

BONN-AACHEN INTERNATIONAL CENTER FOR INFORMATION
TECHNOLOGY (B-IT)
UNIVERSITY OF BONN

Master's Thesis
Master Programme in Life Science Informatics

Validation of gait parameters for prediction of Parkinson's disease progression

Author: Jackrite To
First Supervisor: Prof. Dr. Holger Fröhlich
Second Supervisor: Dr. Martin Vogt
Advisor: Tamara Raschka
Submission Date: 19th of June 2023

Acknowledgments

I want to express my gratitude to Prof. Dr. Holger Fröhlich for his important guidance and assistance during the working period. His rigorous approach to scientific research definitely benefitted me.

I want to thank Tamara Raschka for her supervision of my thesis, assistance, insightful comments and suggestions during the entire study process.

I want to thank Tom Hähnel for his insightful comments and suggestions during the entire study process.

I want to express my gratitude to Prof. Dr. Martin Hofmann-Apitius and Prof. Dr. Holger Fröhlich for providing me with the great opportunity to work with Fraunhofer SCAI to complete my master's thesis.

Additionally, I want to thank my colleagues at Fraunhofer SCAI for their valuable assistance and feedbacks: Colin Birkenbihl, Manuel Lentzen, Sophia Krix and Danqi Wang.

Abstract

Parkinson's disease (PD) is a disorder that affects the nervous system. In the long term, parts of the brain deteriorate causing involuntary movements of the body. Among the noticeable motor symptoms, the gait of the individual is gradually perturbed over time. The progression and severity of PD symptoms are currently assessed via the Unified Parkinson's Disease Rating Scale (UPDRS) and the Hoehn & Yahr (HY) scale. This UPDRS consists of subjective ratings in questionnaires filled by the patient, their caregivers and the clinician. Since subjective ratings might be subject to inconsistencies, there is a growing demand to provide quantitative assessments of gait. Gait parameters extracted from wearable digital devices worn by the PD patient have the potential to help with the diagnosis, as well as, predict PD progression. The challenge is to accurately validate the gait parameters as digital biomarkers to detect PD and monitor its progression.

Instead of performing multiple single statistical analyses per outcome, the latent time joint mixed effects model (LTJMM) can take into account the correlation among multiple outcomes simultaneously for modeling multiple long-term trajectories. Two cohorts of patients (Erlangen and NCERPD) have been analyzed with the LTJMM. Despite the absence of the HY stage of the patient in the formula of the LTJMM, the latent time ordering of PD patients reflected the HY progression of the severity of the patient.

Moreover, linear modeling techniques testing the association of gait parameters with UPDRS ratings were used in different frameworks. Both the longitudinal analysis (Erlangen) and the cross-sectional one (NCERPD) revealed that Gait speed, Stride length, Stride time and Stance time were associated with the clinical score axial score.

These results have important implications for clinicians and PD patients. They could improve the objective rating accuracy of motor symptoms of PD patients and their progression.

Contents

Acknowledgments	iii
Abstract	v
Abbreviations	viii
1. Introduction	1
1.1. Background	1
1.2. Related work	2
1.3. Contribution of the thesis	3
2. Theoretical background	4
2.1. Parkinson's disease rating	4
2.1.1. Unified Parkinson's Disease Rating Scale (UPDRS)	4
2.1.2. Limitations of UPDRS	4
2.1.3. The Hoehn and Yahr scale (HY)	5
2.1.4. Subtypes Parkinson's Disease score	5
2.2. Measurement of the PD patient movement in clinical settings	5
2.2.1. Gait measurements	5
2.2.2. Tasks performed by the PD patient in clinical settings	6
2.3. Statistical model	7
2.3.1. Linear model	7
2.3.2. Linear mixed effect model	7
2.3.3. Analysis of variance (ANOVA)	9
2.3.4. Latent time joint mixed-effects model (LTJMM)	11
2.3.5. Multiple comparison correction	12
3. Material and methods	14
3.1. Study design	14
3.2. Variables	14
3.3. Preprocessing of data for LTJMM, LM and LME	20
3.4. Equations of models	21
3.5. Level of analysis and research scenarios	24
3.6. Multiple comparison correction	25
3.7. Visualization and similarities between Erlangen and NCERPD analysis	26
4. Results	27
4.1. Dataset description	27
4.2. LTJMM	29
4.2.1. Convergence diagnostics	29
4.2.2. LTJMM outputs	36

Contents

4.3.	ANOVA	41
4.3.1.	Correlation of gait parameters	41
4.3.2.	Significant adjusted p-values	42
4.3.3.	Similarities between Erlangen and NCERPD analysis	50
4.4.	Evaluation of LME and LM	51
4.4.1.	Diagnostics plot	51
4.4.2.	R-squared (R ²)	56
5.	Discussion	59
5.1.	Principle Findings	59
5.2.	Limitations	61
5.3.	Future work	61
6.	Conclusion	63
Bibliography		64
A. Supplementary material		66
List of Figures		69
List of Tables		70

Abbreviations

List of Abbreviations:

- ANOVA: Analysis of Variance
- BY: Benjamin Yekutieli
- diffBLaxialscore: Difference to Baseline of Axial Score
- diffBLupdrs3: Difference to Baseline of UPDRS3
- diffUPDRS_gaitBL: Difference to Baseline of the UPDRS Gait Score
- HY: Hoehn and Yahr
- LM: Linear Model
- LME: Linear Mixed Effect Model
- LTJMM: Latent Time Joint Mixed-Effects Model
- PIGD score: Postural Instability and Gait Disturbance Score
- rdm_AXSC: Random Slope of Axial Score
- rdm_PIGD: Random Slope of Postural Instability and Gait Disturbance
- rdm_TD: Random Slope of Tremor Dominant
- rdm_UPRDS: Random Slope of UPDRS
- REL_axialscore: Relative Score to Baseline at Next Visit of Axial Score
- REL_PIGD score: Relative Score to Baseline at Next Visit of PIGD Score
- REL_Tremor score: Relative Score to Baseline at Next Visit of Tremor Score
- REL_UPRDS: Relative Score to Baseline at Next Visit of UPDRS
- TUG: Time Up and Go
- TD: Tremor Dominant
- UPDRS: Unified Parkinson's Disease Rating Scale

1. Introduction

1.1. Background

Parkinson's disease (PD) is a disorder that affects the nervous system. There are around 108-257 per 100,000 persons in Europe who have PD, and the incidence rate is at around 11-19 per 100,000 every year (Balestrino & Schapira, 2020). The degeneration of dopamine-producing neurons in the brain causes motor and non-motor symptoms. In the long term, parts of the brain deteriorate causing involuntary movements of the body. The order and the severity of the symptoms are heterogeneous among patients. Motor symptoms impede normal walking movement. Hyposmia, insomnia, sadness and constipation are non-motor symptoms. The gait of the individual is gradually perturbed over time. Thus, PD is a progressive disorder. PD is often diagnosed when symptoms are clearly developed. However, the actual disease onset started a long time ago before the diagnosis.

During the early stages of the illness, these symptoms typically affect one side of the body more than the other. The asymmetry of motor symptoms in PD has been studied as a possible indicator of disease severity or progression. Imaging (Zhong, Merkitch, Karaman, et al., 2019), histopathologic analyses (Riederer, Jellinger, Kolber, et al., 2018) that show asymmetric nigrostriatal degeneration corresponding with lateralized severity supported that motor asymmetry occurs in Parkinson's disease. The left-predominant PD patient was linked to lower motor function according to Cubo, Martinez-Martin, Gonzalez-Bernal, et al., 2020. Elkurd, Wang, and Dewey Jr, 2021 discovered that left-side lateralization was linked to quicker symptom development and worse results across several clinical categories. As the illness progress, both sides eventually get afflicted. However, symptoms are less severe on one side compared to the other.

There is currently no known cure for PD. However, drugs, treatments, and therapies can alleviate symptoms. In PD patients, a deficiency of dopamine is observed in the brain. Medication such as Levodopa, a dopamine precursor, is transformed into dopamine in the brain, restoring dopamine levels and easing motor symptoms. When Levodopa is given to PD patients, it helps to improve motor function by decreasing tremors, boosting movement, and regaining balance and coordination.

There is no currently specific biologic test for the diagnosis of PD. A neurologist will diagnose PD using information from medical history, an analysis of the symptoms and a neurological and physical examination.

Moreover, the progression of PD from diagnosis is currently assessed via the Unified Parkinson's Disease Rating Scale (UPDRS). The UPDRS is based on questionnaires that have to be filled independently by the patient, their caregivers and the clinician. However, there is a growing demand to assess the progression of PD more accurately and efficiently. Digital biomarkers could answer this challenge.

Digital biomarkers are objective and measurable physiological and behavioral data

1.2. RELATED WORK

that are collected using digital devices like wearables, smartphones, or sensors. These markers can be used in the context of disease to understand, influence, or predict health outcomes (Fröhlich, Bontridder, Petrovska-Delacréta, et al., 2022). By providing continuous and real-time data, they offer valuable insights into an individual's health status, disease progression, and response to treatment. For instance, Verily Life Sciences has developed a PD motor exam using accelerometer data from smartwatches (Burq, Rainaldi, Ho, et al., 2022). This exam measures various aspects of motor function, including tremors, bradykinesia, and dyskinesia. It was consistent with standard disease severity ratings. Although several studies have been performed to develop and validate digital biomarkers for diagnostic purposes, validation for prognostic and prediction is still lacking. Thus, digital biomarkers require to be compared with the standard UPDRS to be validated and show their added value to current clinical measures.

1.2. Related work

Researchers (Iddi, Li, Aisen, et al., 2018) used the Parkinson's Progression Markers Initiative (PPMI) dataset to track different trajectories using a latent time joint mixed-effects model(LTJMM). The model was shown to properly indicate the temporal ordering of disease state in line with the diagnosis groups and biology of PD at baseline (Healthy control, Prodromal state, PD), even though the diagnostic category was not included in the model. Additionally, outputs from the LTJMM such as the random slope of PIGD and total UPDRS tended to correlate with one another.

A more recent study (Severson, Chahine, Smolensky, et al., 2021) found that there are eight distinct types of Parkinson's disease, each with its specific symptoms (motor and non-motor) and disease progression. The researchers used a new machine-learning method to detect these states. Their approach found non-sequential, overlapping disease progression trajectories, suggesting the use of non-deterministic disease progression models and indicating that static subtype assignment may be insufficient to accurately represent the range of Parkinson's disease progression.

Further, studies explored the feasibility of assessing the gait impairment progression of PD patients with digital sensor devices. Researchers (Rodriguez-Molinero, Sama, Perez-Lopez, et al., 2017) have designed a gait algorithm based on acceleration readings from a sensor at the waist. The purpose was to characterize gait by stride frequency content using accelerometer data to detect motor variations in Parkinson's disease patients. The algorithm outputs were moderately correlated with UPDRS3. They had a higher correlation with several UPDRS3 axial function-related items: walking, rising from a chair, overall bradykinesia, and posture. Items connected to tremors show a weak association.

In another study (Schlachetzki, Barth, Marxreiter, et al., 2017), 63 individuals were examined over around 17 months and classified into stable, better, and worsening gait performance compared to their baseline visit according to the clinician's rating of the UPDRS-III item "gait". Gait parameters were extracted from a wearable sensor. Stride length changes were associated with the clinician's evaluation of the UPDRS-III gait item. Patients who had better gait qualities as judged by the doctor had a 4.4% longer stride. In comparison, individuals whose gait had deteriorated had strides that were

shorter by around 8.3%. These results revealed that wearable sensor-based gait analysis could be applied to identify changes in the gait of the patient.

In a more recent study (Schalkamp, Peall, Harrison, & Sandor, 2022), accelerometry data were extracted from a wrist-worn triaxial accelerometer on the dominant hand. Machine learning models that were trained using accelerometry data outperformed all other models based on genetics, lifestyle, blood biochemistry, and prodromal symptoms data in their ability to distinguish between clinically diagnosed PD and prodromal PD up to seven years before diagnosis from the general population.

1.3. Contribution of the thesis

This thesis aimed to model multiple long-term trajectories of Parkinson's disease (PD) patients using the Latent Time Joint Mixed Effects Model (LTJMM) and investigate the potential of gait measurement extracted from a wearable sensor as digital biomarkers of PD progression.

Longitudinal data from two main sources were collected, the Erlangen and the NCERPD dataset, consisting of demographic data, Unified Parkinson's Disease Rating Scale (UPDRS) scores, and gait parameters. These data were preprocessed with normalization of UPDRS scores, removal of outliers and standardization of gait parameters. PD patients were classified based on the predominant side affected left or right. The classification was made from the pre-existent left and right UPDRS scores.

The LTJMM was applied using Bayesian methods for the estimation of time shift shared across outcomes. It allowed visualizing on a common time scale the temporal order of patients based on their Hoehn and Yahr PD severity rating at baseline. Estimation of random slopes of UPDRS scores was performed with LTJMM. The correlation between these random slopes of UPDRS scores was explored.

With a combination of linear regression model and ANOVA, investigations on how gait parameters were associated with clinical scores were performed. ANOVA aimed to explore whether gait parameters could be associated with UPDRS score. Therefore, a full linear regression model containing latent time predictor, gait parameters predictors and UPDRS outcome dependent variable was compared to a reduced model containing only latent time predictor and UPDRS outcome. Multiple approaches in terms of the number and type of gait parameters involved in the model were performed. Furthermore, a method that took into account the patient's side predominance, as well as the localization of the sensor (left or right) on the foot of the patient, was used. These approaches consist of all gait parameters, individual gait parameter, task, ipsicontralateral, ipsilateral and contralateral. They defined different ways of grouping different gait parameters in the full model and testing different groupings of PD patients. These multiple ANOVA tests were corrected with a False Discovery Rate method. Only significant ANOVAs were reported and processed into a tree graph structure for better visualization. Similarities between the significant ANOVAs in Erlangen and the ones in NCERPD were searched.

Aggregations of multiple significant ANOVAs were performed with the help of count and ratios. The fit performance of LTJMM, LME and LM was analyzed. The goodness of fit of LM and LME was assessed with linearity, influential observations plot, QQ-plot and R² score. The fit of LTJMM was assessed by MCMC convergence diagnostics.

2. Theoretical background

2.1. Parkinson's disease rating

2.1.1. Unified Parkinson's Disease Rating Scale (UPDRS)

The UPDRS is not a method for diagnosing PD but rather a test for assessing the severity and course of PD. It is separated into four sections:

Part 1: Non-motor experiences of daily living

Part 2: Motor experiences of daily living

Part 3: Motor examination

Part 4: Motor complications

In the UPDRS Part 3, a wide range of items is used to evaluate many aspects of motor function, such as tremor, stiffness, bradykinesia, gait, and postural instability. A skilled physician assesses the patient's performance and rates each item according to predetermined standards. The sum of all these items represents the UPDRS3 score, meaning the total motor impairment seen during the examination. More severe motor symptoms are indicated by a higher score and lesser motor impairment by a lower score.

The original UPDRS test was changed by the Movement Disorder Society (MDS) to solve several flaws and enhance consistency and precision. One significant change was to convert all items to a 0-4 rating scale, which was consistent with the overall scale and offered a more reliable framework for scoring. The new UPDRS is named MDS-UPDRS. The MDS-UPDRS has undergone revisions that improve its consistency, standardization, and patient comprehension.

2.1.2. Limitations of UPDRS

Due to its reliance on patients' self-reported symptoms and doctors' subjective assessments, the MDS-UPDRS is considered a subjective scale.

The patient's emotional state, understanding of the questions, and capacity to recall and articulate their symptoms are only a few of the variables that might have an impact on subjective measurements. These variables cause findings to vary and may compromise the scale's accuracy.

Given the significant motor component symptoms of Parkinson's disease, using motion sensors like accelerometers and gyroscopes becomes beneficial in monitoring and analyzing the frequency and severity of these symptoms in patients' daily lives. Compared to conventional questionnaire-based clinical examinations, this technology enables a more objective assessment of illness symptoms.

2.2. MEASUREMENT OF THE PD PATIENT MOVEMENT IN CLINICAL SETTINGS

2.1.3. The Hoehn and Yahr scale (HY)

A common method for describing the progression of Parkinson's disease symptoms and the degree of impairment is the Hoehn and Yahr scale. Margaret Hoehn and Melvin Yahr first published it in 1967. The rating is made based on the physical examination and medical history of the PD patient.

It has been discovered that HY stage progression was correlated with motor impairment, and dopaminergic loss (Riederer, Jellinger, Kolber, et al., 2018).

The following HY stages of PD progression are described:

- Stage 0: No signs of disease
- Stage 1.0: Unilateral involvement only, usually with minimal or no functional disability
- Stage 1.5: Unilateral and axial involvement
- Stage 2.0: Bilateral involvement without impairment of balance
- Stage 2.5: Mild bilateral involvement with recovery on retropulsion (pull) test
- Stage 3.0: Mild to moderate bilateral involvement, some postural instability but physically independent
- Stage 4.0: Severe disability, still able to walk and stand unassisted
- Stage 5.0: Wheelchair-bound or bedridden unless aided

2.1.4. Subtypes Parkinson's Disease score

There are several subtypes or phenotypes of Parkinson's disease that characterize the illness's predominant motor symptoms. Tremor-dominant (TD), postural instability and gait difficulties (PIGD), and axial impairment are three categories frequently identified. In tremor-dominant patients, the main symptom is resting tremor. Hands, arms or legs show noticeable tremors. Compared to other subtypes, these patients have less rigidity and milder bradykinesia meaning slowness of movement. In PIGD, the patient shows severe difficulties with balance, posture and gait. He can have a stumbling walk, difficulties keeping a firm posture, and a propensity to fall. Additionally, he may walk with less arm swing and shows bradykinesia and rigidity. In the axial impairment phenotype, axial symptoms include freezing of gait, postural instability, trunk posture alterations and dysarthrophonia. Additionally, dopaminergic medications and surgical treatments for these symptoms are ineffective.

2.2. Measurement of the PD patient movement in clinical settings

2.2.1. Gait measurements

Gait sensors are equipped with an inertial measurement unit (IMU). It consists of an accelerometer, a gyroscope and a magnetometer. The accelerometer determines the acceleration of the object in three-dimensional space, the gyroscope measures an object angular velocity. The magnetometer measures the orientation of the object in relation to the magnetic fields of the Earth. By integrating these 3 types of information, an IMU can calculate the object's location, velocity, acceleration, and orientation in real-time.

2.2. MEASUREMENT OF THE PD PATIENT MOVEMENT IN CLINICAL SETTINGS

These measurements are further processed in gait parameters that characterize a person's gait. Gait parameters are below described:

- Stride time: The amount of time it takes from the first foot contact to the next foot contact for a full stride to be completed.
- Swing time: The period of time when the foot is traveling toward the next step while not making contact with the ground.
- Stance time: The period of time the foot is firmly planted and supporting weight.
- Stride length: The length of a single stride measured from the first initial contact of one foot to the subsequent first contact of the same foot.
- Gait velocity is the rate at which a person walks, typically expressed in meters per second.
- Maximum lateral excursion: The maximum lateral distance of the body's center of mass during a stride.

2.2.2. Tasks performed by the PD patient in clinical settings

A person's gait, mobility and balance can be assigned by a clinician during a particular task executed by the patient. Examples of tasks are described below :

- The Timed Up and Go (TUG): During the test, the person is timed while they rise from a chair, walk 3.5 meters, turn around, walk back to the chair, and sit down again. Turn, Tray and Count are variations of the TUG task that requires an increasingly challenging additional task for the patient to perform.
 - Turn: TUG and the patient has to turn on himself in the middle of the 3.5m distance one way and back.
 - Tray: Turn and the patient has to carry a tray with a glass on it.
 - Count: Turn and the patient has to execute a mental mathematical calculation.
 - 2x10mPreWithStop: The patient has to walk 2 times 10 meters back and forth with stop.
 - 4x10mPreWithoutStop: The patient has to walk 4 times 10 meters back and forth without stop.

2.3. Statistical model

2.3.1. Linear model

A linear model (LM) is a statistical method used to model the association between a dependent variable and one or more independent variables. The relation between the variables is assumed to be linear, which means that changes in the independent variables will be proportional to the change in the dependent variable. Fitting a linear model to the observed data consists of estimating the regression coefficients that best describe the association between the dependent variable and the independent variables.

The mathematical formula is :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

\mathbf{y} is the vector of dependent variables

\mathbf{X} is the matrix of independent variables

$\boldsymbol{\beta}$ is the vector of coefficients

$\boldsymbol{\epsilon}$ is the vector of errors

2.3.2. Linear mixed effect model

A linear mixed effect model (LME) extends the linear model by inserting both fixed effects and random effects in the model. This model is suitable for situations when there are repeated measurements within an individual and where the observations from the same individual are correlated, such as in longitudinal studies.

The fixed effects show the population-level associations between the dependent variable and the independent variables. These fixed effects represent the average effects of the predictors on the whole population of individuals.

The variance in the data that is unique to each individual is captured by random effects. These random effects capture the heterogeneity of individuals inside a cohort that cannot be only determined by fixed effects. The LME captures both within-population and within-individual variability by estimating fixed and random effects coefficients.

Several works of literature exist on the linear mixed model. The mathematical description of a linear mixed model is derived from Oskolkov, 2020 and Pinheiro and Bates, 2009.

The mathematical formula of a linear mixed model is:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{K}\boldsymbol{u} + \boldsymbol{\epsilon}$$

\mathbf{Y} : the response variable

\mathbf{X} : the design matrix for the fixed effects

\mathbf{X} : a $n \times p$ matrix, where n is the number of observations and p is the number of fixed effects. Each row of \mathbf{X} corresponds to an observation, and each column corresponds to a fixed effect.

