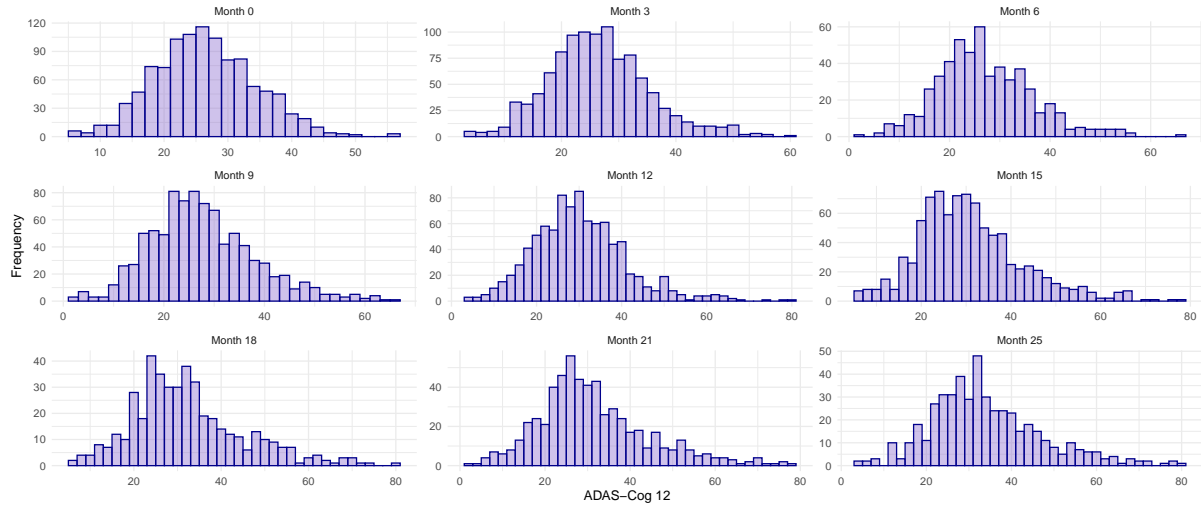


FIGURE B.5: *Distribution of the ADAS-Cog 12 scores at months 3, 6, ..., 21, and 25, as these are the months with the most observations.*

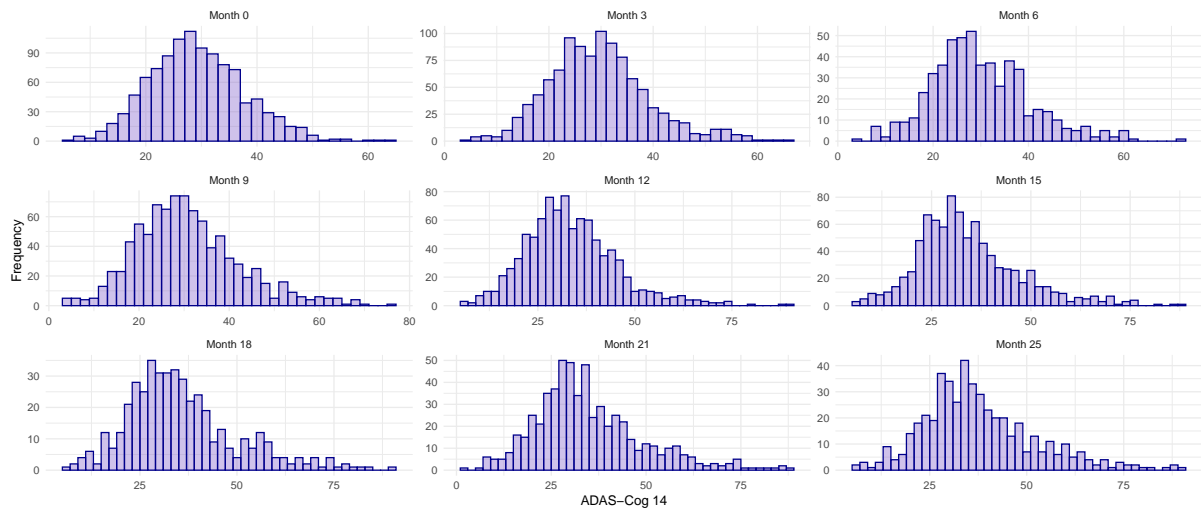


FIGURE B.6: *Distribution of the ADAS-Cog 14 scores at months 3, 6, ..., 21, and 25, as these are the months with the most observations.*

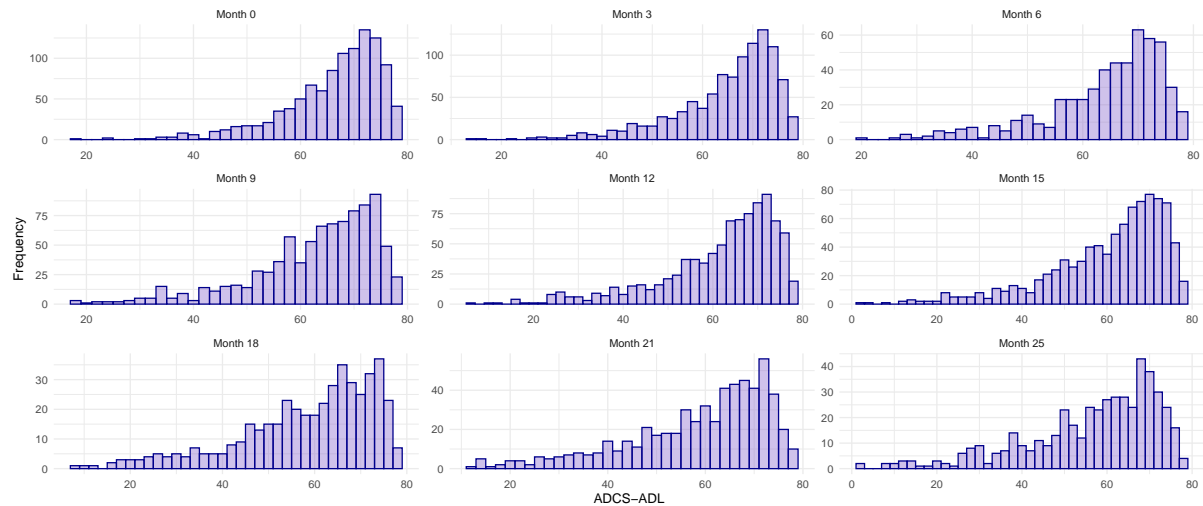


FIGURE B.7: *Distribution of the ADCS-ADL scores at months 3, 6, ..., 21, and 25, as these are the months with the most observations.*

Data Presentation for Subjects with an iADRS Score

The figures and tables in this section are for subjects with an iADRS score.

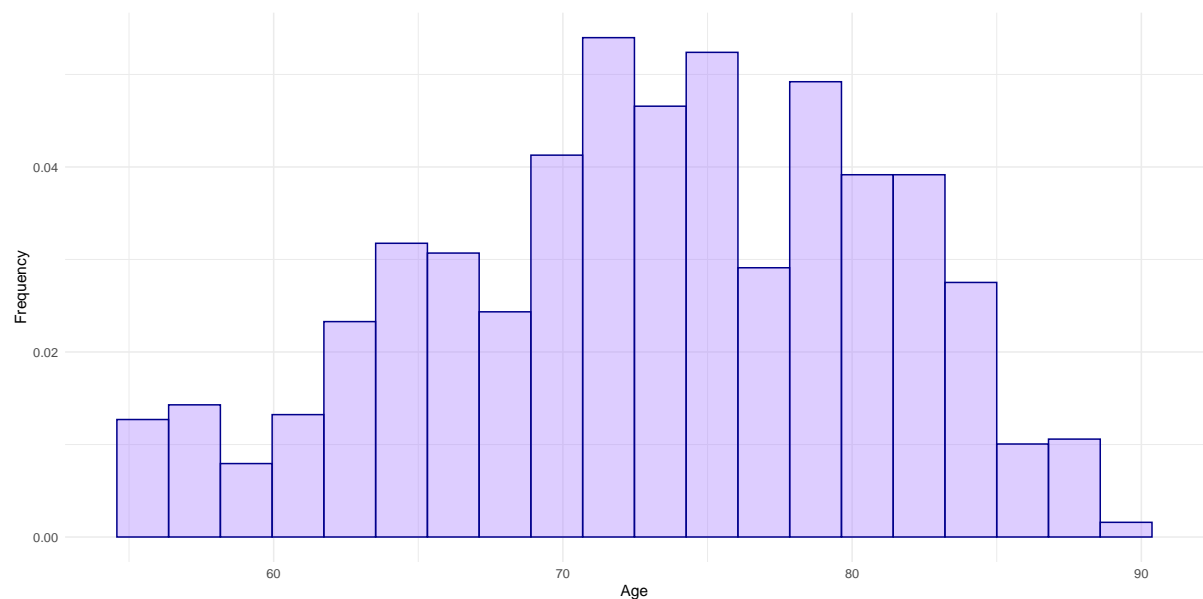


FIGURE B.8: *Histogram of the subjects' ages after sorting for subject with an iADRS score but excluding subjects with ages 999 and 9999.*

Race	%
Not Registered	7.96
Asian	6.65
Black or African American	1.78
Multiple	0.19
White	83.43

TABLE B.9: *The available racial groups and the percentage of subjects within each racial group for subjects with an iADRS score.*

Ethnicity	%
Not Registered	8.29
Hispanic or Latino	4.78
Not Hispanic or Latino	86.93

TABLE B.10: *The available ethnic groups and the percentage of subjects within each ethnic group for subjects with an iADRS score.*

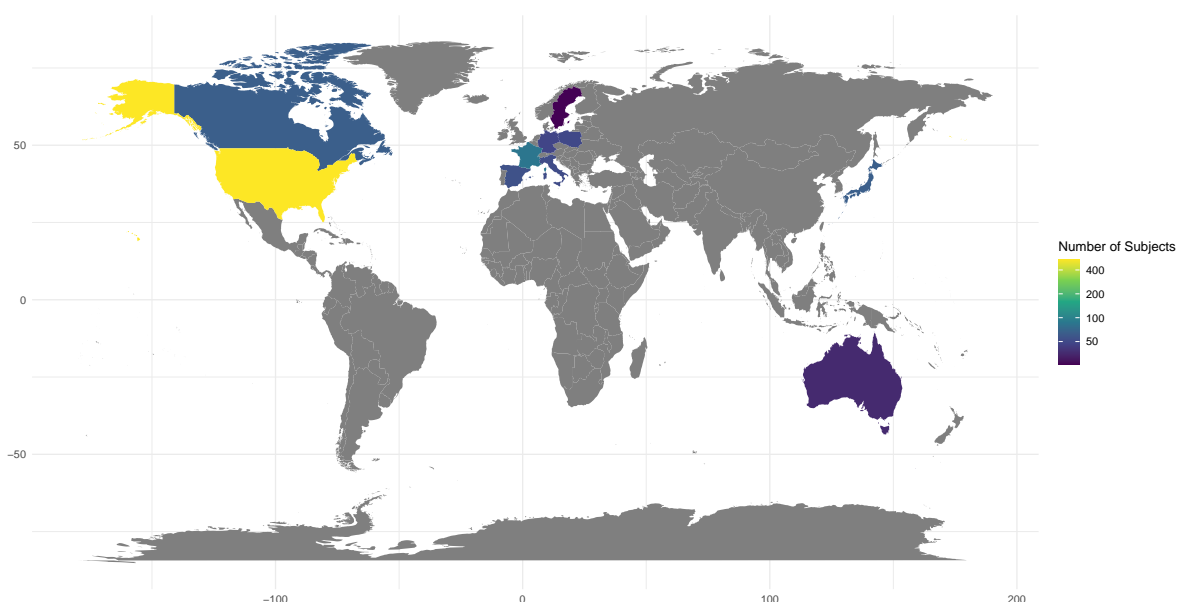


FIGURE B.11: *World map showing the distribution of subjects with an iADRS score. This shows that the majority of countries in North America, and Oceania are represented, whereas only few countries in South America, Africa, Europe, and Asia are represented.*

C Residual Analysis

We start by fitting a PSTPMRM with homoscedastic and non-correlated errors. That is

```
nwcmmmm <- gnls(model = y ~ PSTPMRM(M, v0, v1, v2, v3, v4, v5, v6, v7, v8, b),  
  data = data,  
  params = list(v0 + v1 + v2 + v3 + v4 + v5 + v6 + v7 + v8 ~ 1,  
    b ~ act + 0))
```

Where `data` is the one of the data sets used in the 20% proportional slowing in disease progression scenario. The best way of seeing if there are any indications of heteroskedasticity is by investigating the residuals plotted against either the fitted values or a variance covariate candidate. In our case a variance covariate could for example be the visits. In Figure C.1 the standardized residuals are plotted against the fitted values from `nwcmmrm`.

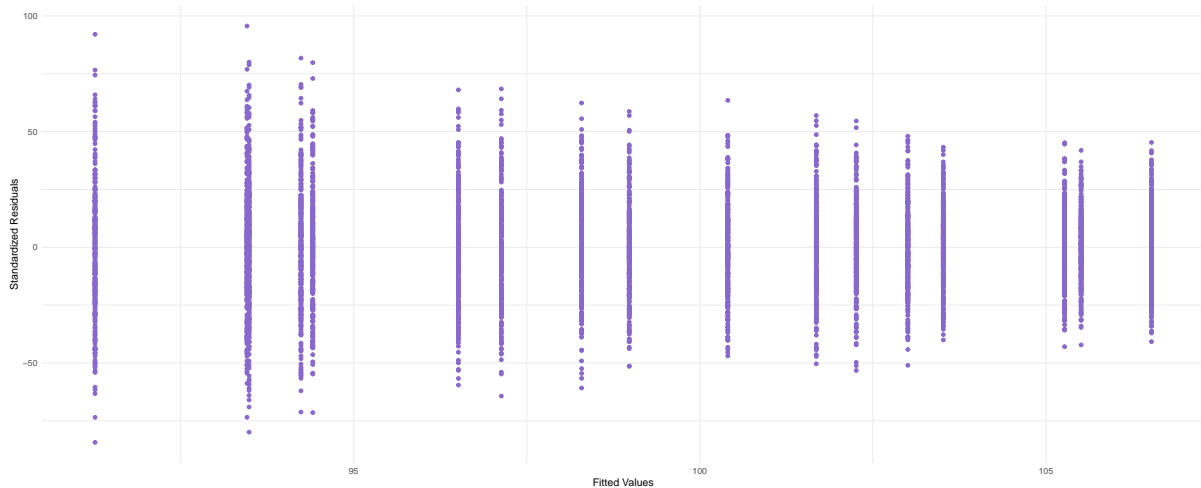


FIGURE C.1: *Standardized residuals plotted against the fitted values, both from the `nwcmmrm`.*

Figure C.1 shows that there is a slight indication that the variance decreases when the values of the fitted values increase. This could potentially be modelled by the `varPower` class of variance functions.

In Figure C.2, we plot the standardized residuals against the possible variance covariate, `visit`.

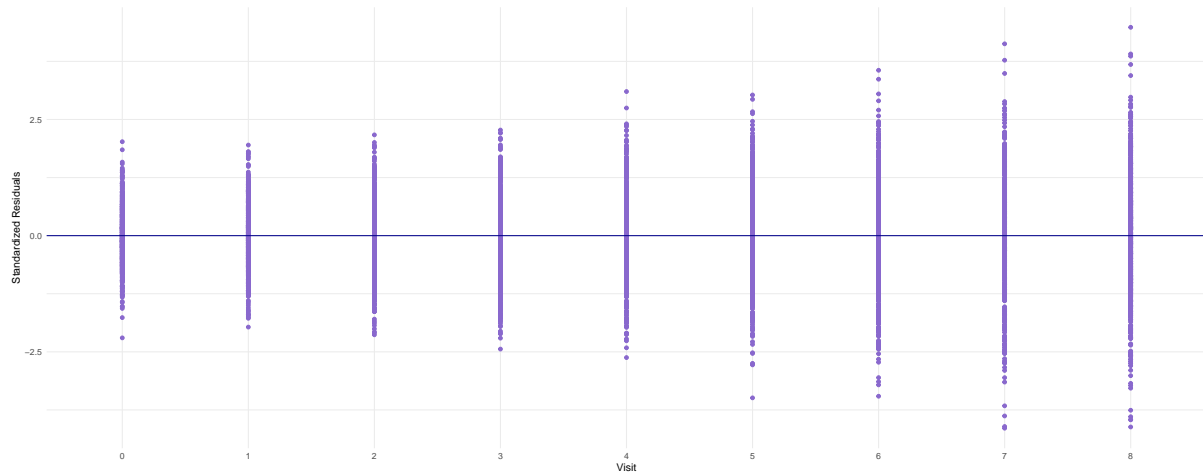


FIGURE C.2: *Standardized residuals plotted from the `nwcmrm` plotted against the variance covariate, visit.*

Here, we see that there is an indication of different variances depending on the visit. This can be modelled using the `varIdent` class of variance functions. To see which of the two classes of functions superior for modelling the potential heteroskedasticity in the model we try both methods, and compare those individually to the `nwcmrm` model, and to each other. This is done by using the `anova` function. The two models are constructed by adding a weight argument to the `gnls` function with `varPower()` and `varIdent()`, respectively. Based on the `anova` function, there is a highly significant decrease in the AIC and BIC when using the `varIdent` variance function. However, this is not seen when using the `varPower` variance function. Furthermore, the use of `varIdent` also shows a significantly decrease in the AIC and BIC compared to the use of `varPower`. Hence, indicating heteroskedasticity depending on the variance covariate "visit".

