2014•2015
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

A randomized comparison of integrated versus consecutive dual-task training in people with Parkinson"s disease: The duality trial

Supervisor :
Prof.dr. Christel FAES

Supervisor :
Prof.dr. ALICE NIEUWBOER

John Epoh Dibato

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

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

universiteit hasselt
KNOWLEDGE IN ACTION

Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt
Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek

universiteit hasselt | Maastricht University

2014•2015
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

A randomized comparison of integrated versus consecutive dual-task training in people with Parkinson"s disease: The duality trial

Supervisor :
Prof.dr. Christel FAES

Supervisor :
Prof.dr. ALICE NIEUWBOER

John Epoh Dibato

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

universiteit hasselt | Maastricht University

FACULTY OF SCIENCE :
Master Thesis in Biostatistics, 2014-2015

A randomised comparison of integrated versus consecutive dual-task training in people with Parkinson's disease:
The DUALITY trial

Submitted by:

John Epoh DIBATO

Master of Statistics, Hasselt University, Martelarenlaan 42, 3500 Hasselt, Belgium

Supervised by:

Prof. dr. Christel Faes, Universiteit Hasselt
Prof. dr. Alice Nieuwboer, Katholieke Universiteit Leuven

Thesis submitted in partial fulfillment of the requirements for the degree of Master in Statistics: Biostatistics

September, 2015

Acknowledgements

I would like to thank God Almighty for seeing me through up till this moment and making it possible for me to achieve this prestigious academic subject in my educational career.

I would also like to express my profound gratitude to my supervisors; Prof. dr. Christel Faes and Prof. dr. Alice Nieuwboer for their continuous guidance and direction through insightful suggestions and the provision of some scientific journals geared towards the realization of this project.

Special thanks to Ms Carolien Strouwen of Katholieke Universiteit, Leuven for providing the data for this project. I would also wish to appreciate the one-on-one consistent interaction and discussion we had on this project from its inception to the final end.

A very big thanks to the University of Hasselt in association with IBIOSTAT in general and the staff of the Statistics department in particular for imparting and boosting my career positively. "*I will ever live to think of this great opportunity*".

Most importantly, I appreciate my family and friends who contributed in one way or the other till this final moment. For this reason, I hereby dedicate this thesis to the following: Mrs ENNI Evelyn, Dibato Clarisse, Ako Damasius, Gilbert Tebah, and Dibato Heldrine.

Principal Abbreviations

GEE: Generalized Estimating Equations

LMM: Linear Mixed Model

MAR: Missing At Random

MCAR: Missing Completely At Random

MI: Multiple Imputation

MNAR: Missing Not At Random

NBM: Negative Binomial Model

PM: Poisson Model

PMM: Pattern Mixture Model

RCTs: Randomized Clinical Trials

Abstract

Introduction: This project involved Parkinson's Disease (PD) patients who were randomized to two dual-task training interventions (Integrated and Consecutive) with the aim of improving their dual-task performances and achieving better lifestyles. The main objectives of this study were: 1) to check whether dual-task training could result in improvements of dual-task performances among PD patients; 2) to assess the more efficient training strategies employed in terms of dual-task performance and safety.
Methods: LMM fitted on Gait velocity and NBM fitted on Fall counts were used to assess the efficacy and safety respectively of the training exercises.
Results: Results confirmed that patients under the Integrated dual-task training had higher rate of increase of Gait velocity and lower fall rate compared to patients under the Consecutive dual-task training but statistically the differences were not significant (efficacy: est=0.63cm/s pvalue=0.4147, safety: est=-0.03 RR=0.97 pvalue=0.8916). Further result checks revealed that there was highest significant increase of Gait velocity (about 11cm/s increase, pvalue$<$0.0001) at six weeks after training interventions compared to measurements at the beginning of the study. On the other hand, greatest drop in Fall counts (about 21% drop, pvalue=0.4218) occurred six weeks after training interventions compared to Fall counts at baseline though it was not significant.
Conclusion: In this study, it was concluded that the training practices actually improved on the performance of the dual-task among the PD patients. Though Integrated task seemed better than Consecutive in improving dual-task performances, the two approaches did not differ much from each other and can be employed to improve the dual-tasking abilities of PD patients. In addition to improved dual-task performances achieved, better lifestyle was also realised since the fall rate among the patients dropped as a result of performing the training practices. In the light of the presence of missing responses common to most longitudinal clinical trials data, accounting for missing data can be further addressed with the use of Selection Models methodology to give additional confidence in the findings gained so far from the models already applied in this project.

Keywords:Dual-task, Missing At Random, Parkinson's Disease, Sensitivity Analysis.

Softwares: SAS 9.4; R Studio; laTex, SPSS 21

Contents

List of Tables

List of Figures

1 Introduction and Literature Review

Parkinson's Disease (PD), is a progressive disease of the nervous system marked by tremor, postural instability, muscular rigidity and slow, imprecise movement (bradykinesia), chiefly affecting middle-aged and elderly people [30, 34]. Symptoms of Parkinson's disease can appear at any age, but the average age of onset is 60 [30]. One safety challenge many patients with advanced PD face is "freezing", which is the temporary, involuntary inability to move [31]. The cause of freezing is unknown and freezing creates a danger of falling because the beginning and end of a freezing episode are unpredictable [31]. Approximately 60 percent of people with PD fall each year and about two-thirds fall recurrently [2]. Falls in PD occur mostly when turning or changing directions and is often related to a "freezing episode". Not all people with PD experience freezing episodes, but those who do are at a much higher risk of falling [31]. Apart from frequent falls experienced by PD patients, they also face daily normal challenges of activities involving multitasking.

In several activities of daily living, more than one task is executed at the same time. The ability to execute dual tasks is highly advantageous and is a pre-requisite to normal life. Walking around, for example, allows for communicating with someone else, transporting of objects from one place to another and monitoring of the environment to avoid accidents [33,42]. In normal circumstances, the concomitant execution of motor and cognitive tasks is common, and in such situations, motor activities are performed "automatically", with little or no effort in conscious attention required [12]. Such autonomous stage of performance of a motor ability [14] is achieved through a process of motor learning in which practice and its variability bring about the formation of action programs [40].

Action programs are controlled by open circuit, with little interference of feedback [40]. Therefore, demands on attention mechanisms necessary to efficient performance are very low, facilitating attention to focus on other items relevant to task performance. From this point on, it is possible for an individual to execute a second task simultaneously with the first, without any interference in performance. Dual-task is the analysis of the performance cost of a task executed simultaneously with a secondary task [19]. This involves the execution of a primary task, which is the main focus of attention, and a secondary task, performed simultaneously [33] but with different goals. The execution of two tasks at the same time demands a high level of information processing and thus performance of one or both tasks is deteriorated [4, 16]. The negative influence on the primary and/or secondary task occurs because both tasks compete with similar processing demands [4].

The influence of cognition or motor control alterations (or both) in the performance of dual-task can be an important indicator of the functional status of a patient during illness or during rehabilitation. Such alteration is usually regarded as motor-cognitive interference [11]. After a cerebral lesion for example, motor-cognitive interference can occur, causing activities which were previously automatic to require a controlled process, with increased attention demands. This in turn deteriorates performance in dual tasks

[16]. Such is common in patients with Parkinson's Disease with a progressive degenerative disorder of the central nervous system and as a consequence results in impairments of the basal nuclei, with progressive loss of neurons from the substantia nigra pars compacta. As a result of impairments in the basal ganglia, automatic task performance in people with Parkinson's disease is affected.

Patients with Parkinson's Disease can generate normal motor patterns when they focus their attention on performance, that is, when they think on the execution of movements [47]. This way they activate the intact pre-motor cortex (Figure 1) area and avoid relying on the impaired circuitry of the basal nuclei to assist in the production of movements. In dual task situations, the use of these cortical resources to execute motor tasks can restrict performance in both tasks. This is because secondary cognitive tasks rely on these brain resources and when performing the two tasks together, it is more likely that residual capacity will not be sufficient.

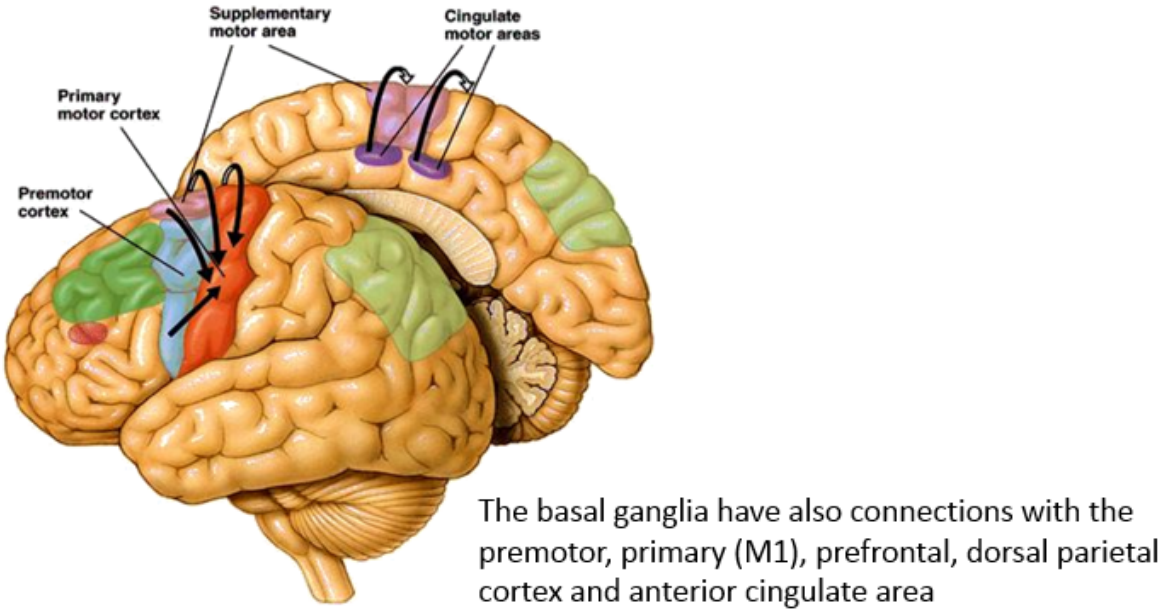

Figure 1: *Overview of structures and functions of major components of brain (Motor circuit of Basal ganglia)*
Source: Sherwood (2011)

Wu and Hallett (2007) carried on a study to assess the correlation between neurons and dual tasks performance in PD patients and healthy volunteers. They realised that after practice, the dual tasks activated similar brain regions in both groups. It was realised also that patients had greater activity in the cerebellum, premotor area and prefrontal cortex compared to normal individuals. They concluded that difficulty in performing two tasks simultaneously in PD patients is probably due to limited attentional resources, defective central executive function and less automaticity in performing the tasks. From

their findings, they recommended that practice can diminish dual-task interference and improve performance in patients with PD.

In another study to assess the safety of exercises on PD patients conducted by Colleen et al. [9], it was confirmed that fewer falls in exercise group were recorded compared to patients who did not carry on the exercise.

However, in physiotherapeutic guidelines it is often advised to avoid dual-task training in people with Parkinson's disease because of an increased risk for adverse events such as falls [18]. In recent years, there is however a tendency of supporting dual-task training in people with PD since recent pilot studies showed improvements in clinical outcomes (gait, balance) after dual-task training. In addition, dual-task training might make patients aware of their limitations causing them to be more cautious when performing these kind of tasks. As of now, large RCTs confirming these pilot study results are lacking and it is still not known whether it is safe to use these kinds of interventions (independently of possible clinical benefits these interventions might show).

To address these issues, in this project PD patients were trained under two dual-task training strategies as a means of focusing their attention and improving in their performances. The first strategy involved consecutive task training (CTT), whereby each task was trained separately. The other strategy involved the integrated dual-task training (IDT) in which the two tasks were performed simultaneously. Various gait parameters were measured at four different sessions for each patient. Being a longitudinal study involving patients, it is prone to missingness.

Missing observations frequently occur in all types of clinical trials especially in a repeated measure data such as longitudinal study involving humans. Reasons for missing data in a longitudinal trial study can range from causes such as accidents, patients' refusal to continue and illness of partner. A major problem in the analysis of clinical trials is missing data caused by patients dropping out of the study before completion. This problem can result in biased treatment comparisons and also impact the overall statistical power of the study [26]. It can become more problematic if the rate, time to, and reason for withdrawal differ widely among treatment groups. For these reasons, data analyses therefore should pay special attention to incomplete data sets.

According to Little and Rubin [24], missing data have been classified into three different types based on the possible reasons for the missingness: 1) missing completely at random (MCAR) for which the missingness is not related to any observed and unobserved factors. Examples here could be domestic relocation, suffering from an accident, or unrelated illness. 2) missing at random (MAR) for which the missingness depends on observed factors but is independent on the unobserved data. An example could be lack of efficacy of treament. 3) missing not at random (MNAR) for which the missingness depends on unobserved data as well as some observed factors. Previous analyses pointed out that the MAR assumption may be more plausible in practice than that of MCAR [10,21]. Nevertheless, a MCAR missing mechanism is also a MAR but not every MAR is a MCAR [44]. MNAR is useful in assessing the sensitivity of the results that are not MAR

[8, 26].

There are several statistical approaches used to analyse longitudinal data with missing values [7, 8, 23, 27, 41]. The methods chosen is based on the data missing mechanism, since different statistical methods are valid only under certain missingness mechanism with specified missing rates. This means there is no unique "best" method available for all situations. Missingness mechanism is difficult to test in a longitudinal clinical study and there is no clear definition into how much is considered as too much missing data [27]. Sprint and Dupin-Sprint [41] pointed out that the tolerable amount of missing data is that which will not hide an effect in the opposite direction. Two commonly used simple methods of analysing missing data [26] are **Complete Case Analysis** (CCA) and **Last Observation Carried Forward** (LOCF). CCA involves only individuals with data at all time points. With LOCF, every missing value is replaced by the last observed value from the same individual. CCA has some major shortcomings being that it suffers from loss of information and consequently ineffiecient estimators and will produce bias results when the data are MAR [26]. The LOCF technique can severely affect the features of a Linear Mixed Model (Verbeke and Molenberghs, 1997, chapter 5). The LOCF, CCA and other ad hoc approahes to deal with missing data can lead to serious bias. As a result, simple methods of analysing missing data have been replaced recently by approaches including Direct Likelihood (DL) and Multiple Imputation (MI) which are fully consistent with Intention To Treat (ITT) paradigm and valid under MAR [26, 43]. Assuming data is MAR, DL such as random effects model, and bayesian methods are valid procedures for analysis [6, 15].

