2014•2015
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Patterns in physical activity during a 5km start-to-run program for persons with Multiple Sclerosis

Supervisor :
Prof. dr. Anneleen VERHASSELT

Supervisor :
Prof. dr. PETER FEYS

Alfred Dusabimana
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.

KNOWLEDGE IN ACTION

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

2014•2015

FACULTY OF SCIENCES

Master of Statistics

Master's thesis

Patterns in physical activity during a 5km start-to-run program for persons with Multiple Sclerosis

Supervisor :
Prof. dr. Anneleen VERHASSELT

Supervisor :
Prof. dr. PETER FEYS

Alfred Dusabimana

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

Acknowledgement

To submit this master thesis is possible for the support of others. Therefore, I would like to get an opportunity for acknowledging them.

I would like to acknowledge my supervisor Prof. Dr. Anneleen VERHASSELT, Hasselt University, Belgium, for her extensive and continuous guidance and support during the preparation of this thesis.

I would also like to show my greatest gratitude to my co-supervisors, Prof. Dr. Peter Feys, Dr. Florian Van Halewyck, Hasselt University, Belgium and Paul Van Asch Move to sport project, Belgium, for their support and help.

To my family and those not mentioned by names, thank you for your unconditional support and encouragement.

Finally, I would like to acknowledge my Uncle Jean de Dieu DUSINGIZIMANA for giving me the opportunity to come to study in Hasselt University and for his support at any time in any way.

Abstract

Multiple sclerosis (MS) is an inflammatory disease in which the immune system attacks the protective myelin that covers the nerves of the central nervous system. In this study, 42 adults MS patients with mild symptoms entry were divided into two groups, an intervention group and control group. The intervention group performed a 12 week training program preparing them for a 5km run, whereas the control group continued their regular therapy. This thesis focused on investigating the evolution of the total number of steps per day and total active time per day during 12 weeks of the training periods in two groups. The total number of steps and total active time per day were recorded repeatedly across the training periods for each MS patient in the training. A linear mixed model was used in the analysis as the method that provides a general, flexible approach in the situations of the correlated data. The first autoregressive covariance structure was found to be plausible for the model of total number of steps per day and total active time per day outcomes. The training time and quadratic training time had a statistical significant effect on to total number of steps per day at the baseline. It has shown that an increase in training time increases the total number of steps per day at the baseline. The evolution of total number of steps per day over the training period in the intervention group was found to be the same as in control group. There was no evidence for the difference between the intervention group and the control group for any week of the training periods. For the total active time per day, the training time and quadratic training time had no significant effect at the baseline on the total active time per day. As expected, the participants in the intervention group were significantly more active per day compared to the participants of the control group.

Key Words: *Genaralized Linear Mixed Model (GLMM), Likelihood Ration Test(LRT), Linear Mixed Model(LMM), Multiple Sclerosis (MS), Restricted Maximum Likelihood(REML)*

Contents

1 INTRODUCTION

1.1 Background of the Study

Multiple Sclerosis (MS) is a chronic, often disabling disease that affects the central nervous system. When myelin is damaged or destroyed, the ability of nerves to conduct electrical impulses to the brain and from the brain is disrupted, thereby producing the various symptoms of MS such as sensory problems, difficulty walking and poor coordination (Janet and Johnston, 2008). Destruction of the nerve fibers themselves is believed to cause the permanent disability that many people with MS experience (National MS, 2009). More than 2 million people are estimated for having MS worldwide and the disease is among the most common causes of neurological disability in young adults (World Health Organization, 2008). The World Health Organization, (2008) reports that the proportion of new cases during a defined time period and the proportion of the population that has the disease at or during a specified time vary considerably between regions and populations. The median estimated prevalence of MS seemed to be the greatest in Europe (80 per 100 000), followed by the Eastern Mediterranean (14.9 per 100 000), the Americas (8.3 per 100 000), the Western Pacific (5 per 100 000), South-East Asia (2.8 per 100 000) and Africa (0.3 per 100 000) (World Health Organization, 2008). Results of meta-analyses suggest that the incidence of MS has increased over time and provide some evidence that this increase has occurred primarily in women (Koch-Henriksen and Sørensen, 2010). In Belgium, the prevalence in southern Flanders was 88 per 100 000 in 1991, for the study done on 74 men and 101 for women (Van Ooteghem *et al.*, 1995). In the survey study, Van Ooteghem *et al.*, (1995) and Pugliatti et al., (2006) had shown that the highest total rates were estimated for the age groups of 35–49 and 50–64 years, with women contributing mostly in the first and men in the second group. The distribution of prevalent cases by disability was 54%, 23% and 23% for mild, moderate and severe MS, respectively (Carton, 1997, Van Ooteghem et. al, 1995).

Individuals with the diagnosis of multiple sclerosis require a range of rehabilitation services due to the chronic and progressive nature of the disease (Medical Disability Advisor, 2012). The frequency and duration of rehabilitation is contingent upon the severity of the symptoms. Because this disease is chronic and progressive, individuals generally require rehabilitation at regular intervals throughout their lives. Physical activity has recently been considered as a rehabilitation

task known as physical therapy (Medical Disability Advisor, 2012), as it seems to slow down the progression of the MS among persons afflicted with this disease (Bennett, 2013).
Unfortunately people with MS have lower levels of physical activity than the general population (National MS, 2009). Patients with MS who are involved in physical exercise program are reported to experience an increased sense of well-being in addition to other benefits such as increased muscle strength and endurance, maintained and improved joint range of motion and flexibility, improved coordination, improved balance and reduced fatigue level (National MS, 2009 and Motl and Gosney, 2008). Exercise training is thus a promising strategy for increasing quality of life in the people with MS.