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
<code>nwcmrm</code>	1	11	48542.63	48615.16	-24260.31			
<code>varIdent_model</code>	2	19	48026.18	48151.47	-23994.09	1 vs 2	532.4483	<.0001

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
<code>nwcmrm</code>	1	11	48542.63	48615.16	-24260.31			
<code>varPower_model</code>	2	13	48546.63	48632.35	-24260.31	1 vs 2	1.74623e-10	1

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
<code>varPower_model</code>	1	19	48026.18	48151.47	-23994.09			
<code>varIdent_model</code>	2	13	48546.63	48632.35	-24260.31	1 vs 2	532.4483	<.0001

Of course, there are many different possibilities of variance structures, however, this variance structure seems to model the heteroskedasticity well. Furthermore, this is also the structure used by [Raket, 2022], and therefore the model is now given as

```
nwcmrm <- gnls(model = y ~ TPMRM_br(M, v0, v1, v2, v3, v4, v5, v6, v7, v8, b, b, b,
  b, b, b, b, b),
  data = data,
  params = list(v0 + v1 + v2 + v3 + v4 + v5 + v6 + v7 + v8 ~ 1,
    b ~ act + 0),
  weights = varIdent(form = ~ 1 | visit))
```


Now that we have checked if there is any heteroskedasticity in the model's errors we would want to test if we also need to include correlation in the errors. Here, it could be evident to look at the within-subject errors, which the autocorrelation function (ACF) can be used for. The ACF for each subject is presented in Figure C.3 which is made by plotting `ACF(ncmmrm, form = ~ 1 | id, maxLag = 8)`. The ACF function computes the autocorrelation function for the residuals of the specified model (`ncmmrm`). The formula `~ 1 | id` indicates that the autocorrelation should be computed separately for each level of the grouping variable `id`.

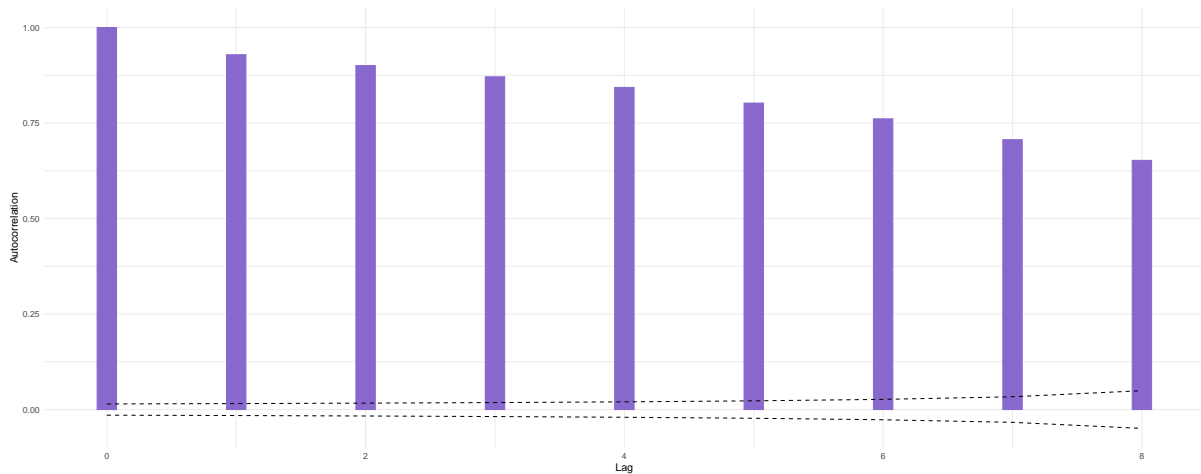


FIGURE C.3: ACF of the model `ncmmrm` computed on the subject-level.

Figure C.3 shows a slow decline in the ACF indicating that there is some autocorrelation present. Hence, we try to fit some correlation structures including the classes `corSymm`, `corCompSymm`, `corAR`, and `corARMA`. We test the models including a correlation structure against the simpler model without correlation included in the model using the `anova` function. Doing so, we see that there is a significant decrease in the AIC and BIC when including a correlation structure. The outputs from the `anova` are presented in Listing 17.

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
<code>ncmmrm</code>	1	19	48026.18	48151.47	-23994.09			
<code>corSymm_model</code>	2	55	37909.70	38272.38	-18899.85	1 vs 2	10188.48	<.0001

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
<code>ncmmrm</code>	1	19	48026.18	48151.47	-23994.09			
<code>corCompSymm_model</code>	2	20	39847.88	39979.76	-19903.94	1 vs 2	8180.301	<.0001

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
<code>ncmmrm</code>	1	19	48026.18	48151.47	-23994.09			
<code>corAR1_model</code>	2	20	38370.70	38502.58	-19165.35	1 vs 2	9657.483	<.0001

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
<code>ncmmrm</code>	1	19	48026.18	48151.47	-23994.09			
<code>corARMA_model(q=1,p=1)</code>	2	21	38079.30	38217.78	-19018.65	1 vs 2	9950.875	<.0001

LISTING C.4: Results from comparing the different models introduced to determine which is best using AIC.

Furthermore, we also see that the correlation structure `corSymm` is preferred compared to

the other correlation structures. This is also the most complex type of correlation structure we can implement in the models, and it can be seen in the code below, where df is 55 when including the `corSymm` correlation structure compared to 20 and 21 in the others. This is caused by the fact that there in this structure adds 36 parameters compared to one and two in the other structures.

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
<code>corCompSymm_model</code>	1	20	39847.88	39979.76	-19903.94			
<code>corSymm_model</code>	2	55	37909.70	38272.38	-18899.85	1 vs 2	2008.176	<.0001

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
<code>corAR1_model</code>	1	20	38370.7	38502.58	-19165.35			
<code>corSymm_model</code>	2	55	37909.7	38272.38	-18899.85	1 vs 2	530.9936	<.0001

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
<code>corARMA_model(q=1,p=1)</code>	1	21	38079.3	38217.78	-19018.65			
<code>corSymm_model</code>	2	55	37909.7	38272.38	-18899.85	1 vs 2	237.6014	<.0001

Hence, the model that should be used to determine initial values is the following:

```
PSIPMRM <- gnls(model = y ~ TPMRM_br(M, v0, v1, v2, v3, v4, v5, v6, v7, v8, b, b, b,
  b, b, b, b, b),
  data = data,
  params = list(v0 + v1 + v2 + v3 + v4 + v5 + v6 + v7 + v8 ~ 1,
    b ~ act + 0),
  weights = varIdent(form = ~ 1 | visit),
  correlation = corSymm(value = cor_vec, form = ~ as.numeric(visit)
    | id))
```

D Simulation Study Results

This chapter expands the results of the power analyses conducted in Chapter 5. The chapter will therefore primarily only present the results without any associated text, apart from the captions of the different tables and figures. The data, tables, and figures can also be found in <https://github.com/loprtq/Master-s-Thesis-Results>.

D.1 Optimisation Algorithm for True Effects

As described in Chapter 5, the true effects are derived from the mean trajectories of the placebo- and active arm. The placebo- and active arm from each simulation in each scenario is inserted into an optimisation algorithm, which either uses SANN or Nelder-Mead. SANN is used to determine the proportional slowing and proportional reduction in decline treatment effects, whereas Nelder-Mead is used to determine the time-based changes. For this we use `optim()` in a similar fashion to how the different active arms were derived. Take for example that we wish to determine the proportional reduction in decline of a data set, then we determine which ζ that minimises the mean square error (MSE) problem:

$$(\beta_{j,\text{active}} - (\zeta(\beta_{j,\text{placebo}} - \beta_{0,\text{placebo}}) + \beta_{0,\text{placebo}}))^2,$$

for all j . $\beta_{j,\text{active}}$ and $\beta_{j,\text{placebo}}$ are the mean trajectory at visit j , for the active- and placebo arm, respectively. Similarly can be done for the PSTPMRM and TPMRM where the MSE problem is $(\beta_{j,\text{active}} - \beta_{j,\zeta_j,\text{placebo}})^2$ where ζ_j can either vary (TPMRM) or be constant, $\zeta_j = \zeta$ (PSTPMRM).

D.2 Convergence of Estimates

300 Subjects in each arm

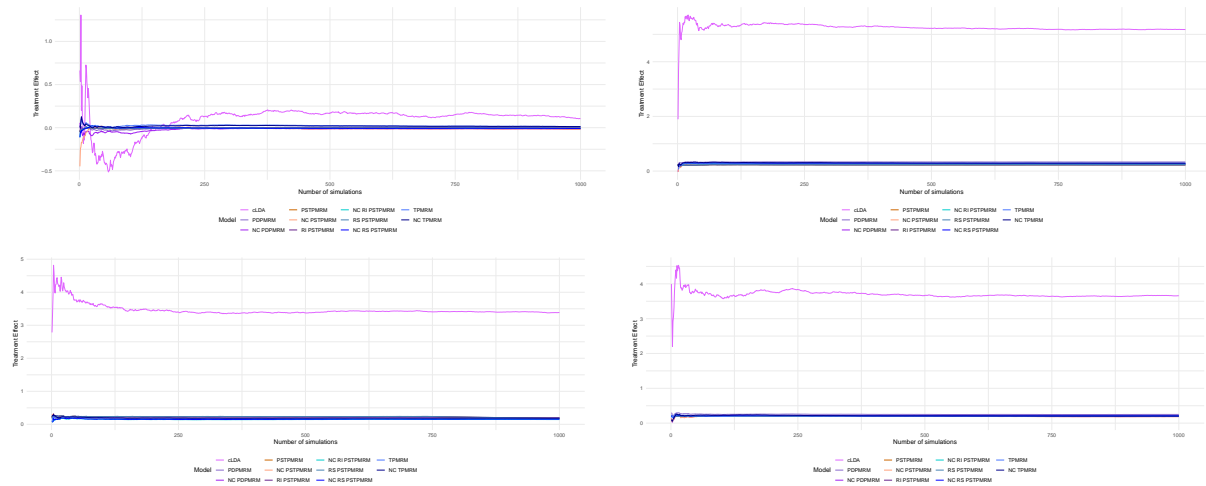


FIGURE D.1: Mean of estimates of the progression models for repeated measures and constrained longitudinal data analysis model throughout 1000 simulations for 300 subjects in each arm.

500 Subjects in each arm

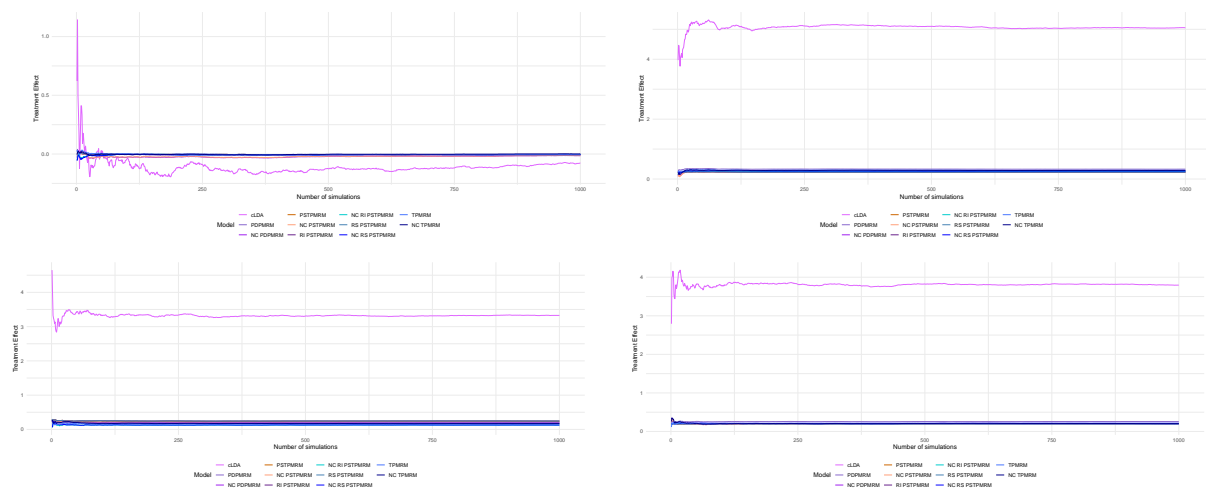


FIGURE D.2: Mean of estimates of the progression models for repeated measures and constrained longitudinal data analysis model throughout 1000 simulations for 500 subjects in each arm.

1000 Subjects in each arm

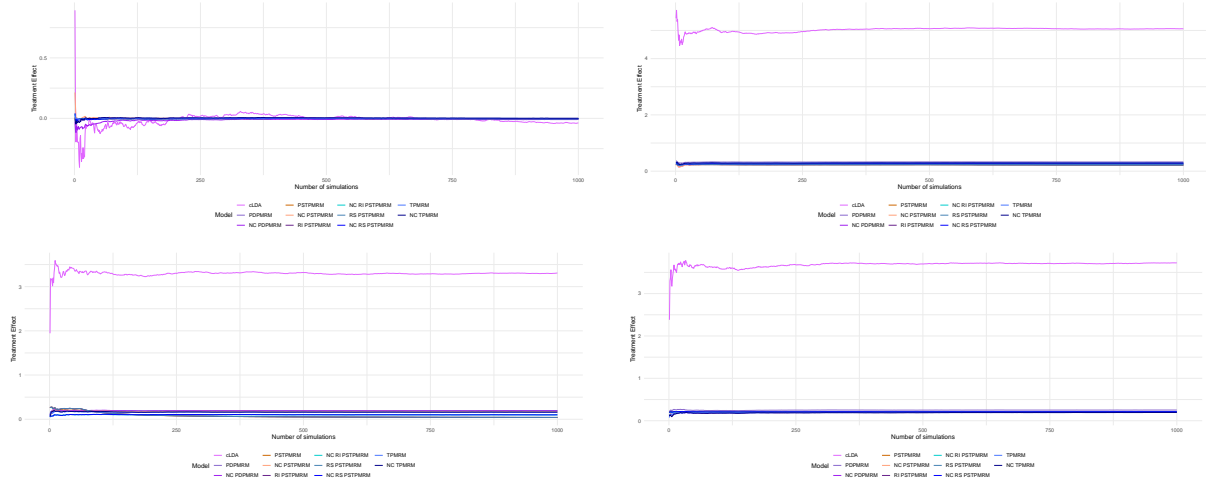


FIGURE D.3: Mean of estimates of the progression models for repeated measures and constrained longitudinal data analysis model throughout 1000 simulations for 1000 subjects in each arm.

D.3 Convergence of Statistical Power

300 Subjects in each arm

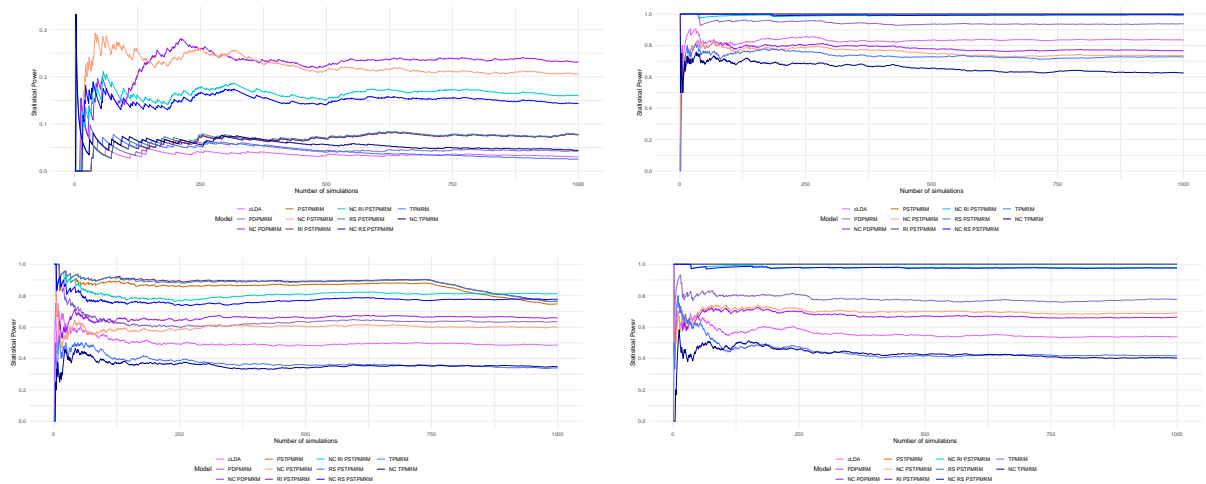


FIGURE D.4: Statistical power of the progressions model for repeated measures and constrained longitudinal data analysis model throughout 1000 simulations for 300 subjects in each arm.

500 Subjects in each arm

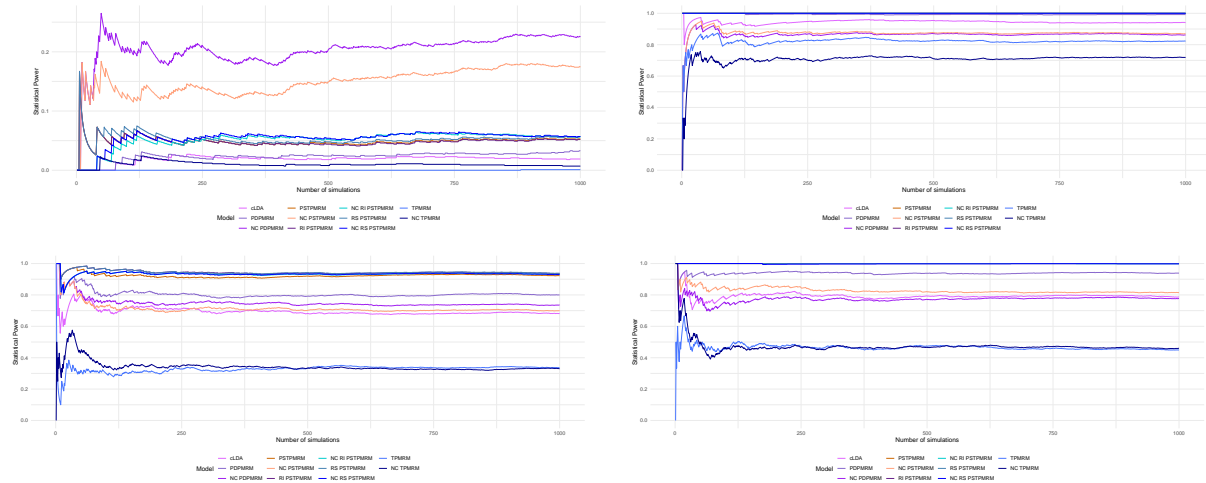


FIGURE D.5: Statistical power of the progressions model for repeated measures and constrained longitudinal data analysis model throughout 1000 simulations for 500 subjects in each arm.