Several researchers have considered and constructed simulation studies for the proof of strong consistency of imputation methods and to check the efficiency of the imputation methods. For example, Myers [28] compared the results of two imputation methods (that is, the complete case method and the multiple imputation method) based on simulated data sets of three hundred patients with a dropout rate ranging from 20% to 60%. He concluded that results from complete case method provided estimates and inference which differ from the simulated complete data set. On the other hand, multiple imputation provided mean values and results very close to those of the complete data set.

In a related study conducted by Ali, et al. [1], a simulated survival data analysis under MCAR and MAR were compared for four imputation methods: complete case analysis (CCA), means substitution (MS), and multiple imputation (MI) with and without the inclusion of the outcome (MI+ and MI- respectively). The simulation results suggested that in general MI+ was likely to be the best method.

Nakai et al. [29] concluded based on a simulation study that MI was the most effective imputation method in longitudinal data analysis under MCAR.

Despite the positive views of the previous researchers towards MI, others preferred DL over MI. For example Allison [35], stated that DL is more efficient than MI, for fully efficiency of MI can only be achieved after producing and analysisng an infinite number of data sets which is impossible.

In addition to the view of the latter researcher, Allison and Schafer [17,35] both acknowledged the fact that there is always a conflict between the imputation model and the analysis model. Such can happen when the imputation model has predictors which are absent in the analysis model. In such a situation some parameter estimates and standard errors may differ between the DL and MI [17].

Although DL can be more efficient than MI, there are situations where MI has advantages over DL. Molenberghs and Kenward [26] found that MI can work better in cases of missing predictors. It was further discussed that in fitting population-averaged models to repeated discrete outcomes with non-likelihood methods such as GEE, leads to results which are valid under MCAR. They recommended the use of MI then GEE to produce results which are valid under MAR. Finally, as a sensitivity analysis to assess the MAR assumptions of the primary analyses, MI provide a convenient approach [26] by which MNAR pattern mixture models can be used at the imputation step.

1.1 Objective and Hypothesis

Objectives: The objectives of this multicentered, single blind study were to check if dual-task training could result in improvements in dual-task performance; to distinguish between the two training strategies, the more efficient in terms of better dual-task outcomes and more safe with respect to reduce fall rates. This can be re-expressed as follows:

1. Do the training groups differ with respect to change in dual-task outcomes from baseline (0 and 6 weeks before intervention=T1 and T2 respectively) to post intervention (6 weeks after randomization=T3) controlling for other important variables.
2. Do the training groups differ with respect to change in dual-task outcomes from baseline (0 week before intervention=T1) to 12 weeks following intervention (T4) controlling for other important variables.

Hypothesis: It is hypothesized that Integrated Dual Task (IDT) practice will lead to better dual-task outcomes particularly in patients without cognitive impairment.

2 Data Description

The data set consists of PD patients of both sexes, between 40 to 80 years of age and who are at the Hoehn and Yahr stage II to III of the disease status. The patients were randomly allocated to two training groups: G=1 involved 65 patients who carried on six weeks of integrated dual task training (IDT); or G=0 which involved 56 patients who did six weeks of consecutive task training (CTT). Both training methods involved a walking and cognitive practice which were carried on at the patients' premises at the same rate with a 12 supervised sessions by a physiotherapist and 12 unsupervised sessions. Preceding the randomised period was a 6 week control period for all the patients. A total of 121 patients were recruited which was based on a pilot study previously conducted that aimed for a sample size of 120 with a power set at 80% after subjects who withdrew or lost-to-follow-up. The 120 sample size resulted in an expectation of 54 patients in each arm. A drop out rate of 5% was realised. The two outcome variables of particular interest were Gait Velocity (primary) in centimeter per second (V in cm/s) and Fall counts (secondary). Gait velocity represents the distance covered in a second by a patient. It was measured with the use of a GaitRite mat placed in quiet laboratories at Radboud University Nijmegen Medical Centre, in the Netherlands and at the Catholic University of Leuven in Belgium (Figure 2). Measurements were collected in four clinic visits including baseline (T1), 6 weeks before training (T2), 6 weeks after training (T3) and after 12 weeks follow-up periods (T4). Fall counts (FC) represent the total number of falls an individual can experience within a period of time. It was recorded through a weekly phone call made to patients by a specialist. Total Fall counts were then collected at four measurement visits including first 6 weeks before training (T1), last 6 weeks before training (T2), first six weeks after training (T3) and last six weeks after training (T4). Possible predictors recorded for each patient at baseline included Hoehn and Yahr stages II and III (H=2 were 82 patients with stage II and H=3 were 39 patients of stage III), Freezing of Gait (F=0 for 52 nonfreezers and F=1 for 69 freezers) and patient's country (C=1 were 62 patients from Belgium and C=0 were 59 patients from the Netherlands).

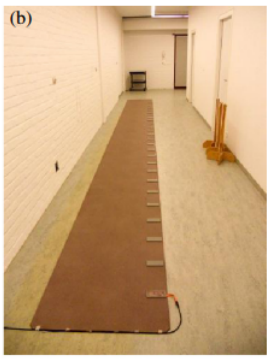

Figure 2: *GaitRite mat*
Source: Strouwen (2014)

3 Methodology

This was divided into the exploratory and analysis parts. For efficacy testing, analysis of the dual-task was performed under three different measurements of Gait velocity based on the three secondary tasks (a backward Digit Span task, an auditory Stroop task and a Mobile Phone task) done during the testing sessions. This gave rise to three Gait velocities: V1, V2 and V3 respectively. Results from V2 were of primary interest because the cognitive task was not practised during the training session. To reinforce on results and possible transfer of performance gained under trained tasks to untrained tasks, results from V2 were compared to those of V1 and V3.

3.1 Exploratory Data Analysis (EDA)

Exploratory analysis of longitudinal data was done to discover patterns of systematic variation across groups of patients, as well as aspects of random variation that distinguish individual patients. In this project, the number of patients and their corresponding mean velocity alongside the number of falls recorded at each test session were computed and compared for each training group and country. This was further substantiated with mean and patient profile plots between the two interventions on both response types. The variability and correlation between responses at the different test sessions were assessed both graphically and with some summary statistics. In addition to this, Gait velocity and Fall counts from each patient were assumed to be normally and poisson distributed respectively, and so, histograms and quantile plot were produced for each response per treatment arm to illustrate the nature of the assumed distributions. The exploratory approaches mentioned above, served as a guide in the modeling techniques applied in the subsequent sections.

Before analysis of the data, reasons for dropout from the patients with missing values were highlighted with some corresponding plausible missingness mechanism stated. This served as a first exploratory guide to missingness mechanism (which is normally assumed during data analysis) plausible with the data.

More formally, logistic regression (Dropout Model) analysis was used to establish key independent predictors of withdrawal. **MAR** is often considered more plausible if missing data are mainly due to monotone, with a scattering of nonmonotone missing data [26].

3.2 Accounting for missing data: Direct Likelihood (DL) and Multiple Imputation (MI) Approaches.

Many techniques have been developed to handle missing data in longitudinal clinical trials. However, there are few methods that are actually used in real trials with missing data. In this project, primary analysis was based on the DL and MI under the MAR

assumptions. Further analyses were also done using CCA and MI under the MCAR and MNAR assumptions respectively as part of sensitivity analyses.

Direct Likelihood: This involves fitting a model with likelihood estimation facilities which can handle incompletely observed data. DL approach can be applied to both continuous and discrete repeatedly measured outcomes. The principle and approach to DL in missing data analysis is based on the following: For every independent subject i=1,...,N in a study, let Y_{ij} be a set of planned measurements for subject i at measurement time j, and R_{ij} be missing data indicator (1 if Y_{ij} is observed and 0 if Y_{ij} is missing). The full data (Y_i R_{ij}) with Y_{ij} ($Y_i^o = observed$ or $Y_i^m = missing$) is modeled as illustrated below:

$$f(Y_i^o, R_i/\theta, \psi) = f(Y_i^o/X_i, \theta) f(R_i/Y_i^o, \psi) \tag{1}$$

where θ and ψ are the parameters for the measurement and missingness processes respectively and X_i the matrix of covariates. Within the likelihood framework, if the observed data and missingness process are independent of each other, the separability condition is fulfilled [26]. Under this condition, inference can be made solely under the marginal observed data, $f(Y_i^o/X_i, \theta)$, while ignoring the missingness process (unless it is of scientific interest). According to Rubin (1976) and Molenberghs and Kenward [26], likelihood based methods and inference with ignorability condtion are valid under the MAR assmptions. Two commonly used likelihood methods are the linear mixed model (LMM) for continuous outcome and generalized linear mixed model (GLMM) for discrete outcomes. Parameters under the LMM model have both marginal and subject-specific interpretations whereas those of the GLMM have subject-specific interpretions. Either of them can be applied with respect to the research question of interest.

Since Gait velocity is a continuous response, the LMM [26, 43] was used. Given V_i being the vector of measurements of Gait velocity available for patient i, a general linear mixed model can be expressed as:

$$V_i = X_i\beta + Z_i b_i + \epsilon_i \tag{2}$$

where β is a vector of population-averaged regression coefficients representing the fixed effects which are often of scientific interest, b_i is a vector of patient-specific regression coefficients which measure how a PD patient deviates from the average in the population of interest. X_i and Z_i are the matrices of known covariates (G, T, F, H, C and interactions). The random effects (b_i) and the residual components (ϵ_i) are assumed to be independent and normally distributed expressed as N(0, G) and N(0, Σ_i) respectively. Inference in LMM, is usually based on maximum likelihood (ML) or restricted maximum likelihood (REML) for both the fixed effect parameters and variance components. The fitting of a Linear Mixed Model usually lead to an implied marginal model with mean $X_i\beta$ and covariance $V_i(\alpha) = Z_i G Z_i' + \Sigma_i$, with α being the 'variance components' in the matrices of D and Σ_i. This can be accomplished by fitting a LMM for incomplete longitudinal

gait velocity measures using the SAS procedure, *PROC MIXED* and will lead to valid estimates and likelihood ratios under MAR [26].

If interest on inference is based on β, t and F distributions are often used with denominator degrees of freedom estimated either by Satterthwaite or Kenward and Roger approximation methods [Satterthwaite 1941; Kenward and Roger 1997]. On the other hand inference for some of the α_s can be done by using classical wald, score and likelihood ratio test. But the null hypothesis of some of the parameters (variance of random effects) usually have boundary problems and as a consequent, the mixture of chisquare are used to get an appropriate likelihood ratio test and p values [Stram and Lee (1994, 1995), Molenberghs and Verbeke (2006)].

Multiple Imputation: Multiple Imputation proposed by Rubin [37] to analyze incomplete data under the MAR missing mechanism is a method whereby missing values in the data set are filled multiple times with plausible estimates to generate multiple (m) data sets. Then, these imputed data sets are analyzed by standard procedures that are commonly used in analyzing complete data sets. Finally, the results of analyses are combined to provide a single estimate of parameters of interest, together with standard errors that reflect the uncertainty common to imputed data set [3, 13]. In principle, the m data sets will form m different estimates of the regression parameters, β (i.e $\hat{\beta}^{(k)}$ for k=1,...,m) and the respective estimates for variance-covariance. The final estimates are then calculated thus:

$$\hat{\beta} = \frac{1}{m}\sum_{k=1}^{m}\hat{\beta}^{(k)} \tag{3}$$

and

$$\widehat{Cov}(\hat{\beta}) = W + (1 + m^{-1})B \tag{4}$$

where W and B are the within and between imputation variability respectively.

Under the MAR assumption, MI is based on the expression given below:

$$f(Y_i^m/Y_i^o, X_i, R_i) = \underbrace{f(Y_i^o/X_i, \theta)}_{predictive, distribution} \tag{5}$$

From the expression, imputation of Y_i^m is obtained by randomly drawing values from the predictive distribution. In this setting, the predictive distribution of the missing data given the observed data does not depend on the missing data indicator, R_i. When the data missing pattern is monotone, regression method can be applied to impute the missing values. This requires the data to be in the "horizontal format" and the first response is always observed. Subsequent responses have missing data due to attrition and the amount of missingness increases as the response number increases. The appropriate regression

model first predict Y_{i2} from Y_{i1} and X_i from which random samples are drawn to fill the missing values in the second response. This is done sequentially to the last response. For nonmonotone missingness Schafer (1997) proposed draws using the Markov chain Monte Carlo (MCMC) methods. In the MCMC approach, the Markov chains are drawn repeatedly and long enough until convergence. Once target distribution is attained, draws can be made from the posterior distribution (multivariate normal) which are used to fill the missing responses. Both MCMC and regression methods will lead to similar results for monotone missingness pattern and can be applied in SAS using the procedure PROC MI. A few assumptions and constraints of MI are: 1) missing data mechanism should be MAR; 2) the imputation model must match the analysis model [6]; and 3) the algorithm used to generate imputed values must accommodate the variables associated with the missingness of the data as well as other related variables. Two major advantages of MI are that it allows the use of complete data methods for data analysis and it incorporates random errors in the imputation process. MI can accommodate any model with any data and does not require specialized software. In addition, MI increases efficiency of the estimates through minimizing the standard errors [37]. However, Rubin [37] pointed out that the three disadvantages of MI are more effort to create the multiple imputations, more time to run the analyses, and more computer storage space for the imputed data sets.