$\boldsymbol{\beta}$: the vector of fixed effect coefficients

\mathbf{K} : the design matrix for the random effects

K : a $n \times q$ matrix, where n is the number of observations and q is the number of random effects. Each row of K corresponds to an observation, and each column corresponds to a random effect.

u : the vector of random effect coefficients

ϵ : the vector of residuals

The random-effects vector, u , and the error vector, ϵ , are assumed to follow an independent prior normal distribution with mean = 0, where K is a symmetric and positive semidefinite matrix, I is an $n \times n$ identity matrix, and σ^2 is the error variance.:.

$$u \sim N(0, \sigma^2 K)$$

$$\epsilon \sim N(0, \sigma^2 I)$$

Definition of the expected value of Y and variance of Y

$E[Y]$: the expected value of Y

$$Y = X\beta + Ku + \epsilon$$

$$E[Y] = X\beta + 0 + 0$$

$\text{var}[Y]$: variance of Y

$\sigma^2 s$: random effect variance

σ^2 : residual variance

Σ_y : variance covariance matrix filled with $\sigma^2 s$ and σ^2

$$\text{var}[Y] = \Sigma_y = 0 + \text{var}(Ku) + \text{var } \epsilon = \sigma^2 s K K^T + \sigma^2 I$$

The aim of fitting a linear mixed model is an optimization problem that maximizes the likelihood of a multivariate Gaussian distribution function with respect to parameters β , $\sigma^2 s$, σ^2 .

The likelihood function L for the linear mixed model with respect to parameters β , Σ_y is:

$$L(\beta, \sigma^2 s, \sigma^2) = L(\beta, \Sigma_y)$$

$$L(\beta, \Sigma_y) = \frac{1}{\sqrt{2\pi} |\Sigma_y|} \exp \left(\frac{1}{2} (Y - X\beta)^T \Sigma_y^{-1} (Y - X\beta) \right)$$

Maximizing the likelihood of the data is equivalent to minimizing the log-likelihood.

2 methods exist for the optimization problem: the estimation via Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML).

Maximum Likelihood

The maximum likelihood estimates both fixed-effects regression coefficients β and the variance components in Σ_y .

$$\log(L(\beta, \Sigma_y)) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(|\Sigma_y|) - \frac{1}{2} (\mathbf{Y} - \mathbf{X}\beta)^T \Sigma_y^{-1} (\mathbf{Y} - \mathbf{X}\beta)$$

Restricted Maximum Likelihood

Restricted maximum likelihood estimates only the variance components in Σ_y . β is estimated in a second step.

$$\log \left[\int L(\beta, \Sigma_y) d\beta \right] = -\frac{1}{2} \log(|\Sigma_y|) - \frac{1}{2} (\mathbf{Y} - \mathbf{X}\hat{\beta})^T \Sigma_y^{-1} (\mathbf{Y} - \mathbf{X}\hat{\beta}) - \frac{1}{2} \log(|\mathbf{X}^T \Sigma_y^{-1} \mathbf{X}|)$$

When the random effects are of importance, the REML technique is chosen over the maximum likelihood method because it offers unbiased estimates of the random effects.

The estimated coefficients $\hat{\beta}$ and \hat{u} are obtained once the optimization has been solved.

2.3.3. Analysis of variance (ANOVA)

A good model must be effective in fitting the data. It should have an adequate level of complexity to represent the data. To support its greater complexity, the complex model must offer a better fit to the data compared to a simpler model. ANOVA is used to determine if including more variables in a model significantly enhances its ability to fit data as opposed to a model with fewer variables. The ANOVA examines whether the full model with added variables explains more variability than the simpler one. To compare the full model with the reduced model, the reduced model should be nested in the full model. The fit of the full model is compared with the fit of the reduced model. There are 2 ways to perform the comparison: with the F statistic or with the likelihood ratio test.

- Comparison of models with the F statistic

The null model should be a nested version of the full model. The sum of squared residuals (SSR) for both the reduced model and the full model is computed using the ANOVA test for model comparison. If the difference in SSR is statistically significant, it implies that the full model fits the data better than the reduced model, showing that at least one of the additional variables significantly contributes to explaining the variance in the result.

Here is described the mathematical process of the ANOVA with F statistic.

Fit the full model to the data and compute the error sum of squares SSEF

Fit the reduced model to the data and compute the error sum of squares SSER

SSEF (Sum of Squares Error Full Model) = $\sum_{k=1}^p (y - \hat{y})^2$

SSER (Sum of Squares Error Reduced Model) = $\sum_{k=1}^p (y - \hat{y})^2$

where:

y : the observed response

\hat{y} : the fitted response

then compute

$$\text{MSR (Mean Square Regression)} = \frac{SSER - SSEF}{df_R - df_F}$$

$$\text{MSE (Mean Square Error)} = \frac{SSEF}{df_F}$$

where:

2.3. STATISTICAL MODEL

The degrees of freedom df_R and df_F are the ones with the reduced and full model error sum of squares, respectively.

then compute the F statistic :

$$F = \frac{MSR}{MSE}$$

In the F statistic,

the null hypothesis is: the parameter coefficient associated with the added variable is equal to 0

the alternative hypothesis is: the parameter coefficient associated with the added variable is different from 0.

The P-value asks whether the probability that we would have obtained an F statistic as large as we did if the null hypothesis were true. F is compared to an F distribution that has $df_R - df_F$ numerator degrees of freedom and df_F denominator degree of freedom to calculate the P-value.

- Comparison of models with the likelihood ratio test statistic

The likelihood ratio test statistic formula is used to compare the two models. The null model should be a nested version of the full model.

$$\text{Deviance} = -2 \ln \left(\frac{\mathcal{L}(\theta_R)}{\mathcal{L}(\theta_F)} \right)$$

$\mathcal{L}(\theta_R)$ is the likelihood of the null model (reduced model) and $\mathcal{L}(\theta_F)$ is the likelihood of the alternative model (full model).

The deviance test statistic is compared to a chi-squared distribution with degrees of freedom equal to the difference in the number of parameters between the two models.

- The ANOVA's type III

There exist 3 different types of ANOVA: the ANOVA's type I, the ANOVA's type II, and the ANOVA's type III. Their difference relies on the way the sum of squares is calculated in the full and reduced model. In the F statistic, SSER-SSEF is defined as the sequential sum of squares. SSER-SSEF is equivalent to the reduction in the error sum of squares (or increase in the regression sum of squares). ANOVA type III determined the reduction in the error sum of squares when the reduced model, which includes existent predictors, is expanded to include one or more predictor variables. It implies that the order of specification of predictors does not matter.

The Type III Sums of Squares (SS) do indicate an interaction effect and are mathematically described here:

- $SS(A | B, AB)$ for independent variable A, the variation assigned to independent variable A is accounting for B and interaction effect AB
- $SS(B | A, AB)$ for independent variable B, the variation assigned to independent variable B is accounting for A and interaction effect AB

2.3.4. Latent time joint mixed-effects model (LTJMM)

The LTJMM builds further on LME by adding a latent time element in the model. Li, Iddi, Thompson, et al., 2019 introduced the LTJMM as a way to model multivariate longitudinal data together. It estimates the underlying time structure of the data by analyzing time-varying covariates. In the context of time-related diseases such as PD, it is able to model multiple long-term trajectories taking into account correlation among outcome measures. This model allows computation of intra and inter-individual variability across a common disease time scale derived from a mean disease progression trajectory. A subject-specific latent time shift parameter allows time estimation of the progression of the subject into disease relative to the population. Moreover, a patient-specific random slope parameter defines the rate of change of evolution of the disease compared to the average individual.

LTJMM equation

$$y_{ijk} = x t_{ijk} \beta_k + \gamma_k(t_{ijk} + \delta_i) + \alpha_{0ik} + \alpha_{1ik} t_{ijk} + \epsilon_{ijk}$$

i: individual subject

j: individual time per subject and outcome

k: outcome

t: time

X : vector of covariates ; fixed effects

y : vector of outcomes

δ_i : time shift progression of the subject relative to the average PD population; random effects

α_{0ik} : subject outcome specific random intercept ; random effects

α_{1ik} : subject outcome specific random slope ; random effects

ϵ_{ijk} : error term

The LTJMM estimation of a parameter is based on the principle of Bayesian linear regression. The goal of Bayesian inference is to determine the posterior distribution of the model parameters.

$$p(\beta, \sigma^2 | y, X) = \frac{p(y|X, \beta, \sigma^2) \cdot p(\beta, \sigma^2)}{p(y|X)}$$

$p(\beta, \sigma^2 | y, X)$: posterior distribution of the parameters β (regression coefficients) and σ^2 (error variance) given the observed data y and the design matrix X

$p(y|X, \beta, \sigma^2)$: likelihood function, which is the probability of observing the data y given the parameters β and σ^2 and the design matrix X .

$p(\beta, \sigma^2)$: prior distribution of β and σ^2 , which represents the initial belief about the parameters before observing the data

$p(y|X)$: marginal likelihood which is a normalization constant that enables the integration of the posterior distribution to 1.

The marginal likelihood is not of interest. Thus, the posterior distribution is a proportionality:

$$p(\beta, \sigma^2 | y, X) \propto p(y | X, \beta, \sigma^2) p(\beta, \sigma^2)$$

However, the posterior distribution is often intractable. Therefore, it is approximated with the Markov Chain Monte Carlo (MCMC) method.

The MCMC method requires a sampler, a model, the data and the priors.

The sampler will explore the parameter space available. It will generate random different parameters.

The prior probability of each parameter is determined and returns to the sampler.

Each parameter is used jointly with the data to determine the likelihood.

The likelihood is multiplied by the prior to output the posterior probability.

Numerous individual samples of the unknown posterior distribution are computed following a specific algorithm such as the Metropolis-Hastings algorithm. These samples are connected and form a chain from the term "Markov chain". Every sample in the chain corresponds to an estimated potential state of the parameter. The term Monte Carlo refers to the method of sampling from a probability distribution. The Metropolis-Hastings algorithm is a way to sample the target posterior distribution. It iteratively selects samples from a proposal distribution and accepts or rejects them following a predetermined acceptance ratio to build the chain.

From all these samples linked into a chain, the final posterior distribution of the parameter of interest is approximated.

From Taboga, 2021, the Metropolis-Hastings algorithm is described below with θ the parameters of interest:

1. Initialize the chain with an initial state $\theta^{(0)}$.
2. For $t = 1, 2, \dots, T$:
 - a) Sample a parameter θ^* from a proposal distribution $q(\theta^* | \theta^{(t-1)})$.
 - b) Calculate the acceptance ratio with f a stationary distribution:
$$r = \frac{f(\theta^*) q(\theta^{(t-1)} | \theta^*)}{f(\theta^{(t-1)}) q(\theta^* | \theta^{(t-1)})}$$
 - c) Generate an uniform number $u \sim \text{Uniform}(0, 1)$.
 - d) If $u \leq \min(1, r)$, accept the proposal and set $\theta^{(t)} = \theta^*$. Otherwise, set $\theta^{(t)} = \theta^{(t-1)}$.
 - e) Continue until desired samples θ^* are generated

2.3.5. Multiple comparison correction

Multiple comparison correction is a statistical technique used to minimize the increased risk of false positives while doing numerous hypothesis tests. As the number of tests rises, the likelihood of finding at least one significant result by chance alone increases, which raises the likelihood of incorrectly rejecting a null hypothesis. To account for the number of tests runs, multiple comparison correction methods adjust the p-values obtained from individual tests. The False Discovery Rate (FDR) correction is one technique employed. FDR correction controls the expected proportion of false

2.3. STATISTICAL MODEL

discoveries among all significant results. More specifically, the Benjamini-Yekutieli FDR correction is particularly useful when the tests and the features in the dataset are not independent.

The BY formula is drawn from “GeneTrail: P-value Adjustments,” 2023 :

$$\gamma = \sum_{i=1}^n \frac{1}{i}$$
$$\bar{p}_i = \begin{cases} \gamma p_i & \text{for } i = n \\ \min\left(\bar{p}_{(i+1)}, \gamma \frac{n}{i} p_i\right) & \text{for } i = n-1, \dots, 1 \end{cases}$$

Explanation of the formula BY:

First, order all p-values from small to large. Associate the rank i of the p-value from small to large. Then multiply each p-value by the total number of tests n and divide by its rank order i and remultiply by the γ (the sum of the inverse of the rank). It results in an adjusted p-value. An adjusted p-value is given the value of its predecessor if it is less than that value. This guarantees the order's monotonicity.

3. Material and methods

3.1. Study design

In Erlangen, the participant was assigned the assessment of gait by a clinician based on clinical examination. Therefore, a gait assessment of an individual with perfect mobility would not be recorded in the Erlangen study. At baseline, relevant measures of the patients were assessed. To track the participants' development, follow-up visits were planned at different intervals to assess clinical scores and gait features.

In NCERPD, no specific criteria were applied for the selection of participants in the study. It consisted of a group of PD participants and a control group. Clinical scores were measured on a longitudinal scale. Gait parameters were only assessed at a single visit.

In both Erlangen and NCERPD, the IMU sensor was placed on the feet of the patient. The IMU sensor was made by Portables. The features were described below.
Accelerometer: 3-axis, adjustable range: $\pm 2 - \pm 16$ g

Gyroscope: 3-axis, adjustable range: $\pm 125 - \pm 2000$ °/s

Barometer: Measure atmospheric air pressure

Synchronization: Wirelessly synchronize multiple sensors

From the IMU sensor raw signals, gait parameters were extracted and computed via dedicated algorithms such as mgl-algo by another scientific team and provided to this study in a precalculated form.

3.2. Variables

Erlangen

Selection of variables

Relevant variables in the study were the following:

Clinical scores:

- UPDRS_motor_scale
- UPDRS_speech
- UPDRS_facial_expr
- UPDRS_arise_from_chair
- UPDRS_gait
- UPDRS_postural_stability

- UPDRS_posture

31 gait parameters :

- stride time [s]_2x10mPrefWithStop_left_sensor
- swing time [s]_2x10mPrefWithStop_left_sensor
- stance time [s]_2x10mPrefWithStop_left_sensor
- stride length [m]_2x10mPrefWithStop_left_sensor
- gait velocity [m/s]_2x10mPrefWithStop_left_sensor
- stride time [s]_2x10mPrefWithStop_right_sensor
- swing time [s]_2x10mPrefWithStop_right_sensor
- stance time [s]_2x10mPrefWithStop_right_sensor
- stride length [m]_2x10mPrefWithStop_right_sensor
- gait velocity [m/s]_2x10mPrefWithStop_right_sensor
- stride time [s]_4x10mPrefWithoutStop_left_sensor
- swing time [s]_4x10mPrefWithoutStop_left_sensor
- stance time [s]_4x10mPrefWithoutStop_left_sensor
- stride length [m]_4x10mPrefWithoutStop_left_sensor
- gait velocity [m/s]_4x10mPrefWithoutStop_left_sensor
- stride time [s]_4x10mPrefWithoutStop_right_sensor
- swing time [s]_4x10mPrefWithoutStop_right_sensor
- stance time [s]_4x10mPrefWithoutStop_right_sensor
- stride length [m]_4x10mPrefWithoutStop_right_sensor
- gait velocity [m/s]_4x10mPrefWithoutStop_right_sensor
- stride time [s]_TUG_left_sensor
- swing time [s]_TUG_left_sensor
- stance time [s]_TUG_left_sensor
- stride length [m]_TUG_left_sensor
- gait velocity [m/s]_TUG_left_sensor
- stride time [s]_TUG_right_sensor
- swing time [s]_TUG_right_sensor

3.2. VARIABLES

- stance time [s]_TUG_right_sensor
- stride length [m]_TUG_right_sensor
- gait velocity [m/s]_TUG_right_sensor
- max. lateral excursion [m]_TUG_right_sensor

Age at diagnosis, medication status, sex, Hoehn and Yahr (HY) scale, and participation date were also included in the relevant variables.

Medication status defined the status of the patient. It was a categorical variable that defined whether the patient took or not regularly drugs that could alleviate his motor symptoms.

Many UPDRS scores contain missing data (95% of NaN), so they were not usable for analysis. Therefore, UPDRS_motor_scale and the components of the axial score (speech, facial expression, arising from chair, gait, postural stability, and posture), and gait were considered relevant for further analysis. UPDRS_motor_scale was considered as the UPDRS3 score. Axial impairment which is the lack of control over the trunk and posture, is quantified by the axial score. The severity of axial symptoms is evaluated using the axial score. The axial score is normally evaluated based on the sum of specific sub-items of the UPDRS3: speech, facial expression, arising from a chair, gait, freezing of gait, postural stability, and posture.

Processing of variables

Further variables were created from the previously selected variables.

- Time since diagnosis

Time since diagnosis in years was calculated from the difference in years between age at diagnosis and the participation date.

- Axial score

The evaluation of axial motor symptoms in Parkinson's disease is represented by the axial score. The axial score is normally evaluated based on the sum of specific sub-items of the UPDRS3: speech, facial expression, arising from chair, gait, freezing of gait, postural stability, and posture. However, in the Erlangen dataset, the sub-item freezing of gait was not available. Thus, the axial score was evaluated on the sum of the above sub-items without freezing of gait.

- Difference of Score from Baseline:

The change in the score of an outcome variable from the baseline assessment to a later assessment was referred to as the difference of score from the baseline. This measurement was created by subtracting the score from the baseline from the score at the visit of interest. This outcome was processed from each of the clinical scores (UPDRS3, axial score and UPDRS_gait) and labeled respectively diffBLupdrs3, diffBLaxialscore and diffBLgait.

- PD side predominance

In the dataset, there was no information on the side predominance where the PD patient was mostly affected. Thus, the assignation of the PD patient to the left or the right predominant side was calculated from the sum of the following variables.

3.2. VARIABLES

For the left side:

- UPDRS_rigidity_LUE
- UPDRS_rigidity_LLE
- UPDRS_action_tremor_LUE
- UPDRS_rest_tremor_LUE
- UPDRS_rest_tremor_LLE
- UPDRS_movements_fingertabs_L
- UPDRS_movements_handMov_L
- UPDRS_movements_altMov_L
- UPDRS_movements_legAgility_L

For the right side:

- UPDRS_rigidity_RUE
- UPDRS_rigidity_RLE
- UPDRS_action_tremor_RUE
- UPDRS_rest_tremor_RUE
- UPDRS_rest_tremor_RLE
- UPDRS_movements_fingertabs_R
- UPDRS_movements_handMov_R
- UPDRS_movements_altMov_R
- UPDRS_movements_legAgility_R

The highest score among the left sum and the right sum decided which predominant side (left or right) the PD patient was assigned to.

NCERPD

Selection of variables

Relevant variables in the study were the following:

Clinical scores:

- MDS-UPDRS1
- MDS-UPDRS2
- MDS-UPDRS3

3.2. VARIABLES

- Tremor score (TD)
- MDS-UPDRS3.1 speech, MDS-UPDRS3.2 facial expression, MDS-UPDRS3.9 arising from chair, MDS-UPDRS3.10 gait, MDS-UPDRS3.11 freezing of gait, MDS-UPDRS3.12 postural stability, and MDS-UPDRS3.13 posture
- PIGD score

48 Gait parameters:

- Stance.Time..s._count_left
- Swing.Time..s._count_left
- Gait.Speed..m.s._count_left
- Stride.Length..cm._count_left
- Max..Lateral.Excursion..cm._count_left
- Stride.Time..s._count_left
- Stance.Time..s._count_right
- Swing.Time..s._count_right
- Gait.Speed..m.s._count_right
- Stride.Length..cm._count_right
- Max..Lateral.Excursion..cm._count_right
- Stride.Time..s._count_right
- Stance.Time..s._tray_left
- Swing.Time..s._tray_left
- Gait.Speed..m.s._tray_left
- Stride.Length..cm._tray_left
- Max..Lateral.Excursion..cm._tray_left
- Stride.Time..s._tray_left
- Stance.Time..s._tray_right
- Swing.Time..s._tray_right
- Gait.Speed..m.s._tray_right
- Stride.Length..cm._tray_right
- Max..Lateral.Excursion..cm._tray_right
- Stride.Time..s._tray_right

3.2. VARIABLES

- Stance.Time..s._tug_left
- Swing.Time..s._tug_left
- Gait.Speed..m.s._tug_left
- Stride.Length..cm._tug_left
- Max..Lateral.Excursion..cm._tug_left
- Stride.Time..s._tug_left
- Stance.Time..s._tug_right
- Swing.Time..s._tug_right
- Gait.Speed..m.s._tug_right
- Stride.Length..cm._tug_right
- Max..Lateral.Excursion..cm._tug_right
- Stride.Time..s._tug_right
- Stance.Time..s._turn_left
- Swing.Time..s._turn_left
- Gait.Speed..m.s._turn_left
- Stride.Length..cm._turn_left
- Max..Lateral.Excursion..cm._turn_left
- Stride.Time..s._turn_left
- Stance.Time..s._turn_right
- Swing.Time..s._turn_right
- Gait.Speed..m.s._turn_right
- Stride.Length..cm._turn_right
- Max..Lateral.Excursion..cm._turn_right
- Stride.Time..s._turn_right

The PIGD score and the TD score were more specifically the mean PIGD score and the mean TD score.

Age at diagnosis, LDD (Levodopa Daily Dose), sex, Hoehn and Yahr scale, and participation date were also included in the relevant variables.

Processing of variables

Further variables were created from the previously selected variables.

- Time since diagnosis

Time since diagnosis in years was calculated from the difference in years between age at diagnosis and the participation date.

The variable LDD (Levodopa Daily Dose) was processed into a categorical variable which defined if the patient had a value greater than zero or not.