1000 Subjects in each arm

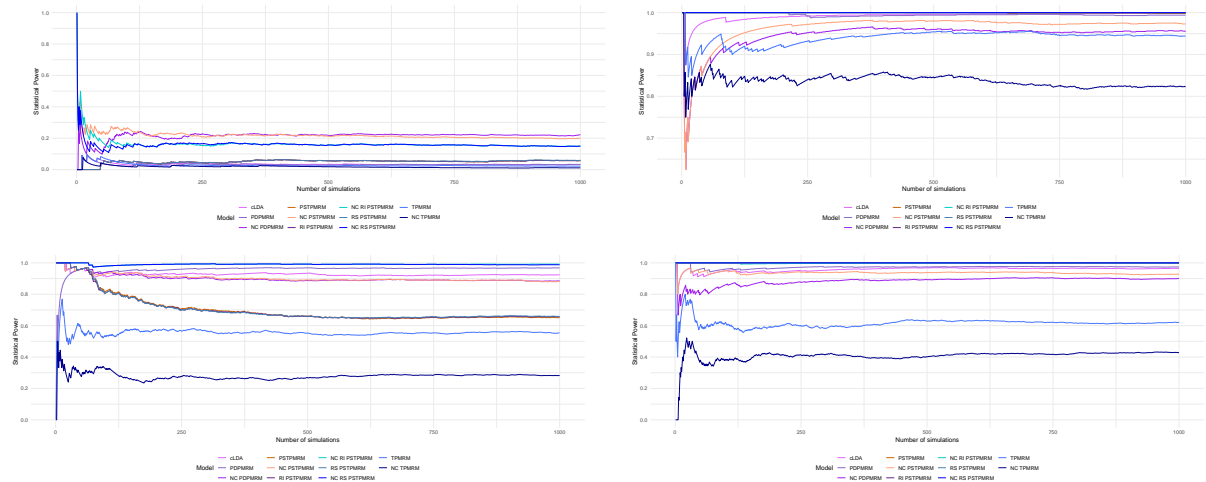


FIGURE D.6: Statistical power of the progression models for repeated measures and constrained longitudinal data analysis model throughout 1000 simulations for 1000 subjects in each arm.

D.4 Parameter Estimates

300 Subjects in each arm

Scenario	Model	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
No effect	cLDA	0.046	0.060	130.435	3.934
	PSTPMRM	0.941	0.063	6.694	0.092
	NC PSTPMRM	0.941	0.078	8.289	0.187
	RI PSTPMRM	0.941	0.063	6.694	0.092
	NC RI PSTPMRM	0.941	0.067	7.122	0.093
	RS PSTPMRM	0.941	0.063	6.694	0.092
	NC RS PSTPMRM	0.941	0.068	7.226	0.095
	PDPMRM	1.012	-0.011	-1.087	0.183
	NC PDPMRM	1.012	0.006	0.593	0.030
	TPMRM	0.943	0.044	4.665	0.088
	NC TPMRM	0.943	0.045	4.772	0.074
20% Proportional reduction in decline	cLDA	3.457	-0.071	-2.053	1.534
	PSTPMRM	0.799	0.018	2.253	0.154
	NC PSTPMRM	0.799	0.016	2.003	0.057
	RI PSTPMRM	0.799	0.014	1.751	0.150
	NC RI PSTPMRM	0.799	0.056	7.009	0.091
	RS PSTPMRM	0.799	0.014	1.751	0.152
	NC RS PSTPMRM	0.799	0.045	5.631	0.102
	PDPMRM	0.797	0.004	0.502	0.169
	NC PDPMRM	0.797	0.008	1.004	0.041
	TPMRM	0.797	0.031	3.889	0.114
	NC TPMRM	0.797	0.018	2.259	0.076

Continued on next page

Table D.7: Mean true treatment effect, mean bias, relative bias, and estimation standard error of the progression models for repeated measures and constrained longitudinal data analysis model in different scenarios for 300 subjects in each arm.

Scenario	Model	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
20% Proportional slowing	cLDA	3.684	-0.027	-0.733	1.621
	PSTPMRM	0.802	-0.004	-0.499	0.130
	NC PSTPMRM	0.802	0.012	1.497	0.056
	RI PSTPMRM	0.802	-0.004	-0.499	0.130
	NC RI PSTPMRM	0.802	-0.004	-0.499	0.098
	RS PSTPMRM	0.802	-0.004	-0.499	0.130
	NC RS PSTPMRM	0.802	-0.013	-1.621	0.097
	PDPMRM	0.802	-0.055	-6.860	0.181
	NC PDPMRM	0.802	0.006	0.748	0.043
	TPMRM	0.788	0.017	2.157	0.108
	NC TPMRM	0.788	0.011	1.396	0.058
10 to 30% Time based changes	cLDA	5.135	0.041	0.798	1.615
	PSTPMRM	0.750	0.023	3.067	0.141
	NC PSTPMRM	0.750	0.021	2.800	0.061
	RI PSTPMRM	0.750	0.024	3.200	0.141
	NC RI PSTPMRM	0.750	-0.004	-0.533	0.097
	RS PSTPMRM	0.750	0.023	3.067	0.142
	NC RS PSTPMRM	0.750	-0.024	-3.200	0.098
	PDPMRM	0.759	-0.092	-12.116	0.177
	NC PDPMRM	0.759	-0.001	-0.132	0.043
	TPMRM	0.696	0.015	2.155	0.114
	NC TPMRM	0.696	0.016	2.299	0.069

Table D.7: Mean true treatment effect, mean bias, relative bias, and estimation standard error of the progression models for repeated measures and constrained longitudinal data analysis model in different scenarios for 300 subjects in each arm.

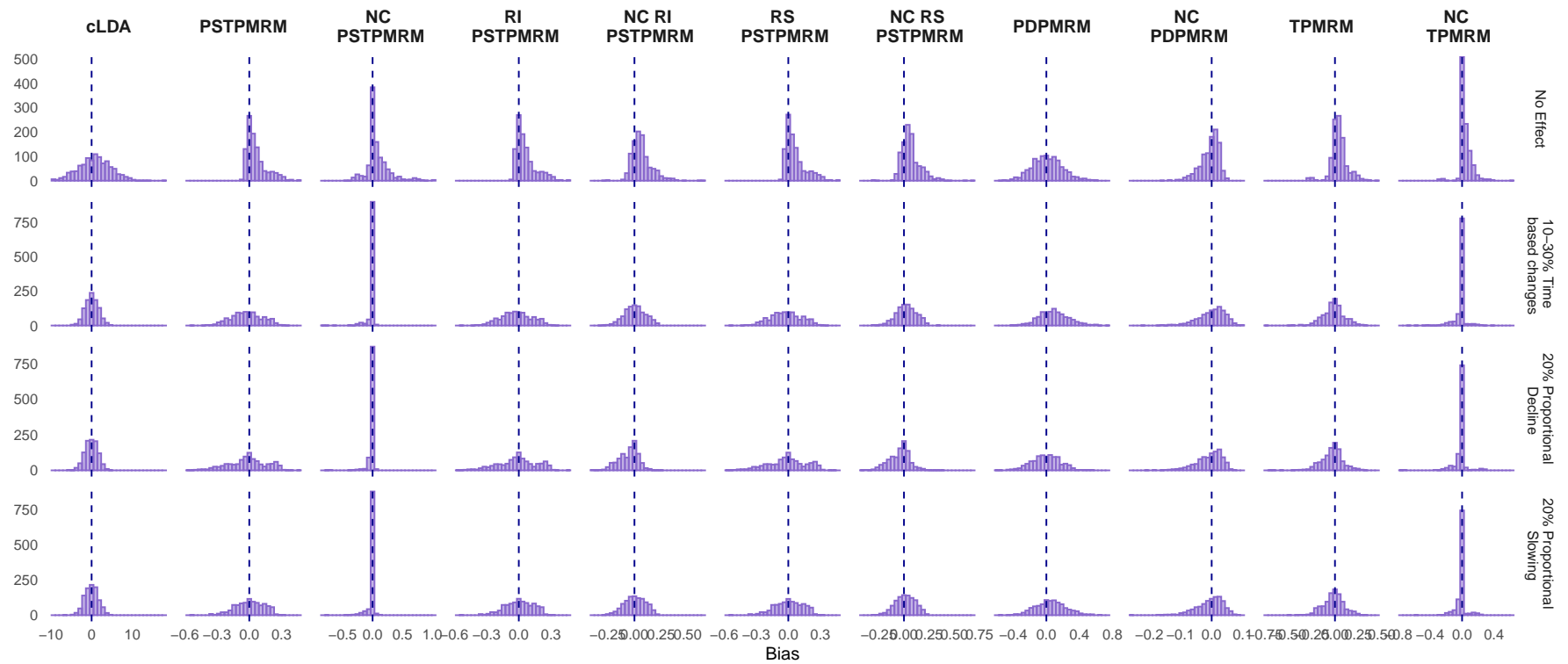


FIGURE D.8: Bias for each of the 1000 effect estimates from each simulated data set. Estimates are from data sets with 300 subjects in each arm.

500 Subjects in each arm

Scenario	Model	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
No effect	cLDA	-0.080	0.003	-3.750	2.935
	PSTPMRM	0.958	0.046	4.802	0.068
	NC PSTPMRM	0.958	0.055	5.742	0.091
	RI PSTPMRM	0.958	0.045	4.696	0.068
	NC RI PSTPMRM	0.958	0.047	4.906	0.082
	RS PSTPMRM	0.958	0.045	4.696	0.068
	NC RS PSTPMRM	0.958	0.048	5.009	0.084
	PDPMRM	1.010	-0.004	-0.396	0.144
	NC PDPMRM	1.010	0.004	0.396	0.024
	TPMRM	0.961	0.043	4.475	0.062
	NC TPMRM	0.961	0.038	3.954	0.049
20% Proportional reduction in decline	cLDA	3.289	0.035	1.064	1.310
	PSTPMRM	0.813	-0.051	-6.272	0.140
	NC PSTPMRM	0.813	0.013	1.599	0.028
	RI PSTPMRM	0.813	-0.058	-7.134	0.134
	NC RI PSTPMRM	0.813	0.061	7.506	0.152
	RS PSTPMRM	0.813	-0.057	-7.011	0.134
	NC RS PSTPMRM	0.813	0.058	7.135	0.161
	PDPMRM	0.810	-0.008	-0.988	0.144
	NC PDPMRM	0.810	0.003	0.370	0.034
	TPMRM	0.812	0.018	2.216	0.095
	NC TPMRM	0.812	0.017	2.094	0.051

Continued on next page

Table D.9: Mean true treatment effect, mean bias, relative bias, and estimation standard error of the progression models for repeated measures and constrained longitudinal data analysis model in different scenarios for 500 subjects in each arm.

Scenario	Model	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
20% Proportional slowing	cLDA	3.850	-0.056	-1.455	1.282
	PSTPMRM	0.796	0.004	0.502	0.113
	NC PSTPMRM	0.796	0.004	0.502	0.032
	RI PSTPMRM	0.796	0.005	0.628	0.113
	NC RI PSTPMRM	0.796	0.000	0.000	0.120
	RS PSTPMRM	0.796	0.005	0.628	0.113
	NC RS PSTPMRM	0.796	-0.010	-1.257	0.122
	PDPMRM	0.790	-0.047	-5.949	0.142
	NC PDPMRM	0.790	0.005	0.633	0.034
	TPMRM	0.781	0.021	2.688	0.091
	NC TPMRM	0.781	0.014	1.793	0.035
10 to 30% Time based changes	cLDA	5.078	-0.024	-0.473	1.247
	PSTPMRM	0.751	0.020	2.663	0.115
	NC PSTPMRM	0.751	0.012	1.598	0.027
	RI PSTPMRM	0.751	0.020	2.663	0.114
	NC RI PSTPMRM	0.751	-0.003	-0.399	0.123
	RS PSTPMRM	0.751	0.020	2.663	0.114
	NC RS PSTPMRM	0.751	-0.019	-2.530	0.131
	PDPMRM	0.754	-0.084	-11.141	0.137
	NC PDPMRM	0.754	-0.002	-0.265	0.034
	TPMRM	0.700	0.010	1.429	0.087
	NC TPMRM	0.700	0.012	1.714	0.039

Table D.9: Mean true treatment effect, mean bias, relative bias, and estimation standard error of the progression models for repeated measures and constrained longitudinal data analysis model in different scenarios for 500 subjects in each arm.

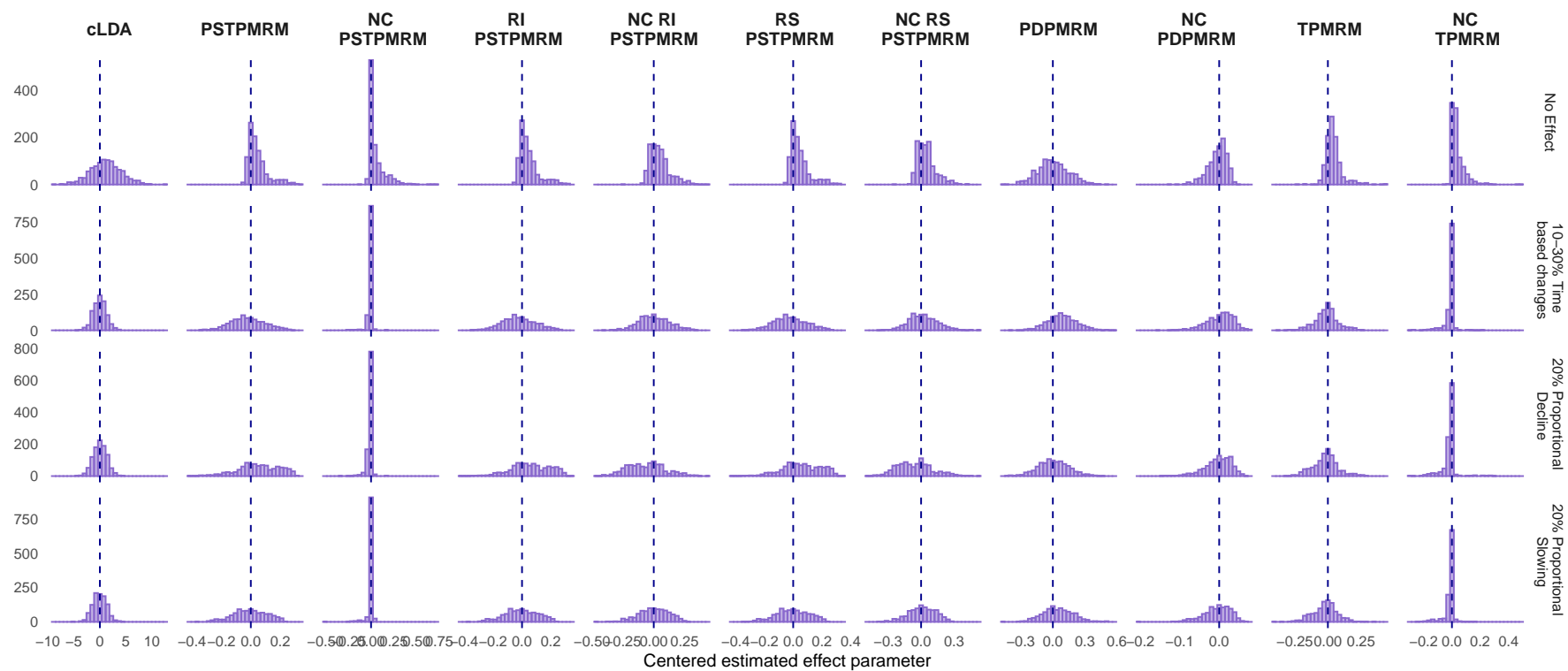


FIGURE D.10: Bias for each of the 1000 effect estimates from each simulated data set. Estimates are from data sets with 500 subjects in each arm.

1000 Subjects in each arm

Scenario	Model	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
No effect	cLDA	-0.047	0.010	-21.276	2.277
	PSTPMRM	0.969	0.033	3.404	0.052
	NC PSTPMRM	0.969	0.036	3.713	0.080
	RI PSTPMRM	0.969	0.033	3.404	0.052
	NC RI PSTPMRM	0.969	0.035	3.616	0.051
	RS PSTPMRM	0.969	0.033	3.404	0.052
	NC RS PSTPMRM	0.969	0.035	3.616	0.051
	PDPMRM	1.007	-0.003	-0.298	0.107
	NC PDPMRM	1.007	0.002	0.199	0.018
	TPMRM	0.970	0.031	3.196	0.044
	NC TPMRM	0.970	0.029	2.990	0.047
20% Proportional reduction in decline	cLDA	3.281	0.026	0.791	0.893
	PSTPMRM	0.814	0.046	5.648	0.103
	NC PSTPMRM	0.814	0.011	1.352	0.019
	RI PSTPMRM	0.814	0.045	5.526	0.103
	NC RI PSTPMRM	0.814	0.008	0.983	0.074
	RS PSTPMRM	0.814	0.045	5.526	0.104
	NC RS PSTPMRM	0.814	0.008	0.983	0.077
	PDPMRM	0.807	-0.004	-0.496	0.100
	NC PDPMRM	0.807	0.002	0.248	0.026
	TPMRM	0.814	0.026	3.195	0.073
	NC TPMRM	0.814	0.025	3.072	0.041

Continued on next page

Table D.11: Mean true treatment effect, mean bias, relative bias, and estimation standard error of the progression models for repeated measures and constrained longitudinal data analysis model in different scenarios for 1000 subjects in each arm.