The MI method was used for the Fall count data. In line with the response being discrete, likelihood methods appropriate for it under the MAR is GLMM. Since research questions are population based, appropriate models which are valid under MAR would be the MI using MCMC method at the imputation step and later fitting the imputed data using appropriate Count regression model under the GEE approach. Assuming a Poisson model for the Fall counts, after filling the data multiple times each imputed data is then fitted using the following marginal model formulation:

$$log(E[FC_{ij}]) = \beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip} \tag{6}$$

where β_i are the fixed effects and X_i the covariates and $Var(Fc_{ij}) = \phi v(E[FC_{ij}])$ where ϕ is the scale parameter that needs to be estimated. Correlation in the data, $Cor(Y_{ij}, Y_{ik}) = \alpha$ also needs to be estimated. The GEE approach to estimate the parameters, solves the following equations:

$$\sum_{i=1}^{n} D_i' V_i^{-1} (y_i - E[FC]) = 0 \tag{7}$$

where V_i is the "working" covariance matrix, D_i the "derivative" matrix of expected FC with respect to β. These equations are solved iteratively where β, α and ϕ are updated until convergence is attained.

As an approach to sensitivity analysis under MNAR, MI was used for both FC and V models. In this approach, the predictive distributions from which Y_i^m values can be filled

were modified to depend on R_i or other missing pattern specified. This changed the missingness mechanism from MAR to MNAR and draws from the modified predictive distributions were done using the MI methodology described above.

3.3 Statistical Analysis

The presence of missing responses in the data require statistical modeling strategies to account for the missingness in the data as elaborated above.

First analysis was to explore the dependence of the missing responses on previous observed responses (P), and other covariates. This was done by fitting a logistic dropout model [26] as stated in equation (8).

$$logit[p(dropout)] = \psi_0 + \psi_1 P_i + \psi_2 G_i + \psi_3 F_i + \psi_4 H_i + \psi_5 C_i + \psi_6 T_i + \psi_7 T_i^2 + \psi_8 G_i * T_i \quad (8)$$

Fitting this model helped to have an idea of the type of assumptions of missingness (MAR or MCAR) prominent with respect to the data. Nevertheless, a data which is MCAR is also MAR but not the reverse [26, 44]. By this setting, the proceeding modeling for both responses was done under MAR assumptions and sensitivity of results were assessed by further model fitting under the MNAR assumptions.

3.3.1 Efficacy

For efficacy testing, a saturated linear mixed model (LMM) for missing Gait velocity was fitted as illustrated below:

$V_{ij} = (b_{0i} + \beta_0) + \beta_1 Group_i + (b_{1i} + \beta_2) Time_{ij} + (b_{2i} + \beta_3) Time_{ij}^2 + \beta_4 Freezing_i + \beta_5 Hoehn\_Yahr_i + \beta_6 Country_i + \beta_7 Group_i * Time_{ij} + \beta_8 Group_i * Time_{ij}^2 + \epsilon_{ij}$

where b_{0i}, b_{1i}, b_{2i}, and ϵ_{ij} are the random patient's effects for the intercept, linear time, quadratic time and error terms respectively. The random effects were introduced into the model to account for the correlations of repeated Gait velocity measurements for each patient.

According to Verbeke & Molenberghs [43], such complex structure will favour getting consistent estimators for both the mean and covariance structures during model building. The initial model structure was then reduced to a parsimonious structure through the *top down* model building approach [43]. Similar analysis was done through multiple imputation (Rubin 1987) by which the missing responses were filled 50 times (using PROC MI in SAS) from a predictive multivariate normal distribution using the MCMC method as described under multiple imputation methodology above. Each of the filled data were then fitted using the LMM and with the use of PROC MIANALYZE in SAS, the

results were combined to produce single estimates for both the fixed effects and covariance parameters. Both analyses are valid under the MAR [26] and were compared to each other for consistency of results and inference. Based on some literature as discussed earlier, results from LMM (MAR1) was finally preferred to MI then LMM (MAR2) [17, 35].

3.3.2 Safety

In this analysis, Count regression models such as Poisson or Negative Binomial regression are much more appropriate [13]. Models on Poisson distribution assume that the mean and variance of fall counts are equal. For the Negative Binomial Models (NBM), this assumption is relaxed since it can handle data with variance greater than the means (overdispersion) by correcting for overdispersion which very often occurs in real data setting [13]. The choice between the two models was done both graphically and with the use of Pearson Chisquare goodness-of-fit statistics, BIC and AIC values. In Count regression models, the predictors are connected to the outcomes (Fall counts), via a natural logarithmic link function. Just like in logistic regression, the coefficients are usually exponentiated and interpreted as incidence rate ratios (**IRR**).

For the safety testing, an initial saturated Poisson model (PM) [3,13] was fitted based on Generalized Estimating Equations [22] with an initial unstructured (UN) covariance matrix to cater for the correlation of the repeated Fall counts per patient:

$$log[E(FC_{ij})] = \beta_0 + \beta_1 Group_i + \beta_2 Time_{ij} + \beta_3 Time_{ij}^2 + \beta_4 Freezing_i + \beta_5 Hoehn\_Yahr_i + \beta_6 Country_i + \beta_7 Group_i * Time_{ij} + \beta_8 Group_i * Time_{ij}^2$$

where the β_s are interpreted as the log of the incidence rate of Fall counts for each population of interest.

Reduction of the model was done by first choosing an appropriate covariance structure. This was achieved by comparing the empirical (EBSE) and model based (MBSE) standard errors. The model whose EBSE and MBSE were closed to each other was considered appropriate. The fixed effects were then reduced by removing insignificant terms and also keeping some terms based on the research questions. Since the response is a count data and there can be a possibility of overdispersion, a Negative Binomial Model (NBM) was also fitted under the GEE approach with the unstructure covariance matrix as an initial choice [3,13].

The two initial models fitted using the GEE approach to the missing Fall counts was done under the nonlikelihood approach and parameters estimated iteratively through the generalized estimating equations until convergence is reached. These models are valid under MCAR [26].

Since inference with respect to the research questions is population based, a similar model (either NBM or PM) was fitted after filling in the missing Fall counts 50 times using the MCMC methods applied in multiple imputation as described above. The 50

imputed data sets were then modelled under GEE approach and the results from each data combined using PROC MIANALYZE in SAS to get a parsimonious final model. This final model (MI then GEE) is valid under MAR and was chosen for the primary analysis of the safety testing because of its broader validity and also based on what was observed in the literature from the papers of other researchers [26]. The model under MCAR was considered for sensitivity analysis.

3.3.3 Sensitivity Analysis (SA)

The missing data model of the MAR assumption used as the primary analysis whose assumptions cannot be checked still needs to be assessed for the sensitivity of results to change in assumptions. The purpose of an SA is to analyse the missing data under different assumptions regarding missingness in order to ascertain a range of conditions under which the study conclusions are robust as well as conditions under which conclusions can be overruled [26]. This is the principle behind the sensitivity analysis approach used in this project.

Having a thorough knowledge for the reasons of missing data from the EDA, a systematic approach begins here with accounting for the reasons (mechanism) which caused the data to be missing. Three prominent models for such analysis are the Pattern Mixture Models, Selection Models and Shared Parameter Models [26]. In this project focus was based on MNAR models obtained using the standard SAS procedures for MI (PROC MI and MIANALYZE) based on PMMs. In this approach, patients are put into groups based on common patterns of missing data (such as: time of dropout, reasons for discontinuation, treatment arms to which patients are randomised. The full data model under PMM is expressed as below:

$$f(Y_i^o, Y_i^m, R_i/X_i) = \underbrace{f(R_i/X_i,)}_{MIM} * \underbrace{f(Y_i^o/X_i, R_i)}_{ADM} * \underbrace{f(Y_i^m/Y_i^o, X_i, R_i)}_{PDM} \tag{9}$$

where MIM=missingness indicator model, ADM=available data model within each pattern, PDM=predictive distribution model within each pattern.

Due to missing values in some patterns, some parameters in PDM cannot be computed and as a consequence, "identifying restrictions" (ADM) are explicitly imposed or equated to the PDM. Imputation is then followed using sequence of calls to PROC MI with a MONOTONE statement (if data is not monotone). This is in contrast to a single call typically done to fill values under the MAR assumptions [26]. In this project two pattern-mixture MNAR models were constructed using the multiple imputation approach as discussed below:

MNAR1: Adjusted Imputed responses: For MNAR1 model analysis, PD patients were grouped according to having or not having missing responses within each training group. The focus here was to check if the pattern of the data was the same in patients

who do and do not have missing responses. In this analysis the distributions of the missing responses (PDM) were filled from the corresponding observed responses (ADM) 50 times using the MCMC method. Each imputed value were then adjusted with some arbitrary parameters at the different sessions. For the Gait velocity, it was assumed that the distribution of missing Gait velocity at T1, T2, T3 and T4 had expected values that were 2, 4, 6 and 8 respectively, lower than those of the corresponding distribution of the observed Gait velocity at T1, T2, T3 and T4. On the other hand, it was assumed that the distribution of missing Fall counts at T1, T2, T3 and T4 had expected values that were 0.1 higher than those of the corresponding distribution of the observed fall counts at T1, T2, T3 and T4 [9, 38]. The 50 imputed data sets were then analysed using LMM for Gait velocity or an appropriate Count regression model under GEE for Fall counts. The results from each data set were then combined using PROC MIANALYZE to produce final estimates for the MNAR1 model.

MNAR2: Complete Case Missing Value Restrictions: For MNAR2 model analysis, PD patients were grouped with respect to time of discontinuation of studies. The missing responses for each pattern of missingness (PDM) were filled in from the observed data from the completers (ADM) using the regression method approach. This was done 50 times based on the standard MI approached discussed earlier. The 50 imputed data sets were then analysed using LMM for Gait velocity or an appropriate Count regression model under GEE for Fall counts. The results from each data set were then combined using PROC MIANALYZE to produce final estimates for the MNAR2 model.

The results and estimates were assessed to check for sensitivity of results to assumptions made on the data. The final model was checked with respect to fits and assumptions levied on the residuals and random effects. This was done with the use of diagnostic plots including histogram, QQ plot, and plots between predicted and observed responses.

4 Results and Discussion

4.1 Efficacy

4.1.1 Exploratory Data Analysis

Table 1 shows the mean and standard deviations of V2 (Gait velocity from walking and untrained stroop tasks) for each training group and country. From Test session 1 to 4, the mean values seem not to differ much between the intervention groups (CTT and IDT). On average, there seem to be an increasing Gait velocity trend in both treatment groups. At the end of the study, 5 patients dropped out from the CTT group as compared to 7 patients in IDT. Patients from Belgium (BE) seem to have higher Gait velocity as compared to patients from Netherlands (NL) but the difference is not much. At the end of the study, 3 patients were missing for Belgium compared to 9 patients from the Netherlands. The standard deviation of the mean Gait velocity across the different Test sessions seem fairly constant hence depicting the presence of constant variance.

Table 1: *Mean(standard deviation) V2 per Training Group and Country*

	CTT			IDT			BE			NL		
T	**N**	**n**	**mean(sd)**	**N**	**n**	**mean(sd)**	**N**	**n**	**mean(sd)**	**N**	**n**	**mean(sd)**
1	65	65	92.63(19.95)	56	56	91.43(21.46)	62	62	93.55(19.41)	59	59	90.53(22.28)
2	65	62	92.60(18.04)	56	53	94.60(20.26)	62	62	94.34(18.55)	59	53	92.57(19.71)
3	65	61	102.32(20.18)	56	53	105.28(18.62)	62	60	103.06(21.71)	59	54	104.40(16.75)
4	65	60	98.98(21.75)	56	49	103.45(19.49)	62	59	102.87(21.52)	59	50	98.77(19.89)

N=number of patients at beginning of study,
n=number of patients remaining at each test session, T=test session

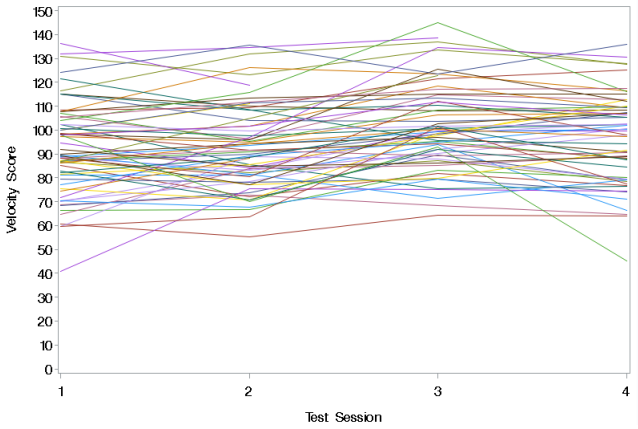

Figure 3: *Patient V2 profile for CTT*

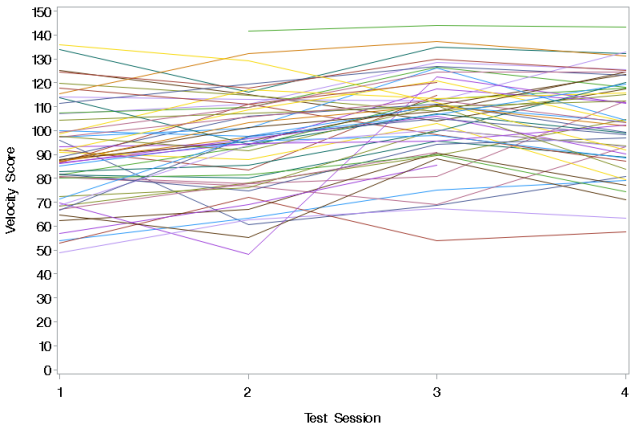

Figure 4: *Patient V2 profile for IDT*

Figures 3 and 4 illustrate the patients Gait velocity profile plots in the CTT and IDT

groups respectively. There seem to be much between patient variability compared to within patient variability. This suggests the presence of random effects and strong correlation between Gait velocity measurements within patients compared to Gait velocity values between patients. Studying the average Gait velocity evolution over time (Figure 5), indicates no much difference between the IDT and CTT training practices but the trend seem to present some downward curvature after Test session 3. This kind of trend suggest a quadratic evolution of V2 with time. Figure 6 and Table 2 show the relation between Gait velocity values at each Test session. The values show significant raw association in the data that need to be accounted for during modeling. Similar trend in the patient profile, mean structure and associations were observed with V1 and V3 (Appendices A and B).