- Axial score

The axial score was calculated as the sum of these MDS-UPDRS subscores: 3.1 speech, 3.2 facial expression, 3.9 arising from chair, 3.10 gait, 3.11 freezing of gait, 3.12 postural stability, and 3.13 posture.

- Relative score at next visit

In this dataset, the gait assessment for gait parameters was performed during a single visit. The relative score at the next visit was calculated by subtracting the baseline clinical score from the clinical score at the next visit time point. This outcome was processed from each of the clinical scores (MDS-UPDRS1, MDS-UPDRS2, MDS-UPDRS3, Tremor score, Axial score and PIGD) and labeled respectively REL-UPDRS1, REL-UPDRS2, REL-UPDRS3, REL-Tremorscore, REL-Axialscore and REL-PIGD.

- Predominant side

The features Tremor-predominant side, Rigidity-predominant side, Bradykinesia-predominant side were present in the dataset. Their corresponding values were either Left, Right, or No predominant side. The Parkinson's predominant side of the patient was the one where all values of these features agreed on the same side.

3.3. Preprocessing of data for LTJMM, LM and LME

Normalization

In the Erlangen dataset, the UPDRS 3 score and the axial score were normalized by using the theoretical maximum reachable score as a normalizing factor. In the UPDRS3 questionnaire, there were 27 questions with a maximum score of 4 points each. For the axial score, there were 6 questions with a maximum score of 4 points each. Thus the normalizing factor for the UPDRS3 score was $27 \times 4 = 108$, and for the axial score, it was $6 \times 4 = 24$. After normalization, these scores were converted to a range from 0 to 1 values.

Normalization for the NCERPD dataset was performed in the same manner as in the Erlangen dataset.

Outlier removal

In Erlangen :

Outliers were removed by selecting the variables of interest:

- Stride length [m]_2x10mPrefWithStop_left_sensor
- Stride length [m]_2x10mPrefWithStop_right_sensor

- Stride length [m]_4x10mPrefWithoutStop_left_sensor
- Stride length [m]_4x10mPrefWithoutStop_right_sensor
- Stride length [m]_TUG_left_sensor
- Stride length [m]_TUG_right_sensor

Thus, for each variable, mean and standard deviation were calculated. Subsequently, upper and lower thresholds were determined by multiplying the standard deviation by 5 and adding/subtracting them from the mean, respectively. Values in the variable that surpassed these thresholds were identified as outliers. Outliers were stride length values above 2 meters. Finally, these outliers were excluded from the dataset.

In NCERPD :

The gait parameter variables selected for outlier removal were Max. Toe Clearance [cm] and Max. Foot Clearance [cm]. Values of Max. Toe Clearance above 25 cm were removed. Values of Max. Foot Clearance above 20 cm were removed. Thus, patients with these values were excluded from the dataset.

Standardization

In Erlangen and NCERPD datasets, all gait parameter variables were standardized. First, the mean and standard deviation were calculated for each variable. Subsequently, the mean was subtracted from each observed value of the variable and divided by the standard deviation. Thus, the standardized variable had a mean of 0 and a standard deviation of 1.

3.4. Equations of models

LTJMM

The LTJMM was used to analyze the progression of disease severity over time and investigate the relationship between random slopes of clinical scores.

Bayesian methods were used for estimation and inference, posterior means in the 95% credible intervals were reported for parameter estimates.

The code for LTJMM was used from an R package from
<https://bitbucket.org/mdonohue/ltjmm>.

$$y_{ijk} = x t_{ijk} \beta_k + \gamma_k (t_{ijk} + \delta_i) + \alpha_{0ik} + \alpha_{1ik} t_{ijk} + \epsilon_{ijk}$$

Input in the model :

i: individual patient

j: individual time per patient and outcome

k: clinical outcome

t: time since diagnosis

\mathcal{X} : vector of covariates (intercept, age at diagnosis, sex, medication status)

\mathcal{Y} : vector of clinical outcomes

3.4. EQUATIONS OF MODELS

For Erlangen, the k clinical outcomes were UPDRS3 and axial score.

For NCERPD the k clinical outcomes were MDS-UPDRS1, MDS-UPDRS2, MDS-UPDRS3, PIGD, axial score and TD.

Selection of Outputs of the model:

δ_i : time shift progression for each patient relative to the average PD population and shared among outcomes.

α_{1ik} :individual-patient-outcome specific random slope

The predictors included the time_since_diagnosis variable, which captured the time elapsed since the Parkinson's diagnostic, as well as fixed effects such as Sex, Age_of_onset (age at diagnosis), and medication status. These fixed effects directly influenced the outcome variable. The model also included a random effect on the intercept, specifying a unique intercept for each patient.

The model was fit assuming a multivariate Gaussian distribution for the random effects. Four parallel Markov chains were run for 50000 iterations and the first 25000 warm-up iterations were discarded.

The δ_i was used to calculate the latent time for further analysis:

Latent time = Time since diagnosis + δ_i

The random slopes of clinical scores were further compared between each other and tested for association with gait parameters.

ANOVA

ANOVA was used to determine whether including more gait parameter variables in a model significantly enhanced its ability to fit data as opposed to a model with fewer variables. The ANOVA examined whether the complex model (full model) with added gait parameter variables explained more variability than the simpler one (null model). Thus, gait parameters would show a significant association with the outcomes. The basic structure of the ANOVA is described below including the type of model (LM or LME) and variables used. 2 different ANOVAs were applied in the subsequent level of analysis and research scenario described further.

In Erlangen:

First ANOVA

ANOVA (full model, null model)

Full model:

$$y_{ij} \text{outcomes} = \beta_0 + \beta_1 \cdot \text{latent_time}_{ij} + \beta_2 \cdot \text{gait_parameter}_{ij} + \beta_3 \cdot \text{gait_parameter}_{ij} \cdot \text{latent_time}_{ij} + b_{0i} + e_{ij}$$

y_{ij} is the outcome variable for the jth observation of the ith patient

β_0 is the intercept

β_1 is the coefficient of latent time

β_2 is the coefficient of the gait parameter

β_3 is the interaction effect coefficient between the gait parameter and latent time

latent_time_{ij} is the latent time variable for the jth observation of the ith patient

3.4. EQUATIONS OF MODELS

gait_parameter_ij is the gait parameter variable for the jth observation of the ith patient

b0_i is the random intercept for the ith patient

e_ij is the residual error for the jth observation of the ith patient

Null model:

$$y_{ij} \text{outcomes} = \beta_0 + \beta_1 \cdot \text{latent_time}_{ij} + b_{0i} + e_{ij}$$

In the full model, the dependent variable outcomes was associated to the fixed effects of latent-time, gait parameter, and their interaction (gait parameter * latent-time), as well as a random intercept for each level of the grouping variable Patient. Each patient had its own intercept.

In the null model, the dependent variable outcomes was associated to the fixed effects of latent-time, as well as a random intercept for each level of the grouping variable Patient.

For the first ANOVA, the outcomes were UPDRS3, axial score, rdm_UPDRS3 (random slope of UPDRS3), rdm_AXSC (random slope of axial score), diffBLupdrs3 (difference to baseline of UPDRS3), diffBLaxialscore (difference to baseline of the axial score), diffUPDRS_gaitBL (difference to baseline of the UPDRS gait score)

Second ANOVA

ANOVA (full model, null model)

Full model:

$$y_{ij} \text{latenttime} = \beta_0 + \beta_1 \cdot \text{sex}_i + \beta_2 \cdot \text{age_at_diagnosis}_i + \beta_3 \cdot \text{gait_parameter}_{ij} + b_{0i} + e_{ij}$$

Null model:

$$y_{ij} \text{latenttime} = \beta_0 + \beta_1 \cdot \text{sex}_i + \beta_2 \cdot \text{age_at_diagnosis}_i + b_{0i} + e_{ij}$$

In the full model, the dependent variable latent-time was associated with the fixed effects of gait parameters, sex and age at diagnosis, as well as a random intercept for each level of the grouping variable Patient. Each patient had its own intercept.

In the null model, the dependent variable latent-time was associated with the fixed effects of sex and age at diagnosis, as well as a random intercept for each level of the grouping variable Patient.

In NCERPD:

First ANOVA

ANOVA (full model, null model)

Full model:

$$y_{\text{outcomes}} = \beta_0 + \beta_1 \cdot \text{latent time} + \beta_2 \cdot \text{gait parameter} + \beta_3 \cdot (\text{gait parameter} \times \text{latent time}) + e$$

Null model:

$$y_{\text{outcomes}} = \beta_0 + \beta_1 \cdot \text{latent time} + e$$

In the full model, the dependent variable Outcomes was associated to the fixed effects of latent-time, gait parameter, and their interaction (gait parameter * latent-time).

In the null model, the dependent variable outcomes was associated with the fixed effects of latent-time.

For the first ANOVA, the outcomes were: UPDRS1, UPDRS2, UPDRS3, Tremor score, axial score, PIGD score, REL_UPRDS1 (relative score to baseline at next visit of UPDRS1), REL_UPRDS2, REL_UPRDS3, REL_Tremor score, REL_axialscore, REL_PIGD score, rdm_UPRDS1 (random slope of UPDRS1), rdm_UPRDS2, rdm_UPRDS3, rdm_TD, rdm_axialscore, rdm_PIGD

Second ANOVA

ANOVA (full model, null model)

Full model:

$$y_{\text{latent time}} = \beta_0 + \beta_1 \cdot \text{sex} + \beta_2 \cdot \text{age at diagnosis} + \beta_3 \cdot \text{gait parameter} + e$$

Null model:

$$y_{\text{latent time}} = \beta_0 + \beta_1 \cdot \text{sex} + \beta_2 \cdot \text{age at diagnosis} + e$$

In the full model, the dependent variable latent-time was associated with the fixed effects of gait parameter, sex and age at diagnosis.

In the null model, the dependent variable latent-time was associated with the fixed effects of sex and age at diagnosis.

3.5. Level of analysis and research scenarios

The basic structure of ANOVA was used in different levels and research scenarios described below. In the full model equation inside ANOVA described above, the gait parameter variable consisted of a selection of gait parameters consistent with the applicable level and scenario.

Level of analysis

Allgait: In the full model, all gait parameters were used as independent variables on all patients.

1 gait: In the full model, only one gait parameter is used as an independent variable. There were as many ANOVAs as the number of gait parameters available on all patients.

Task: the gait parameters belonging to the same task were used together on all patients.

Research scenario

In this setup, this analysis aimed to emphasize the asymmetrical component of PD symptoms.

Side: the gait parameters belonging to the same labeled side of the parameter were used together

Task-side: the gait parameters belonging to the same task and to the same labeled side of the parameter were used together

The collection of gait parameters by side and task-side side were tested in three scenarios Ipsicontralateral, Ipsilateral, and Contralateral.

Ipsicontralateral: the collection of gait parameters by side and task-side were tested on all patients

3.6. MULTIPLE COMPARISON CORRECTION

Ipsilateral: the collection of gait parameters by side and task-side were tested on the same side of the predominant side patients

Contralateral: the collection of gait parameters by side and task-side were tested on the opposite side of the predominant side patients.

A typical explanatory example of analysis is shown here for one outcome:

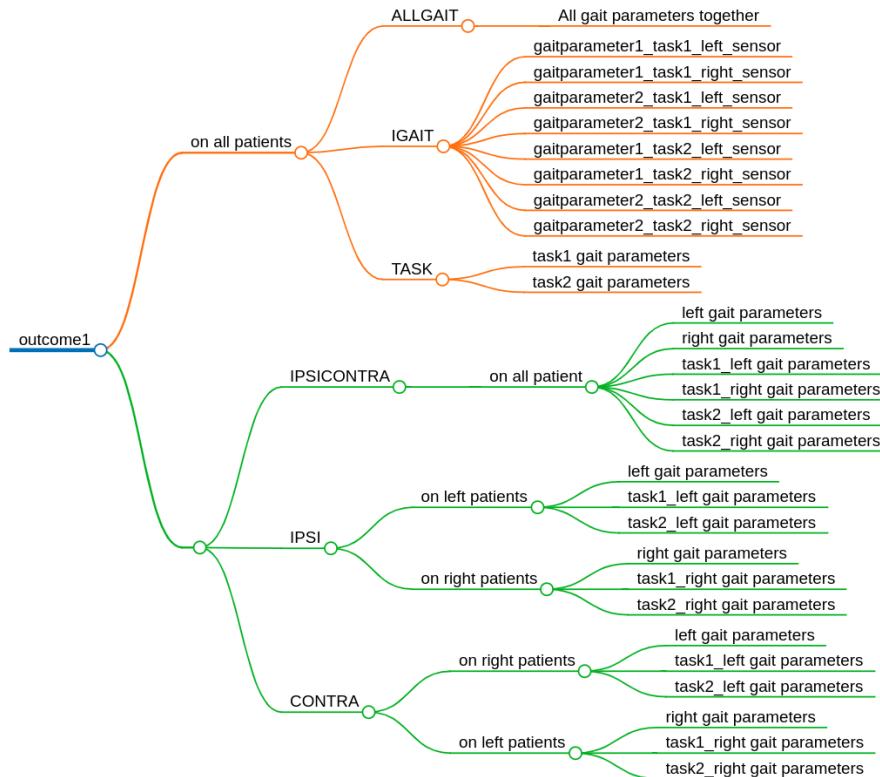


Figure 3.1.: Level of analysis and research scenarios

3.6. Multiple comparison correction

The Benjamin-Yekutieli (BY) correction was employed to adjust the p-values of the multiple ANOVA tests while considering various outcomes. It was used within the same research scenario (individual gait parameter, gait parameter by task, gait parameter by task-side) for each outcome. Here is an example of the outcome UPDRS3:

Individual gait parameter: BY correction of the 31 p-values of ANOVA

Gait parameter by task: BY correction of the 3 p-values of ANOVA of the gait parameters by tasks (2x10mPrefWithStop, 4x10mPrefWithoutStop, TUG)

Within an ipsicontralateral analysis, an ipsilateral analysis, and a contralateral analysis:

Gait parameter by task-left : BY correction of the 3 p-values of ANOVA of the gait parameters by tasks and by left side (2x10mPrefWithStop_left, 4x10mPrefWithoutStop_left, TUG_left)

Gait parameter by task-right : BY correction of the 3 p-values of ANOVA of the gait parameters by tasks and by right side
(2x10mPrefWithStop_right, 4x10mPrefWithoutStop_right, TUG_right)

3.7. Visualization and similarities between Erlangen and NCERPD analysis

Visualization

All adjusted p-values of all ANOVAs were collected in a table. Only adjusted p-values under the level of significance of 0.05 were further collected and processed into a tree graph for visualization. The tree was readable in this manner:

Parent node—child—child—child = Name of dataset — outcome — level of analysis and research scenario — gait parameters used in the full model of ANOVA of which the adjusted p-value was significant

Similarities between Erlangen and NCERPD analysis

The same gait parameters within the same level of analysis or research scenario that showed the association to the same score in each of the tested cohorts were reported. Although the processings of the variables Difference of Score from Baseline and Relative score at next visit were different, they pointed to the same concept Difference of score to baseline. Therefore, for this comparison process, these two variables were considered equivalent.

4. Results

4.1. Dataset description

A variety of people from both sexes and varied age groups were included in the Erlangen study. Participants had a mean age of study of 63 at baseline. They were 284 participants who attended at least 2 visits as shown in Figure 4.1. 172 were male and 112 were female. They were used for the LTJMM.

Eventually, the dataset included data from 80 participants, providing both UPDRS scores and gait parameters with at least 2 visits for the ANOVA's. Each patient had multiple visits with corresponding assessments of clinical scores and gait parameters. In this dataset, 43 patients were classified as Left predominant PD and 37 as Right predominant.

In the NCERPD study, participants had a mean age of study of 66 at baseline. They were 595 participants who attended at least 2 visits as shown in Figure 4.2. 401 were male and 194 female. They were used for the LTJMM.

Eventually, the dataset included data from 79 participants, providing both MDS-UPDRS scores and gait parameters for the ANOVAs. Each patient had only the visit corresponding to the assessment of clinical scores and gait parameters. In this dataset, 26 patients were classified as Left predominant PD and 28 as Right predominant.

4.1. DATASET DESCRIPTION

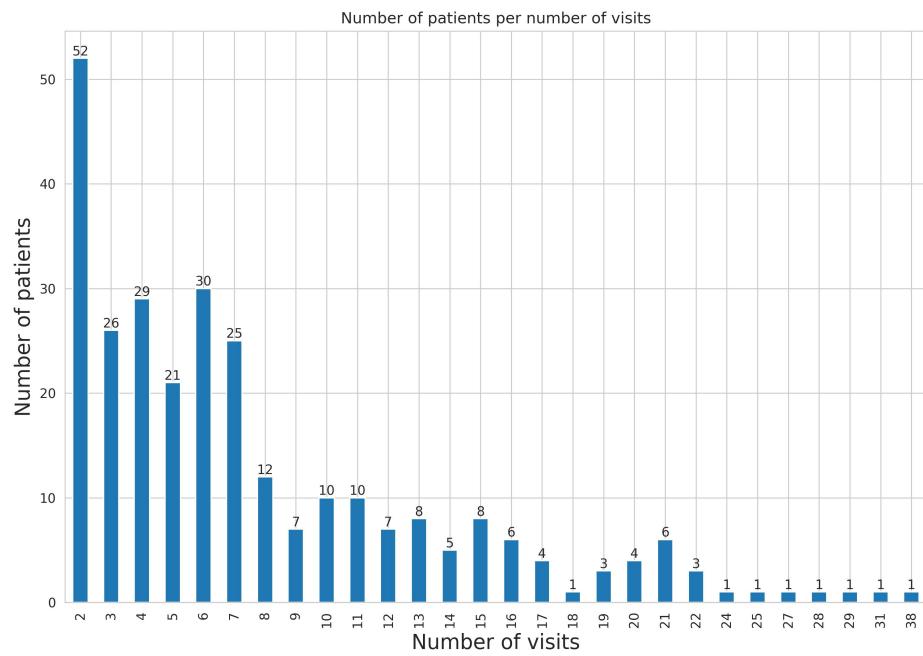


Figure 4.1.: Number of patients per visits in Erlangen

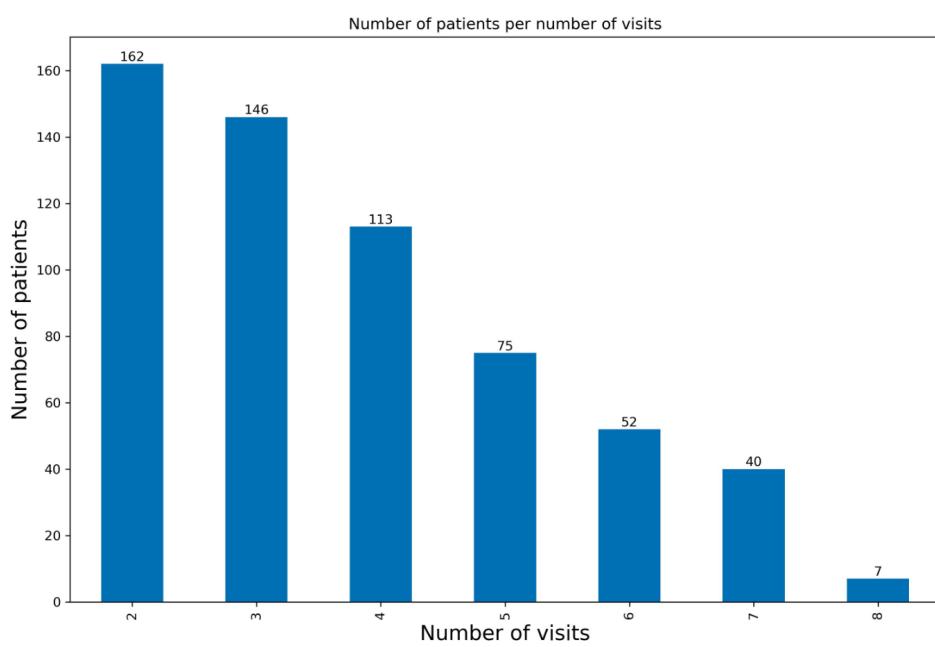


Figure 4.2.: Number of patients per visits in NCERPD

4.2. LTJMM

In NCERPD, the MDS-UPDRS score was designated simply as the UPDRS score for a better understanding of the results.