Scenario	Model	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
20% Proportional slowing	cLDA	3.729	-0.051	-1.369	0.887
	PSTPMRM	0.804	-0.009	-1.119	0.084
	NC PSTPMRM	0.804	0.001	0.014	0.022
	RI PSTPMRM	0.804	-0.010	-1.529	0.082
	NC RI PSTPMRM	0.804	-0.008	-2.868	0.063
	RS PSTPMRM	0.804	-0.010	-1.529	0.082
	NC RS PSTPMRM	0.804	-0.017	-2.207	0.062
	PDPMRM	0.797	-0.049	-5.912	0.104
	NC PDPMRM	0.797	0.003	-0.910	0.029
	TPMRM	0.788	0.018	2.284	0.073
	NC TPMRM	0.788	0.018	2.871	0.037
10 to 30% Time based changes	cLDA	5.045	0.013	0.258	0.895
	PSTPMRM	0.753	0.030	3.984	0.087
	NC PSTPMRM	0.753	0.012	2.001	0.020
	RI PSTPMRM	0.753	0.030	4.347	0.087
	NC RI PSTPMRM	0.753	-0.004	0.218	0.057
	RS PSTPMRM	0.753	0.030	4.345	0.086
	NC RS PSTPMRM	0.753	-0.019	-2.453	0.055
	PDPMRM	0.755	-0.084	-12.881	0.099
	NC PDPMRM	0.755	-0.003	-0.079	0.031
	TPMRM	0.700	0.004	1.753	0.060
	NC TPMRM	0.700	0.010	2.194	0.034

Table D.11: Mean true treatment effect, mean bias, relative bias, and estimation standard error of the progression models for repeated measures and constrained longitudinal data analysis model in different scenarios for 1000 subjects in each arm.

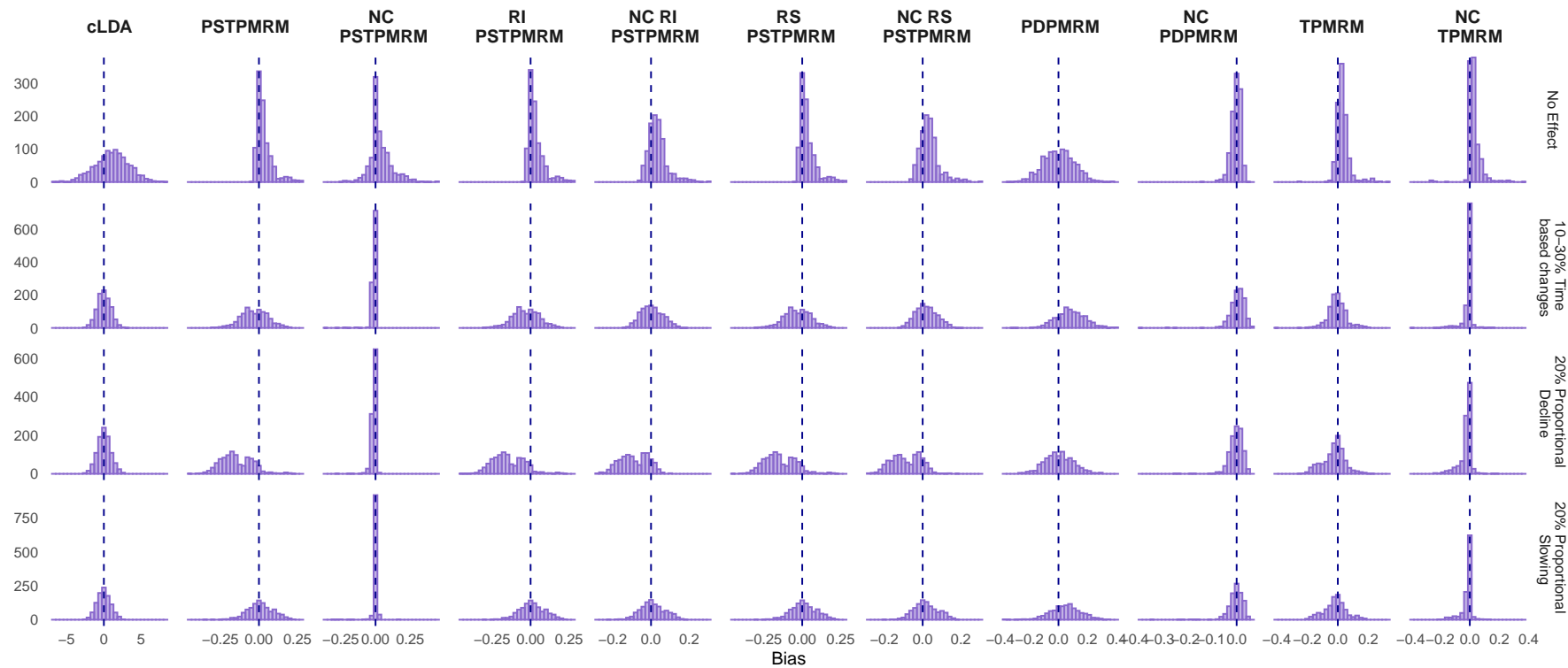


FIGURE D.12: Bias for each of the 1000 effect estimates from each simulated data set. Estimates are from data sets with 1000 subjects in each arm.

D.5 Statistical Power

Subjects per arm	Scenario	cLDA	PDPMRM	NC PDPMRM	PSTPMRM	NC PSTPMRM	RI PSTPMRM	NC RI PSTPMRM	RS PSTPMRM	NC RS PSTPMRM	TPMRM	NC TPMRM
300	No effect	0.030	0.043	0.231	0.077	0.205	0.077	0.161	0.078	0.144	0.025	0.044
	20% Proportional reduction in decline	0.486	0.634	0.659	0.748	0.600	0.765	0.813	0.765	0.778	0.340	0.348
	20% proportional slowing	0.538	0.776	0.664	1.000	0.690	1.000	0.979	1.000	0.976	0.417	0.403
	10→30% time based changes	0.834	0.938	0.765	1.000	0.735	1.000	0.995	1.000	0.995	0.726	0.624
500	No effect	0.019	0.033	0.225	0.053	0.174	0.052	0.056	0.057	0.057	0.001	0.007
	20% Proportional reduction in decline	0.681	0.799	0.735	0.922	0.699	0.937	0.931	0.938	0.930	0.336	0.330
	20% Proportional slowing	0.787	0.939	0.777	1.000	0.815	1.000	0.999	1.000	0.998	0.449	0.460
	10 to 30% Time based changes	0.941	0.992	0.862	1.000	0.870	1.000	1.000	1.000	1.000	0.823	0.719
1000	No effect	0.032	0.033	0.222	0.057	0.201	0.057	0.150	0.059	0.149	0.022	0.011
	20% Proportional reduction in decline	0.924	0.968	0.885	0.651	0.881	0.657	0.992	0.659	0.988	0.554	0.283
	20% proportional slowing	0.965	0.973	0.900	0.996	0.928	1.000	0.999	1.000	1.000	0.621	0.428
	10→30% time based changes	0.998	0.994	0.955	0.998	0.973	1.000	1.000	1.000	1.000	0.944	0.823

cLDA: Constrained longitudinal data analysis, **PDPMRM:** Proportional reduction in decline progression model for repeated measures, **PSTPMRM:** Proportional slowing of disease progression progression model for repeated measures, **TPMRM:** Time based changes in disease progression model for repeated measures, **RI:** Random intercept, **RS:** Random scaling factor, **NC:** Not correlated error terms.

TABLE D.13: Statistical power of progression models for repeated measures and the constrained longitudinal data analysis model in different scenarios, and for a different number of subjects in each arm.

D.6 Time Homogeneity

300 Subjects in each arm

In Figure D.14 78% of the null hypotheses are not rejected for NE, 82% for 10 to 30% time based changes, 40% are not rejected for 20% PD, and 93% are not rejected for 20% PS.

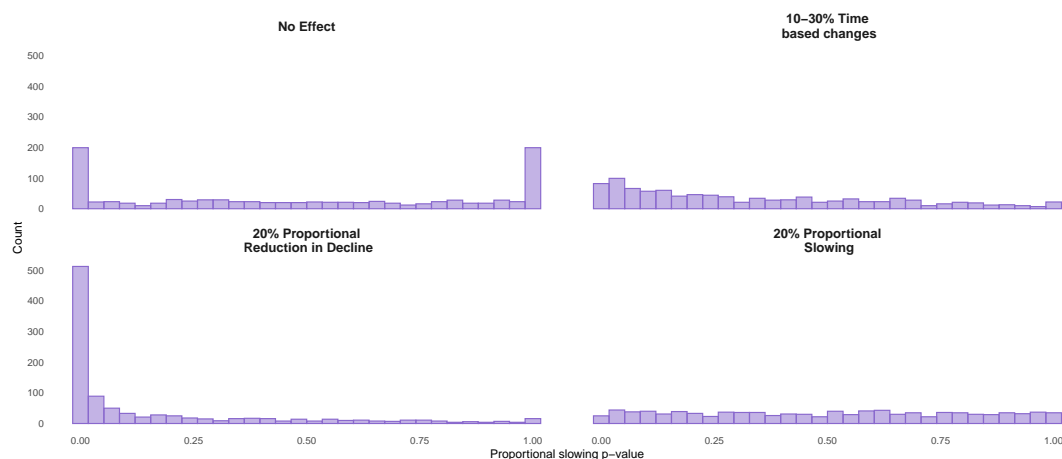


FIGURE D.14: Time homogeneity p -values testing the assumption of proportional slowing in the scenarios; no effect, 10 to 30% time based changes, 20% proportional reduction in decline, 20% proportional slowing, where there are 300 subjects in each arm.

1000 Subjects in each arm

In Figure D.15 76% of the null hypotheses are not rejected for NE, 34% for 10 to 30% time based changes, 34% are not rejected for 20% PD, and 90% are not rejected for 20% PS.

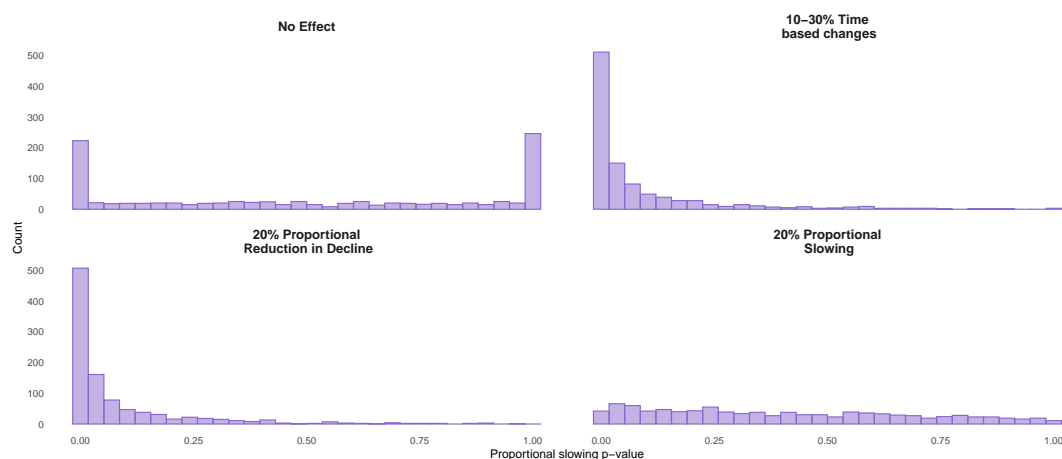


FIGURE D.15: Time homogeneity p -values testing the assumption of proportional slowing in the scenarios; no effect, 10 to 30% time based changes, 20% proportional reduction in decline, 20% proportional slowing, where there are 1000 subjects in each arm.

E Results from the Subgroup Analysis

E.1 Convergence of Estimates

300 Subjects in each arm

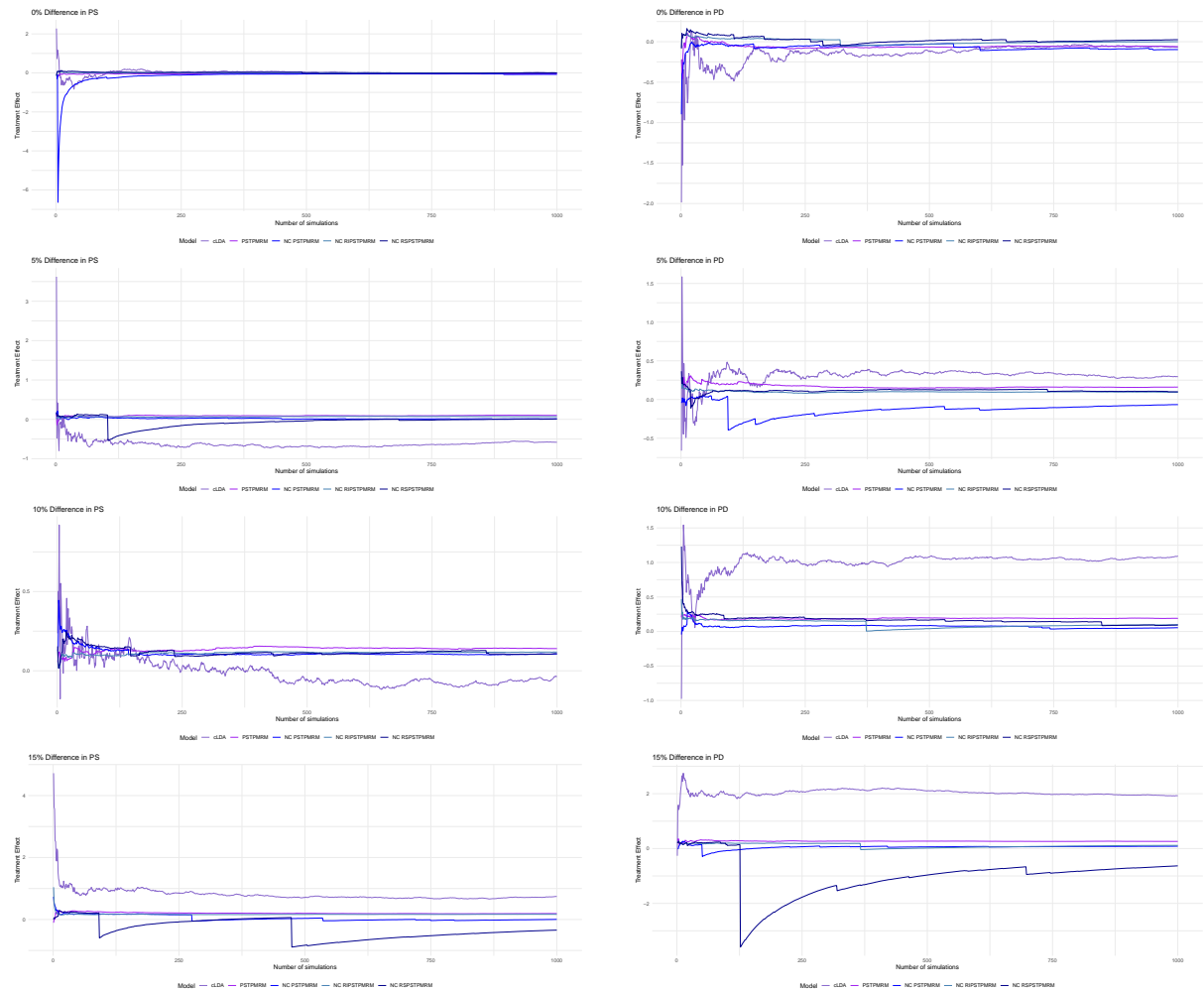


FIGURE E.1: Mean of estimates of the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model throughout 1000 simulations for 300 subjects in each arm.