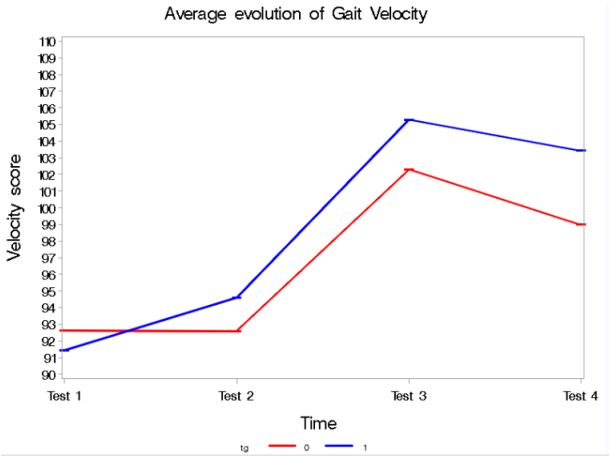

Figure 5: *Mean Evolution of V2*

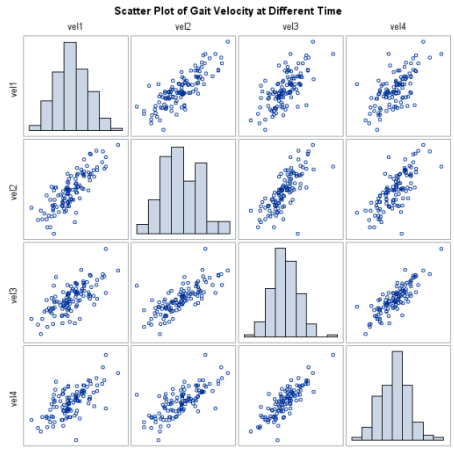

Figure 6: *Scatter Plot for V2*

Table 2: *Pearson correlation between V2 values at the test sessions*

	vel1	**vel2**	**vel3**	**vel4**
vel1	1	0.81364	0.68378	0.649
		<.0001	<.0001	<.0001
vel2		1	0.7605	0.73337
			<.0001	<.0001
vel3			1	0.83376
				<.0001
vel4				1

vel=Gait velocity

Analysis of LMMs are carried out under the assumption that missing data are **missing at random** (*MAR*). Under MAR, the probability of having missing data on Gait velocity may depend on the previous observed Gait velocity but not on missing Gait Velocity and according to Verbeke and Molenberghs (2000), inference based on maximum likelihood methods are valid. Apart from MAR, Little and Rubin (2002) and Allison (2001)

highlighted two other assumptions of missing data patterns: **missing completely at random** (*MCAR*) in which the probability of missing Gait velocity does not depend on previous observed Gait velocity and **missing not at random**, MNAR, which assumes that the chance of missingness depends on the value of the missing Gait velocity as well as the observed Gait velocity. In this analysis, the two main types of missingness pattern were explored. These include the **monotone** missingness where Gait velocity from a certain test session onward will be unobserved and **nonmonotone** missingness in which the unobserved Gait velocity values are being observed at certain test session later in the study. From Table 3, **nonmonotone** missingness accounted for about 2% in CTT and 2% in IDT group, while **monotone** missingness accounted for about 6% of the missing Gait velocity in CTT and 12% in IDT group. The remaining are completers who had measurements at all the test sessions. Missingness pattern for V1 and V3 are found in Tables 16 and 19 of Appendix B. Results from all Gait velocity values, revealed higher missing values were recorded for patients in the IDT group compared to patients in the CTT group.

Table 3: *Missingness pattern based on V2 per treatment group*

	CTT						**IDT**					
Pattern	**T1**	**T2**	**T3**	**T4**	**n**	**%**	**T1**	**T2**	**T3**	**T4**	**n**	**%**
1	O	O	O	O	60	92.3	O	O	O	O	48	85.7
2	O	O	M	M	2	3.1	O	O	O	M	4	7.1
3	O	M	M	M	2	3.1	O	O	M	M	1	1.8
4	O	M	O	M	1	1.5	O	M	M	M	2	3.6
5							O	M	O	O	1	1.8

n=number of patients for each pattern of missingness

In the analysis, the reasons for missing Gait velocity as stipulated in the design study and missingness mechanism were explored with the use of a dropout model. This actually helped in the plausibility of the assumed missingness mechanism in the data. These reasons were recorded both after Randomisation and at Follow-up periods:

Randomisation

- Five patients did not receive the allocated interventions with reasons such as: time consuming (MCAR); illness of partner (MCAR); relapse of disease (MAR); operation of knee (MAR); testing too intense (MCAR).

Follow-up

For the patients that were followed up and had missingness,

- The reasons included: discontinued intervention because of hospital stay (MAR); operation of back (MAR); cerebrovascular accident after test 3 (MNAR); illness of partner (MCAR); operation after test 3 (MNAR).

The dropout models results in Table 20 Appendix B, revealed insignificant effect of all covariates and including previous Gait velocity . There was a lone significant effect of country for V1. It can be concluded that missingnesss does not depend on the previous Gait velocity and other covariates except country. This means there is high chance of valid analysis based on MCAR for V2 and V3 which implied MAR and a higher chance for valid results under MAR for V1. Not withstanding the plausibility of results under MCAR and MAR, sensitivity of results still need to be confirmed under MNAR.

4.1.2 Statistical Analysis

To get a parsimonious model for Gait velocity, the initial saturated LMM highlighted at the methodology, was reduced using the "top-down" model building approach. Results led to a significant random quadratic term for time for both V1 (LRT=12.5, pvalue=0.0039) and V2 (LRT=9.2, pvalue=0.0184) meanwhile only the random patient intercept was left for V3 since the other two were insignificant. For V2, the maximised REML loglikelihood involving all three random effects was 3651.3 while the REML loglikelihood involving the random intercept and linear slope was 3660.5. This gave a gain in loglikelihood of 9.2 and based on a mixture of chisquare led to a significant p value of 0.0184. Variance variation was also assessed for the error terms by allowing for a gaussian serial correlation but the difference in REML loglikelihood was small for all three Gait velocities. Details can be found in Table 21, Appendix B. Consequently, the independent identity variance-covariance structure was adopted with no serial correlation. The fixed effects from the loaded initial model were reduced with the use of a t-test under Restricted Maximum Likelihood (REML) estimation approach. Based on the research question, a reduced model was formed having all significant main effects, linear and quadratic terms for Time and the interaction between Time and Training Group. A further test for the need of a quadratic Time was done by comparing the models with linear and quadratic Time. This was done with the use of Likelihood ratio Test under the Maximum Likelihood (ML) estimation approach. The result was further confirmed with the significance of the t-test for the quadratic time effect (Tables 4, 5, 6). Both results confirmed significant quadratic effect for the Time for all the Gait velocities and led to the final reduced model illustrated below:

$$V_{ij} = (b_{0i} + \beta_0) + \beta_1 G_i + \beta_2 H_i + \beta_3 F_i + \beta_4 C_i + (b_{1i} + \beta_5) T_{ij} + (b_{2i} + \beta_6) T_{ij}^2 + \beta_7 G_i * T_{ij} + \epsilon_{ij}$$

where $i = 1, 2, ..., 121$, $j = 1, 2, 3, 4$, V_{ij}=Gait velocity of patient i at Test session j, b_{0i}, b_{1i}, b_{2i}, and ϵ_{ij} are the random patient intercepts, linear slopes, quadratic slopes and error term respectively.

Tables 4, 5 and 6 show the model parameter estimates, standard errors and pvalues

obtained from the three different Gait velocity measurements. All results from different missingness mechanisms were compared to each other and summarised on the Tables as sensitivity analysis tools. The different assumptions being tested included MCAR (for LMM under CCA analysis), MAR1 (for LMM under DL analysis), MAR2 (for LMM under MI), MNAR1 (for PMM with Shift-parameter approach) and MNAR2 (for PMM under CCMV approach). With respect to the research question, some insignificant effects were kept including training group, freezing and interaction term between training group and time.

Examining these results, it was observed that the estimates for the corresponding parameters are comparable and their numerical values were indeed very close to each other. It can be seen from the analyses that the associated p-values for the fixed effect assessments of group, freezing and interaction between group and time were all insignificant, their p-values being all greater than 5%. However, the associated p-values for the intercepts, Hoehn & Yahr, Time and Time^2 were significant and an additional significant Country effect was realised with V3 (pvalues < 0.05). The parameter estimates for Hoehn & Yahr based on CCA and DL are almost twice those obtained from MI which is in conformity to what was discussed by Allison and Schafer [17, 35]. In addition to this the variance components estimates did not differ much from each other. As observed from the EDA, estimates for the between variance were much higher compared to estimates for the within variance (for V2: $\text{Var}(b_{0i}) + \text{Var}(b_{1i}) + \text{Var}(b_{2i})$=332.6 » $\text{Var}(\epsilon)$=87.08). In summary, the results and inference based on the different assumptions did not change significantly and as a consequent give more weight to conclusions under the MAR assumptions which was deemed plausible for the data. The assumed normal distribution for the error terms and model fit of the data was explored through some diagnostic plots (Figure 19 Appendix A). The plots based on studentized residuals confirmed the fit and assumptions were fulfilled.

Table 4: *Model parameter estimates(standard errors) and pvalues under the MCAR, MAR and MNAR assumptions for V2*

	MCAR		**MAR1**		**MAR2**		**MNAR1**		**MNAR2**	
Effects	**Est(se)**	**P value**	**Est(se)**	**P value**	**Est(se)**	**P value**	**Est(se)**	**P value**	**Est(se)**	**P value**
Intercept	94.15(3.02)	<.0001*	96.14(2.92)	<.0001*	97.47(3.03)	<.0001*	96.77(2.99)	<.0001*	97.61(3.02)	<.0001*
Group	1.96(3.47)	0.5726	0.74(3.31)	0.8230	-1.09(3.42)	0.0750	-0.67(3.37)	0.8437	-1.17(3.42)	0.7333
Hoehn & Yahr	-8.18(1.82)	<.0001*	-8.28(1.75)	<.0001*	-14.14(3.73)	0.0002*	-13.34(3.67)	0.0003*	-14.24(3.71)	0.0001*
Freezing	-3.51(3.30)	0.2888	-5.17(3.15)	0.1035	-4.63(3.24)	0.1523	-4.09(3.20)	0.2011	-4.77(3.22)	0.1387
Time	7.48(1.25)	<.0001*	7.47(1.21)	<.0001*	7.84(1.24)	<0.0001*	7.06(1.29)	<.0001*	7.70(1.23)	<.0001*
Time^2	-1.28(0.25)	<.0001*	-1.28(0.24)	<.0001*	-1.34(0.24)	<0.0001*	-1.26(0.25)	<.0001*	-1.33(0.24)	<.0001*
Group*Time	0.60(0.78)	0.4444	0.63(0.77)	0.4147	0.75(0.79)	0.3451	0.87(0.84)	0.2969	0.88(0.79)	0.2646
Variance Components										
Var(b0i)	314.47		315.43							
Cov(boi,b1i)	-48.75		-48.35							
Var(b1i)	19.50		17.92							
Cov(boi,b2i)	7.27		7.02							
Cov(b1i,b2i)	0.08		0.31							
Var(b2i)	-0.73		-0.75							
Var(ϵ)	88.18		87.08							

*=significant at 5%, Est=estimate, se=standard error, Var=variance, Cov=covariance

Results from V2 based on MAR1, show a positive effect of interaction between Time

and Training Groups keeping other variables constant. This means the rate of change of Gait velocity for patients in IDT is more compared to patients in CTT but the difference was insignificant (Est=0.63, pvalue=0.4147). On the other hand, there was a significant common positive effect of Time in the Training Groups. Hence, the Gait velocity in both training groups was increasing significantly (Est=7.47, pvalue<.0001). Though there was a significant increasing Gait velocity, there was a subsequent significant drop at some point in time of the study. This was because there was a significant downward curvature for quadratic Time (Est=-1.28, pvalue<.0001). Similar trend of Gait velocity evolution was observed with results from V1 and V3 in Tables 5 and 6 respectively. Comparing the results from the three Gait velocities, it was realised that patients increased their velocities fastest when performing walking and trained Digit Span (WDS) compared to the other dual-tasks. Walking and Mobile phone tasks (WMT) resulted in the slowest increase in patients' velocities. Among all three dual-tasking procedures tested, walking and Stroop task (WST) had the steepest drop in Gait velocity while WMT had the slowest drop.