4.2.1. Convergence diagnostics

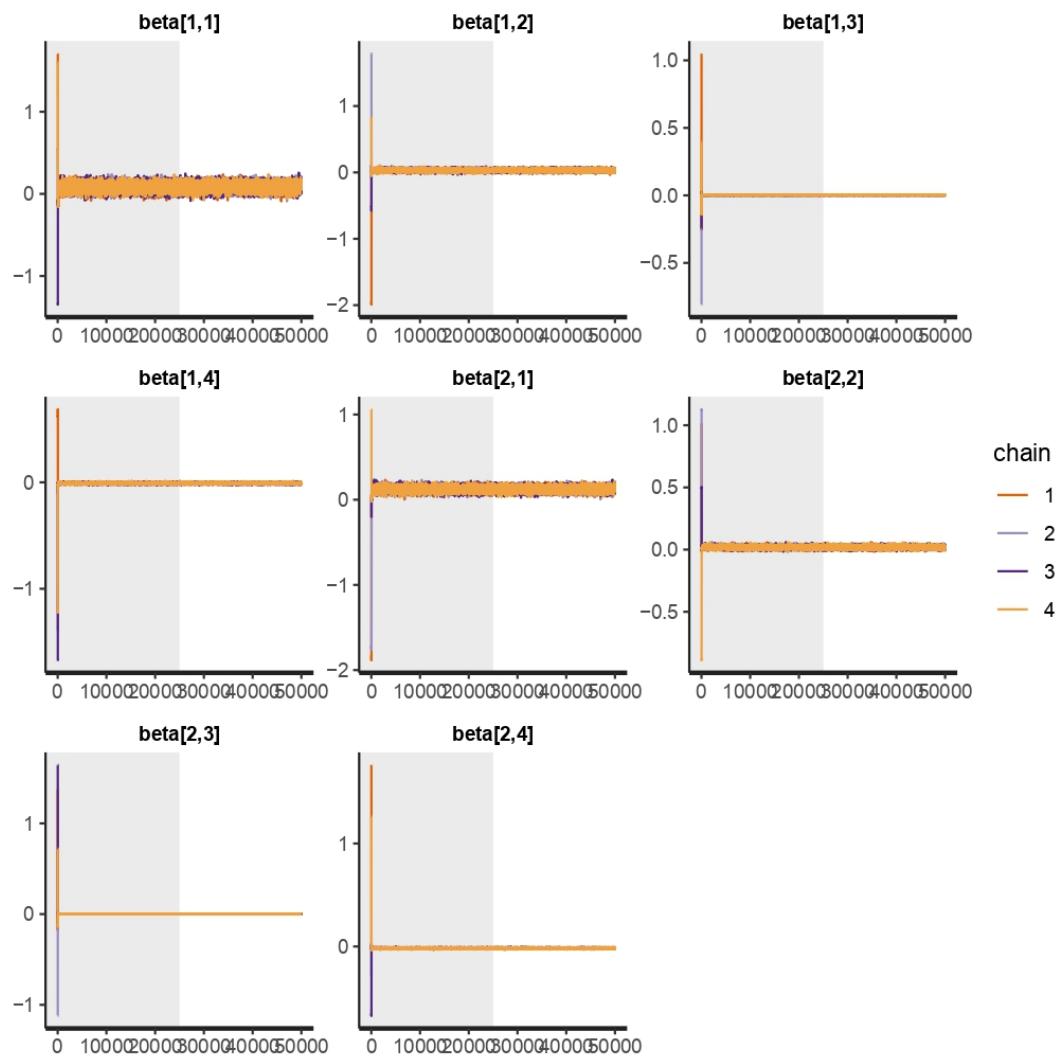


Figure 4.3.: Trace plot of beta in Erlangen

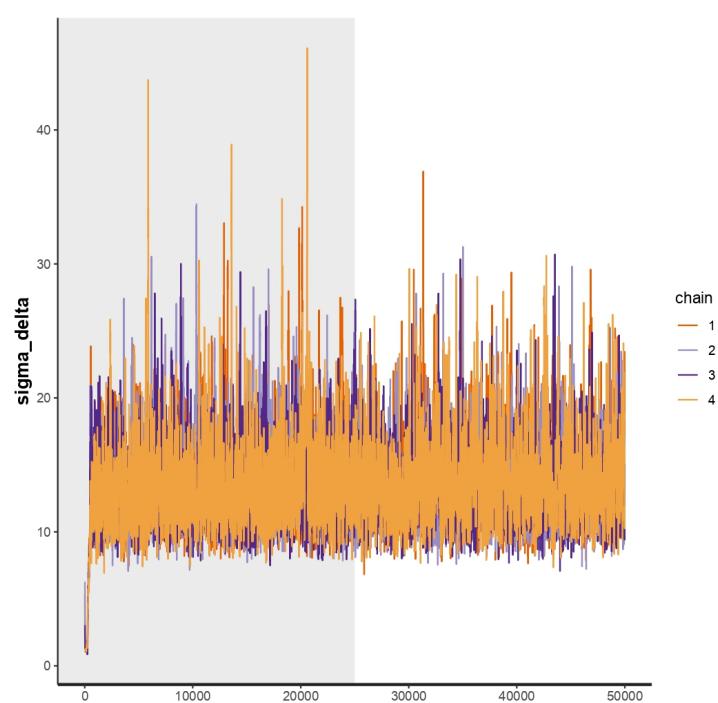
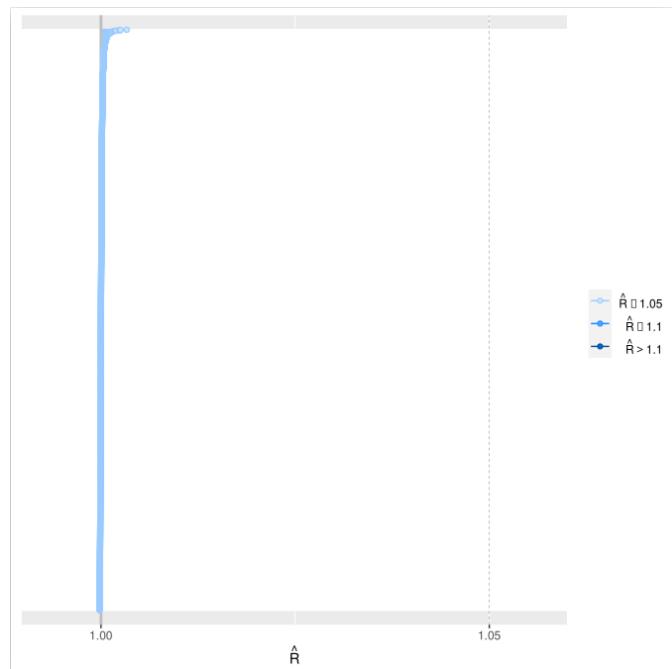
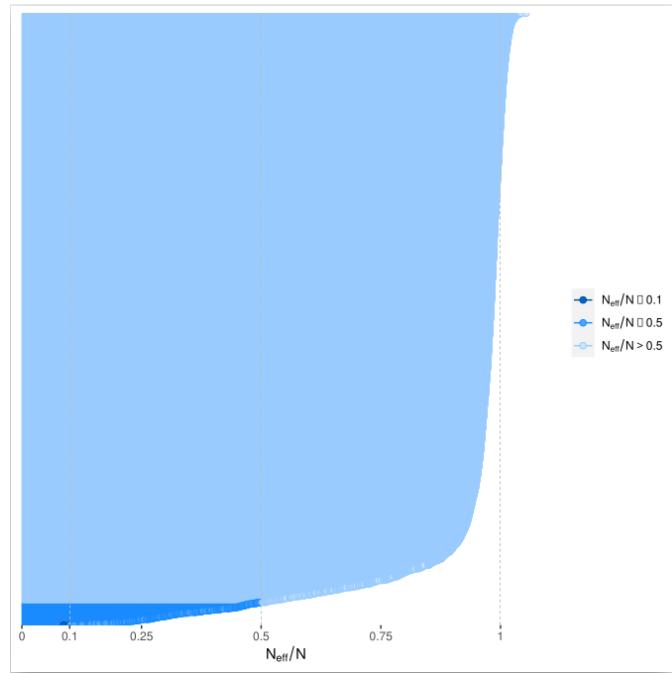


Figure 4.4.: Trace plot of sigma-delta in Erlangen



(a) $R\hat{a}$: potential scale reduction factor for all parameters



(b) N_{eff}/N : effective sample size/total sample size

Figure 4.5.: Erlangen LTJMM convergence diagnostics

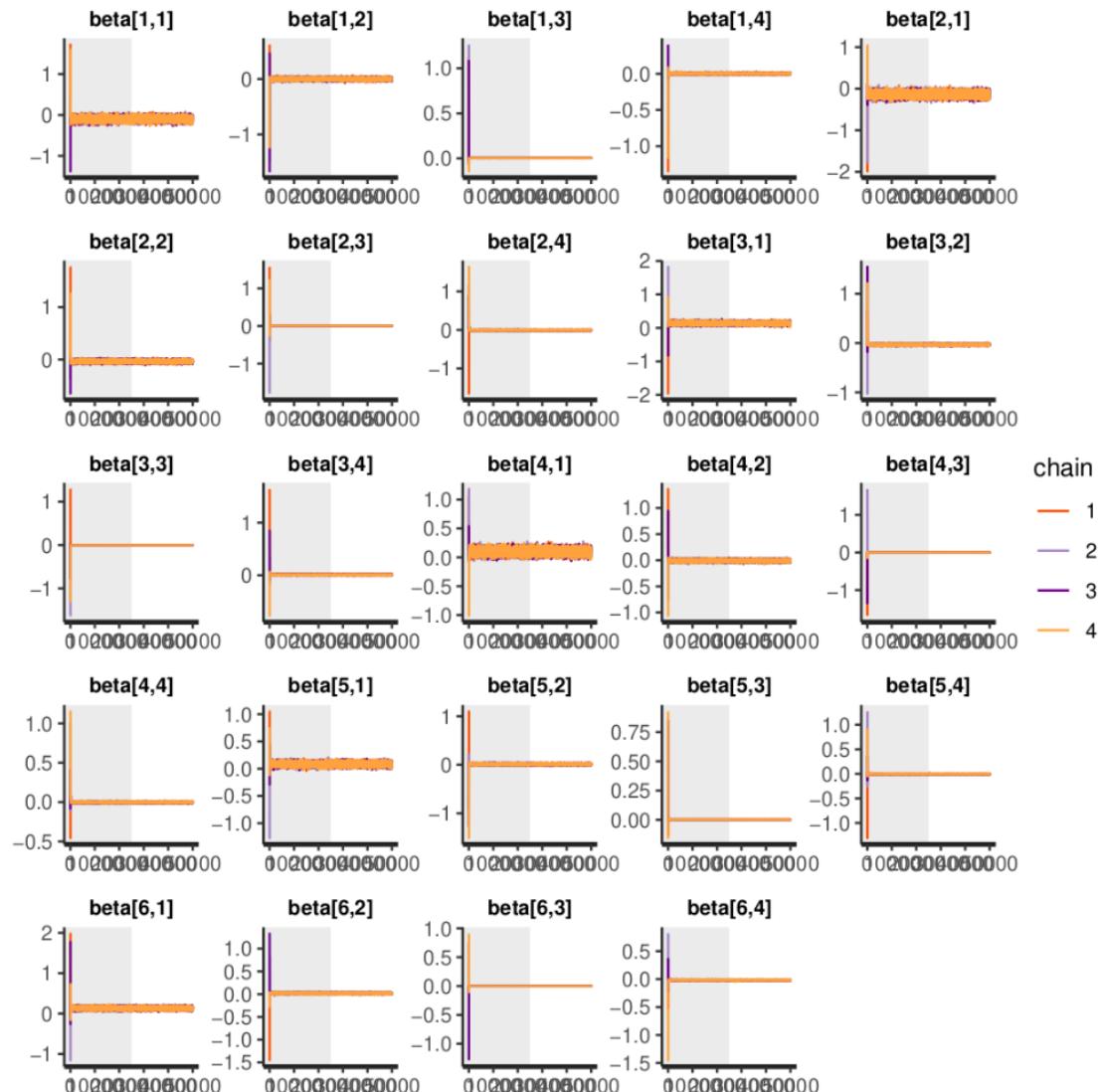


Figure 4.6.: Trace plot of beta in NCERPD

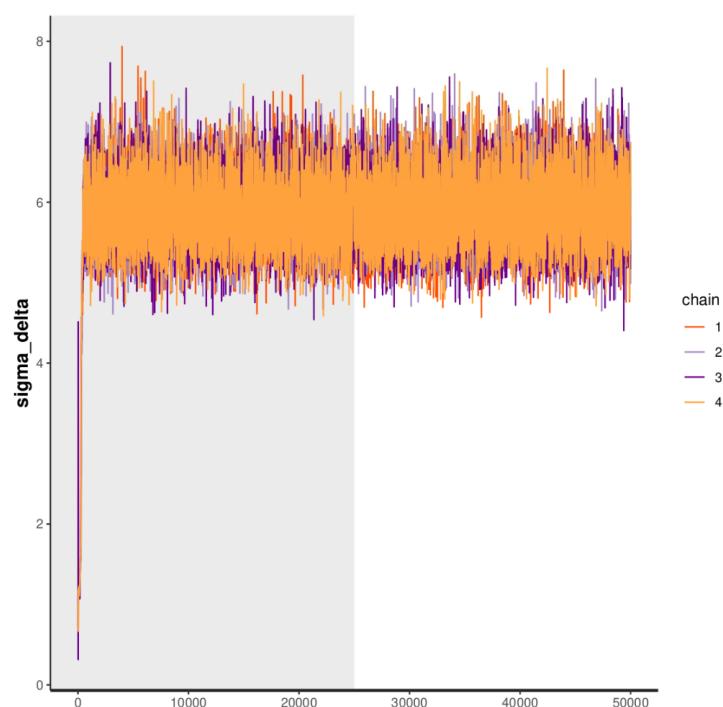
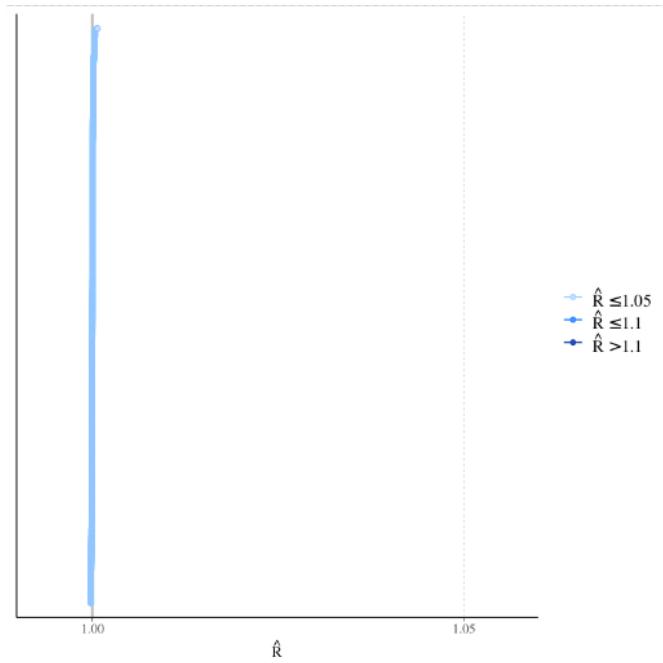
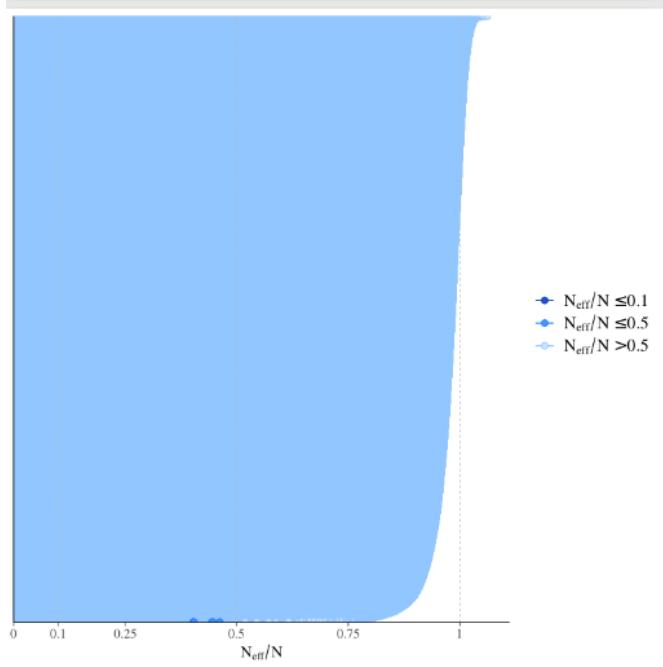


Figure 4.7.: Trace plot of sigma-delta in NCERPD



(a) Rhat: potential scale reduction factor for all parameters



(b) Neff/N: effective sample size/total sample size

Figure 4.8.: NCERPD LTJMM convergence diagnostics

Trace plots

In Figures 4.3, 4.4, 4.6, 4.7, the trace plots of the beta coefficient parameter and the standard deviation of delta (σ_{δ}) of the LTJMM estimation are shown. $\text{Beta}[i,j]$ is the coefficient for the i -th outcome clinical score and the j -th covariate in the LTJMM model.

The trace plot depicted the evolution of a parameter over 4 Markov chains. With the iteration number on the x-axis and the parameter value on the y-axis, it was a plot of the parameter values sampled at each iteration of the MCMC algorithm. The chains consisted of a series of 50000 connected samples. Each sample in the chain depended on the sample before it. The chains as a whole explored the posterior distribution of the parameters. The trace plots could be analyzed on three components:

Stationarity: The absence of systematic patterns of the chains suggested that the chains had attained a stable distribution.

Mixing: The samples from the 4 chains were well-mixed, indicating that they sufficiently explored the target distribution without getting trapped in any specific areas.

Randomness: The samples in the chain were randomly shown on the plot without following patterns.

Estimated potential scale reduction factor and effective sample size/total sample size ratio (Neff/N)

In Figures 4.5 and 4.8, in the R-hat plot, each point on the y-axis represents a particular parameter for which the Estimated potential scale reduction factor value is computed. The x-axis represented the magnitude of that value.

In Figures 4.5 and 4.8, in the Neff/N plot, each point on the y-axis represents a particular parameter for which the effective sample size/total sample size ratio is computed. The x-axis represented the magnitude of that value.

The description of the Estimated potential scale reduction factor and effective sample size/total sample size was drawn from (Gabry & Modrák, 2021).

The estimated potential scale reduction factor compares the within-chain variation to the between-chain variance. The chains may not have converged or mixed well if the R-hat value is larger than 1.1.

The number of independent draws from the posterior distribution is a measure of the effective sample size (Neff). It is based on the drawings' capacity to determine the parameter's true mean value. In the case of auto-correlation, the draws in the Markov chain are not independent resulting in a smaller Neff. The Neff/N ratio is a heuristic rather than a statistical value. It should not be below 0.1.

As shown in Figure 4.5, in Erlangen LTJMM diagnostics, the estimated potential scale reduction factors \hat{R} were below 1.1 for all parameters and the effective sample size/total sample size (Neff/N) were above 0.1 for all parameters except 1.

As shown in Figure 4.8, in NCERPD LTJMM diagnostics, the estimated potential scale reduction factors \hat{R} were below 1.1 and the effective sample size/total sample size (Neff/N) were above 0.1 for all parameters.

The LTJMM diagnostics results of Erlangen and NCERPD demonstrated successful convergence of the models using Bayesian inference.

4.2.2. LTJMM outputs

Erlangen

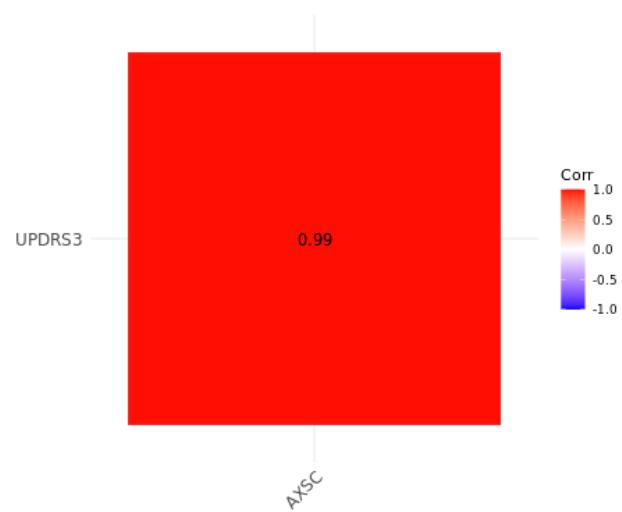


Figure 4.9.: Correlation of random slopes in Erlangen

As shown in Figure 4.9, the correlation of the posterior means of random slopes of UPDRS3 and the axial score was 99%.

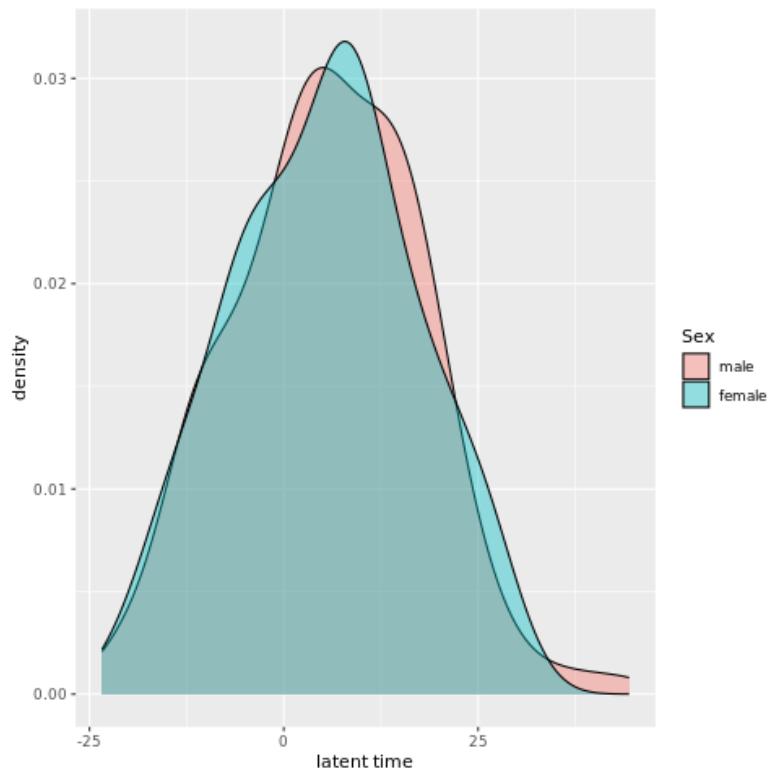


Figure 4.10.: Distribution of latent time by sex group in Erlangen

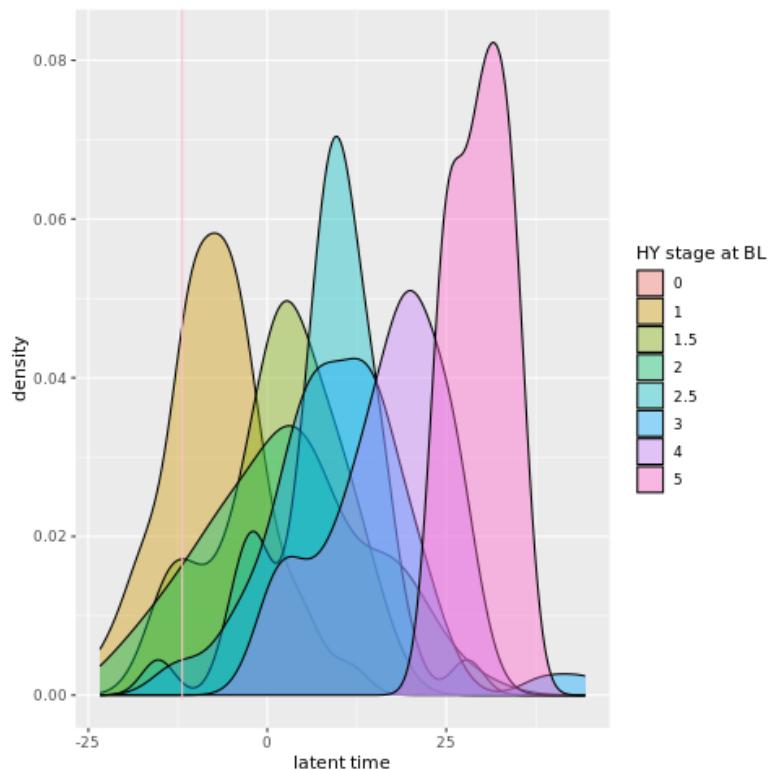


Figure 4.11.: Distribution of latent time by HY group in Erlangen

Table 4.1.: Number of patients at baseline by HY status in Erlangen

HY scale	0	1	1.5	2	2.5	3	4	5
Number of patients at baseline	1	46	28	68	39	62	36	3

Figure 4.10 and Figure 4.11 showed density plots for the latent time by sex group and by HY group at baseline at first visit respectively. Latent time was calculated as follows: Latent time = Time since diagnosis + δ_i estimates from the LTJMM.