500 Subjects in each arm

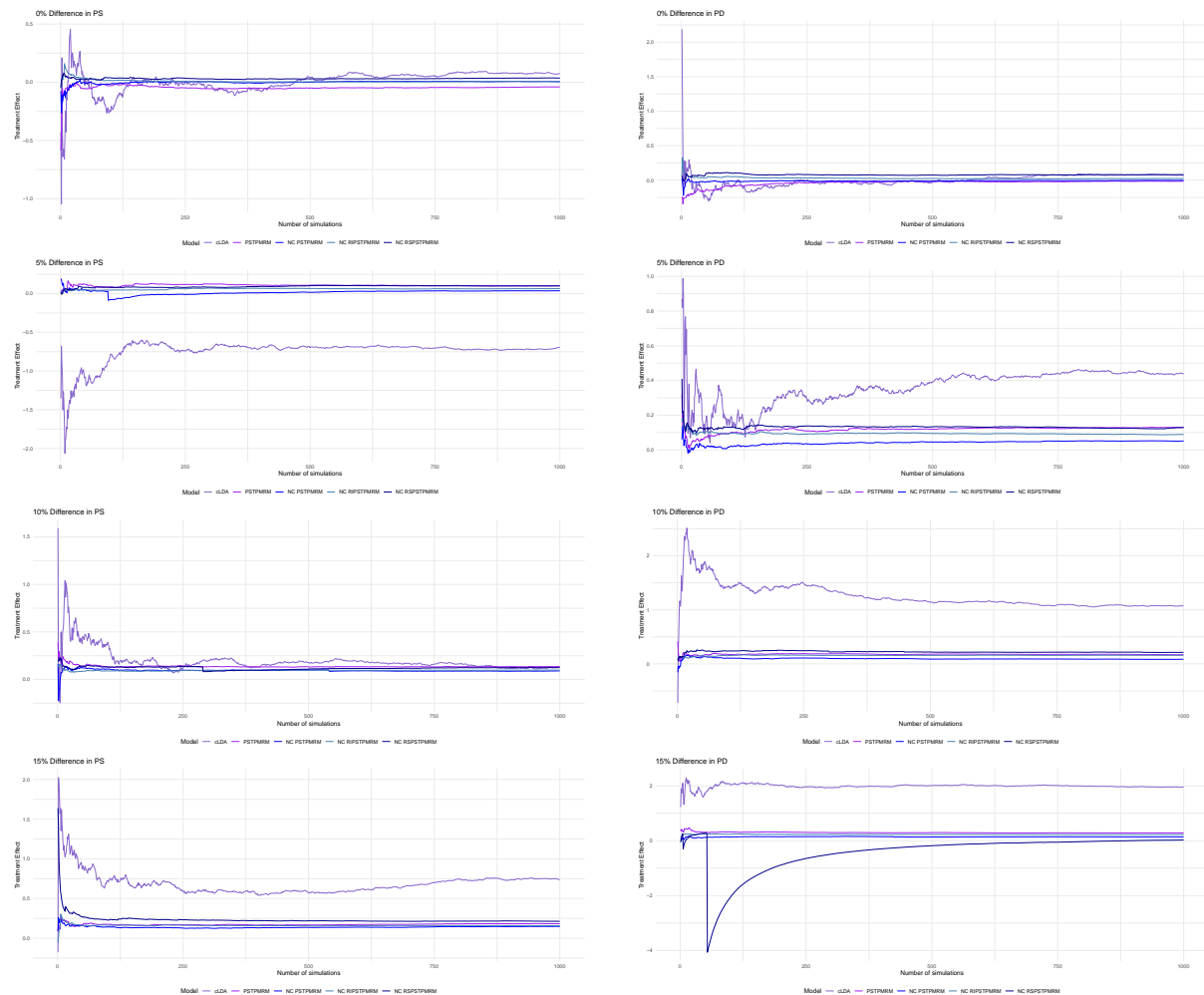


FIGURE E.2: Mean of estimates of the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model throughout 1000 simulations for 500 subjects in each arm.

1000 Subjects in each arm

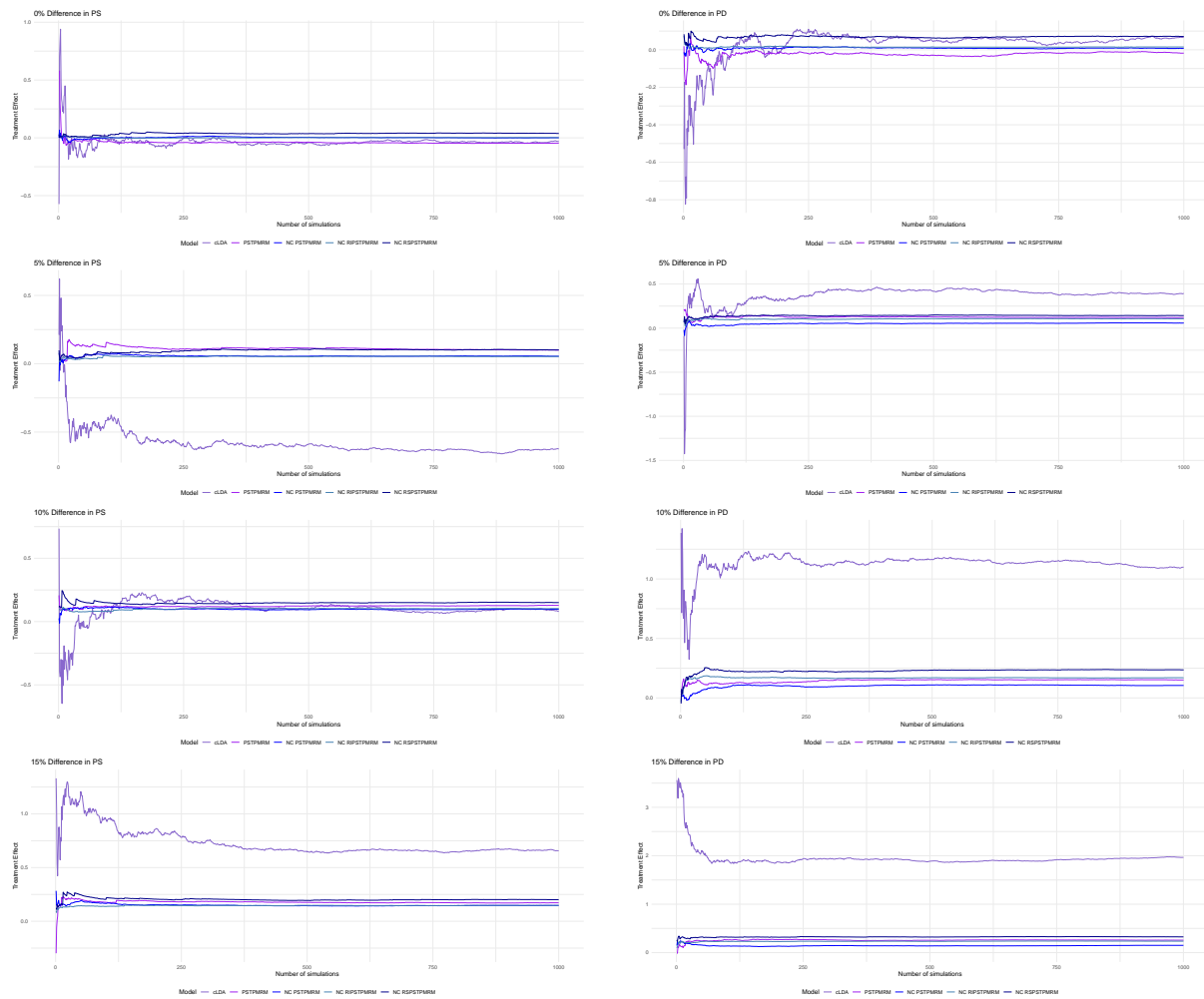


FIGURE E.3: Mean of estimates of the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model throughout 1000 simulations for 1000 subjects in each arm.

E.2 Convergence of Statistical Power

300 Subjects in each arm

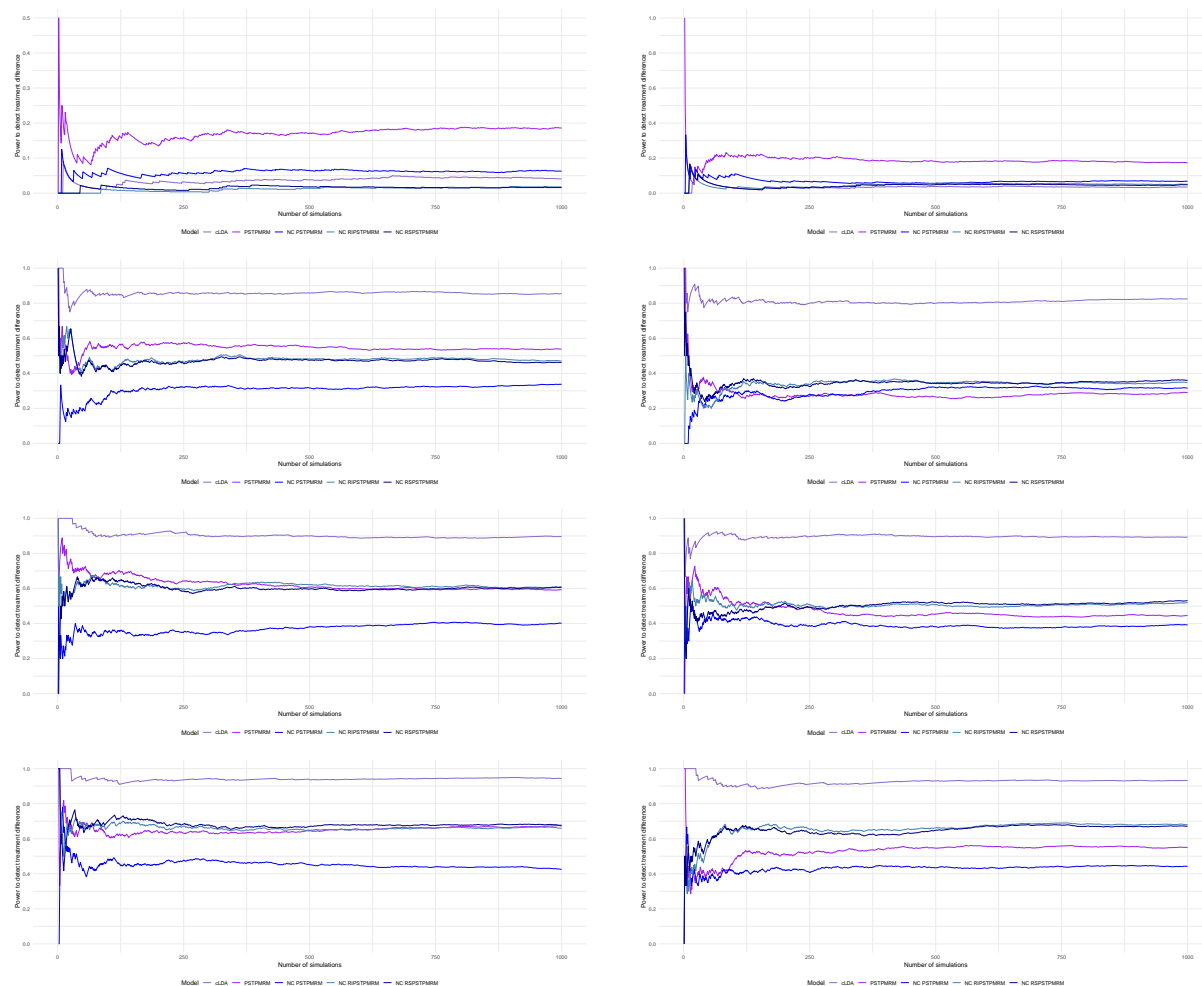


FIGURE E.4: Statistical power of the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model throughout 1000 simulations for 300 subjects in each arm.

500 Subjects in each arm

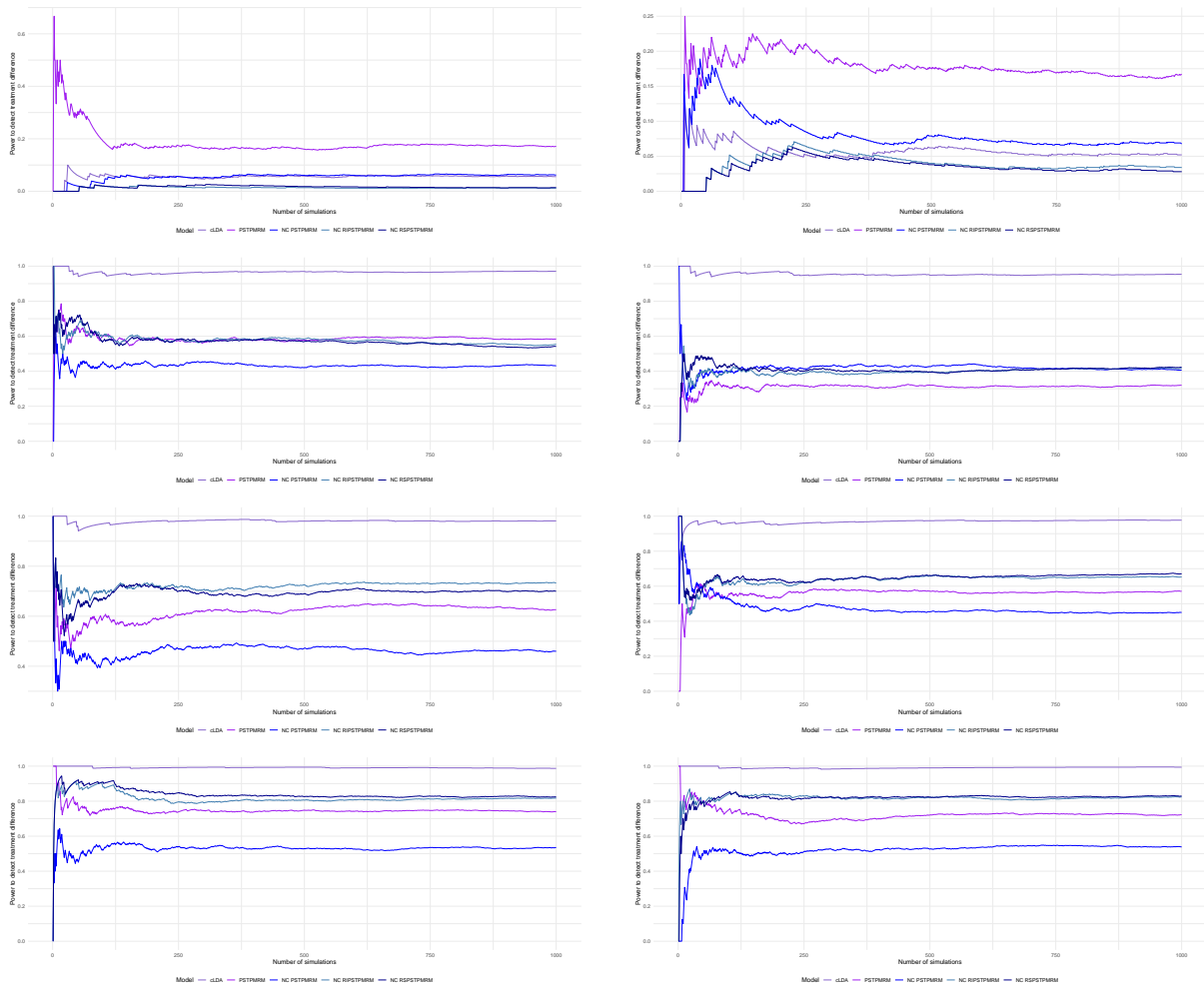


FIGURE E.5: Statistical power of the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model throughout 1000 simulations for 500 subjects in each arm.

1000 Subjects in each arm

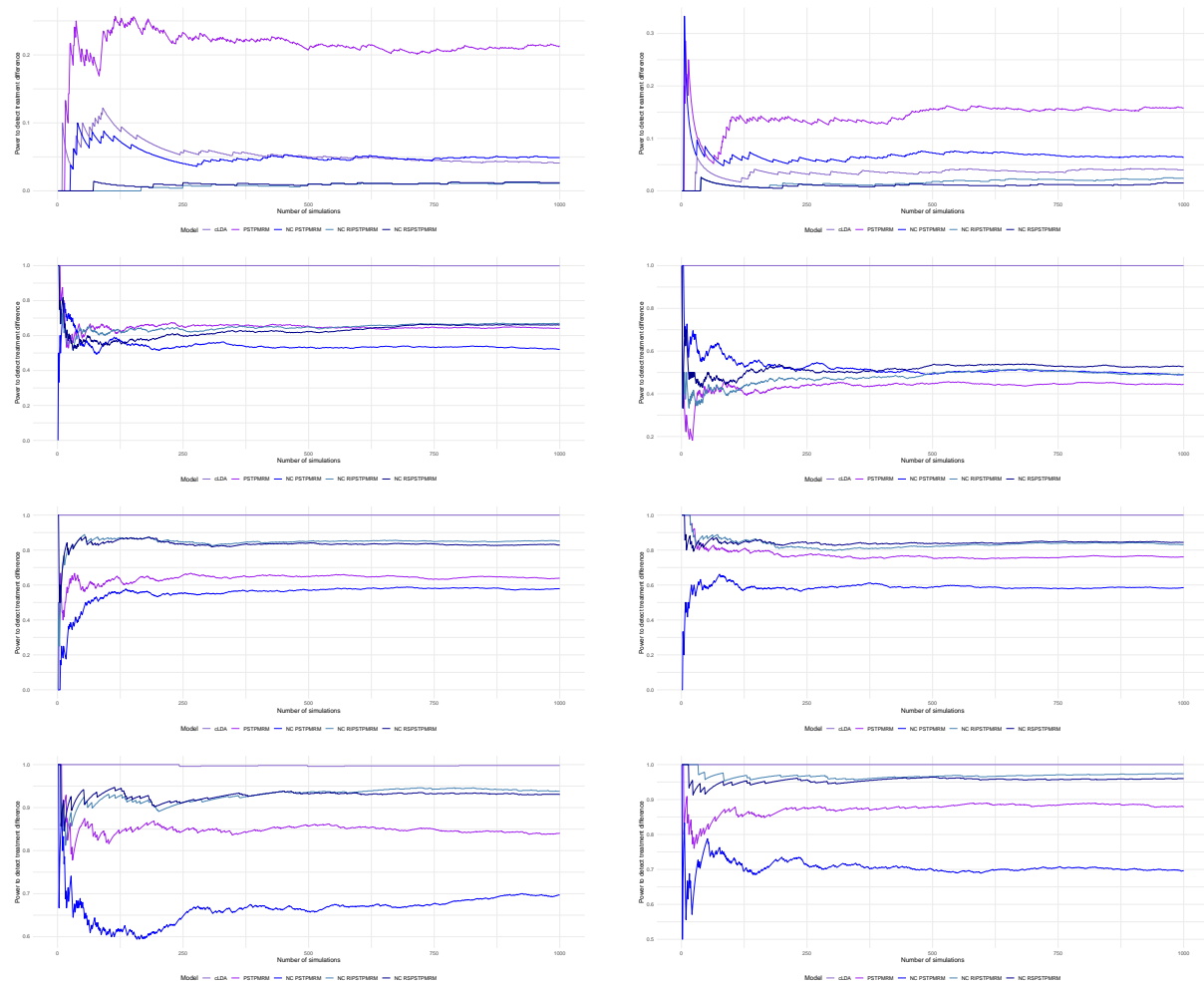


FIGURE E.6: *Statistical power of the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model throughout 1000 simulations for 1000 subjects in each arm.*