Table 5: *Model parameter estimates (standard errors) and pvalues under the MCAR, MAR and MNAR assumptions for V1*

	MCAR		**MAR1**		**MAR2**		**MNAR1**		**MNAR2**	
Effects	**Est(se)**	**P value**	**Est(se)**	**P value**	**Est(se)**	**Pvalue**	**Est(se)**	**P value**	**Est(se)**	**P value**
Intercept	92.74(3.18)	<.0001*	94.89(2.91)	<.0001*	96.36(3.02)	<.0001*	95.67(3.03)	<.0001*	96.45(3.01)	<.0001*
Group	1.54(3.57)	0.6667	0.14(3.29)	0.9672	-2.03(3.40)	0.5497	-1.44(3.38)	0.6709	-2.04(3.38)	0.5453
Hoehn & Yahr	-7.81(1.84)	<.0001*	-7.38(1.74)	<.0001*	-14.04(3.73)	0.0002*	-13.14(3.77)	0.0005*	-14.02(3.72)	0.0002*
Freezing	-2.79(3.44)	0.4184	-5.30(3.18)	0.0976	-4.45(3.22)	0.1669	-3.77(3.27)	0.248	-4.53(3.23)	0.1606
Time	7.94(1.33)	<.0001*	7.59(1.26)	<.0001*	8.03(1.29)	<0.0001*	6.54(1.40)	<.0001*	8.02(1.29)	<.0001*
Time^2	-1.24(0.27)	<.0001*	-1.18(0.25)	<.0001*	-1.26(0.26)	<0.0001*	-1.09(0.27)	<.0001*	-1.27(0.26)	<.0001*
Group*Time	-0.04(0.73)	0.9570	0.02(0.72)	0.9809	0.27(0.75)	0.7198	0.41(0.82)	0.6155	0.29(0.74)	0.6987
Variance Components										
Var(b0i)	325.62		322.47							
Cov(boi,b1i)	-59.51		-61.35							
Var(b1i)	47.79		46.54							
Cov(boi,b2i)	10.22		10.52							
Cov(b1i,b2i)	-6.51		-6.38							
Var(b2i)	0.7		0.72							
Var(ϵ)	81.48		78.59							

*=significant at 5%, Est=estimate, se=standard error, Var=variance, Cov=covariance

Controlling for other factors in the model, Hoehn and Yahr stage III (H3) patients did worse compared to stage II (H2) patients in terms of Gait velocities gotten from the different dual-task performances. H3 patients on average had significantly less Gait velocities compared to stage II patients during the study period. Results from WST, WDS and WPT testing procedures showed that the Gait velocities for H3 patients were respectively 8.28cm/s, 7.38 cm/s and 7.41cm/s less than those of H2 patients. In the same scenario, freezers of gait experienced less Gait velocities compared to non-freezers though the difference was not significant statistically from all three approaches of dual practices.

Table 6: *Model parameter estimates (standard errors) and pvalues under the MCAR, MAR and MNAR assumptions for V3*

	MCAR		MAR1		MAR2		MNAR1		MNAR2	
Effects	**Est(se)**	**P value**	**Est(se)**	**P value**	**Est(se)**	**P value**	**Est(se)**	**P value**	**Est(se)**	**P value**
Intercept	76.77(3.72)	<.0001*	78.29(3.61)	<.0001*	79.49(3.64)	<.0001*	78.65(3.66)	<.0001*	79.46(3.65)	<.0001*
Group	1.51(3.78)	0.6905	0.36(3.68)	0.9221	-1.44(3.71)	0.6975	-1.20(3.73)	0.7479	-1.51(3.72)	0.6851
Hoehn & Yahr	-8.15(1.93)	<.0001*	-7.41(1.88)	<.0001*	-14.49(4.22)	0.0006*	-13.85(4.23)	0.0011*	-14.39(4.23)	0.0007*
Freezing	-2.38(3.65)	0.5157	-4.00(3.59)	0.2664	-2.85(3.66)	0.4363	-2.71(3.68)	0.4614	-2.88(3.67)	0.4325
Country	9.63(3.59)	0.0085*	9.76(3.52)	0.0065*	10.05(3.51)	0.0042*	11.05(3.52)	0.0017*	10.13(3.52)	0.0040*
Time	5.92(1.24)	<.0001*	5.61(1.23)	<.0001*	5.91(1.27	<0.0001*	5.27(1.30)	<.0001*	5.99(1.24)	<.0001*
Time^2	-0.87(0.28)	0.0021*	-0.80(0.28)	0.0039*	-0.85(0.28)	0.0023*	-0.77(0.29)	0.0078*	-0.88(0.28)	0.0016*
Group*Time	0.27(0.68)	0.6900	0.23(0.67)	0.7285	0.35(0.69)	0.6108	0.22(0.72)	0.7540	0.35(0.70)	0.6125
Variance Components										
Var(b0i)	318.64		333.79							
Var(ϵ)	108.42		109.23							

*=significant at 5%, Est=estimate, se=standard error, Var=variance

Since the study took place at two different trial centers (Belgium=BE and Netherlands=NL) with the hope of patients being exposed to same conditions of treatment, country was not expected to have significant effect. On the contrary, results from WMT testing procedure depicted that patients from Belgium on average portrayed a significant 9.76cm/s higher Gait velocity than patients from the Netherlands which remained constant over time.

Further analysis was carried on with respect to some specific questions of interest as stated by the hypothesis. Based on that, the above model was reparameterized and the insignificant interaction term was removed. Due to some computational problems realised from the output in SAS, the *nobound* option was used alongside the *PROC MIXED* statement and the random effects for linear and quadratic time were omitted to have a positive definite variance component in the implied marginal model [5,32,43]. The *nobound* option actually helped to remove constraints levied on some variance components (such as negative variance for random slope realised). In a related note, removing some higher order terms for the random effects solved the problem of "infinite likelihood" realised from the analysis. Added to this, Verbeke and Molenberghs [45] stated that different hierarchical models can lead to the same marginal model but the reverse is not true. This was realised from this analysis since estimates and inference did not change from the model that was not reparameterized. The final model was thus expressed as follows:

$$V_{ij} = \beta_0 + \beta_1 G_i + \beta_2 T_{i2} + \beta_3 T_{i3} + \beta_4 T_{i4} + \beta_5 H_i + \beta_6 F_i + \beta_7 C_i + \beta_8 T_{ij}^2 + b_{0i} + \epsilon_{ij}$$

where β_3, β_4, and $\beta_3 - \beta_2$ are the differences in estimated Gait velocities between time points T1 vs T3, T1 vs T4 and T2 vs T3 respectively. These differences were common in both intervention groups since there was no significant interaction with time points. Figure 7 shows the mean predicted Gait velocity for V2. From the plot, highest value (104cm/s) was reached at test session 2 (T3) which was the period of 6 weeks after training practices. Similar trend was observed for V1 and V3 (Figure 20 Appendix A). Results in Table 7 below shows a summary of Gait velocity differences at the test sessions

for V1, V2 and V3. Details of the results are found in Table 22 Appendix B.

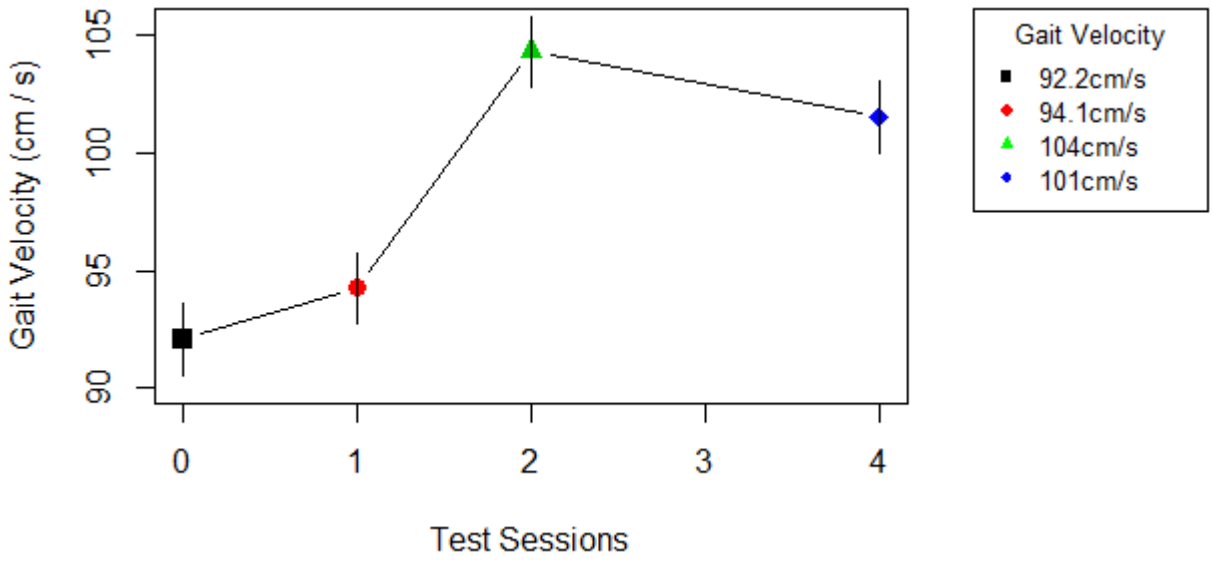

Figure 7: Predicted Mean Profile Plot for V2

Table 7: *Gait velocity comparison at each time point*

	V1		**V2**		**V3**	
T	**Est(se)**	**P value**	**Est(se)**	**P value**	**Est(se)**	**P value**
T1 vs T3	11.45(1.34)	<.0001	11.78(1.32)	<.0001	8.94(1.37)	<.0001
T1 vs T4	10.40(1.34)	<.0001	8.79(1.34)	<.0001	9.18(1.39)	<.0001
T2 vs T3	8.31(1.35)	<.0001	9.95(1.33)	<.0001	6.39(1.39)	<.0001

T=Test session, V=Gait velocity, Est=Estimate, se=standard errors

Results from Table 7 showed highest difference in Gait velocity between T1 (0 week before training practice) and T3 (after 6 weeks of training practice). There was a significant overall increase in Gait velocity after intervention of dual-task practice compared to baseline Gait velocities before patients were randomized. This implies the dual-tasking practices in general improved the dual-task performances of the PD patients and can easily be extended to tasks that were not trained. Although patients who were in the IDT training group performed better than those in the CTT training group in terms of Gait velocity, the difference was not significant.

4.2 Safety

4.2.1 Exploratory Data Analysis

To get a view of the distribution of the Fall counts, frequency plots were produced per Training group (Figure 8) . The figures suggest the distributions are strongly skewed with a mode of zero counts. This suggests that regression of Fall counts will rarely meet the distributional assumptions of LMM. Moreover, Fall counts will also violate the equal variances assumptions of linear models, as higher values of Fall counts will certainly have higher variances [3, 13].

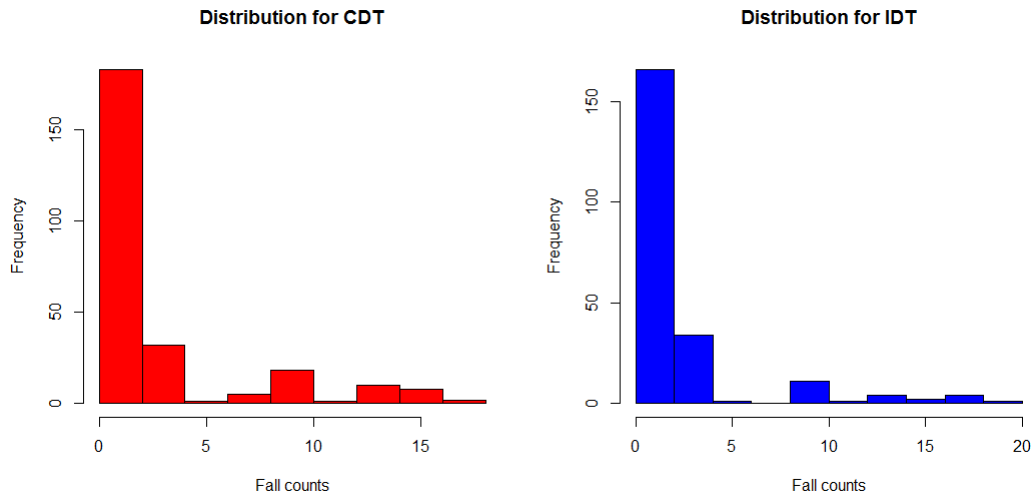

Figure 8: *Distribution of the number of Falls per training group*

Considering how the Fall counts are related to the Training group of patients, the mean Fall counts alongside the corresponding variances at each time point were computed and presented in Table 8. CTT on average had higher number of falls than IDT which experienced a slight drop from 1.49 to 1.36 whereas IDT experienced an increase in the number of fall counts from 0.75 to 0.92 on average. The trend of evolution between the training groups suggested a fairly constant evolution (Figure 9). Comparing the mean and the variance, it was realized that the variances were higher compared to the means hence suggesting the presence of overdispersion. This is a very important issue in count data modeling and needs to be taken into account by fitting a parsimonious model that can accommodate such.

Table 8: *Average fall counts(variance) at each time point*

	CTT			IDT		
T	**N**	**n**	**mean(var)**	**N**	**n**	**mean(var)**
1	65	65	1.49(19.69)	56	55	0.75(4.16)
2	65	62	1.24(12.22)	56	53	0.70(2.33)
3	65	61	1.15(10.63)	56	51	0.57(2.85)
4	65	61	1.36(10.83)	56	50	0.92(9.42)

N=number of patients at beginning of study,
n=number of patients remaining at each test session, T=test session

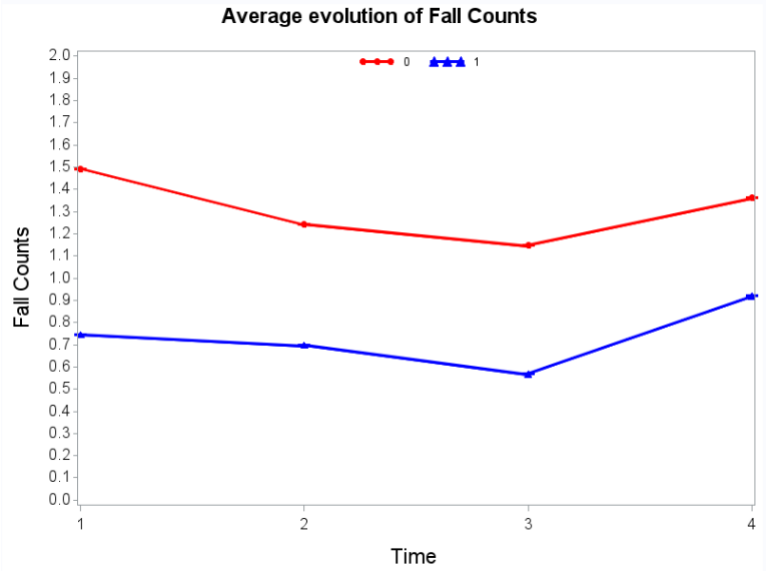

Figure 9: *Average evolution of the number of Falls per Group (O=CTT, 1=IDT)*

Being a longitudinal data involving patients, missingness is paramount in such study and needs to be explored for proper modeling. The presence of missing Fall counts was explored and with the results from the dropout model (Table 9), it was realised that there was a plausibility of valid analysis under MCAR or in the broader sense MAR. This was because all effects including previous fall counts were unrelated to missing fall counts (all pvalues are bigger than 5%).