As shown in Figure 4.10, the density plot of the male was not different from the female one. Their peaks were approximately at the same position. The latent time did not help in discriminating between male and female PD patients.

On Table 4.1, the number of patients at baseline by HY status was recalled here to get a better understanding of the density plot on Figure 4.11. It explained the vertical pink line on the density plot for HY =0 at latent time = -11. There was only one patient in HY=0 status. As shown in Figure 4.11, the latent time estimates were time-dependently grouping people in a way that was consistent with the progression of the severity of the disease rated with the HY scale. The latent time was approximately in the range of -11 years to +37 years between the peak of HY=0 and the peak of HY=5. For a patient in the HY=5 stages, he was more likely that his time elapsed since diagnosis was approximately 37 years compared to the general population of PD patients.

However, a high degree of overlap of curves was observed for similar HY categories (3 and 2.5, 1.5 and 1). Three distinct clusters of patients could be visually discriminated with latent time: the first cluster containing patients with HY status 0 and 1, the second cluster containing HY status 3 and 2.5 and the third patients with HY status 5.

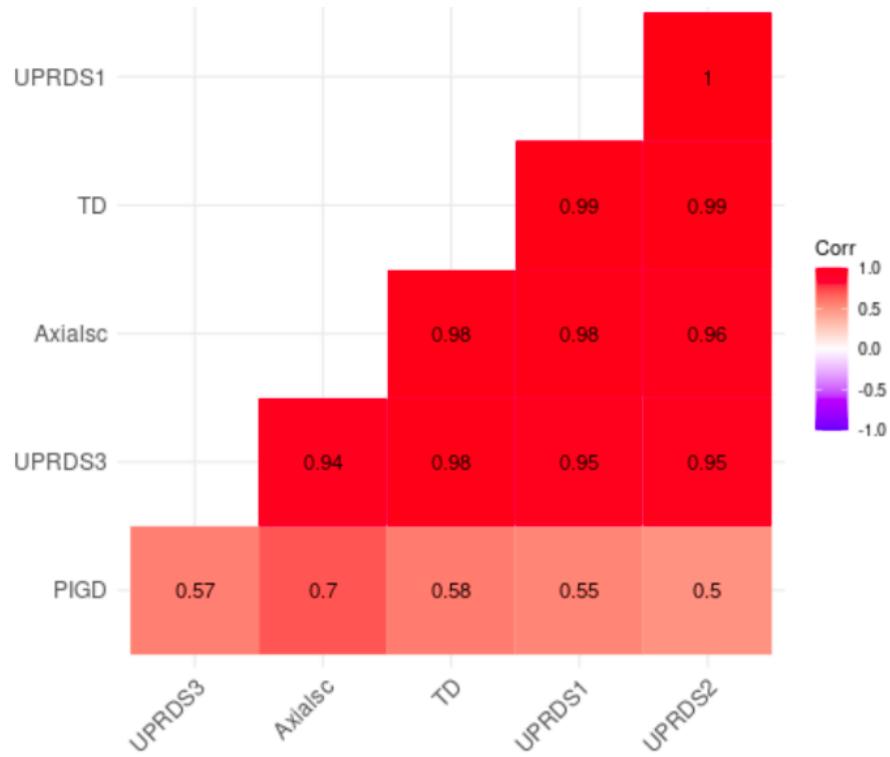
NCERPD

Figure 4.12.: Correlation of random slopes in NCERPD

As shown in Figure 4.12, the correlations of the posterior means of random slopes of UPDRS1, UPDRS2, UPDRS3, TD, Axial score, and PIGD were moderate to strong. The highest correlation was for the pair UPDRS1-UPDRS2. The correlations between Tremor dominant score and UPDRS 1,2,3 were respectively 0.98, 0.99, 0.99. The PIGD score showed a lower correlation with any of the other scores. The correlations of PIGD with UPDRS 1,2,3 were respectively 0.55, 0.5, 0.57.

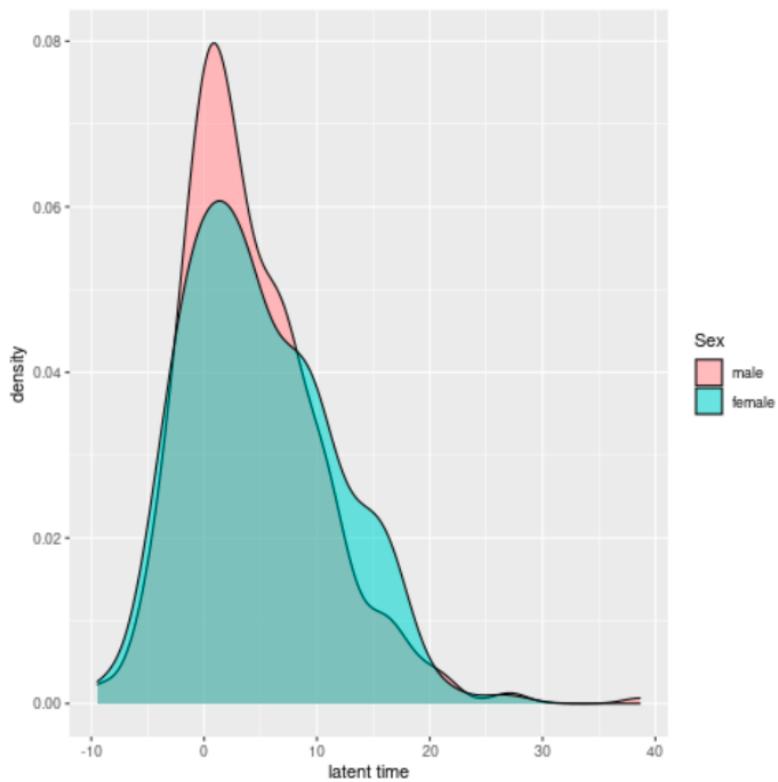


Figure 4.13.: Distribution of latent time by sex group in NCERPD

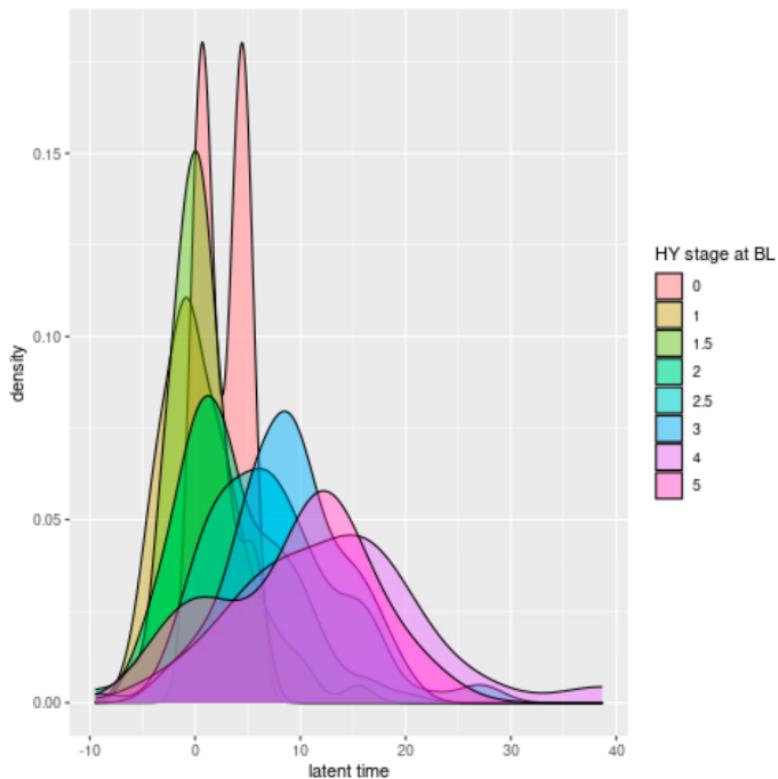


Figure 4.14.: Distribution of latent time by HY group in NCERPD

4.3. ANOVA

Table 4.2.: Number of patients at baseline by HY status in NCERPD

HY scale	0	1	1.5	2	2.5	3	4	5
Number of patient at baseline	2	61	45	314	80	57	24	12

As shown in Figure 4.13, the density plot of the male was not different from the female one. Their peaks were approximately at the same position on the x-axis. On the y-axis, the male peak was higher accounting for a male latent time data more tightly clustered around 1. The latent time did not help in discriminating between male and female PD patients.

On Table 4.2, the number of patients at baseline by HY status was retrieved here to get a better understanding of the density plot on Figure 4.14.

As shown in Figure 4.14, the latent time estimates were time-dependently grouping people in a way that was consistent with the progression of the severity of the disease rated with the HY scale. Nevertheless, 2 peaks of the density plot at HY=0 corresponding to one individual each were slightly before the peak of the density at HY=1. The latent time was approximately in the range of 0 years to 17 years between the peak of the HY=0 group and the peak of the HY=5 group. For a patient in the HY=5 stage, he was more likely that his time elapsed since diagnosis was approximately 17 years compared to the general population of PD patients. Moreover, among all the curves, a high degree of overlap was observed. The latent time did not help in discriminating between the different HY stages.

4.3. ANOVA

4.3.1. Correlation of gait parameters

Table 4.3.: Spearman rank correlation of Erlangen gait parameters of TUG left side

gait parameter pair 1	gait parameter pair 2	spearman rank
stride.time..s._TUG_left_sensor	stance.time..s._TUG_left_sensor	0.944115
stride.length..m._TUG_left_sensor	gait.velocity..m.s._TUG_left_sensor	0.860460
stride.time..s._TUG_left_sensor	swing.time..s._TUG_left_sensor	0.698396
stance.time..s._TUG_left_sensor	gait.velocity..m.s._TUG_left_sensor	-0.645866
stride.time..s._TUG_left_sensor	gait.velocity..m.s._TUG_left_sensor	-0.578326
swing.time..s._TUG_left_sensor	stance.time..s._TUG_left_sensor	0.462314
stance.time..s._TUG_left_sensor	stride.length..m._TUG_left_sensor	-0.244086
swing.time..s._TUG_left_sensor	gait.velocity..m.s._TUG_left_sensor	-0.221109
stride.time..s._TUG_left_sensor	stride.length..m._TUG_left_sensor	-0.143462

4.3. ANOVA

Table 4.4.: Spearman rank correlation of NCERPD gait parameters of TUG left side

gait parameter pair 1	gait parameter pair 2	spearman rank corr.
Stance.Time..s._tug_left	Stride.Time..s._tug_left	0.9761
Gait.Speed..m.s._tug_left	Stride.Length..cm._tug_left	0.8507
Swing.Time..s._tug_left	Stride.Time..s._tug_left	0.7294
Stance.Time..s._tug_left	Gait.Speed..m.s._tug_left	-0.6341
Stance.Time..s._tug_left	Swing.Time..s._tug_left	0.5939
Gait.Speed..m.s._tug_left	Stride.Time..s._tug_left	-0.569

Within the same task for the same gait concept such as gait speed, gait parameter values on the left side were assumed to be highly correlated with gait parameter values on the right side. In the Tables 4.3, 4.4, Spearman rank correlations were performed for gait parameter values recorded on the left sensor in the context of the task present in both Erlangen and NCERPD study, which was TUG. The correlations were made for all visits for the gait parameters measured at the same visit. Both tables showed consistent correlations. The pairs Stance Time – Stride Time and Gait speed – Stride Length were highly correlated.

4.3.2. Significant adjusted p-values

All ANOVA adjusted p-values were gathered in a colored table in Figures A.1 and A.2. They are shown in the Supplementary material.

From these tables only adjusted p-values under the level of significance of 0.05 were further collected and processed into a tree graph for visualization in Figures 4.15, 4.16, 4.17, 4.18, 4.19, 4.20.

Erlangen

As shown in Figures 4.15, 4.16, 4.17, 4.18, the variables UPDRS3, rdm_UPDRS3, axial score, and latent time were the outcomes for which a high number of adjusted p-values were significant. Among the clinical scores, the axial score was the outcome with the highest number of significant adjusted p-values in all levels of analysis and research scenarios. The outcomes (difference to baseline, random slope) derived from the clinical scores showed a lower number of significant adjusted p-values.

Significant gait parameters in individual gait analysis did not always lead to a significant multivariate model when these parameters were combined in the full model. As an example, for the outcome rdm_UPDRS3, 15 individual gait parameters were significant and contained gait parameters related to the tasks 2x10mPrefWithStop and 4x10mPrefWithoutStop. However, for the level of analysis Task, no multivariate models for the tasks 2x10mPrefWithStop and 4x10mPrefWithoutStop were significant.

NCERPD

As shown in Figures 4.19, 4.20, the variables axial score and PIGD score were the outcomes for which a high number of adjusted p-values were significant. The clinical

4.3. ANOVA

scores UPDRS 1,2,3 and the outcomes (difference to baseline, random slope) derived from the clinical scores showed a lower number of significant adjusted p-values.

Significant gait parameters in individual gait analysis did not always lead to a significant multivariate model when these parameters were combined in the full model.

In contradiction with Erlangen analysis, the outcome latent time was rarely associated with gait parameters, only once with Turn-right related gait parameters in ipsicontralateral scenario.

Aggregation of results from Erlangen and NCERPD

As shown in Table 4.5, when the level Task, the research scenarios (ipsicontralateral, ipsilateral and contralateral) and all the clinical outcomes (except latent time) were considered, the task (or task-side) for which there was the highest amount of significant adjusted p-values was TUG for Erlangen and Count for NCERPD.

Table 4.5.: Count of tasks in the total of the significant adjusted p-values

Erlangen task	count	NCERPD task	count
2x10mPreWithStop	10	Turn	10
4x10mPreWithoutStop	8	Tray	10
TUG	21	TUG	7
		Count	14

Legend: Inside colored Tables A.1 and A.2 and in Task, Ipsicontralateral, Ipsilateral, Contralateral, for each type of tasks, the number of times the task (or task-side) of interest was counted in the total of the significant adjusted p-values considering all outcomes except latent time.

Table 4.6.: Ratio of significant adjusted p-values (%)

	Erlangen	rank	NCERPD	rank
allgait	71.42		0	
1gait	21.19		5.43	
Task	38.09		12.5	
Ipsicontralateral	33.92	2	21.11	1
Ipsilateral	39.28	1	2.78	2
Contralateral	14.28	3	1.66	3

To aggregate the numerous results, a ratio of all significant first ANOVAs (dependent variables were the clinical scores and derived scores) adjusted p-values over the total number of ANOVAs grouped by level of analysis and research scenarios was computed. These ratios were shown in Table 4.6. The absolute values were not of interest but rather their rank. For both datasets, ipsicontralateral scenario ratio values were approximately the highest; ipsilateral ratios were higher than contralateral ones; the contralateral ratios were the lowest.

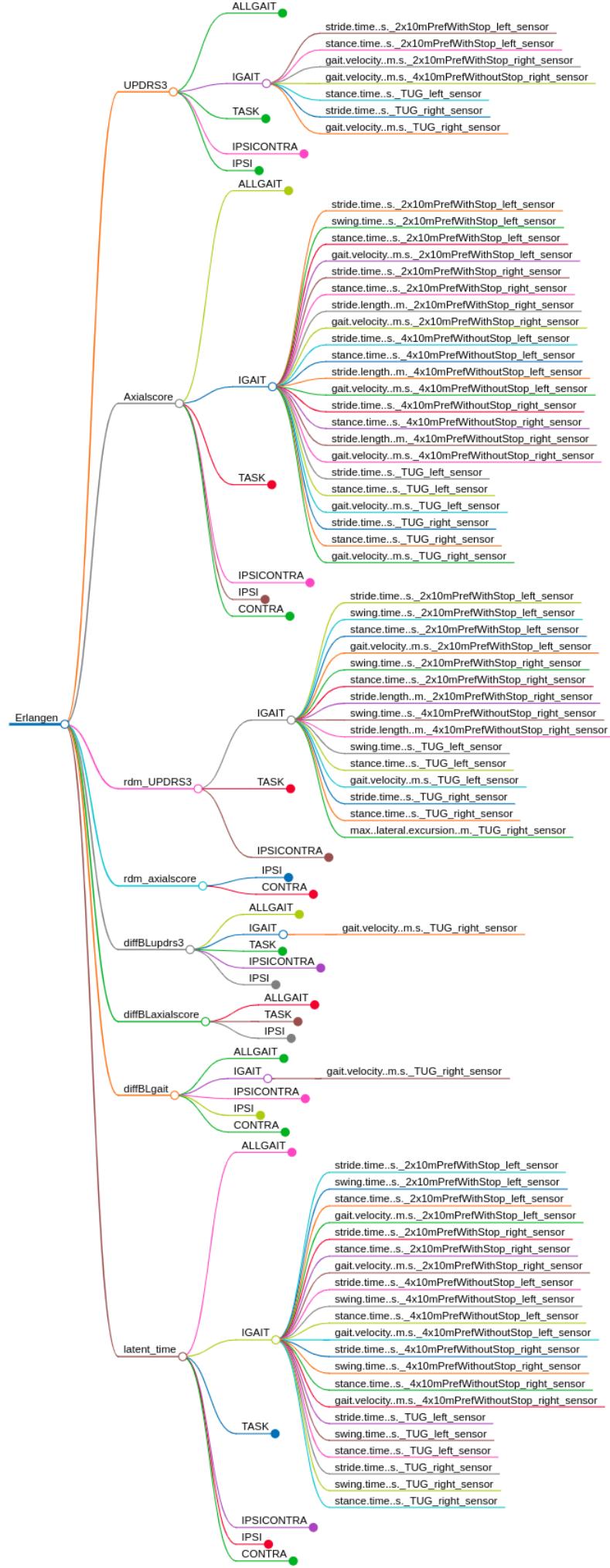


Figure 4.15.: Erlangen significant adjusted p-values of ANOVAs, view by 1gait

4.3. ANOVA

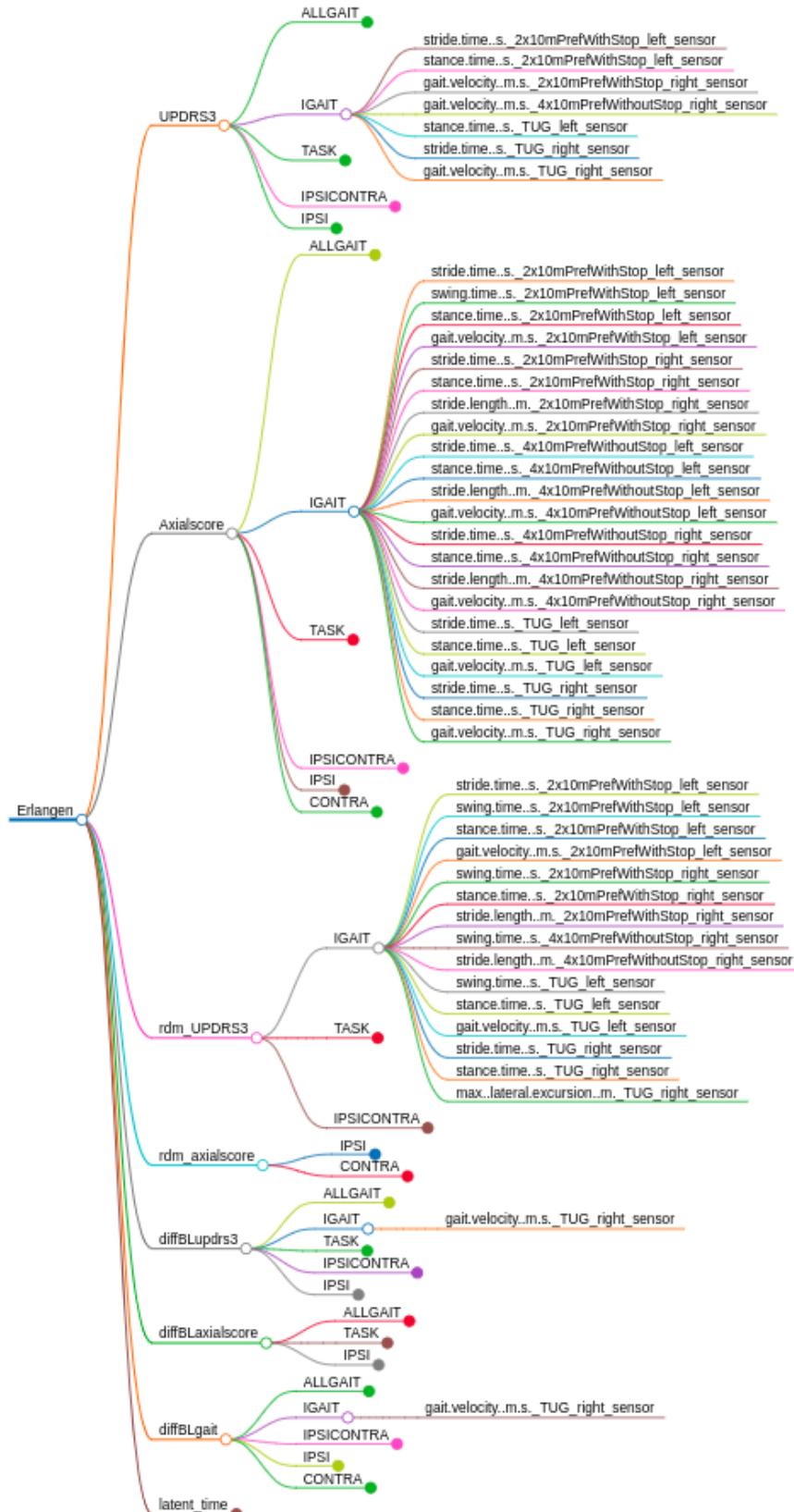


Figure 4.16.: Erlangen significant adjusted p-values of ANOVAs, view by 1gait (without latent time)