E.3 Subgroup Parameter Estimates

300 Subjects in each arm

Model	Scenario	Mean True Treatment Difference	Mean Bias	Relative Bias (%)	Estimation Standard Error
cLDA	0% difference in PS	-0.003	0.003	-100.000	2.271
	0% difference in PD	-0.043	-0.025	58.140	2.337
	5% difference in PS	-0.631	0.050	-7.924	2.243
	5% difference in PD	0.316	-0.019	6.013	2.313
	10% difference in PS	0.147	-0.182	-123.810	2.293
	10% difference in PD	1.020	0.073	7.157	2.317
	15% difference in PS	0.782	-0.043	-5.498	2.269
	15% difference in PD	2.069	-0.142	-6.863	2.337
PSTPMRM	0% difference in PS	0.014	-0.068	-485.714	0.304
	0% difference in PD	0.003	-0.062	-2066.667	0.378
	5% difference in PS	0.047	0.054	114.893	0.360
	5% difference in PD	0.042	0.121	288.095	0.359
	10% difference in PS	0.103	0.037	35.922	0.326
	10% difference in PD	0.085	0.106	124.706	0.381
	15% difference in PS	0.141	0.053	7 37.588	0.318
	15% difference in PD	0.146	0.115	78.767	0.339
NC PSTPMRM	0% difference in PS	0.014	-0.088	-628.571	1.278
	0% difference in PD	0.003	-0.100	-3333.333	1.300
	5% difference in PS	0.047	-0.033	-70.213	0.478
	5% difference in PD	0.042	-0.106	-252.381	1.592
	10% difference in PS	0.103	0.004	3.883	0.187
	10% difference in PD	0.085	-0.029	-34.118	0.608
	15% difference in PS	0.141	-0.135	-95.745	2.415
	15% difference in PD	0.146	-0.060	-41.096	1.002

Continued on next page

Table E.7: Mean true treatment difference, mean bias, relative bias, and estimation standard error of the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios for 300 subjects in each arm.

Model	Scenario	Mean True Treatment Difference	Mean Bias	Relative Bias (%)	Estimation Standard Error
NC RI PSTPMRM	0% difference in PS	0.014	0.000	0.000	0.130
	0% difference in PD	0.003	-0.009	-300.000	0.859
	5% difference in PS	0.047	0.021	44.681	0.214
	5% difference in PD	0.042	0.049	116.667	0.165
	10% difference in PS	0.103	0.025	24.272	0.217
	10% difference in PD	0.085	0.008	9.411	1.700
	15% difference in PS	0.141	0.036	25.532	0.191
	15% difference in PD	0.146	-0.014	-9.589	2.647
NC RS PSTPMRM	0% difference in PS	0.014	-0.006	-42.857	0.996
	0% difference in PD	0.003	0.016	533.333	0.960
	5% difference in PS	0.047	-0.037	-78.723	2.241
	5% difference in PD	0.042	0.055	130.952	0.759
	10% difference in PS	0.103	0.011	10.680	0.732
	10% difference in PD	0.085	0.002	2.353	1.658
	15% difference in PS	0.141	-0.474	-336.170	14.489
	15% difference in PD	0.146	-0.286	195.890	6.623

Table E.7: Mean true treatment difference, mean bias, relative bias, and estimation standard error of the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios for 300 subjects in each arm.

Model	Scenario	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
CLDA	0% PS	-0.011	-0.077	700.000	1.604
	0% PD	-0.045	0.000	0.000	1.637
	20% PS	3.712	-0.006	-0.162	1.651
	20% PD	3.372	0.045	1.335	1.643
	20% PS	3.740	0.058	1.551	1.621
	20% PD	3.367	-0.047	-1.396	1.616
	20% PS	3.712	0.029	0.781	1.688
	20% PD	3.209	0.122	3.802	1.655
PSTPMRM	0% PS	0.940	0.012	1.277	0.283
	0% PD	0.939	0.005	0.532	0.365
	20% PS	0.798	0.062	7.769	0.324
	20% PD	0.802	0.084	10.474	0.327
	20% PS	0.798	0.047	5.890	0.283
	20% PD	0.805	0.069	8.571	0.338
	20% PS	0.798	0.052	6.516	0.283
	20% PD	0.814	0.035	4.300	0.277
NC PSTPMRM	0% PS	0.940	0.008	0.851	1.275
	0% PD	0.939	-0.010	-1.065	1.291
	20% PS	0.798	-0.020	-2.506	0.474
	20% PD	0.802	-0.088	-10.973	1.590
	20% PS	0.798	0.011	1.378	0.181
	20% PD	0.805	-0.017	-2.112	0.604
	20% PS	0.798	-0.129	-16.165	2.415
	20% PD	0.814	-0.052	-6.388	1.000

Continued on next page

Table E.8: Mean true treatment effect, mean bias, relative bias, and estimation standard error of ζ for the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios for 300 subjects in each arm.

Model	Scenario	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
NC RI PSTPMRM	0% PS	0.940	0.074	7.872	0.097
	0% PD	0.939	0.063	6.709	0.851
	20% PS	0.798	0.028	3.509	0.186
	20% PD	0.802	0.072	8.978	0.127
	20% PS	0.798	0.021	2.632	0.191
	20% PD	0.805	0.010	1.242	1.703
	20% PS	0.798	0.030	3.759	0.163
	20% PD	0.814	-0.017	-2.088	2.652
NC RS PSTPMRM	0% PS	0.940	0.072	7.660	0.989
	0% PD	0.939	0.089	9.478	0.957
	20% PS	0.798	-0.055	-6.892	2.243
	20% PD	0.802	0.047	5.860	0.747
	20% PS	0.798	-0.022	-2.757	0.734
	20% PD	0.805	-0.041	-5.093	1.652
	20% PS	0.798	-0.520	-65.163	14.492
	20% PD	0.814	-0.820	-100.737	16.381

Table E.8: Mean true treatment effect, mean bias, relative bias, and estimation standard error of ζ for the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios for 300 subjects in each arm.

Model	Scenario	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
CLDA	0% PS	-0.014	-0.074	528.571	1.638
	0% PD	-0.088	-0.025	28.409	1.677
	25% PS	3.082	0.043	1.395	1.584
	25% PD	3.688	0.026	0.705	1.620
	30% PS	3.887	-0.124	-3.190	1.616
	30% PD	4.387	0.026	0.593	1.659
	35% PS	4.494	-0.013	-0.289	1.587
	35% PD	5.278	-0.020	-0.379	1.673
PSTPMRM	0% PS	0.926	0.080	8.639	0.114
	0% PD	0.936	0.067	7.158	0.103
	25% PS	0.751	0.008	1.065	0.152
	25% PD	0.761	-0.037	-4.862	0.180
	30% PS	0.695	0.010	1.439	0.156
	30% PD	0.720	-0.037	-5.139	0.183
	35% PS	0.658	-0.001	-0.152	0.157
	35% PD	0.668	-0.080	-11.976	0.197
NC PSTPMRM	0% PS	0.926	0.096	10.367	0.154
	0% PD	0.936	0.089	9.509	0.153
	25% PS	0.751	0.013	1.731	0.051
	25% PD	0.761	0.018	2.365	0.073
	30% PS	0.695	0.007	1.007	0.042
	30% PD	0.720	0.013	1.806	0.054
	35% PS	0.658	0.007	1.064	0.036
	35% PD	0.668	0.008	1.198	0.036

Continued on next page

Table E.9: Mean true treatment effect, mean bias, relative bias, and estimation standard error of γ for the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios for 300 subjects in each arm.

Model	Scenario	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
NC RI PSTPMRM	0% PS	0.926	0.074	7.991	0.086
	0% PD	0.936	0.071	7.585	0.083
	25% PS	0.751	0.007	0.932	0.104
	25% PD	0.761	0.023	3.022	0.107
	30% PS	0.695	-0.003	-0.431	0.105
	30% PD	0.720	0.002	0.278	0.111
	35% PS	0.658	-0.006	-0.912	0.107
	35% PD	0.668	-0.003	-0.449	0.108
NC RS PSTPMRM	0% PS	0.926	0.077	8.315	0.094
	0% PD	0.936	0.073	7.799	0.089
	25% PS	0.751	-0.017	-2.264	0.103
	25% PD	0.761	-0.008	-1.051	0.123
	30% PS	0.695	-0.036	-5.180	0.104
	30% PD	0.720	-0.043	-5.972	0.124
	35% PS	0.658	-0.046	-6.991	0.106
	35% PD	0.668	-0.061	-9.132	0.118

Table E.9: Mean true treatment effect, mean bias, relative bias, and estimation standard error of γ for the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios for 300 subjects in each arm.

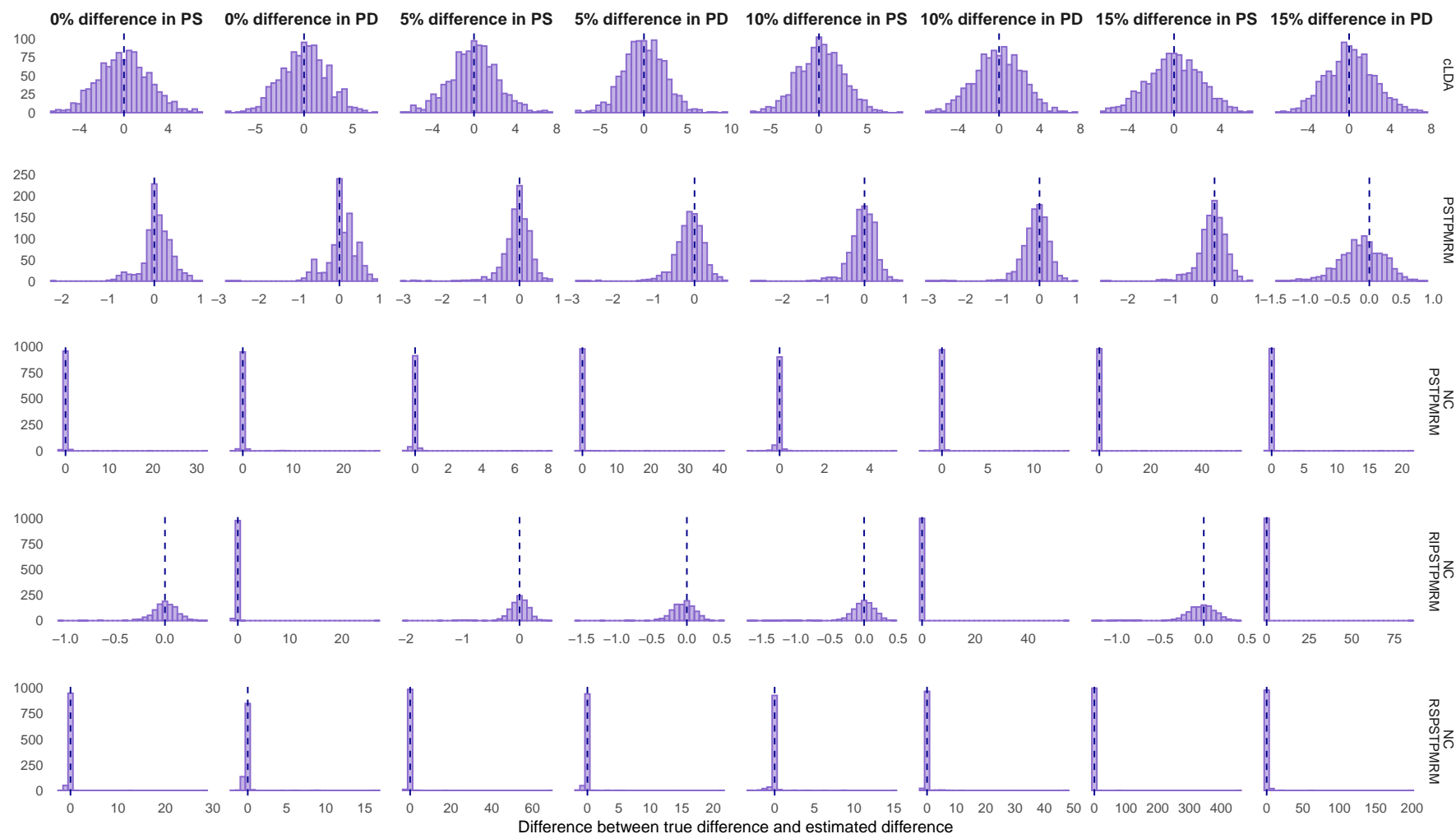


FIGURE E.10: Bias for each of the 1000 effect estimates from each simulated data set. Estimates are from data sets with 300 subjects in each arm.

500 Subjects in each arm

Model	Scenario	Mean True Treatment Difference	Mean Bias	Relative Bias (%)	Estimation Standard Error
cLDA	0% difference in PS	0.089	-0.015	-16.854	1.863
	0% difference in PD	0.043	0.022	51.163	1.880
	5% difference in PS	-0.702	0.007	-0.997	1.834
	5% difference in PD	0.417	0.024	5.755	1.822
	10% difference in PS	0.111	0.025	22.523	1.858
	10% difference in PD	1.042	0.034	3.263	1.820
	15% difference in PS	0.761	-0.020	2.638	1.836
	15% difference in PD	1.918	0.037	1.929	1.797
PSTPMRM	0% difference in PS	0.016	-0.055	-343.750	0.267
	0% difference in PD	0.011	-0.025	-227.273	0.381
	5% difference in PS	0.048	0.054	112.500	0.306
	5% difference in PD	0.048	0.082	170.833	0.281
	10% difference in PS	0.098	0.037	37.755	0.279
	10% difference in PD	0.084	0.088	104.762	0.276
	15% difference in PS	0.146	0.036	24.658	0.239
	15% difference in PD	0.141	0.144	102.128	0.308
NC PSTPMRM	0% difference in PS	0.016	-0.011	-68.750	0.153
	0% difference in PD	0.011	-0.014	-127.273	0.147
	5% difference in PS	0.048	-0.010	-20.833	0.363
	5% difference in PD	0.048	0.002	4.167	0.045
	10% difference in PS	0.098	-0.007	-7.142	0.257
	10% difference in PD	0.084	0.002	2.381	0.049
	15% difference in PS	0.146	0.002	1.370	0.038
	15% difference in PD	0.141	-0.002	-1.418	0.155

Continued on next page

Table E.11: Mean true treatment difference, mean bias, relative bias, and estimation standard error of the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios for 500 subjects in each arm.

Model	Scenario	Mean True Treatment Difference	Mean Bias	Relative Bias (%)	Estimation Standard Error
NC RI PSTPMRM	0% difference in PS	0.016	0.000	0.000	0.099
	0% difference in PD	0.011	0.013	118.182	0.114
	5% difference in PS	0.048	0.009	18.750	0.181
	5% difference in PD	0.048	0.050	104.167	0.129
	10% difference in PS	0.098	0.002	2.041	0.145
	10% difference in PD	0.084	0.065	77.381	0.133
	15% difference in PS	0.146	0.009	6.164	0.155
	15% difference in PD	0.141	0.085	60.284	0.146
NC RS PSTPMRM	0% difference in PS	0.016	0.031	193.750	0.173
	0% difference in PD	0.011	0.069	627.273	0.214
	5% difference in PS	0.048	0.044	91.667	0.232
	5% difference in PD	0.048	0.089	185.417	0.198
	10% difference in PS	0.098	0.027	27.551	0.505
	10% difference in PD	0.084	0.119	141.667	0.320
	15% difference in PS	0.146	0.065	44.521	0.236
	15% difference in PD	0.141	0.121	85.816	0.390

Table E.11: Mean true treatment difference, mean bias, relative bias, and estimation standard error of the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios for 500 subjects in each arm.

Model	Scenario	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
cLDA	0% PS	-0.042	0.018	42.857	1.287
	0% PD	-0.005	-0.015	300.000	1.268
	20% PS	3.716	-0.002	-0.054	1.323
	20% PD	3.371	-0.070	-2.077	1.261
	20% PS	3.643	0.031	0.851	1.271
	20% PD	3.427	-0.063	-1.838	1.318
	20% PS	3.700	0.005	0.135	1.258
	20% PD	3.341	-0.049	-1.467	1.312
PSTPMRM	0% PS	0.957	0.004	0.418	0.250
	0% PD	0.956	0.030	3.138	0.377
	20% PS	0.803	0.060	7.472	0.294
	20% PD	0.807	0.055	6.815	0.249
	20% PS	0.806	0.037	4.591	0.251
	20% PD	0.802	0.057	7.107	0.243
	20% PS	0.804	0.036	4.478	0.209
	20% PD	0.809	0.058	7.169	0.260
NC PSTPMRM	0% PS	0.957	0.058	6.061	0.101
	0% PD	0.956	0.052	5.439	0.086
	20% PS	0.803	-0.006	-0.747	0.362
	20% PD	0.807	0.010	1.239	0.028
	20% PS	0.806	-0.004	-0.496	0.256
	20% PD	0.802	0.009	1.122	0.037
	20% PS	0.804	0.004	0.498	0.032
	20% PD	0.809	0.004	0.494	0.153

Continued on next page

Table E.12: Mean true treatment effect, mean bias, relative bias, and estimation standard error of ζ for the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios for 500 subjects in each arm.