Table 9: *Dropout model for Fall count*

Effect	**Est**	**se**	**P value**
Intercept	2.85	3.28	0.3853
Group	-2.85	2.86	0.3187
Previous FC	-0.30	0.49	0.5340
Time	-1.56	0.99	0.1174
Hoenh & Yahr	-1.17	1.11	0.2906
Freezing	0.89	0.75	0.2328
Country	-1.16	0.82	0.1600
Group*Time	1.22	1.15	0.2862

Est=estimates, se=standard error

Table 10, shows the missingness pattern observed from the data. Missing data patterns were all due to dropouts. There were more missing data for patients in the IDT group (about 9%) compared to patients in the CTT group (about 6%). One patient had missing data all through the entire study for IDT group since Fall counts were not recorded for that patient.

Table 10: *Missingness pattern for fall counts*

	CTT						**IDT**					
Pattern	**T1**	**T2**	**T3**	**T4**	**n**	**%**	**T1**	**T2**	**T3**	**T4**	**n**	**%**
1	O	O	O	O	61	93.85	O	O	O	O	50	89.29
2	O	O	M	M	1	1.54	O	O	O	M	1	1.79
3	O	M	M	M	3	4.62	O	O	M	M	2	3.57
4							O	M	M	M	2	3.57
5							M	M	M	M	1	1.79

O=observed, M=missing

Although valid analysis can be done under the MAR assumptions as confirmed from the dropout model results, deviations of results under the MNAR assumptions was later on confirmed through fitting of two Pattern Mixture MNAR models as seen in subsequent analyses.

4.2.2 Statistical Analysis

Two competing count models were fitted with a loaded mean structure and UN correlation structure using GEE approach as illustrated in section 3.2.2 in the Methodology. The Compound symmetry structure was finally chosen since the EBSE and MBSE were closed

to each other compared to the independent structure (Table 11). The models with the UN structure did not converge. Interaction terms between group and other factors (except linear time) were dropped since they were not significant. All insignificant main effects were retained with respect to the research questions. The final model for fixed effects was expressed thus:

$$log[E(FC_{ij})] = \beta_0 + \beta_1 Group_i + \beta_2 Time_{ij} + \beta_3 Freezing_i + \beta_4 Hoehn\_Yahr_i + \beta_5 Country_i + \beta_6 Group_i * Time_{ij}$$

Plots of observed and fitted fall counts (Figure 10) from the two models did not show major difference in models fit to the observed data. Appropriate model fit statistics were computed (PM: *Pearson chisquare=5.50, BIC=1698.6, AIC=1619.7* ; NBM: *Pearson chisquare=1.27, BIC=1059.3, AIC=1026.3*). The *Pearson chisquare statistic* for NBM was very close to 1 compared to that of PM. In addition to this, the *AIC* and *BIC* values for NBM were much smaller compared to those of PM. This implies NBM fitted the data better than PM and was considered for subsequent analyses.

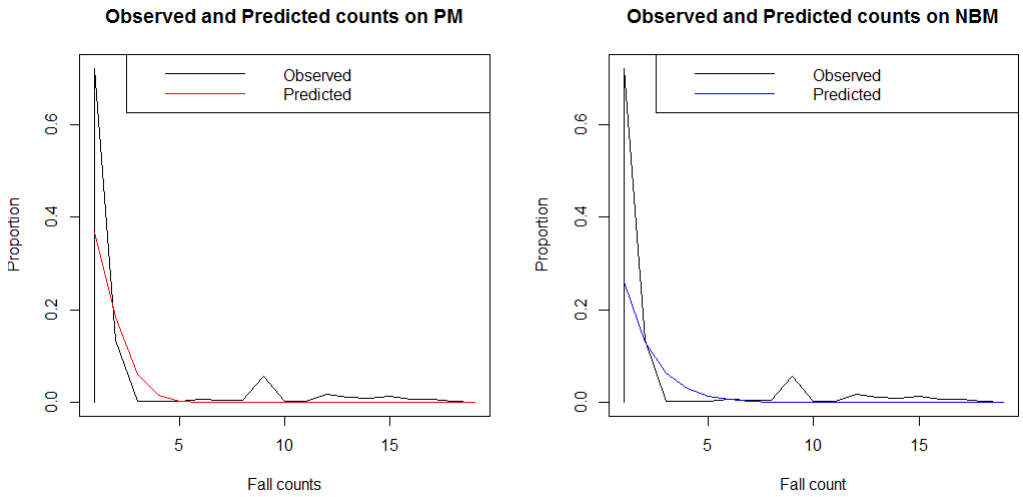

Figure 10: *Observed and predicted Fall counts from PM and NBM*

The presence of overdispersion in the data was also realised from the results since PM had a higher value for the scale parameter (2.39). This overdispersion was taken care of by the NBM because it had a scale parameter of 1.11, which is very close to 1.

Table 11: *Results from Poisson Model(PM) and Negative Binomial Model(NBM)*

	PM		**NBM**	
Effect	**Est(ebse,mbse)**	**P value**	**Est(ebse,mbse)**	**P value**
Intercept	-1.75(0.52,0.70)	<.0001	-1.59(0.49,0.50)	0.0012
Group	-0.75(0.55,0.51)	0.1751	-0.27(0.52,0.51)	0.6001
Hoehn & Yahr	0.21(0.18,0.23)	0.2283	0.17(0.17,0.26)	0.3126
Freezing	1.01(0.56,0.50)	0.0699	0.74(0.45,0.43)	0.1002
Time	-0.05(0.11,0.07)	0.6714	-0.03(0.12,0.09)	0.7661
Country	1.79(0.38,0.60)	<.0001	1.75(0.38,0.43)	<.0001
Group*Time	0.07(0.16,0.12)	0.6405	-0.06(0.17,0.13)	0.7068
Scale parameter	2.39		1.11	

Table 12 shows the results of the NBM fitted under the three missingness assumptions (MCAR, MAR, and MNAR) to assess the sensitivity of results obtained under the MAR assumption. Parameter estimates under the different assumptions are very close to each other. There is a significant effect for the intercept and country under MCAR and an additional significant effects of Freezing and Hoehn and Yahr under MAR and MNAR. Since MAR encompasses MCAR, the MNAR assumptions for the data did not have an influence in terms of inference for the assumed MAR model. Hence more trust can be given to the results under the MAR model for further inference on the research questions.

Table 12: *Results of the NBM under the MCAR, MAR and MNAR assumptions*

	MCAR		**MAR**		**MNAR1**		**MNAR2**	
Effect	**Est(se)**	**P value**	**Est(se)**	**P value**	**Est(se)**	**P value**	**Est(se)**	**P value**
Intercept	-1.59(0.49)	0.0012*	-1.54(0.41)	0.0002*	-1.39(0.41)	0.0007*	-1.54(0.41)	0.0002*
Group	-0.27(0.52)	0.6001	-0.22(0.51)	0.6667	-0.23(0.51)	0.6546	-0.21(0.51)	0.6813
Hoehn & Yahr	0.17(0.17)	0.3126	0.68(0.24)	0.0043*	0.64(0.24)	0.0088*	0.68(0.24)	0.0050*
Freezing	0.74(0.45)	0.1002	0.54(0.24)	0.0249*	0.47(0.24)	0.0559*	0.56(0.25)	0.0267*
Time	-0.03(0.12)	0.7661	0.02(0.12)	0.8569	0.05(0.12)	0.6884	0.02(0.12)	0.8575
Country	1.75(0.38)	<.0001*	1.44(0.25)	<.0001*	1.31(0.25)	<.0001*	1.43(0.25)	<.0001*
Group*Time	-0.06(0.17)	0.7068	-0.03(0.19)	0.8916	-0.03(0.19)	0.8573	-0.02(0.19)	0.9367

*=significant at 5%

The interaction term between the Training groups and Time was -0.03. In terms of incidence rate ratio this value is $e^{-0.03}$=0.97. On average, there is an increase in the number of falls in both Training groups, but patients of the IDT group had about 3% less in the rate of increase in number of falls compared to patients in the CTT group but the difference was not significant (Est=-0.03, IRR=0.97, pvalue=0.8916). Approximately, an IRR of 1.02 ($e^{0.02}$) for Time effect implies there was a weekly 2% increase in the number of Falls in each Trainig group which also was not significant (Pvalue=0.8569). Putting all other factors constant, at baseline, there were about 72% ($e^{0.54}$) more falls among freezers

of Gait compared to non-freezers and 97% ($e^{0.68}$) higher number of falls for Hoehn and Yahr stage III patients compared to stage II patients. Great difference in the number of patients that fell was realised between the patients from the different centres. PD patients from Belgium had on average 4 times ($e^{1.44}$) the number of falls compared to those from the Netherlands.

In order to assess the differences in the average fall counts at the different measurement periods, the interaction term between Time and Training group was dropped since it was insignificant and the model reparameterised. A summary of the results is found in Table 13.

Table 13: *Fall counts comparison at different time points*

T	**Est**	**IRR**	**Pvalue**
T1 vs T3	-0.24	0.79	0.4218
T1 vs T4	0.08	1.08	0.7961
T2 vs T3	-0.13	0.89	0.6705

T=test session

Conditioning on other important variables, moving from Time points 1 to 3 and 2 to 3 with incidence rate ratios of less than 1 implies there was a drop in the average number of falls of about 21% and 11% from both transitions respectively. On the contrary there was an increase of about 8% in the number of falls from Time point 1 to Time point 4. These changes in fall counts were not significant statistically and were common irrespective of the Training groups of patients. From the results, lowest number of falls were recorded at Time point 3 and this period represented the measurements just after the Training practices. In conclusison, the interventions or dual-task training practices had a positive effect in reducing the fall rate of PD patients though not greatly.

5 Conclusion and Recommendation

Dual-task practice involves the execution of a primary task as the main focus of attention and a secondary task executed simultaneously [33] but with different goals. In this study, PD patients of Hoehn and Yahr stages II and III, were randomised to two different dual-task training interventions (Integrated dual-task and Consecutive dual-task) with the aim of improving their dual-task performances.

Gait velocity and Fall counts were recorded from each patient at four test sessions. Though missing values were observed from patients in both training groups, the patients were analyzed as randomized in accordance with the ITT principle which produce valid results under MAR [26]. According to Wu and Hallett [47], patients with PD can diminish dual-task interference and improve performance if they practice such activities. Colleen et al. [9] confirmed fewer falls were reported among patients who did some exercises compared to those who did not. Similar results in improving gait performance and reducing fall rates were obtained from this study.

Results from this study confirmed that integrated dual-task training improved dual-task performance and led to less number of fallers compared to consecutive task training. This was realized because IDT patients had higher rate of increase of Gait velocity and lower rate of increase of fall compared to CTT patients though it was not a significant difference. It is worthy to note that, highest significant increase of Gait velocity of about 11cm/s was realized between measurements at the beginning of study and measurements six weeks after interventions. In addition to this, highest drop in number of falls of about 21% was envisaged between fall counts measurements at baseline and those realized after six weeks of training practices but the difference was not great.

In several activities of daily living, more than one task is executed at the same time. As a result of this, performing more than one task is an indispensable act to normal daily lifestyle. Though multitasking is a major challenge to patients with with PD, such can be enhanced through practice of such activities [47] which was confirmed from this study. This is because results from the trained dual-tasks (walking and digit span) were very similar and in accord to those realised from the untrained dual-tasks (walking and stroop, walking and mobile phone). As a consequent, the training exercises were considered efficient to be employed by PD patients to improve in their daily dual-task performances.

In this study, it was concluded that the training practices actually improved on the performance of the dual-task among the PD patients. Though integrated task seemed better than consecutive in improving the performance, the two approaches did not differ much from each other and can be employed to improve the dual-tasking abilities of PD patients. Gait velocity results involving walking and digit span actually increased the Gait velocity fastest compared to the walking and untrained Stroop and Mobile phone approaches. In addition to improved dual-task performances achieved, better lifestyle was also realised since the fall rate among the patients also dropped as a result of performing

the training practices.

In the light of the presence of missing responses common to most longitudinal clinical trials data, accounting for missing data through MI had one major drawbacks when the imputation and analysis models differ with respect to covariates used in the modeling [17,35]. This was the major challenge faced during fitting models involving MI in this project especially the MNAR Pattern Mixture Models since interaction terms in the analysis model could not be introduced in the imputation model. As part of sensitivity analysis, missing assumptions can be further addressed with the use of Selection Models methodology [25,26] to give additional confidence in the findings gained so far from the models already applied in this project. In addition to this, the presence of many zeros for fall counts, can be addressed by the use of Zero-inflated Negative Binomial Models [13] since it was also a challenge in fitting it for this project.

6 References

[1] Ali, A.M.G., Dawson, S.J., Blows, F.M., Provenzano, E., Ellis, I.O., Baglietto, L., Huntsman, D., Caldas, C. and Pharoah, P.D. (2011) Comparison of Methods for Handling Missing Data on Immunohistochemical Markers in Survival Analysis of Breast Cancer. *British Journal of Cancer*. 104. p.693-699.

[2] Allen, NE. Schwarzel, AK. (2013) Canning CG (2013) Recurrent falls in Parkinson's disease: a systematic review.*Parkinsons Dis* .906274.