4.3. ANOVA

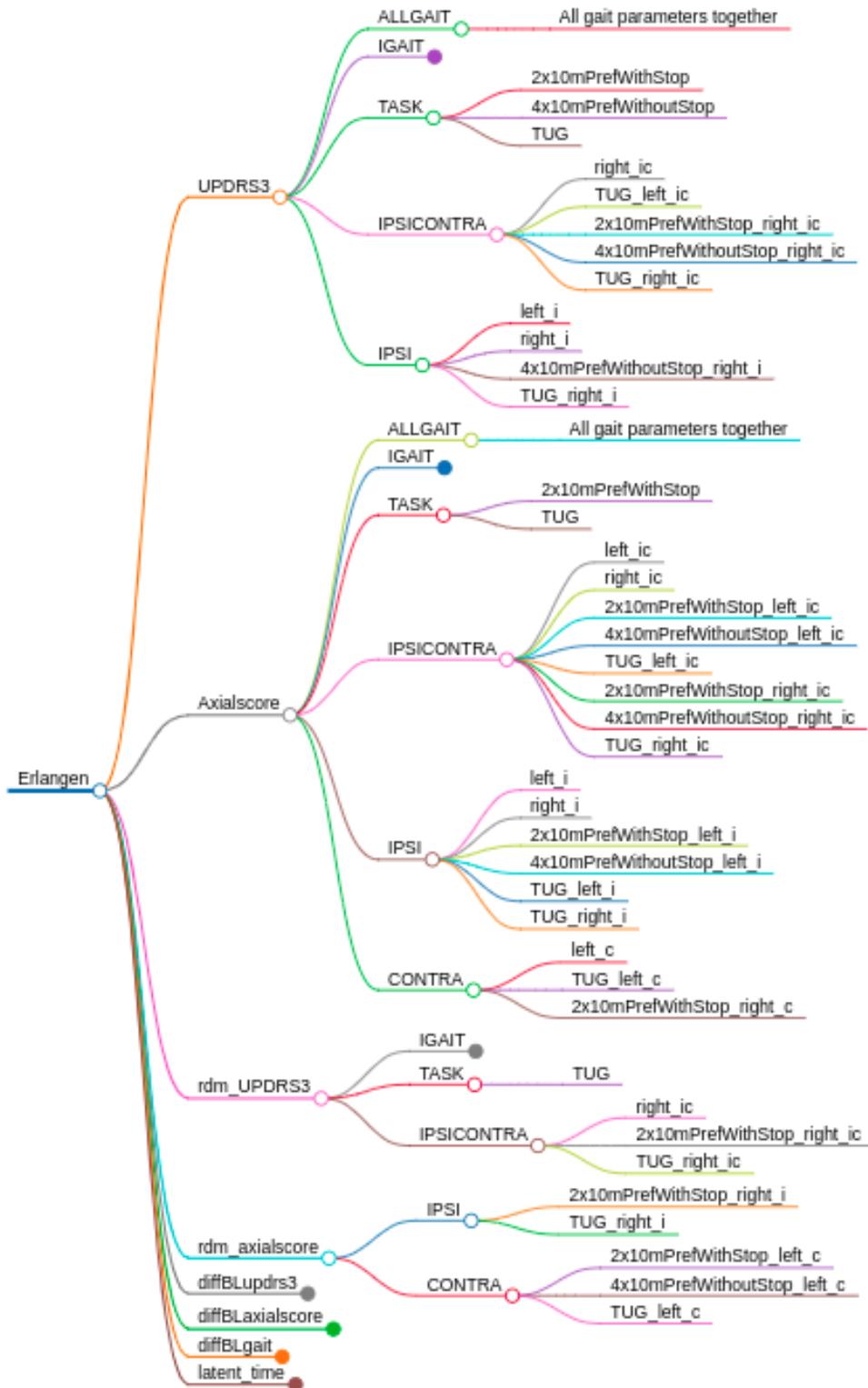


Figure 4.17.: Erlangen significant adjusted p-values of ANOVAs,
view by UPDRS3, axial score, rdm-UPDRS3, rdm-axialscore, (without
1gait)

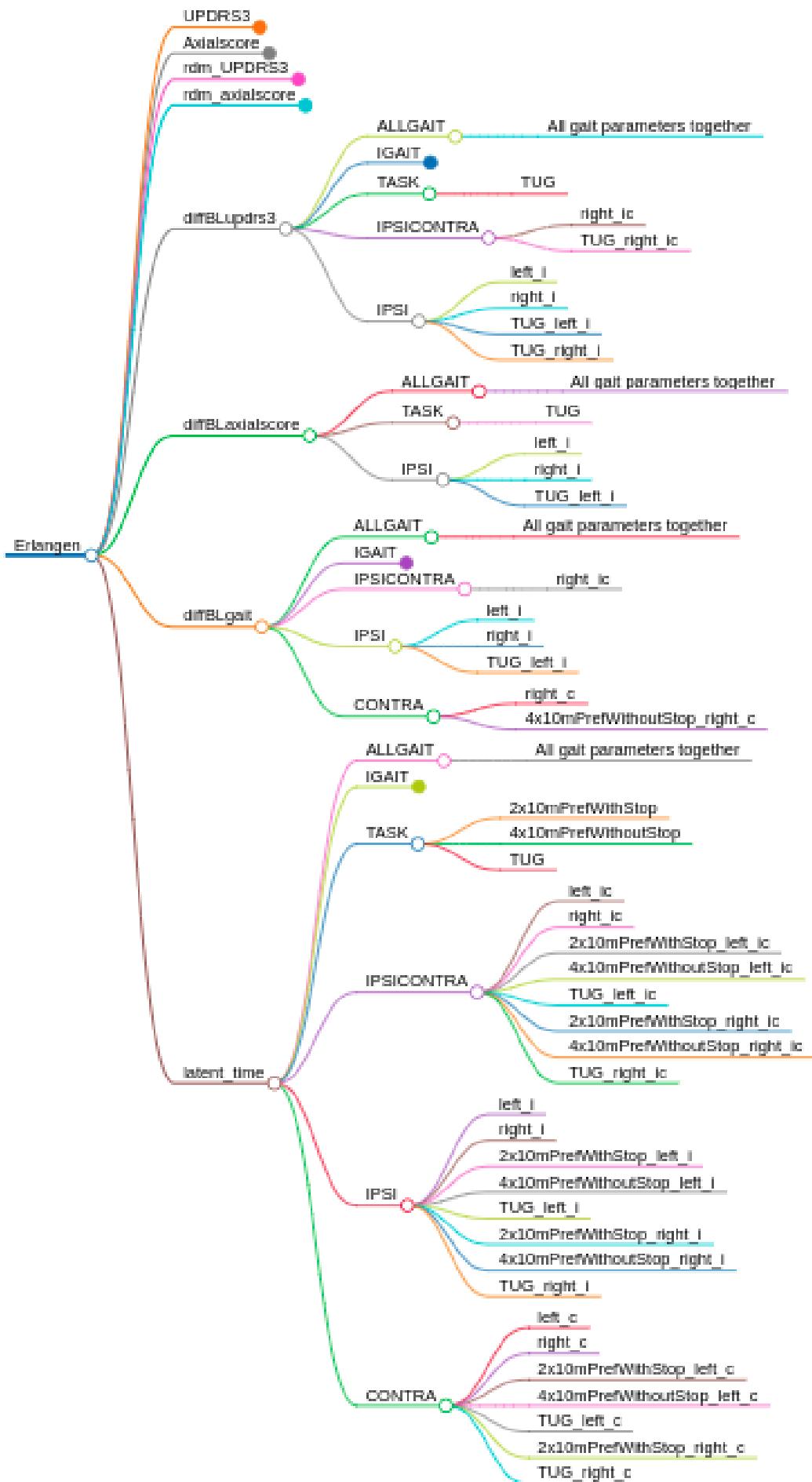


Figure 4.18.: Erlangen significant adjusted p-values of ANOVAs, view by diffBL-UPDRS3, diffBL-axialscore, diffBL-gait, latent time, (without 1gait)

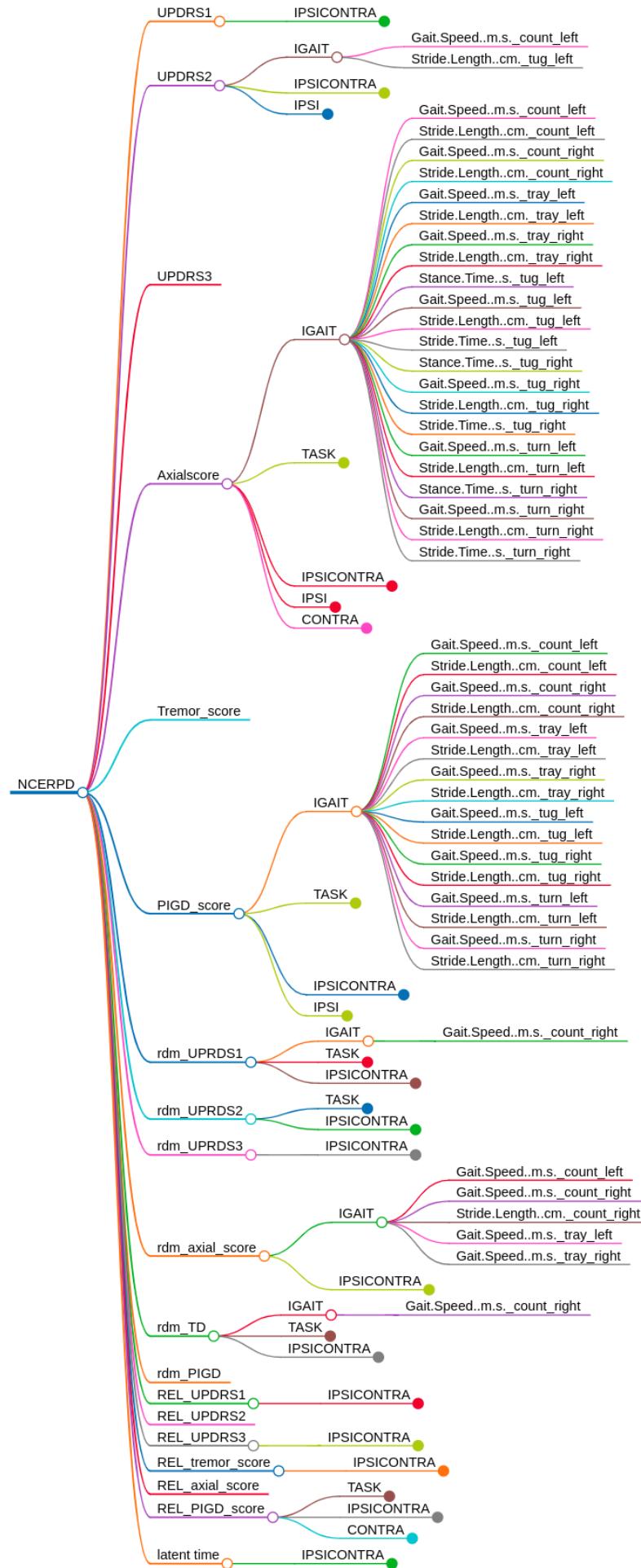


Figure 4.19.: NCERPD significant adjusted p-values of ANOVA's, view by individual gait

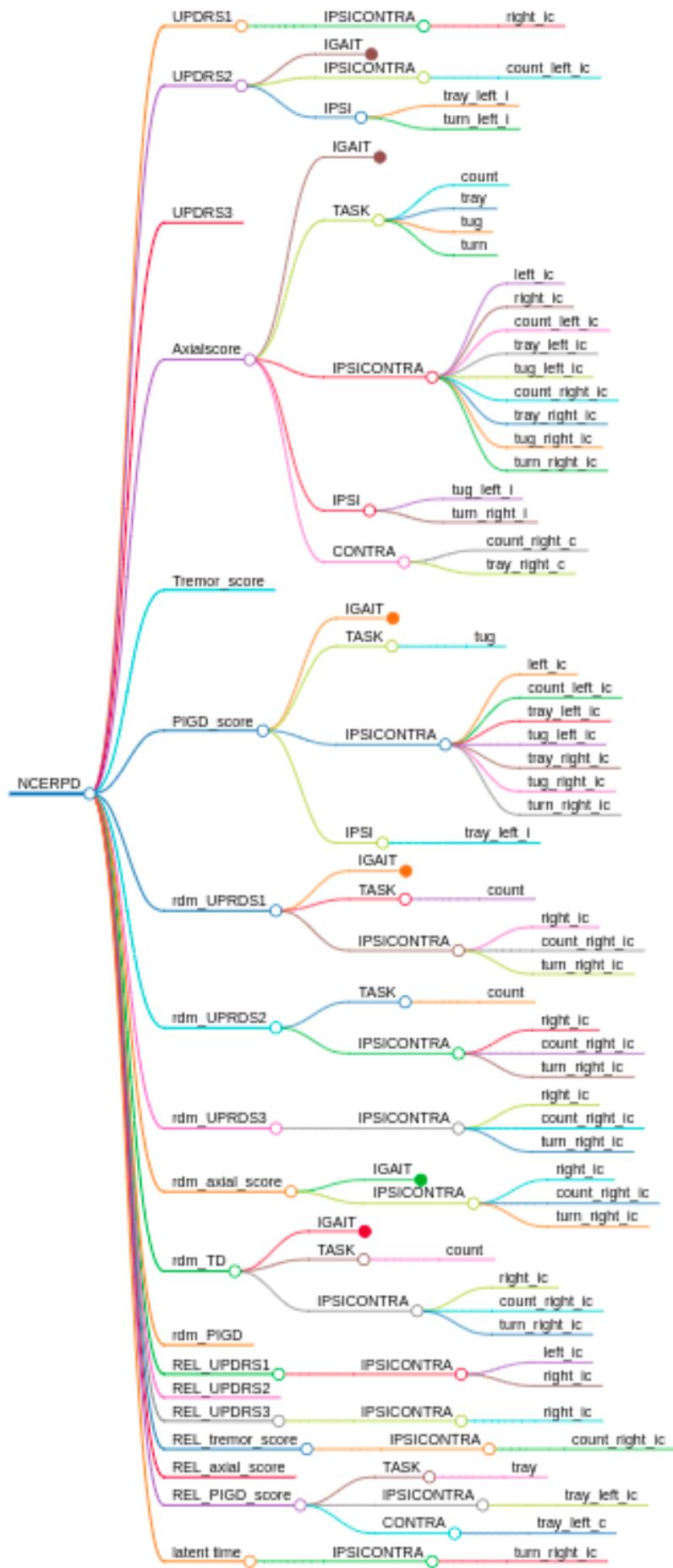


Figure 4.20.: NCERPD significant adjusted p-values of ANOVA's,
view by all gait, task, ipsicontralateral, ipsilateral, contralateral

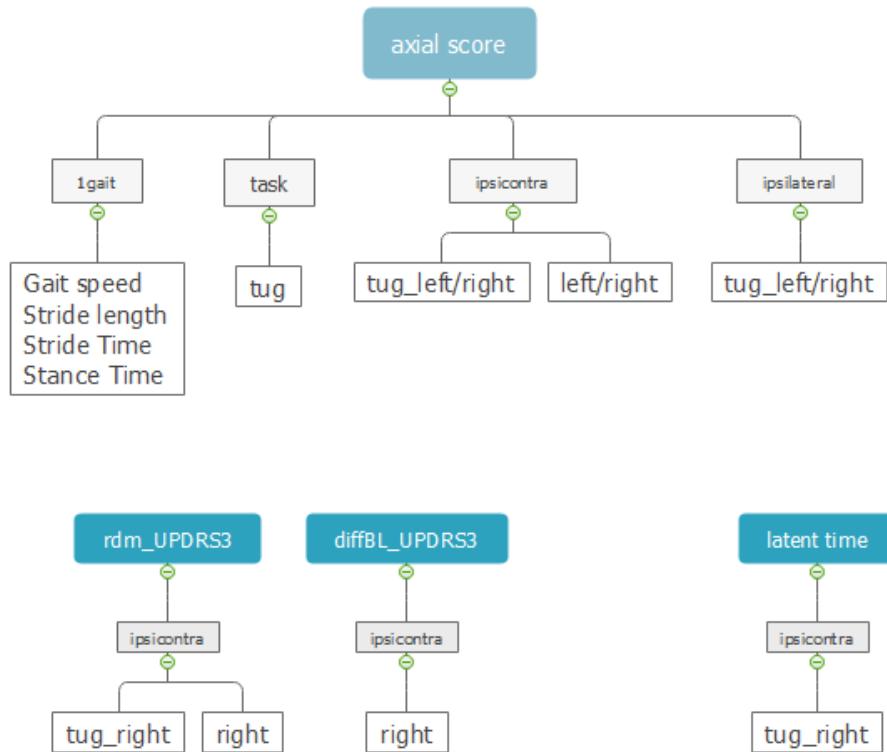


Figure 4.21.: Similarities between Erlangen and NCERPD analysis

4.3.3. Similarities between Erlangen and NCERPD analysis

From the collection of the tree graphs, comparison and search for similarities were performed. The result was finally shown in a tree graph in Figure 4.21.

The outcomes axial score, rdm_UPDRS3, diffBLupdrs3 and latent time were present in both the Erlangen dataset and NCERPD. They were responsible for outputting the same associated gait parameters under the same level and research scenario.

In multiple levels and research scenarios, the axial score was the clinical score for which gait parameters were the most associated.

In the level of analysis individual gait (1gait), the gait parameters were summarized under the same gait concept. The side and task where the gait parameters belonged were stripped out for a better understanding. Gait speed, Stride length, Stride time and Stance time were associated with the clinical score axial score.

This similarity between both datasets indicated that the gait parameters association with clinical outcomes was consistent and reliable across different groups of subjects.

4.4. Evaluation of LME and LM

4.4.1. Diagnostics plot

From the significant ANOVA's reported in Figure 4.21, LM and LME of the full model in the ANOVAs were computed only for the outcomes axial score, the level of analysis 1gait, ipsicontra. In the level 1gait, only the gait parameters related to TUG were computed.

To evaluate the goodness of fit of the different linear models LM and LME, the assumptions of linearity, homoscedasticity, normality and independence of errors have to be checked. Moreover, the presence of outliers which can violate the assumption of homoscedasticity has to be also checked. 3 types of plots are shown to test the assumptions: Linearity, Influential observations and QQ-plot.

On figures 4.22, 4.23, 4.24, 4.25, each column represented a type of plot, while each row showed the gait parameters involved in the linear models. The gait parameter or task-side of interest was shown on the left of the row.

In Erlangen and NCERPD plots :

The Linearity plot displayed the residuals (the difference between the observed values and the predicted values) on the y-axis and the fitted values (the predicted values) on the x-axis. It helped to check for linearity, homoscedasticity and independence of errors. The green regression line was approximately linear and flat at $y=0$ which confirmed a linear relationship between the predictors and the outcome. The residual points were equally randomly spread around the regression line located at $y=0$ which confirmed homoscedasticity. No pattern in the points suggested that the points followed a particular direction which confirmed the independence of residuals. However, parallel straight lines were noticeable in all linearity plots. They reflected the nature of the outcome variable axial score. This axial score variable was closer to a categorical variable than a true continuous variable since the score was rated on a scale of 1 to 4 with only 0.5 as an intermediate point.

The QQ-plot tested the assumption of normality of the residuals. It compared the quantiles of the sample residuals on the y-axis against the quantiles of a standard normal distribution on the x-axis. The points on the QQ plot approximately followed a straight line (except 3 points in Erlangen plots), thus the residuals had a normal distribution.