Model	Scenario	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
NC RI PSTPMRM	0% PS	0.957	0.057	5.956	0.075
	0% PD	0.956	0.067	7.008	0.094
	20% PS	0.803	0.012	1.494	0.160
	20% PD	0.807	0.074	9.170	0.093
	20% PS	0.806	0.001	0.124	0.119
	20% PD	0.802	0.074	9.227	0.092
	20% PS	0.804	0.002	0.249	0.124
	20% PD	0.809	0.077	9.518	0.113
NC RS PSTPMRM	0% PS	0.957	0.090	9.404	0.159
	0% PD	0.956	0.125	13.075	0.207
	20% PS	0.803	0.024	2.989	0.217
	20% PD	0.807	0.087	10.781	0.172
	20% PS	0.806	-0.008	-0.993	0.503
	20% PD	0.802	0.082	10.224	0.306
	20% PS	0.804	0.019	2.363	0.223
	20% PD	0.809	-0.182	22.497	7.409

Table E.12: Mean true treatment effect, mean bias, relative bias, and estimation standard error of ζ for the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios for 500 subjects in each arm.

Model	Scenario	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
cLDA	0% PS	0.047	0.002	4.255	1.295
	0% PD	0.039	0.007	17.949	1.332
	25% PS	3.014	0.006	0.199	1.281
	25% PD	3.787	-0.046	-1.215	1.246
	30% PS	3.754	0.056	1.492	1.291
	30% PD	4.469	-0.029	-0.648	1.290
	35% PS	4.460	-0.016	-0.359	1.313
	35% PD	5.259	-0.012	-0.228	1.221
PSTPMRM	0% PS	0.941	0.060	6.376	0.088
	0% PD	0.946	0.054	5.708	0.079
	25% PS	0.755	0.006	0.795	0.127
	25% PD	0.759	-0.026	-3.426	0.143
	30% PS	0.708	0.000	0.000	0.122
	30% PD	0.718	-0.031	-4.318	0.156
	35% PS	0.657	0.001	0.152	0.124
	35% PD	0.668	-0.086	-12.874	0.167
NC PSTPMRM	0% PS	0.941	0.069	7.333	0.112
	0% PD	0.946	0.067	7.082	0.120
	25% PS	0.755	0.005	0.662	0.024
	25% PD	0.759	0.007	0.922	0.033
	30% PS	0.708	0.004	0.565	0.021
	30% PD	0.718	0.007	0.975	0.028
	35% PS	0.657	0.003	0.457	0.017
	35% PD	0.668	0.006	0.898	0.023

Continued on next page

Table E.13: Mean true treatment effect, mean bias, relative bias, and estimation standard error of γ for the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios for 500 subjects in each arm.

Model	Scenario	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
NC RI PSTPMRM	0% PS	0.941	0.057	6.057	0.064
	0% PD	0.946	0.055	5.814	0.063
	25% PS	0.755	0.004	0.530	0.086
	25% PD	0.759	0.024	3.162	0.092
	30% PS	0.708	-0.002	-0.282	0.084
	30% PD	0.718	0.009	1.253	0.094
	35% PS	0.657	-0.007	-1.065	0.086
	35% PD	0.668	-0.007	-1.048	0.086
NC RS PSTPMRM	0% PS	0.941	0.059	6.270	0.071
	0% PD	0.946	0.055	5.814	0.065
	25% PS	0.755	-0.020	-2.649	0.084
	25% PD	0.759	-0.002	-0.264	0.106
	30% PS	0.708	-0.037	-5.226	0.083
	30% PD	0.718	-0.037	-5.153	0.108
	35% PS	0.657	-0.046	-7.002	0.083
	35% PD	0.668	-0.069	-10.329	0.097

Table E.13: Mean true treatment effect, mean bias, relative bias, and estimation standard error of γ for the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios for 500 subjects in each arm.

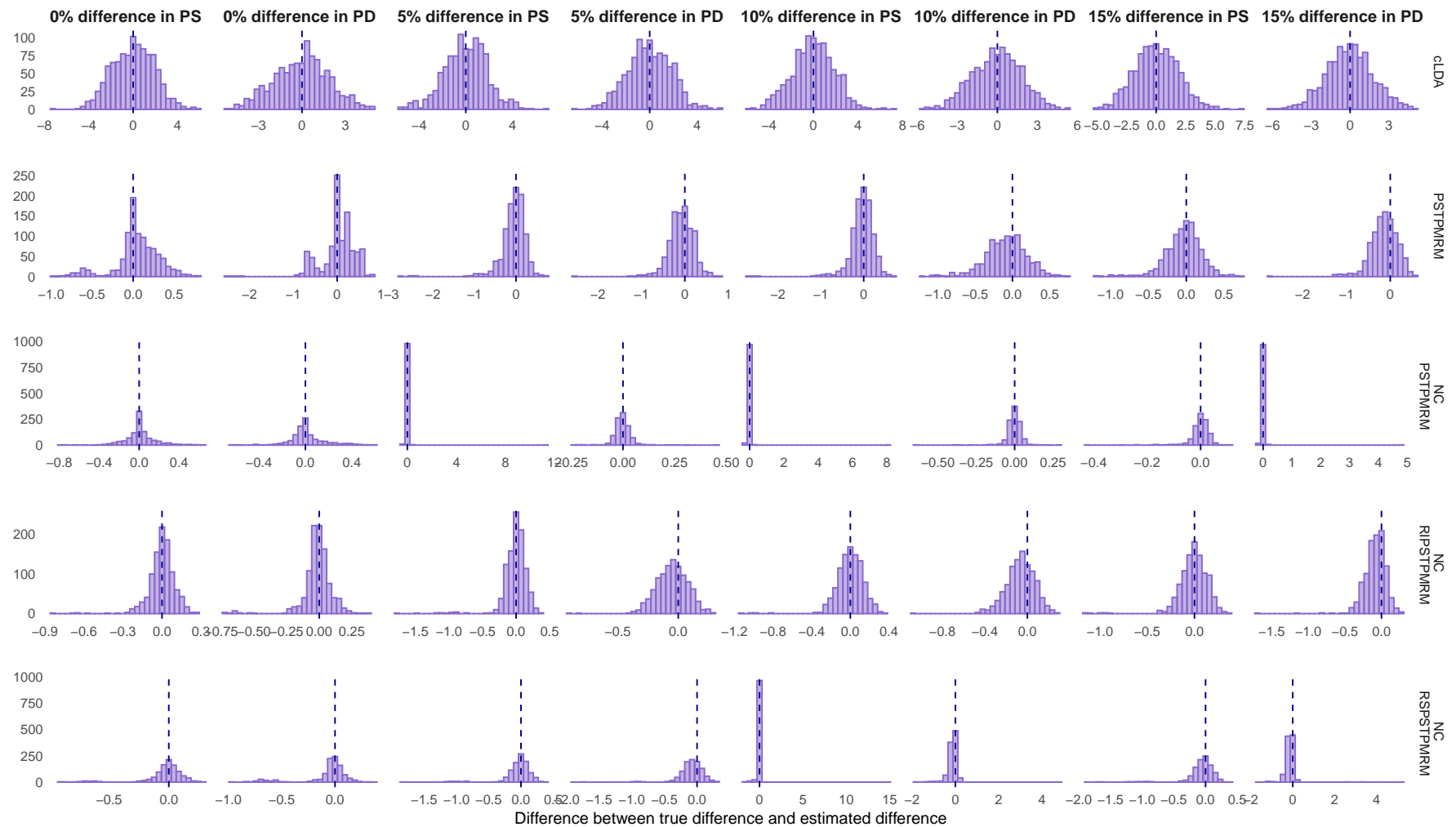


FIGURE E.14: Bias for each of the 1000 effect estimates from each simulated data set. Estimates are from data sets with 500 subjects in each arm.

1000 Subjects in each arm

Model	Scenario	Mean True Treatment Difference	Mean Bias	Relative Bias (%)	Estimation Standard Error
cLDA	0% difference in PS	-0.023	-0.010	43.478	1.262
	0% difference in PD	0.042	0.025	59.524	1.264
	5% difference in PS	-0.575	-0.048	8.348	1.307
	5% difference in PD	0.431	-0.040	-9.281	1.265
	10% difference in PS	0.648	0.034	5.247	1.314
	10% difference in PD	1.163	-0.064	-5.503	1.282
	15% difference in PS	0.716	-0.058	-8.101	1.255
	15% difference in PD	1.918	0.047	2.450	1.322
PSTPMRM	0% difference in PS	0.012	-0.056	-466.667	0.219
	0% difference in PD	0.002	-0.020	-1000	0.366
	5% difference in PS	0.059	0.042	71.186	0.269
	5% difference in PD	0.052	0.070	134.615	0.205
	10% difference in PS	0.100	0.027	27.000	0.194
	10% difference in PD	0.099	0.054	54.545	0.211
	15% difference in PS	0.150	0.022	14.667	0.195
	15% difference in PD	0.144	0.114	79.167	0.252
NC PSTPMRM	0% difference in PS	0.012	-0.011	-91.667	0.088
	0% difference in PD	0.002	0.005	250.000	0.090
	5% difference in PS	0.059	-0.004	-6.780	0.020
	5% difference in PD	0.052	0.006	11.538	0.028
	10% difference in PS	0.100	-0.003	-3.000	0.018
	10% difference in PD	0.099	0.006	6.061	0.026
	15% difference in PS	0.150	-0.004	-2.667	0.017
	15% difference in PD	0.144	0.006	4.167	0.029

Continued on next page

Table E.15: Mean true treatment difference, mean bias, relative bias, and estimation standard error of the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios for 1000 subjects in each arm.

Model	Scenario	Mean True Treatment Difference	Mean Bias	Relative Bias (%)	Estimation Standard Error
NC RI PSTPMRM	0% difference in PS	0.012	-0.009	-75.000	0.057
	0% difference in PD	0.002	0.009	450.000	0.056
	5% difference in PS	0.059	0.002	3.390	0.149
	5% difference in PD	0.052	0.059	113.462	0.123
	10% difference in PS	0.100	0.000	0	0.129
	10% difference in PD	0.099	0.075	75.757	0.104
	15% difference in PS	0.150	-0.003	2.000	0.110
	15% difference in PD	0.144	0.085	59.028	0.103
NC RS PSTPMRM	0% difference in PS	0.012	0.028	233.333	0.162
	0% difference in PD	0.002	0.065	3250.000	0.194
	5% difference in PS	0.059	0.050	84.746	0.238
	5% difference in PD	0.052	0.095	182.692	0.176
	10% difference in PS	0.100	0.047	47.000	0.215
	10% difference in PD	0.099	0.139	140.404	0.172
	15% difference in PS	0.150	0.051	34.000	0.201
	15% difference in PD	0.144	0.176	122.222	0.176

Table E.15: Mean true treatment difference, mean bias, relative bias, and estimation standard error of the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios for 1000 subjects in each arm.

Model	Scenario	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
cLDA	0% PS	0.005	-0.006	-120.000	0.898
	0% PD	-0.021	-0.019	90.476	0.929
	20% PS	3.647	0.006	0.165	0.912
	20% PD	3.336	0.012	0.360	0.912
	20% PS	3.693	-0.013	-0.352	0.945
	20% PD	3.363	0.026	0.773	0.936
	20% PS	3.712	0.021	0.565	0.899
	20% PD	3.375	-0.010	-0.296	0.926
PSTPMRM	0% PS	0.971	-0.016	-1.648	0.210
	0% PD	0.969	0.012	1.238	0.360
	20% PS	0.807	0.050	6.196	0.251
	20% PD	0.813	0.045	5.535	0.185
	20% PS	0.805	0.026	3.230	0.177
	20% PD	0.810	0.048	5.926	0.189
	20% PS	0.805	0.026	3.230	0.176
	20% PD	0.808	0.035	4.332	0.215
NC PSTPMRM	0% PS	0.971	0.036	3.708	0.057
	0% PD	0.969	0.041	4.231	0.059
	20% PS	0.807	-0.001	-0.124	0.014
	20% PD	0.813	0.010	1.230	0.016
	20% PS	0.805	-0.001	-0.124	0.014
	20% PD	0.810	0.010	1.235	0.017
	20% PS	0.805	-0.001	-0.124	0.012
	20% PD	0.808	0.010	1.238	0.020

Continued on next page

Table E.16: Mean true treatment effect, mean bias, relative bias, and estimation standard error of ζ for the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios for 1000 subjects in each arm.

Model	Scenario	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
NC RI PSTPMRM	0% PS	0.971	0.032	3.296	0.036
	0% PD	0.969	0.042	4.334	0.040
	20% PS	0.807	0.004	0.496	0.138
	20% PD	0.813	0.089	10.947	0.100
	20% PS	0.805	-0.001	-0.124	0.115
	20% PD	0.810	0.080	9.877	0.077
	20% PS	0.805	-0.003	-0.373	0.089
	20% PD	0.808	0.080	9.901	0.079
NC RS PSTPMRM	0% PS	0.971	0.068	7.003	0.152
	0% PD	0.969	0.099	10.217	0.193
	20% PS	0.807	0.029	3.594	0.236
	20% PD	0.813	0.103	12.669	0.162
	20% PS	0.805	0.014	1.739	0.212
	20% PD	0.810	0.103	12.716	0.167
	20% PS	0.805	0.012	1.490	0.195
	20% PD	0.808	0.098	12.129	0.148

Table E.16: Mean true treatment effect, mean bias, relative bias, and estimation standard error of ζ for the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios for 1000 subjects in each arm.

Model	Scenario	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
cLDA	0% PS	-0.017	-0.017	100.000	0.869
	0% PD	0.022	0.006	27.273	0.884
	25% PS	3.072	-0.042	-1.367	0.936
	25% PD	3.768	-0.028	-0.743	0.918
	30% PS	3.740	0.021	0.561	0.885
	30% PD	4.526	-0.037	-0.817	0.905
	35% PS	4.428	-0.037	-0.835	0.926
	35% PD	5.293	0.036	0.680	0.904
PSTPMRM	0% PS	0.959	0.040	4.171	0.065
	0% PD	0.967	0.033	3.413	0.051
	25% PS	0.748	0.009	1.203	0.093
	25% PD	0.761	-0.025	-3.285	0.112
	30% PS	0.705	-0.001	-0.142	0.085
	30% PD	0.711	-0.006	-0.844	0.114
	35% PS	0.655	0.004	0.611	0.088
	35% PD	0.664	-0.078	11.747	0.146
NC PSTPMRM	0% PS	0.959	0.047	4.901	0.066
	0% PD	0.967	0.036	3.723	0.069
	25% PS	0.748	0.003	0.401	0.012
	25% PD	0.761	0.004	0.526	0.020
	30% PS	0.705	0.002	0.284	0.010
	30% PD	0.711	0.003	0.422	0.016
	35% PS	0.655	0.002	0.305	0.010
	35% PD	0.664	0.003	0.452	0.016

Continued on next page

Table E.17: Mean true treatment effect, mean bias, relative bias, and estimation standard error of γ for the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios for 1000 subjects in each arm.

Model	Scenario	Mean True Treatment Effect	Mean Bias	Relative Bias (%)	Estimation Standard Error
NC R I PSTPMRM	0% PS	0.959	0.039	4.067	0.041
	0% PD	0.967	0.030	3.102	0.037
	25% PS	0.748	-0.001	-0.134	0.063
	25% PD	0.761	0.031	4.074	0.079
	30% PS	0.705	-0.009	-1.277	0.064
	30% PD	0.711	0.009	1.266	0.072
	35% PS	0.655	0.012	1.832	0.054
	35% PD	0.664	-0.003	-0.452	0.067
NC R S PSTPMRM	0% PS	0.959	0.036	3.754	0.044
	0% PD	0.967	0.032	3.309	0.039
	25% PS	0.748	-0.025	-3.342	0.062
	25% PD	0.761	0.007	0.920	0.089
	30% PS	0.705	-0.042	-5.957	0.063
	30% PD	0.711	-0.036	-5.063	0.087
	35% PS	0.655	-0.027	-4.122	0.055
	35% PD	0.664	-0.077	-11.596	0.084

Table E.17: Mean true treatment effect, mean bias, relative bias, and estimation standard error of γ for the extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios for 1000 subjects in each arm.