[3] Atkins, DC. Gallop, RJ .(2007 Dec). Rethinking how family researchers model infrequent outcomes: a tutorial on count regression and zero-inflated models. *J Fam Psychol*; 21(4). p.726-35.

[4] Bowen, A. Wenman, R. Mickelborough, J. Foster, J. Hill, E. Tallis, R. (2001) Dual-task effects of talking while walking on velocity and balance following a stroke. *Age and Ageing*. 30. p.319-23.

[5] Brady, T. Kathleen, B. and Andrzej, T. (2007) *Linear Mixed Models. A Practical Guide Using Statistical Software*. Chapman and Hall.

[6] Carpenter, JR. Kenward, MG. (2008) *MAR methods for quantitative data. In: Missing data in randomised controlled trials— a practical guide*: National Institute for Health Research, Birmingham.

[7] Carpenter, J., Kenward, M.G., Evans, S. and White, I. (2004) Last Observation Carry-Forward and Last Observation Analysis. *Statistics in Medicine*.23. p.3241-3242. Available from: http://dx.doi.org/10.1002/sim.1891. [Accessed: July 30, 2015].

[8] Carpenter, J.R., Kenward, M.G. and Vansteelandt, S. (2006) A Comparison of Multiple Imputation and Doubly Robust Estimation for Analyses with Missing Data. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*.169. p.571-584. Available from: http://dx.doi.org/10.1111/j.1467-985X.2006.00407.x. [Accessed: August 01, 2015].

[9] Colleen, G. et al (2014) Exercise for falls prevention in Parkinson's disease. A randomized controlled trial. *Australian New Zealand Clinical Trials Registry. Neurology* **84**. p.1-9.

[10] Collins, L.M., Schafer, J.L. and Kam, C.M. (2001) A Comparison of Inclusive and Restrictive Missing-Data Strategies in Modern Missing-Data Procedures. *Psychological Methods*. 6. p.330-351. Available from: http://dx.doi.org/10.1037/1082-989X.6.4.330. [Accessed: July 10, 2015].

[11] Cook, J. Fordham, C. Cockburn, J. Haggard, P. (2003) Who knows best? Awareness of divided attention difficulty in a neurological rehabilitation setting. *Brain Inj*.17(7). p.561-74.

[12] Cookburn, J. Haggard, P. Cook, J. Fordham, C. (2003) Changing patterns of cognitive-motor interference (CMI) over time during recovery from stroke. *Clin Rehabil.* 17(2). p.167-73.

[13] Coxe, S. West, SG. Aiken, LS. (2009 Mar) The analysis of count data: a gentle introduction to poisson regression and its alternatives. *J Pers Assess.* 91(2). p.121-36.

[14] Fitts, PM. Posner, MI. (1967) *Human performance.* Belmont (CA): Brooks/Cole Publishing Company.

[15] Goldstein, H. Carpenter, J. Kenward, MG. Levin, K. *Multilevel models with multivariate mixed response types.* Stat modelling (in press).

[16] Haggard, P. Cockbum, J. Cock, J. Fordham, C. Wade, D. (2000). Interference between gait and cognitive task in a rehabilitating neurological population. *J Neurol Neurosurg Psychiatry.* 69. p.479-86.

[17] Joseph, L. (2003) Multiple Imputation in Multivariate Problems when the Imputation and Analysis Differ. *Statistics Neerlandica* vol.57, nr.1, pp.19-25.

[18] Keus, S. Bloem, B. Hendriks, E. Bredero-Cohen, A. Munneke, M. (2007) Evidence-Based Analysis of Physical Therapy in Parkinson's Disease with Recommendations for Practice and Research. *Movement Disorders.* 22(4). p.451–460.

[19] Ladewig, I. (2000) *A importância da atenção na aprendizagem de habilidades motoras.* Rev Paul Educ Fís. Suppl 3: S62-71.

[20] Lavori, P.W., Dawson, R. and Shera, D. (1995) A Multiple Imputation Strategy for Clinical Trials with Truncation of Patient Data. *Statistics in Medicine.* 14. p.1913-1925. Available from: http://dx.doi.org/10.1002/sim.4780141707. [Accessed: August 15, 2015].

[21] Lavori, P.W., Dawson, R. and Shera, D. (1995) A Multiple Imputation Strategy for Clinical Trials with Truncation of Patient Data. *Statistics in Medicine.* 14. p.1913-1925. Available from: http://dx.doi.org/10.1002/sim.4780141707. [Accessed: August 12, 2015].

[22] Liang, K.-Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrica*, **73**, p.13-22.

[23] Little, R.J.A. (1992) Regression with Missing X's: *A Review. Journal of the American Statistical Association.*87. p.1227-1237.

[24] Little, R.J.A. and Rubin, D.B. (1987) *Statistical Analysis with Missing Data.* John Wiley & Sons, New York.

[25] Molenberghs, G. Michiels, B. and Kenward, G. (1998) Pseudo-likelihood for combined selection and pattern mixture models for missing data problems. *Biometrical Journal*: **40**. p.557-572.

[26] Molenberghs, G. and Kenward, M.G. (2007) *Missing Data in Clinical Studies.* Chich-

ester: John Wiley & Sons, Inc.

[27] Musil, C.M., Warner, C.B., Yobas, P.K. and Jones, S.L. (2002) A Comparison of Imputation Techniques for Handling Missing Data. *Western Journal of Nursing Research*.24. p.815-829. Available from: http://dx.doi.org/10.1177/019394502762477004. [Accessed: August 13, 2015].

[28] Myers, W.R. (2000) Handling Missing Data in Clinical Trials: An Overview. *Drug Information Journal*. 34. p.525-533.

[29] Nakai, M., Chen, D.G., Nishimura, K. and Miyamoto, Y. (2014) Comparative Study of Four Methods in Missing Value Imputations under Missing Completely at Random Mechanism. *Open Journal of Statistics*. 4. p.27-37.

[30] National Institute of Neurological Disorders and Stroke (2015) [Online] Available from: http://www.ninds.nih.gov/disorders/ parkinsons_disease/parkinsons_disease.htm. [Accessed: July 20, 2015].

[31] National Parkinson's Foundation (2015) *Understanding Parkinson's*. [Online] Available from: http://www.parkinson.org/Parkinson-s-Disease/Living-Well/Safety-at-Home/Freezing. [Accessed: August 14, 2015].

[32] Nelder, J.A. (1954). The interpretation of negative components of variance. *Biometrika*. 41. p.544–548.

[33] O'Shea, S. Morris, ME. Iansek, R. (2002) Dual task interference during gait in people with Parkinson disease: effects of motor versus cognitive secondary tasks. *Phys Ther*. 82. p.888-97.

[34] Parkinson's Disease Health Center (2015). *Parkinson's Disease Topic Overview*. [Online] Available from: http://www.webmd.com/parkinsons-disease/tc/parkinsons-disease-topic-overview. [Accessed July 20, 2015].

[35] Paul, D. (2012) *Handling Missing Data by Maximum Likelihood, Statistical Horizons*, Harvard PA, USA, paper 312.

[36] Roberta de, M., Enio, W. Manoela, M. Julia, G. Alberto, C. Elizabeth, M.A (2010) Gait analysis comparing Parkinson's disease with healthy elderly subjects. *Arq Neuropsiquiatr*. 68(1). p.81-86

[37] Rubin, DB. (1978) Multiple imputations in sample surveys-A phenomenological Bayesian approach to nonresponse. *Imputation and Editing of Faulty or Missing Survey Data*. pp.1–23. Washington, DC: U.S. Department of Commerce.

[38] Rybolt, L. and McPartland, M. (2014) A comparison of exercise intervention to standard care in decreasing fall risk for patients with Parkinson's disease . *PT Critically Appraised Topics*. Paper 48.

[39] Schafer JL. (1997) *Analysis of incomplete multivariate data.* London: Chapman and Hall.

[40] Schmidt, RA. Wrisberg, CA. (2001) *Aprendizagem e performance motora: uma abordagem da aprendizagem baseada no problema.* 2nd ed. Porto Alegre: Artmed.

[41] Sprint, A. and Dupin-Sprint, T. (1993) Imperfect Data Analysis. *Drug Information Journal.*27. p.995-994.

[42] Sylwan, RP. Rosin, FM. Galera, C. (1999) Effect of practice and span length on the dual-task coordination executive test. *Braz J Med Biol Res.*32(10). p.1263-8.

[43] Verbeke, G. & Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data.* New York (USA): Springer.

[44] Verbeke, G. & Molenberghs, G. (2015) Londitudinal Data Analysis. Lecture notes: 2015-2016.

[45] Verbeke, G. & Molenberghs, G. (2015) *Introduction to Longitudinal Data Analysis.* Lecture notes: page 115.

[46] Videnovic, A. Marlin, C. Alibiglou, L. et al. (2013) Increased REM sleep without atonia in Parkinson disease with freezing of gait. *Neurology DOI: 10.1212/WNL.0b013e3182a4a408.* Available from: http://dx.doi.org/10.1212/WNL.0b013e3182a4a408. [Accessed: July 18, 2015].

[47] Wu, T. and Hallett, M. (2007) Neural correlates of dual-task performance in patients with Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry.*79. p.760-766.

[48] Yogev, G., Giladi, N., Peretz, C., Springer, S., Simon, E. S. and Hausdorff, J. M. (2005), Dual tasking, gait rhythmicity, and Parkinson's disease: Which aspects of gait are attention demanding?. *European Journal of Neuroscience.*22. p.1248–1256.