The Influential observations plot helped to check the homoscedasticity and the presence of outliers. The standardized residuals (residuals divided by their standard deviation) were plotted on the y-axis, and the leverage values (a measure of how far the values of an observation's independent variables differ from those of the other observations) were plotted on the x-axis. The 2 green dashed lines represented the Cook's distance cutoff threshold. The Cook's distance calculates how much predicted values would vary if a certain observation was not included in the dataset. The Cook's distance value of each point was compared to a threshold of 4 divided by the number of points. If the points fell outside the green cutoff dashed lines, they were considered Influential observations. They could significantly affect the estimation of regression coefficients.

4.4. EVALUATION OF LME AND LM

In Erlangen plots:

In Figure 4.22, 3 outliers were noticeable on Influential observations and QQ-plots in the Erlangen dataset. The assumption of normality of residuals might be violated.

In NCERPD plots :

The assumptions of linearity, homoscedasticity, normality and independence of errors were checked and confirmed.

4.4. EVALUATION OF LME AND LM

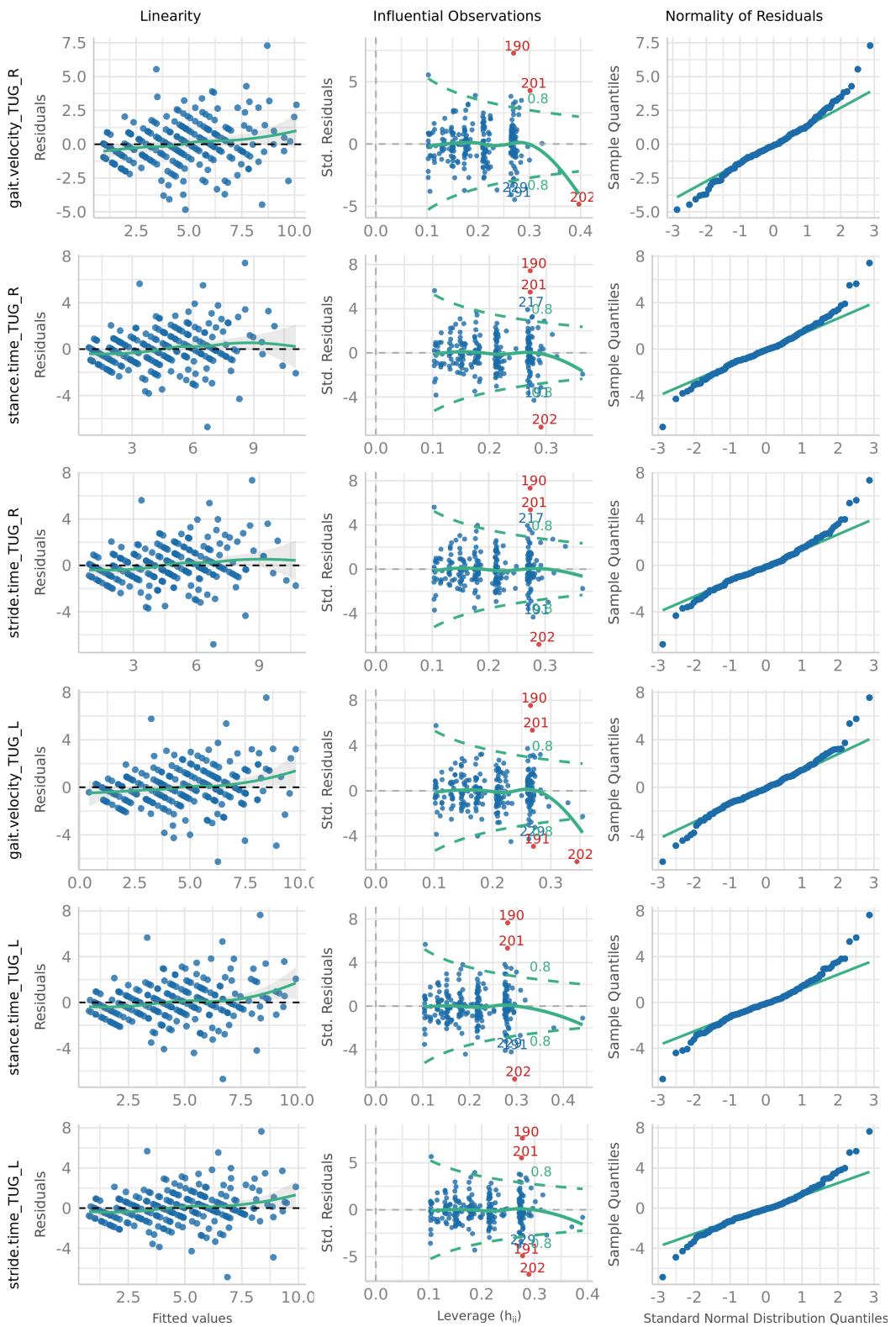


Figure 4.22.: Diagnostics of LME in Erlangen in level 1 gait for the outcome axial score and the predictors related to TUG

4.4. EVALUATION OF LME AND LM

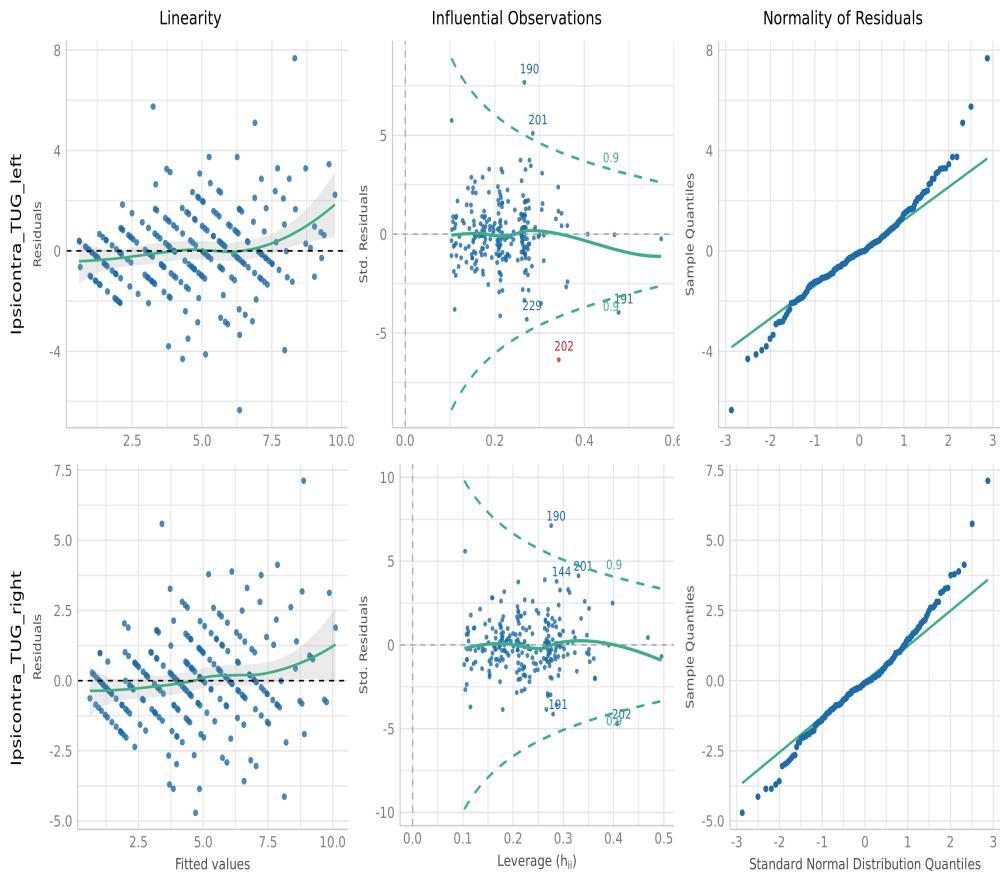


Figure 4.23.: Diagnostics of LME in Erlangen in scenario ipsicontralateral for the outcome axial score and the task-side predictors TUG-left and TUG-right

4.4. EVALUATION OF LME AND LM

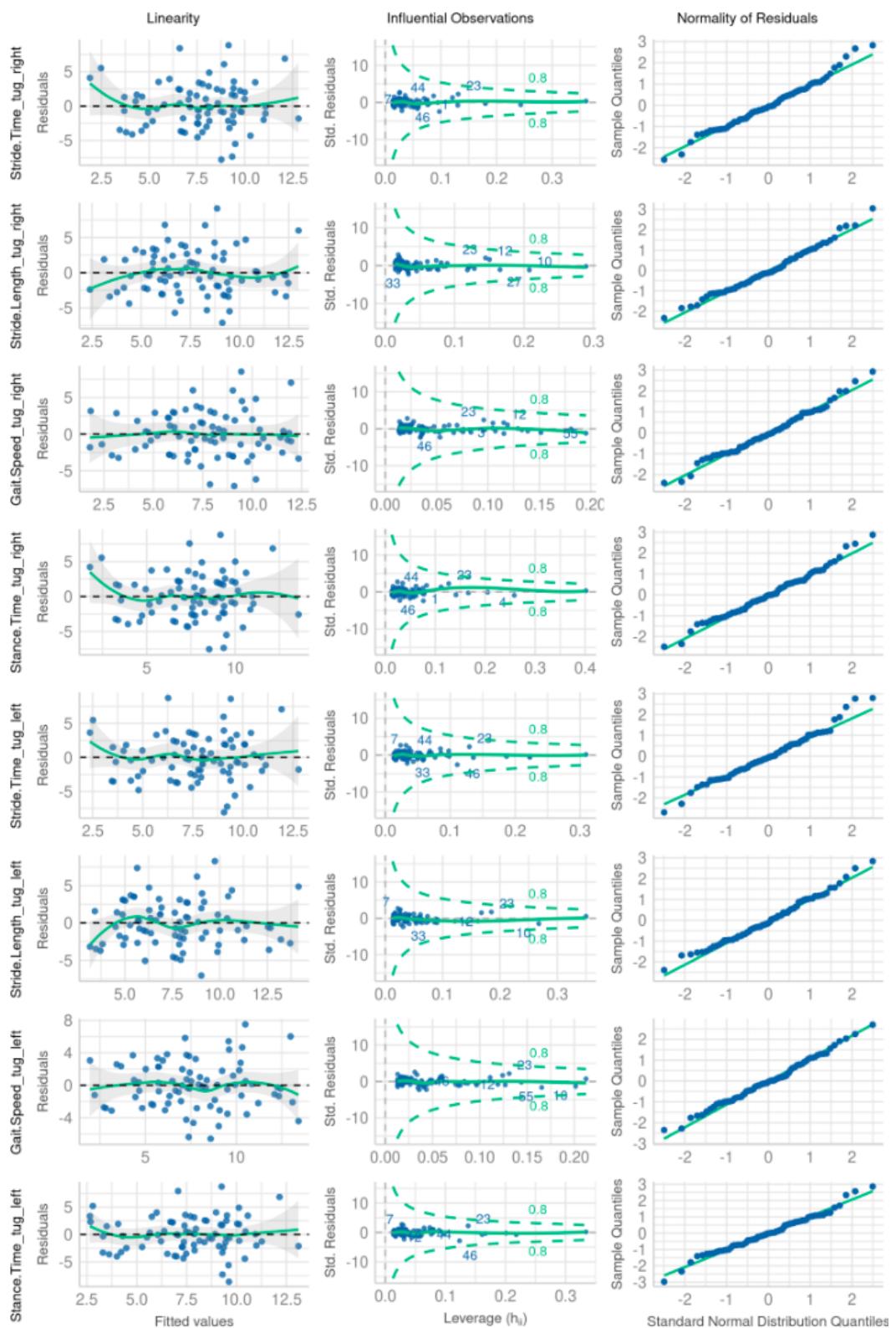


Figure 4.24.: Diagnostics of LM in NCERPD in level 1 gait for the outcome axial score and the predictors related to TUG

4.4. EVALUATION OF LME AND LM

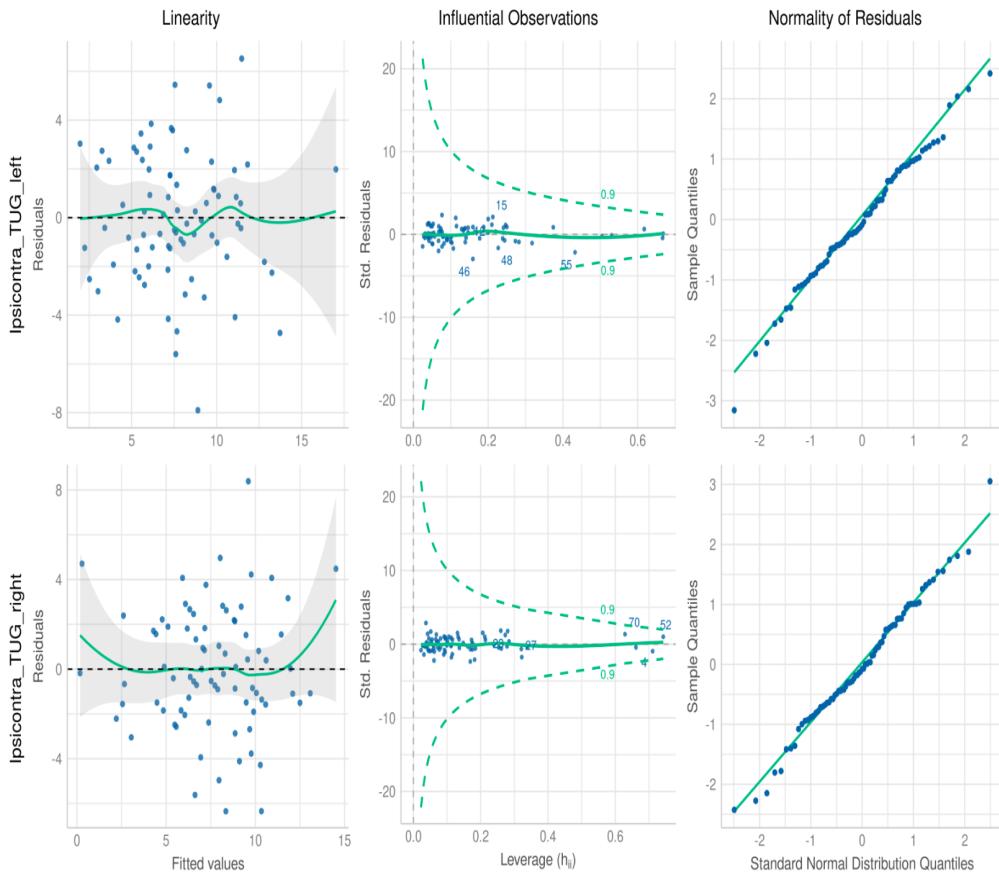


Figure 4.25.: Diagnostics of LM in NCERPD in scenario ipsicontralateral for the outcome axial score and the task-side predictors TUG-left and TUG-right

4.4.2. R-squared (R²)

From the full model in the significant ANOVA's reported in Figure 4.21, LM, LME, adjusted R², conditional R² and marginal R² were computed only for the outcomes axial score, the level of analysis 1gait, ipsicontralateral, ipsilateral and contralateral. In the level 1gait, only the gait parameters related to TUG were computed. In ipsicontralateral, ipsilateral and contralateral, only the collection of parameters related to the task TUG jointly with the side (left or right) was computed that was named TUG-left or TUG-right respectively. The results were shown in the tables 4.7 and 4.8.

R-squared is a metric that shows the extent to which the variance in the dependent variable can be explained by the independent variables in a regression model.

Adjusted R-squared is a variant of R-squared that takes into account the number of predictors in a model.

Conditional R-squared and marginal R-squared are used in mixed-effect models.

Conditional R² quantifies the proportion of the total variance that can be explained by both fixed and random effects.

Marginal R² quantifies the proportion of the total variance that can be explained by fixed effects only.

One similarity between Erlangen and NCERPD could be of interest was that the

4.4. EVALUATION OF LME AND LM

ipsilateral-TUG-left R2 score was substantially higher than all the other scores. In Erlangen-ipsilateral-conditional R2, TUG-left was the highest (0.685). In NCERPD-ipsilateral-adjusted R2, TUG-left was the highest (0.8184).

However, there were no obvious other similarities between the R2 in Erlangen and NCERPD in terms of values, relationship and rank order. The methods, the cohorts, and the metrics involved were too different to make a comparison.

In Erlangen, the conditional R-squared values were approximately higher than the marginal R-squared values, indicating that the random effects contributed to the model's ability to explain the variance in the data. They were approximately all similar around 0.6. In ipsilateral and contralateral sections, the marginal R2 values were different indicating different levels of goodness of fit.

In NCERPD, in the Individual gait section, the adjusted R2 was in the range of 0.29 to 0.44. In the ipsicontralateral and contralateral section, the adjusted R2 values for TUG-left and TUG-right were relatively close, indicating a similar level of model fit for both sides. In the ipsilateral section, the TUG-left was substantially higher than the TUG-right suggesting a better model fit for the left side. The higher values of TUG-left and TUG-right compared to individual gait parameters suggested a better fit of the multivariate models.

Table 4.7.: Conditional R2 and Marginal R2 by level in Erlangen

Individual gait	Conditional R2	Marginal R2
gait.velocity..m.s._TUG_left_sensor	0.575818479059464	0.351220664717734
gait.velocity..m.s._TUG_right_sensor	0.59349770038853	0.371555731869714
stance.time..s._TUG_left_sensor	0.598231109956786	0.358010457589521
stance.time..s._TUG_right_sensor	0.59534214125239	0.36861220397882
stride.time..s._TUG_left_sensor	0.591140592480982	0.355542656717826
stride.time..s._TUG_right_sensor	0.598962122392577	0.371948369331774
Ipsicontralateral	Conditional R2	Marginal R2
TUG_left	0.593380437414836	0.384188255476169
TUG_right	0.614586129362842	0.410295287488943
Ipsilateral	Conditional R2	Marginal R2
TUG_left	0.685122948809238	0.367512742565602
TUG_right	0.60224078230083	0.528159706272611
Contralateral	Conditional R2	Marginal R2
TUG_left	0.547570382316789	0.526382917886841
TUG_right	0.628790770733694	0.376583052765077

Table 4.8.: Adjusted R2 by level in NCERPD

Individual gait	adjusted R2
Stride.Time..s._tug_right	0.2934
Stride.Length..cm._tug_right	0.345
Gait.Speed..m.s._tug_right	0.3833
Stance.Time..s._tug_right	0.3187
Stride.Time..s._tug_left	0.2998
Stride.Length..cm._tug_left	0.3926
Gait.Speed..m.s._tug_left	0.4431
Stance.Time..s._tug_left	0.3213
Ipsicontralateral	adjusted R2
TUG_left	0.4573
TUG_right	0.4332
Ipsi	adjusted R2
TUG_left	0.8184
TUG_right	0.5557
contra	adjusted R2
TUG_left	0.4758
TUG_right	0.4718

5. Discussion

5.1. Principle Findings

In this thesis, an LTJMM was performed to examine to monitor the progression of PD. The HY stage values were not included in the LTJMM to process the resulting time shift. Despite this absence in the model, the latent time ordering of PD patients reflected the progression of the severity of patients rated with the HY scale. The patient at HY=0 who had no signs of Parkinson's disease was more likely to have the earliest latent time compared to the general population. The patient at HY=5 who was wheelchair-bound or bedridden, unless aided, was more likely to have the latest latent time compared to the general population. This result extended previous results in the PD context where the group analyzed at baseline were PD, healthy control, and prodromal (Iddi, Li, Aisen, et al., 2018). In this thesis, the present result deepened the analysis of the PD group by stratifying the PD population by HY stage. It was, therefore, more related in the Alzheimer's disease (AD) context to the ordering of patients by baseline diagnostic group (CN, cognitively normal; SMC, subjective memory concern; EMCI, early mild cognitive impairment; LMCI, late mild cognitive impairment; AD, probable Alzheimer's Disease with mild to moderate dementia) (Li, Iddi, Thompson, et al., 2019).

In both datasets Erlangen and NCERPD, the latent time has appropriately temporally ordered the PD patient on a PD progression severity scale. However, the overlap between different stages of the HY scale was more pronounced in NCERPD but still present in Erlangen. It could hinder the unequivocal interpretation of the HY stage based on the latent of the patient. For one latent time value, several HY stage assignation was possible.

Moreover, in the LTJMM, correlations between random slope estimates of clinical scores were performed. In both datasets, the correlation between random slopes axial score and UPDRS3 was high (above 0.94). This value was expected since the axial score was made up of sub-items of the UPDRS3 score. In LTJMM NCERPD Figure 4.12, the correlations of random slope PIGD with random slopes UPDRS 1,2,3 were respectively 0.55, 0.5, 0.57. These results were approximately similar to the result of the previous study (Iddi, Li, Aisen, et al., 2018) where the correlation between the random slope of PIGD and the random slope total UPDRS(UPDRS1+UPDRS2+UPDRS3) was 0.64. In LTJMM NCERPD Figure 4.12, the correlations between random slopes Tremor dominant score and UPDRS 1,2,3 were respectively 0.98, 0.99, 0.99. These values were higher than the value of the same previous study where the pair of random slopes PIGD-updrs_total was 0.58. This discrepancy was explained by the inclusion in the dataset of healthy control and prodromal patients along with PD patients in the previous study.

In this thesis, it was investigated whether gait parameters were related to specific

5.1. PRINCIPLE FINDINGS

clinical outcomes. In a previous study (Schlachetzki, Barth, Marxreiter, et al., 2017), a cross-sectional analysis revealed that stride length and gait velocity were related to the UPDRS3. Stride length, gait velocity, stride time and stance phase time were also related to the sub-item UPDRS-gait of UPDRS3. PD patients with high PIGD showed also high swing time variation. In the longitudinal analysis of the study, stride length, swing time and stance time were also related to UPDRS-gait.