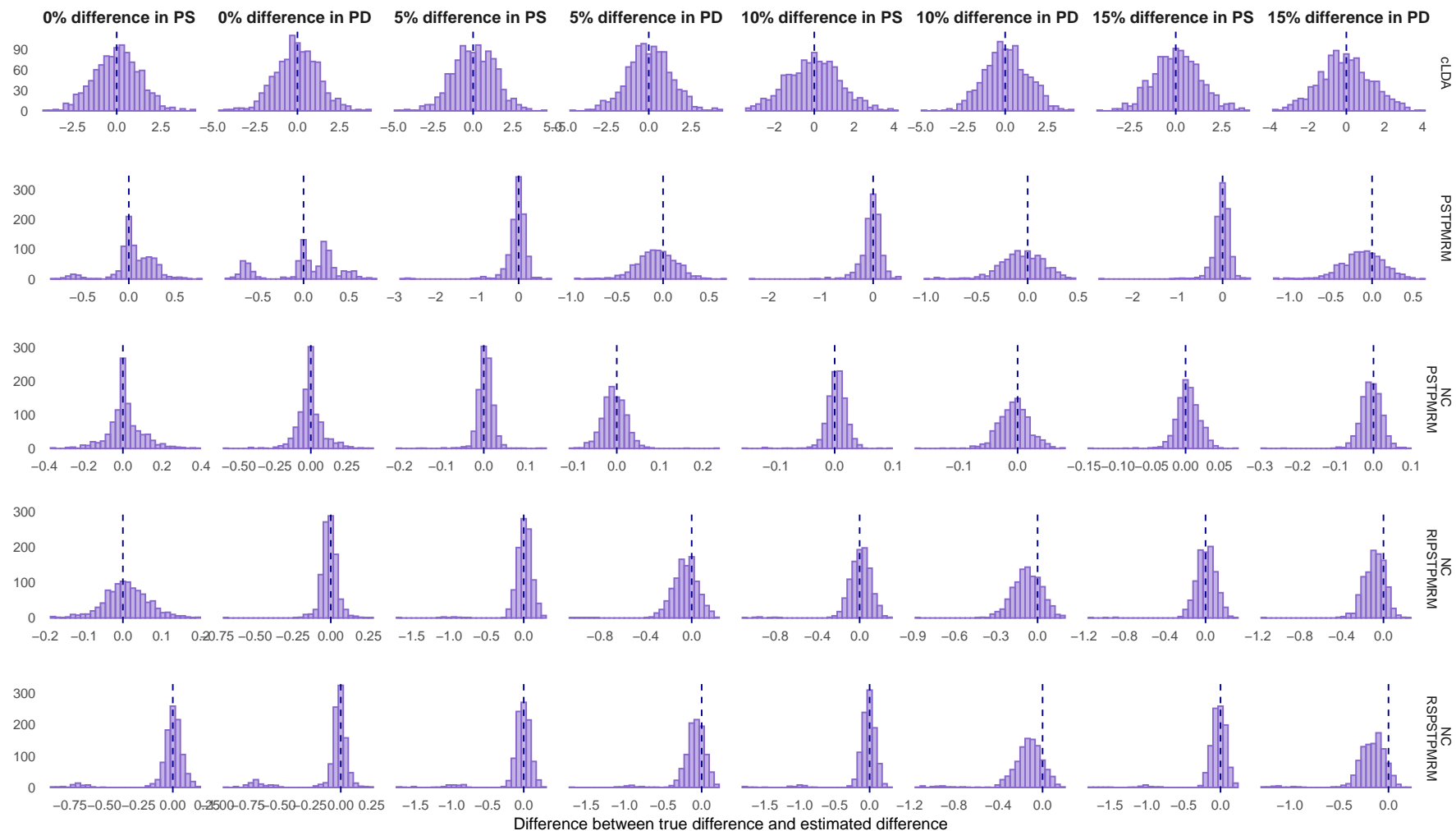


FIGURE E.18: Bias for each of the 1000 effect estimates from each simulated data set. Estimates are from data sets with 500 subjects in each arm.

E.4 Subgroup Statistical Power

Subjects per arm	Scenario	cLDA	PSTPMRM	NC PSTPMRM	NC RI PSTPMRM	NC RS PSTPMRM
300	0% difference in PS	0.035	0.174	0.068	0.051	0.048
	0% difference in PD	0.041	0.185	0.063	0.018	0.016
	5% difference in PS	0.825	0.291	0.315	0.348	0.36
	5% difference in PD	0.854	0.539	0.337	0.471	0.462
	10% difference in PS	0.893	0.446	0.393	0.519	0.53
	10% difference in PD	0.897	0.593	0.402	0.609	0.607
	15% difference in PS	0.932	0.551	0.442	0.681	0.672
	15% difference in PD	0.944	0.673	0.425	0.659	0.676
500	0% difference in PS	0.052	0.167	0.068	0.034	0.028
	0% difference in PD	0.056	0.171	0.062	0.012	0.014
	5% difference in PS	0.954	0.32	0.407	0.417	0.423
	5% difference in PD	0.971	0.584	0.432	0.554	0.542
	10% difference in PS	0.977	0.572	0.45	0.652	0.67
	10% difference in PD	0.981	0.626	0.459	0.733	0.7
	15% difference in PS	0.994	0.723	0.538	0.824	0.829
	15% difference in PD	0.987	0.741	0.535	0.817	0.824
1000	0% difference in PS	0.04	0.158	0.064	0.024	0.015
	0% difference in PD	0.041	0.213	0.049	0.011	0.012
	5% difference in PS	1	0.444	0.49	0.491	0.528
	5% difference in PD	0.999	0.642	0.52	0.67	0.661
	10% difference in PS	1	0.76	0.584	0.831	0.844
	10% difference in PD	1	0.639	0.578	0.85	0.828
	15% difference in PS	1	0.879	0.697	0.974	0.96
	15% difference in PD	0.998	0.841	0.697	0.938	0.93

TABLE E.19: Statistical power of extended progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios, and for a different number of subjects in each arm.

Subjects per arm	Scenario	cLDA	PSTPMRM	NC PSTPMRM	NC RI PSTPMRM	NC RS PSTPMRM
300	0% difference in PS	0.045	0.045	0.046	0.045	0.045
	0% difference in PD	0.045	0.045	0.045	0.045	0.045
	5% difference in PS	0.86	0.085	0.216	0.314	0.347
	5% difference in PD	0.862	0.090	0.235	0.639	0.623
	10% difference in PS	0.932	0.107	0.296	0.496	0.52
	10% difference in PD	0.901	0.111	0.291	0.729	0.728
	15% difference in PS	0.948	0.133	0.331	0.662	0.664
	15% difference in PD	0.948	0.123	0.316	0.774	0.763
500	0% difference in PS	0.045	0.045	0.045	0.045	0.045
	0% difference in PD	0.045	0.045	0.045	0.045	0.045
	5% difference in PS	0.945	0.118	0.323	0.478	0.525
	5% difference in PD	0.965	0.099	0.313	0.687	0.687
	10% difference in PS	0.975	0.154	0.372	0.703	0.743
	10% difference in PD	0.976	0.130	0.363	0.818	0.796
	15% difference in PS	0.994	0.198	0.487	0.851	0.866
	15% difference in PD	0.986	0.155	0.458	0.889	0.89
1000	0% difference in PS	0.045	0.045	0.045	0.045	0.045
	0% difference in PD	0.045	0.045	0.045	0.045	0.045
	5% difference in PS	1	0.146	0.393	0.598	0.7
	5% difference in PD	0.999	0.193	0.482	0.798	0.786
	10% difference in PS	1	0.176	0.493	0.871	0.904
	10% difference in PD	1	0.228	0.555	0.923	0.898
	15% difference in PS	1	0.300	0.623	0.982	0.971
	15% difference in PD	0.998	0.322	0.673	0.975	0.958

TABLE E.20: Calibrated statistical power of progression models for repeated measures and the subgroup extension of the constrained longitudinal data analysis model in different scenarios, and for a different number of subjects in each arm. The highlighted numbers in each row (except the scenario: no effect) highlights the model(s) which has(have) the highest statistical power.

Bibliography

- Alzheimer's Disease International (2022). New study predicts the number of people living with Alzheimer's disease to triple by 2050. <https://www.alzint.org/news-events/news/new-data-predicts-the-number-of-people-living-with-alzheimers-disease-to-triple-by-2050/>.
- Atri, A., Feldman, H. H., Hansen, C. T., Honore, J. B., Johannsen, P., Knop, F. K., Poulsen, P., Raket, L. L., Sano, M., Soininen, H., and Cummings, J. (2022). evoke and evoke+: design of two large-scale, double-blind, placebo-controlled, phase 3 studies evaluating the neuroprotective effects of semaglutide in early Alzheimer's disease. *Alzheimer's & Dementia*, 18(S10). <https://doi.org/10.1002/alz.062415>.
- Basun, H., Ekman, S.-L., Englund, E., Gustafson, L., Lannfelt, L., Terzis, B., Waflund, L.-O., Björilin, G. A., Beck-Friis, B., Anders, W., and Björkstén, K. S. (2006). *DEMENS. JANSSEN-CILAG*, 1st edition.
- Bates, D. M. and Watts, D. G. (1980). Relative curvature measures of nonlinearity. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(1):p. 1–25. <http://www.jstor.org/stable/2984733>.
- Cummings, J. and Fox, N. (2017). Defining disease modifying therapy for Alzheimer's disease. *The journal of prevention of Alzheimer's disease*, 4(2):p. 109–115. <https://doi.org/10.14283/jpad.2017.12>.
- Demidenko, E. (2013). *Mixed Models: Theory and Applications with R*. John Wiley & Sons, 2nd edition.
- Donohue, M. C., Langford, O., Insel, P. S., van Dyck, C. H., Petersen, R. C., Craft, S., Sethuraman, G., Raman, R., Aisen, P. S., and the Alzheimer's Disease Neuroimaging Initiative, F. (2023). Natural cubic splines for the analysis of Alzheimer's clinical trials. *Pharmaceutical Statistics*, 22:p. 508–519. <https://doi.org/10.1002/pst.2285>.
- Honig, L. S., Vellas, B., Woodward, M., Boada, M., Bullock, R., Borrie, M., Hager, K., Andreasen, N., Scarpini, E., Liu-Seifert, H., Case, M., Dean, R. A., Hake, A., Sundell, K., Poole Hoffmann, V., Carlson, C., Khanna, R., Mintun, M., DeMattos, R., Selzler, K. J., and Siemers, E. (2018). Trial of Solanezumab for mild dementia due to Alzheimer's disease. *New England Journal of Medicine*, 378(4):p. 321–330. PMID: 29365294.
- Ito, K., Charpman, R., Pearson, S. D., Tafazzoli, A., Yaffe, K., and Gurwitz, J. H. (2023). Evaluation of the cost-effectiveness of drug treatment for Alzheimer disease in a simulation model that includes caregiver and societal factors. *JAMA network open*, 13(2). <https://doi.org/10.1212/CPJ.0000000000200127>.
- Janelidze, S., Bali, D., Ashton, N. J., Barthélemy, N. R., Vanbrabant, J., Stoops, E., Vanmechelen, E., He, Y., Dolado, A. O., Triana-Baltzer, G., Pontecorvo, M. J., Zetterberg, H., Kolb, H., Vandijck, M., Blennow, K., Bateman, R. J., and Hansson, O. (2022). Head-to-head comparison of 10 plasma phospho-tau assays in prodromal Alzheimer's disease. *Brain*, 146(4):p. 1592–1601. <https://doi.org/10.1093/brain/awac333>.

- Jiang, J. and Nguyen, T. (2021). *Linear and Generalized Linear Mixed Models and Their Applications*. Springer, 2nd edition.
- Joszt, L. (2024). Biogen abandons aducanumab, pivots focus to lecanemab for Alzheimer disease. <https://www.ajmc.com/view/biogen-abandons-aducanumab-pivots-focus-to-lecanemab-for-alzheimer-disease>.
- Jönsson, L., Ivkovic, M., Atri, A., Handels, R., Gustavsson, A., Hahn-Pedersen, J. H., León, T., Lilja, M., Gundgaard, J., and Raket, L. L. (2024). Progression analysis versus traditional methods to quantify slowing of disease progression in Alzheimer's disease. *Alzheimer's Research & Therapy volume*, 16(48):p. 1598–1695. <https://doi.org/10.1186/s13195-024-01413-y>.
- Kornerup, P. and Muller, J.-M. (2006). Choosing starting values for certain Newton–Raphson iterations. *Theoretical Computer Science*, 351(1):p. 101–110. <https://doi.org/10.1016/j.tcs.2005.09.056>.
- Landhuis, E. (2024). Researchers call for a major rethink of how Alzheimer's treatments are evaluated. *Nature*, 627(8003):S18 – S20. <https://doi.org/10.1038/d41586-024-00756-8>.
- Lilly (2023). Lilly's donanemab significantly slowed cognitive and functional decline in phase 3 study of early Alzheimer's disease. <https://investor.lilly.com/news-releases/news-release-details/lillys-donanemab-significantly-slowed-cognitive-and-functional>.
- Lin, G., Whittington, M. D., Wright, A., Agboola, E., Herron-Smith, S., Pearson, S. D., and Rind, D. (2023). Lecanemab for early Alzheimer's disease. https://icer.org/wp-content/uploads/2023/04/ICER_Alzheimers-Disease_Final-Report_For-Publication_04172023.pdf.
- Lindstrom, M. J. and Bates, D. M. (1990). Mixed effects models for repeated measures data. *Biometrics*, 46:p. 673–687. <https://doi.org/10.2307/2532087>.
- Liu, G. F., Lu, K., Mogg, R., Mallick, M., and Mehrotra, D. V. (2009). Should baseline be a covariate or dependent variable in analyses of change from baseline in clinical trials? *Statistics in Medicine*, 28(20):p. 2509–2530. <https://doi.org/10.1002/sim.3639>.
- Lu, K. (2010). On efficiency of constrained longitudinal data analysis versus longitudinal analysis of covariance. *Biometrics*, 66(3):p. 891–896. <https://doi.org/10.1111/j.1541-0420.2009.01332.x>.
- Madsen, H. and Thyregod, P. (2010). *Introduction to General and Generalized Linear Models*. Taylor & Francis Group, 1st edition.
- Mendez, M. F. (2021). *The mental Status Examination Handbook*. Elsevier, 1st edition.
- MIT Critical Data (2016). *Secondary Analysis of Electronic Health Records*. Springer Cham, 1st edition.
- Myeloma Patients Europe (2024). Health technology assessment (HTA). <https://www.mpeurope.org/qas/health-technology-assessment-hta/>.
- National Institute on Aging (2017). What happens to the brain in Alzheimer's disease? <https://www.nia.nih.gov/health/alzheimers-causes-and-risk-factors/what-happens-brain-alzheimers-disease>.

- National Institute on Aging (2022). How biomarkers help diagnose dementia. <https://www.nia.nih.gov/health/alzheimers-symptoms-and-diagnosis/how-biomarkers-help-diagnose-dementia>.
- National Institute on Aging (2023a). Driving safety and Alzheimer's disease. <https://www.nia.nih.gov/health/safety/driving-safety-and-alzheimers-disease>.
- National Institute on Aging (2023b). Nia-funded active alzheimer's and related dementias clinical trials and studies. <https://www.nia.nih.gov/research/ongoing-AD-trials>.
- National Medical Products Administration (2020). Clinical technical requirements for drugs marketed overseas but not marketed in china. https://english.nmpa.gov.cn/2020-11/18/c_568155.htm.
- Nationalt Videnscenter for Demens (2021). Alzheimer's Disease Assessment Scale – Cognitive Section (ADAS-Cog). <https://videnscenterfordemens.dk/da>.
- Newton, W. (2023). Alzheimer's disease trials: How does Lilly's donanemab compare to lecanemab? <https://www.clinicaltrialsarena.com/news/donanemab-lecanemab/?cf-view>.
- Norton, B. J. and Strube, M. J. (2001). Understanding statistical power. *Journal of Orthopaedic & Sports Physical Therapy*, 31(6):p. 307–315. <https://doi.org/10.2519/jospt.2001.31.6.307>.
- Paget, M.-A., Chuang-Stein, C., Fletcher, C., and Reid, C. (2011). Subgroup analyses of clinical effectiveness to support health technology assessments. *Pharmaceutical Statistics*, 10(6):p. 532–538. <https://doi.org/10.1002/pst.531>.
- Pillidge, Z. and Hanschuh, T. (2024). It's time to watch pharma's next move in Alzheimer's disease. <https://www.oliverwyman.com/our-expertise/perspectives/health/2022/dec/its-time-to-watch-pharmas-next-move-in-alzheimers-disease.html>.
- Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer, 1st edition.
- Raket, L. L. (2022). Progression models for repeated measures: Estimating novel treatment effects in progressive diseases. *Statistics in Medicine*, 41:p. 5537–5557. <https://doi.org/10.1002/sim.9581>.
- Rasmussen, J. and Langerman, H. (2019). Alzheimer's disease – Why we need early diagnosis. *Degenerative Neurological and Neuromuscular Disease*, 9:p. 123–130. <https://doi.org/10.2147/DNND.S228939>.
- United States Food and Drug Administration (2023). FDA converts novel Alzheimer's disease treatment to traditional approval. <https://www.fda.gov/news-events/press-announcements/fda-converts-novel-alzheimers-disease-treatment-traditional-approval>.
- U.S. Food and Drug Administration (2022). Multiple endpoints in clinical trials. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/multiple-endpoints-clinical-trials#:~:text=As%20the%20number%20of%20endpoints,not%20appropriate%20adjustment%20for%20multiplicity>.
- Wang, W.-L. (2015). Approximate methods for maximum likelihood estimation of multivariate nonlinear mixed-effects models. *Entropy*, 17:p. 5353–5381. <https://doi.org/10.3390/e17085353>.

- Wessels, A. M., Dennehy, E. B., Dowsett, S. A., Dickson, S. P., and Hendrix, S. B. (2021). Meaningful clinical changes in Alzheimer disease measured with the iADRS and illustrated using the Donanemab TRAILBLAZER-ALZ study findings. *Neurology. Clinical practice*, 4(10). <https://doi.org/10.1001/jamanetworkopen.2021.29392>.
- Whitehead, S. J. and Ali, S. (2010). Health outcomes in economic evaluation: the qaly and utilities. *British Medical Bulletin*, 96:p. 5–21. <https://doi.org/10.1093/bmb/ldq033>.
- World Health Organization (2020). The top 10 causes of death. <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>.
- Zanetti, O., Solerte, S., and Cantoni, F. (2009). Life expectancy in Alzheimer's disease (AD). *Archives of Gerontology and Geriatrics*, 49:p. 237–243. <https://doi.org/10.1016/j.archger.2009.09.035>.