7 Appendix

7.1 A: Graphs

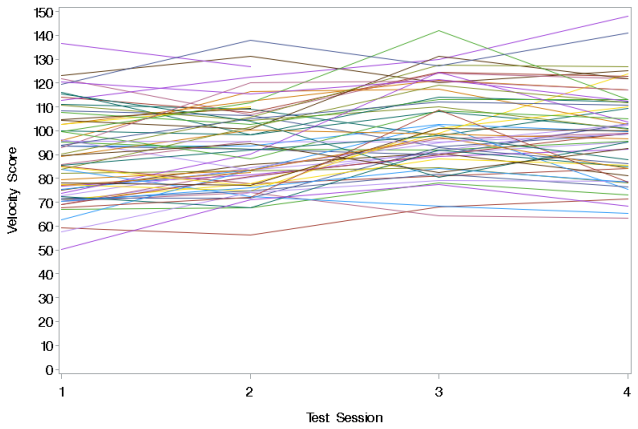

Figure 11: *Patient profile for CTT (V1)*

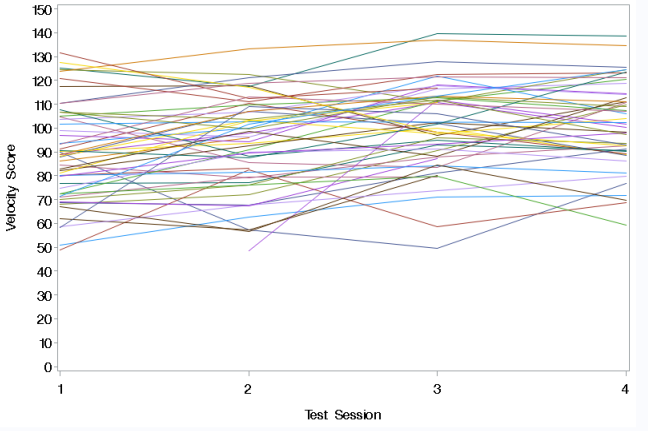

Figure 12: *Patient profile for IDT (V1)*

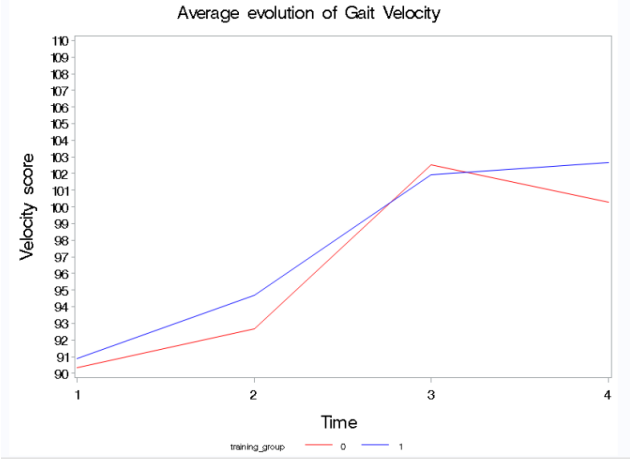

Figure 13: *Mean Evolution of V1*

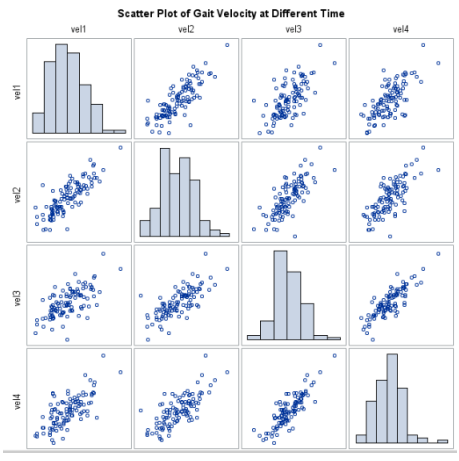

Figure 14: *Scatter Plot for V1*

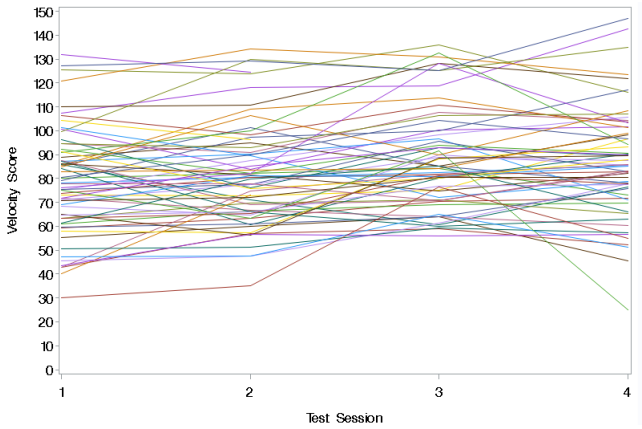

Figure 15: *Patient profile for CTT (V3)*

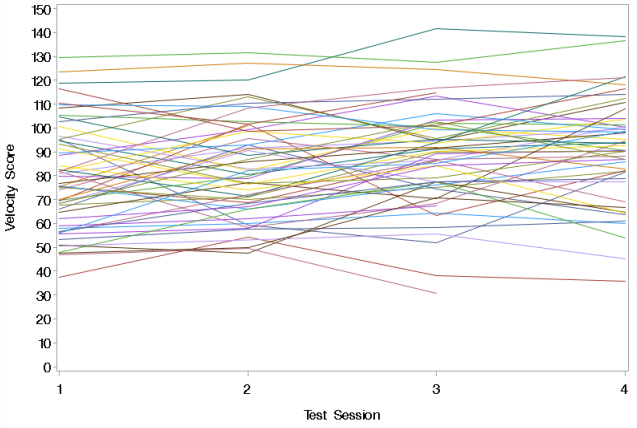

Figure 16: *Patient profile for IDT(V3)*

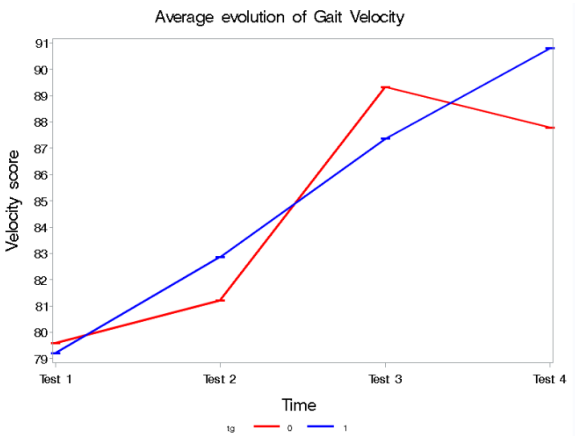

Figure 17: *Mean Evolution for V3*

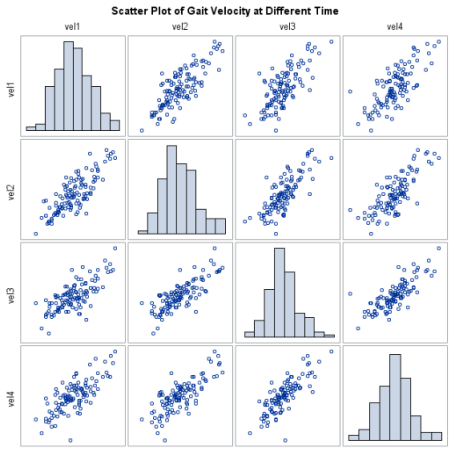

Figure 18: *Scatter Plot for V3*

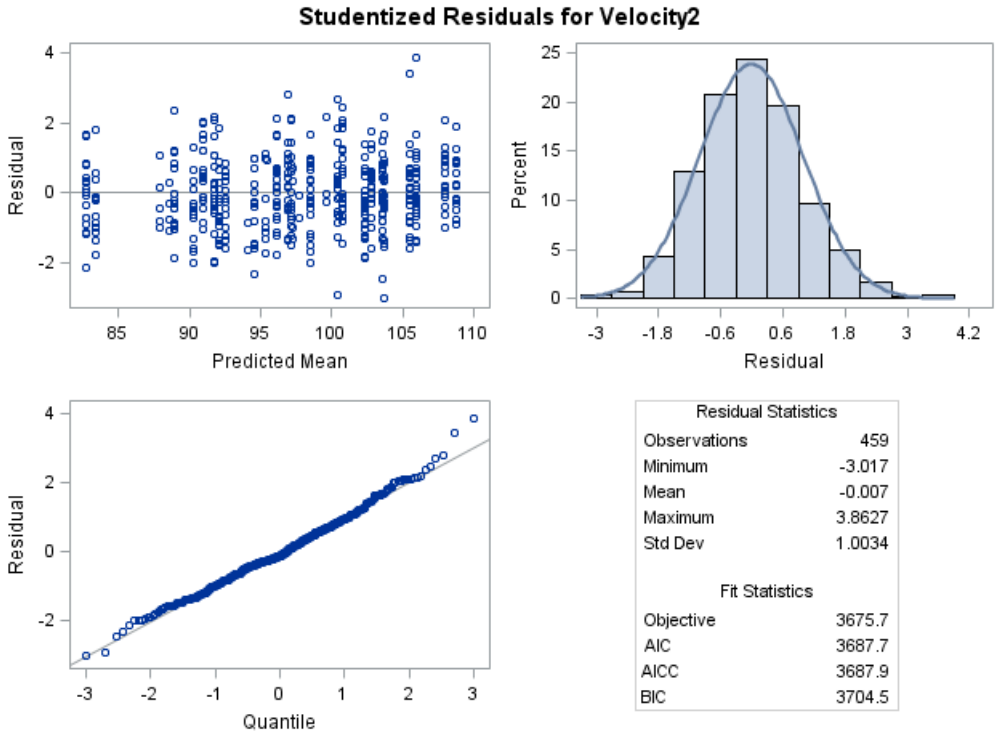

Figure 19: *Diagnostic plots for model fit and assumptions check for V2*

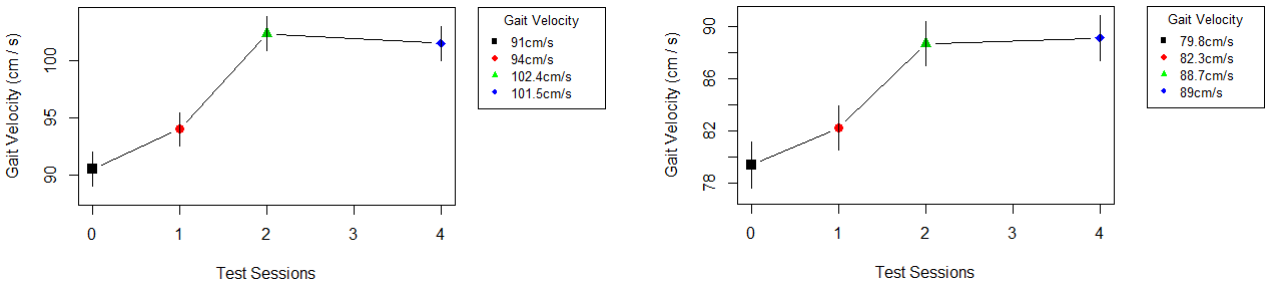

Figure 20: Predicted Mean Profile Plots for V1 (left) and V3 (right)

7.2 B: Tables

Table 14: *Mean Estimates(standard deviation) of V1 at the test sessions*

	CTT			IDT			BE			NL		
Test	**N**	**n**	**Mean(SD)**	**N**	**n**	**Mean(SD)**	**N**	**n**	**Mean(SD)**	**N**	**n**	**Mean(SD)**
1	65	65	90.32(19.14)	56	55	90.87(22.30)	62	62	91.64(19.25)	59	58	89.43(21.99)
2	65	60	92.68(17.72)	56	53	94.67(20.97)	62	62	94.08(17.94)	59	51	93.04(20.90)
3	65	55	102.54(20.19)	56	50	101.92(19.40)	62	58	101.85(21.33)	59	47	102.74(17.75)
4	65	57	100.27(20.87)	56	47	102.68(19.30)	62	59	102.76(20.91)	59	45	99.52(19.10)

N=number of patients at beginning of study,
n=number of patients remaining at each test session, T=test session

Table 15: *Pearson correlation between V1 values at the test sessions*

	vel1	**vel2**	**vel3**	**vel4**
vel1	1	0.81946	0.66823	0.70228
		<.0001	<.0001	<.0001
vel2		1	0.74812	0.74229
			<.0001	<.0001
vel3			1	0.84764
				<.0001
vel4				1

vel=Gait velocity

Table 16: *Missingness Pattern based on V1*

	CTT						IDT					
Pattern	**T1**	**T2**	**T3**	**T4**	**n**	**%**	**T1**	**T2**	**T3**	**T4**	**n**	**%**
1	O	O	O	O	53	81.54	O	O	O	O	43	76.79
2	O	O	M	M	5	7.69	O	O	O	M	5	8.93
3	O	M	M	M	3	4.62	O	O	M	M	2	3.57
4	O	O	M	O	2	3.08	O	M	M	M	2	3.57
5	O	M	O	O	2	3.08	O	O	M	O	2	3.57
6							O	M	O	O	1	1.79
7							M	O	O	O	1	1.79

M=missing, O=observed, n=number of patients

Table 17: *Mean(standard deviation) V3 per treatment group and country*

	CTT			IDT			BE			NL		
Test	**N**	**n**	**mean(sd)**	**N**	**n**	**mean(sd)**	**N**	**n**	**mean(sd)**	**N**	**n**	**mean(sd)**
1	65	65	79.59(22.17)	56	56	79.21(22.18)	62	62	84.35(20.54)	59	59	74.23(22.63)
2	65	63	81.22(21.56)	56	54	82.87(20.86)	62	62	87.00(20.86)	59	55	76.32(20.23)
3	65	61	89.33(22.76)	56	53	87.36(21.04)	62	60	91.20(23.24)	59	54	85.31(20.08)
4	65	61	87.79(24.33)	56	49	90.80(21.67)	62	59	92.63(22.98)	59	51	85.09(22.85)

N=number of patients at beginning of study,
n=number of patients remaining at each test session, T=test session

Table 18: *Pearson correlation based on V3*

	vel1	**vel2**	**vel3**	**vel4**
vel1	1	0.80714	0.72749	0.7486
		<.0001	<.0001	<.0001
vel2		1	0.79706	0.7813
				<.0001
vel3			1	0.79845
				<.0001
vel4				1

vel=Gait velocity

Table 19: *Missingness pattern based on V3 per treatment group*

	CTT						IDT					
Pattern	**T1**	**T2**	**T3**	**T4**	**n**	**%**	**T1**	**T2**	**T3**	**T4**	**n**	**%**
1	O	O	O	O	61	93.9	O	O	O	O	49	87.5
2	O	O	M	M	2	3.1	O	O	O	M	4	7.1
3	O	M	M	M	2	3.1	O	O	M	M	1	1.8
4							O	M	M	M	2	3.6

O=observed, M=missing, n=number of patients

Table 20: Results of Dropout models for Gait velocity

	V1		V2		V3	
Effects	**Est(se)**	**P value**	**Est(se)**	**P value**	**Est(se)**	**P value**
Intercept	-3.72(2.02)	0.0646	-3.10(3.82)	0.4175	0.09(3.40)	0.9796
Group	-2.57(1.42)	0.0705	-0.94(1.33)	0.4768	-1.63(1.53)	0.2853
Previous Velocity	0.01(0.01)	0.5040	0.03(0.02)	0.0942	0.01(0.02)	0.7456
Time	1.15(1.43)	0.4214	-0.85(1.83)	0.6420	-0.90(1.83)	0.6220
Hoehn & Yahr	-0.0004(0.74)	0.9996	-0.98(1.13)	0.3821	-1.24(1.12)	0.2672
Country	-1.78(0.67)	0.0076*	-1.27(0.70)	0.0695	-1.03(0.71)	0.1489
Freezing	0.03(0.54)	0.9553	0.92(0.66)	0.1596	0.63(0.66)	0.3408
$Time^2$	-0.37(0.30)	0.2222	0.11(0.36)	0.7659	0.03(0.38)	0.9363
Group*Time	1.22(0.64)	0.0582	0.53(0.50)	0.2910	1.07(0.75)	0.1528

*=significant at 5%, est=estimate, se=standard error

Table 21: *Results from model building*

Models	**V1**			**V2**			**V3**		
Random effects	$-2l_{REML}$	**ref**	**Pvalue**	$-2l_{REML}$	**ref**	**Pvalue**	$-2l_{REML}$	**ref**	**Pvalue**
Model with intercept, time, $time^2$ (M1)	3504.2		-	3651.3	-	-	3729.7	-	-
Model with intercept, time (M2)	3516.7	M1	0.0039	3660.5	M1	0.0184	3734.3	M1	0.1519
Model with intercept, $time^2$ (M3)	3522.9	M1	0.0002	3670.8	M1	0.0001	3736.1	M1	0.0672
Measurement error									
Model with serial correlation	3496.4			3650.2			3729.7		
Model without serial correlation	3504.2			3651.3			3729.5		

ref=reference

Table 22: *Results of Model Reparameterization for efficacy testing*

		V1		V2		V3	
Effect	**Parameter**	**Est(se)**	**P value**	**Est(se)**	**P value**	**Est(se)**	**P value**
Intercept	β_0	95.93(2.80)	<.0001	97.28(2.80)	<.0001	78.72(3.58)	<.0001
Group	β_1	0.18(3.08)	0.9542	1.38(3.08)	0.6539	0.72(3.5)	0.8365
Time2	β_2	3.14(1.30)	0.0164	1.83(1.31)	0.165	2.55(1.36)	0.061
Time3	β_3	11.45(1.34)	<.0001	11.78(1.32)	<.0001	8.94(1.37)	<.0001
Time4	β_4	10.40(1.34)	<.0001	8.79(1.34)	<.0001	9.18(1.39)	<.0001
Hoehn & Yahr	β_5	-8.45(1.77)	<.0001	-9.13(1.77)	<.0001	-7.48(1.86)	<.0001
Freezing	β_6	-4.92(3.16)	0.1217	-5.10(3.14)	0.1076	-3.95(3.58)	0.273
Country	β_7					9.80(3.52)	0.0062

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
A randomized comparison of integrated versus consecutive dual-task training in people with Parkinson''s disease: The duality trial

Richting: **Master of Statistics-Biostatistics**
Jaar: **2015**

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

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

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

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

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

Voor akkoord,

Dibato, John Epoh

Datum: **1/09/2015**