Although the method used in this thesis was different from the previous study, the results pointed in the same direction. In Erlangen longitudinal analysis, gait velocity, stride time, and stance time were associated with UPDRS3 and rdm-UPDRS3. The diffBLgait was related to gait-velocity. In NCERPD, PIGD was not associated with swing time. This might be due to the increased swing time variability with increased PIGD. A linear model could have difficulties associating variables that have a non-linear relationship. In both datasets, although certain gait parameters might be significant when analyzed individually, combining them in a multivariate model might not always lead to a significant result. This happened for multiple reasons. Gait parameters might be highly correlated as shown in the correlation Tables 4.3 and 4.4. Stride time and stance time as well as stride length and gait velocity were highly correlated (0.94 and 0.86 respectively). It explained also why the two above pairs of gait parameters were together significantly associated with the axial score in the level 1gait. However, when combined in a multivariate model, they might lead to unstable coefficient estimates. Another reason was the negative correlation of a pair of gait parameters such as gait speed and stance time (approximately -0.64). Their respective effects could have canceled each other out.

If the level Task, the research scenarios (ipsicontralateral, ipsilateral and contralateral) and all the clinical outcomes and derived scores were considered, the tasks for which there was the highest amount of significant p-values were TUG and Count. Since Count was a task derived from TUG. The task TUG appeared to be an appropriate task to test for the association between gait parameters and clinical scores. In Table 4.6, the higher ratios of ipsilateral significant p-values compared to contralateral ratios suggested that the ipsilateral research scenario might be a suitable set-up for the association of gait parameters with clinical scores. This analysis reflected the asymmetrical component of the motor symptoms of PD. However, these assumptions on the suitable task TUG and the ipsilateral framework need further statistical analysis to be confirmed.

The latent time was often associated with gait parameters in all conditions and scenarios in Erlangen but rarely in NCERPD. Gait parameters helped to follow the latent time with proper coefficient estimates longitudinally. However, they were not directly associated with latent time in a cross-sectional analysis.

The interpretation of the results must be subject to caution and was relative to the goodness of fit of the models.

In LM, the assumptions of the linear regression were not violated. The majority of the adjusted-R² values were in the range of 0.29 to 0.55 which showed only a moderate explanatory power of the LM. There remained a great portion of unexplained variance. Taking into account that the latent time predictor was also in the model with the gait parameters predictors, the gait parameters might have low explanatory power. The relative contribution of latent time and gait parameters has to be investigated. However,

5.2. LIMITATIONS

the collection of gait parameters in TUG-left in ipsilateral showed an adjusted R² of 0.81 which shows a high explanatory power. This result confirmed the need for a careful selection of appropriate gait parameters in a suitable framework such as ipsilateral or ipsicontralateral.

In LME, the conditional R² and the marginal R² were not adjusted for the number of predictors. The values might have been inflated. The random effects added more complexity to the model. In case of collinearity between predictors and random effect, the R² values might be not reliable. The small sample size in Erlangen might also affect the estimation of the R² values. It was safer not to consider their absolute values but rather their relative order. Moreover, 3 outliers might have violated the assumptions of homoscedasticity and normality of errors leading to unstable coefficient estimates. It would be reasonable to recompute the LME without these outliers or check their clinical significance.

In LTJMM, the goodness of fit was assessed by the MCMC convergence diagnostics. The chains were well mixed. The potential reduction factor and the Neff/N ratio values were reasonable. However, there was no metric assessing the goodness of fit of predictors such as R². R² was based on the sum of squares and was not applicable in LTJMM because it was estimated in the Bayesian approach. The MCMC method for estimating the posterior distribution of estimates and the presence of random effects in the model added layers of complexity to summarize the goodness of fit in a simple metric. A posterior predictive check could be computed. However, the latent time was required to compute the posterior predictive check. The unknown true value of latent time could not be compared to the estimated one from the LTJMM.

5.2. Limitations

The latent time extracted from the LTJMM could moderately serve as a continuous alternative to the HY scale. This moderate imprecision of the latent time to monitor disease progression might be due to the subjective ratings of the clinical scores from which it was processed.

The gait parameters combined in a multivariate linear model were not efficiently selected which might lead to non-significant results. Careful analysis by a clinician of the gait parameters might help to find the right combination of gait parameters.

The limited number of PD patients in both datasets may hinder the identification of subtle relationships between gait parameters and clinical scores. A gait parameter could be associated with a clinical score, but this association might not be detectable due to the small sample size of the dataset.

5.3. Future work

There are further ways to extend this thesis:

In the LTJMM, prodromal patients could be included in the dataset and non-motor symptoms measurements such as REM sleep behavior disorder could be used as outcome. Non-motor outcomes could help identify prodromal PD patients. The time of

5.3. FUTURE WORK

the prodromal state could then be estimated and compared to the different HY stages at baseline.

This work could be continued by testing different combinations of gait parameters in the multivariate linear model. The most efficient combination of parameters would be revealed. The relative contribution of each gait parameter could then be further analyzed. For gait parameters that lead to non-significant results, non-linear techniques could be used to associate them with UPDRS scores.

Since the motor symptoms of PD are asymmetrical and this asymmetry develops over time, the relevance of an ipsilateral analysis could be followed on a longitudinal scale. It is possible, that this ipsilateral set-up is relevant at HY=0 or HY=1 but loses its relevance as the disease further progress to HY=5.

6. Conclusion

This thesis aimed to model multiple long-term trajectories of PD patients using the LTJMM and investigate the potential of gait parameters extracted from a wearable sensor as biomarkers of PD progression.

The findings of this thesis were:

Firstly, despite the absence of HY status in the LTJMM, the latent time ordering of PD patients reflected the progression of the severity of patients rated with the HY scale.

Secondly, Gait speed, Stride length, Stride time and Stance time were associated with the clinical axial score using a linear model.

Moreover, it highlighted the need for further research to explore the right combination of gait parameters in a multivariate linear model, the ipsilateral strategy to analyze the relationship between clinical scores and gait parameters.

The results of this thesis have important implications for clinicians and PD patients. They could improve the objective rating accuracy of motor symptoms of PD patients and their progression.

Overall, this thesis contributed to the growing body of literature on the validation of gait parameters as biomarkers of PD progression.

Bibliography

- Balestrino, R., & Schapira, A. (2020). Parkinson disease. *European Journal of Neurology*, 27(1), 27–42. [https://doi.org/https://doi.org/10.1111/ene.14108](https://doi.org/10.1111/ene.14108)
- Burq, M., Rainaldi, E., Ho, K. C., Chen, C., Bloem, B. R., Evers, L. J., Helmich, R. C., Myers, L., Marks Jr, W. J., & Kapur, R. (2022). Virtual exam for parkinson's disease enables frequent and reliable remote measurements of motor function. *NPJ Digital Medicine*, 5(1), 65.
- Cubo, E., Martinez-Martin, P., Gonzalez-Bernal, J., Casas, E., Arnaiz, S., Miranda, J., Gamez, P., Santos-Garcia, D., et al. (2020). Effects of motor symptom laterality on clinical manifestations and quality of life in parkinson's disease. *Journal of Parkinson's Disease*, 10(4), 1611–1620.
- Elkurd, M., Wang, J., & Dewey Jr, R. B. (2021). Lateralization of motor signs affects symptom progression in parkinson disease. *Frontiers in Neurology*, 1287.
- Fröhlich, H., Bontridder, N., Petrovska-Delacréta, D., Glaab, E., Kluge, F., Yacoubi, M. E., Marin Valero, M., Corvol, J.-C., Eskofier, B., Van Gyseghem, J.-M., et al. (2022). Leveraging the potential of digital technology for better individualized treatment of parkinson's disease. *Frontiers in neurology*, 13, 788427.
- Gabry, & Modrák. (2021). *Visual mcmc diagnostics with bayesplot*. <https://cran.r-project.org/web/packages/bayesplot/vignettes/visual-mcmc-diagnostics.html>
- Genetrail: P-value adjustments [Accessed: 2023-06-16]. (2023). *Center for Bioinformatics, Saarland University*.
- Iddi, S., Li, D., Aisen, P. S., Rafii, M. S., Litvan, I., Thompson, W. K., & Donohue, M. C. (2018). Estimating the Evolution of Disease in the Parkinson's Progression Markers Initiative. *Neurodegenerative Diseases*, 18(4), 173–190. <https://doi.org/10.1159/000488780>
- Li, D., Iddi, S., Thompson, W. K., Donohue, M. C., & Initiative, A. D. N. (2019). Bayesian latent time joint mixed effect models for multicohort longitudinal data. *Statistical methods in medical research*, 28(3), 835–845.
- Oskolkov, N. (2020). *Linear mixed model from scratch*. <https://towardsdatascience.com/linear-mixed-model-from-scratch-f29b2e45f0a4>
- Pinheiro, J., & Bates, D. (2009). 2000. mixed-effects models in s and s-plus. *Statistics and Computing Series*. New York: Springer-Verlag.
- Riederer, P., Jellinger, K., Kolber, P., Hipp, G., Sian-Hulsmann, J., & Kruger, R. (2018). Lateralisation in parkinson disease. *Cell and tissue research*, 373, 297–312.
- Rodriguez-Molinero, A., Sama, A., Perez-Lopez, C., Rodriguez-Martin, D., Alcaine, S., Mestre, B., Quispe, P., Giuliani, B., Vainstein, G., Browne, P., et al. (2017). Analysis of correlation between an accelerometer-based algorithm for detecting parkinsonian gait and updrs subscales. *Frontiers in Neurology*, 8, 431.

Bibliography

- Schalkamp, A.-K., Peall, K. J., Harrison, N. A., & Sandor, C. (2022). Wearable devices can identify parkinson's disease up to 7 years before clinical diagnosis. *medRxiv*, 2022–11.
- Schlachetzki, J. C., Barth, J., Marxreiter, F., Gossler, J., Kohl, Z., Reinfelder, S., Gassner, H., Aminian, K., Eskofier, B. M., Winkler, J., et al. (2017). Wearable sensors objectively measure gait parameters in parkinson's disease. *PloS one*, 12(10), e0183989.
- Severson, K. A., Chahine, L. M., Smolensky, L. A., Dhuliawala, M., Frasier, M., Ng, K., Ghosh, S., & Hu, J. (2021). Discovery of parkinson's disease states and disease progression modelling: A longitudinal data study using machine learning. *The Lancet Digital Health*, 3(9), e555–e564.
- Taboga, M. (2021). *Metropolis-hastings algorithm* [Accessed: 2023-06-15]. <https://www.statlect.com/fundamentals-of-statistics/Metropolis-Hastings-algorithm>
- Zhong, Z., Merkitch, D., Karaman, M. M., Zhang, J., Sui, Y., Goldman, J. G., & Zhou, X. J. (2019). High-spatial-resolution diffusion mri in parkinson disease: Lateral asymmetry of the substantia nigra. *Radiology*, 291(1), 149–157.

A. Supplementary material

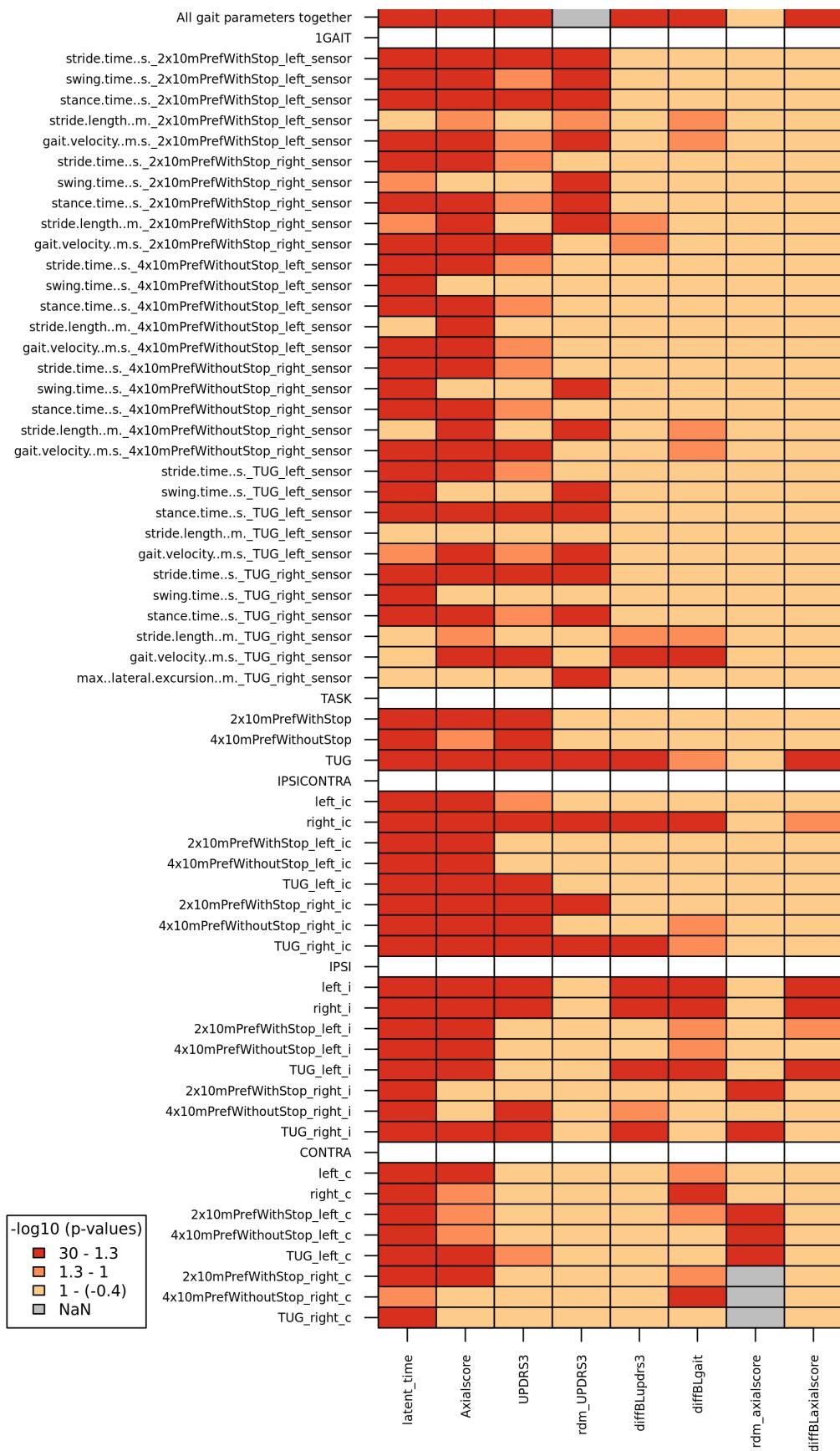


Figure A.1.: All adjusted p-values of all ANOVA's of Erlangen

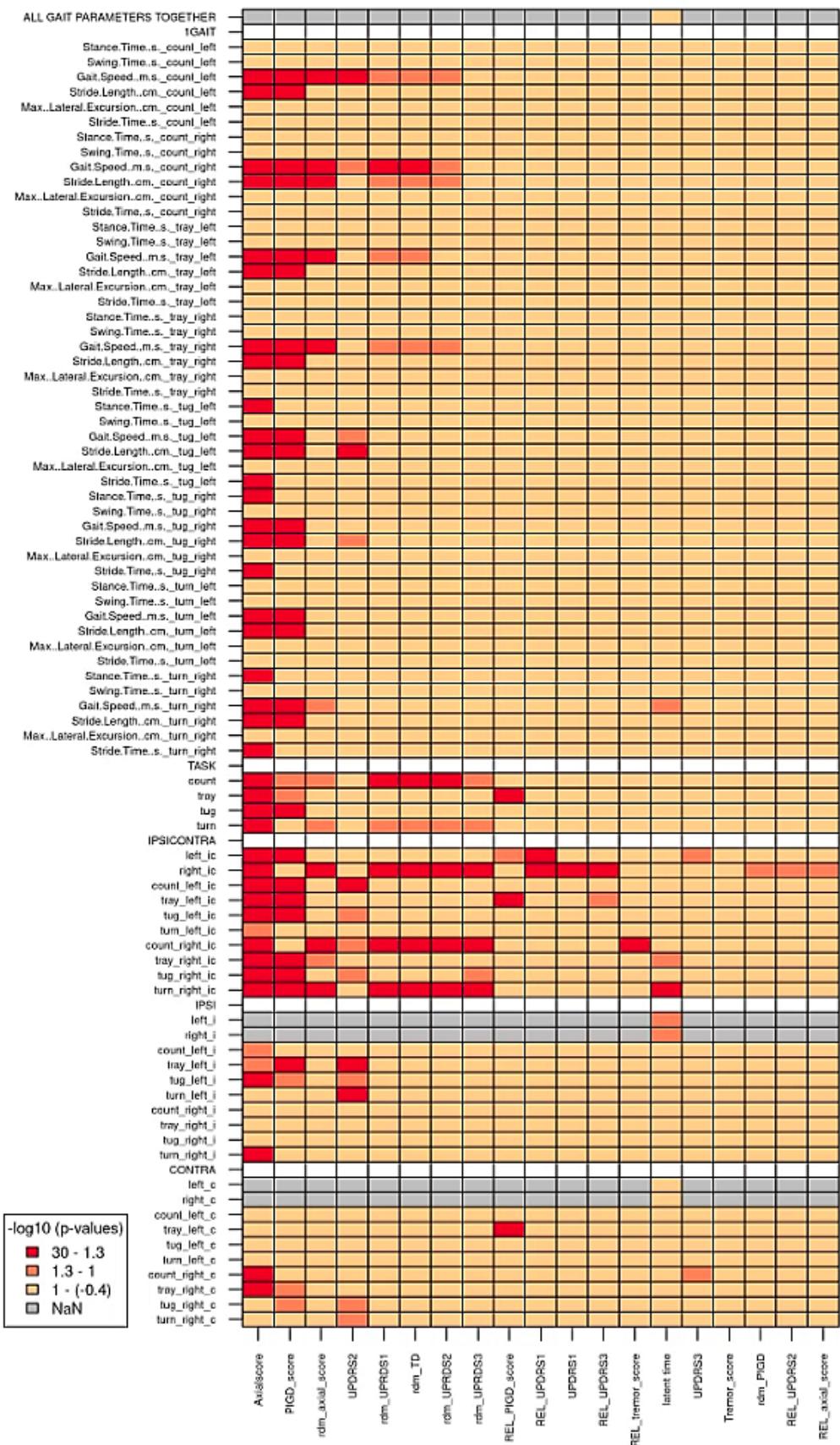


Figure A.2.: All adjusted p-values of all ANOVA's of NCERPD

List of Figures

3.1. Level of analysis and research scenarios	25
4.1. Number of patients per visits in Erlangen	28
4.2. Number of patients per visits in NCERPD	28
4.3. Trace plot of beta in Erlangen	29
4.4. Trace plot of sigma-delta in Erlangen	30
4.5. Erlangen LTJMM convergence diagnostics	31
4.6. Trace plot of beta in NCERPD	32
4.7. Trace plot of sigma-delta in NCERPD	33
4.8. NCERPD LTJMM convergence diagnostics	34
4.9. Correlation of random slopes in Erlangen	36
4.10. Distribution of latent time by sex group in Erlangen	37
4.11. Distribution of latent time by HY group in Erlangen	37
4.12. Correlation of random slopes in NCERPD	39
4.13. Distribution of latent time by sex group in NCERPD	40
4.14. Distribution of latent time by HY group in NCERPD	40
4.15. Erlangen significant adjusted p-values of ANOVAs, view by 1gait	44
4.16. Erlangen significant adjusted p-values of ANOVAs, view by 1gait (without latent time)	45
4.17. Erlangen significant adjusted p-values of ANOVAs, view by UPDRS3, axial score, rdm-UPDRS3, rdm-axialscore, (without 1gait)	46
4.18. Erlangen significant adjusted p-values of ANOVAs, view by diffBL-UPDRS3, diffBL-axialscore, diffBL-gait, latent time, (without 1gait)	47
4.19. NCERPD significant adjusted p-values of ANOVA's, view by individual gait	48
4.20. NCERPD significant adjusted p-values of ANOVA's, view by all gait, task, ipsicontralateral, ipsilateral, contralateral	49
4.21. Similarities between Erlangen and NCERPD analysis	50
4.22. Diagnostics of LME in Erlangen in level 1 gait for the outcome axial score and the predictors related to TUG	53
4.23. Diagnostics of LME in Erlangen in scenario ipsicontralateral for the outcome axial score and the task-side predictors TUG-left and TUG-right	54
4.24. Diagnostics of LM in NCERPD in level 1 gait for the outcome axial score and the predictors related to TUG	55
4.25. Diagnostics of LM in NCERPD in scenario ipsicontralateral for the outcome axial score and the task-side predictors TUG-left and TUG-right	56
A.1. All adjusted p-values of all ANOVA's of Erlangen	67
A.2. All adjusted p-values of all ANOVA's of NCERPD	68

List of Tables

4.1.	Number of patients at baseline by HY status in Erlangen	38
4.2.	Number of patients at baseline by HY status in NCERPD	41
4.3.	Spearman rank correlation of Erlangen gait parameters of TUG left side	41
4.4.	Spearman rank correlation of NCERPD gait parameters of TUG left side	42
4.5.	Count of tasks in the total of the significant adjusted p-values	43
4.6.	Ratio of significant adjusted p-values (%)	43
4.7.	Conditional R ² and Marginal R ² by level in Erlangen	57
4.8.	Adjusted R ² by level in NCERPD	58

I herewith certify that this material is my own work, that I used only those sources and resources referred to in the thesis, and that I have identified citations as such.

Bonn, 19th of June 2023

A handwritten signature in black ink, appearing to read "Jackrite To".

Jackrite To