1.2 Withings Device

Withings Devices are useful in application for monitoring and tracking fitness-related metrics such as walked distance, calorie consumption, heart rate and quality of sleep (Withings Health Institute, 2014). When wore at the wrist, it captures number of steps via the arm swing and movement by knowing where you stand and how fast you are progressing, distance walked, elevation climbed and calories burned. These data empower users to make informed choices and generally motivate them to reach a target amount of physical activity per day.

1.3 Population Of the Study

In this study, 42 MS patients were tested for their walking capacity, physical fitness and cognitive functions and brain imaging before and after the training program. The inclusion criteria for the participants were being at least 18 years old at the beginning of the study, and being able to walk five kilometers without any aid, but have not be able to run 5km in the previous six months. Signed written informed consent was obtained prior to all tests. The current thesis focuses on the measurements collected during the training period between the pre and post-measurements.

1.4 Data Description

The present study was conducted upon on 42 MS patient participated in the study, 21 participants in the intervention group and 21 participants in control group. Among 42 participants in two groups, 26 participant received a withings device and had at least one measurement record. In this thesis, the participants with no records were not considered in the data used in the analysis. 20 participants in intervention group and 6 participants in the control group. Their level of physical activity were monitored in 12 weeks of the training. The withings device was able to track the number of steps taken each minute as well as per day and the total time that the user were active per day. Total active time per day was determined by successive hours that the participant was physically active per day and the total number of steps per day was determined by the overall steps that the participant made per day. Puyau *et al.*, (2004), and Colley, Janssen, and Tremblay, (2012) classified intensity levels of physical activity based on cut points corresponding to each intensity level: sedentary physical activity correspond to the observations that has at most 100 step counts per minute, light physical activity correspond to the observations range from 100 to 1500 step counts per minute, moderate physical activity correspond to the observations range from 1500 to 6500 step counts per minute, and finally the vigorous physical activity correspond to the observations that have more than 6500 counts per minute. While Puyau *et al.*, (2004), and Colley, Janssen, and Tremblay, (2012) classified intensity of physical activity based on minute, Tudor-Locke and Bassett, (2004) proposed a classification of physical activity based on day in adults. Tudor-Locke and Bassett, (2004) classified these intensities based cut point for each level < 5000 steps per day (sedentary), 5000-7499 steps per day as low active, 7500-9999 steps per day as somewhat active, 10000-12499 steps per day as active, and lastly ≤ 12500 steps per day as highly active.

On completion of the pre-measurements examination of the MS patients, eligible patients were asked to wear a withings device over the right hand side for 12 consecutive weeks. The withings device digitalized values are summed on a user-specified interval of 1 min, resulting in a counts value per minute as well as per hour and finally summarized per day. This thesis consisted of analyzing and comparing the daily physical activities profiles over the training period of intervention group and control group. Table 1 present proportion of group information by gender.

Group	Gender		Total	Percent
	Female	Male		
Contol	3	3	6	23.08
Intervention	19	1	20	76.92
Percent	84.62	15.38	26	

Table 1: *Frequency table for the intervention by gender*

From Table 1 nearly 77% participants are in the intervention group and 23% in control group respectively. 85% of participants were female and the male is about 15%. Table 2 present summary of the variables of interest.

Variable	Minimum	Maximum	Mean	Median	Std Dev
age	20	52	40.45	42	8.4
Time	1	140	54.59	53	25.07
Step	24	27228	7608.78	7024	4519.56
Activetime	60	29520	4945.57	4560	3069.42

Table 2: Summary variable of interest

The variable of interest were total number of steps per day which range from 24 to 27228 total number of steps per day, total active time per day is ranging from 60 to 29520 seconds per day, and age at the entry ranged from 20 to 52 years old. The measurement were recorded each day in 12 weeks. There were a total of 1495 number of measurements records on 26 participants. The number of measurements records per participants ranged between 7 and 96 with only one participant having just 7 measurements records.

1.5 Objectives

Scientific research suggests that physical activity can be a useful rehabilitation tool for individuals with MS. Therefore, the overall aim of this research project was to examine the effect of the 12 weeks of the training on physical cognitive and structural parameters in persons with MS. This was achieved by investigating whether the evolution of total number of steps per day and total active time over the training periods depends on the training group of the participant.

This was also achieved by evaluating whether the probability of being intensively active depends on the training group of specific participant. The research question was to investigate whether evolution of physical activities per day over the training period for the participant in the intervention group was equal as the control group.

2 METHODOLOGY

2.1 Linear Mixed Model

In the current study, the measurements for each participant were recorded repeatedly through the duration of the training. This was the source of the correlation among the measurements that was recorded for one participant. This may be due to grouping of repeated measurements on each participants over time. Mixed model analysis provides a general, flexible approach in these situations, because it allows a wide variety of correlation patterns or covariance structures to be explicitly modeled (Fitzmaurice *et al.*, 2004). The mixed model fitted on the current data were given by the equation 1

$$sqrtstep_{ij} = (\beta_0+b_{0i})+\beta_1*Age_i+\beta_2*gder_i+\beta_3*grp_i+(\beta_4*Age_i+\beta_5*grp_i+\beta_6+b_{1i})time_{ij}+\varepsilon_{ij} \tag{1}$$

where $i = 1, ..., 26$ and $j = 1, ..., n_i$ where n_i=total number of measurement records for the i^{th} participant. $sqrtstep_{ij}$ is the square root of total number of steps per day for the i^{th} participant on j^{th} occasion. Age_i is the age of the participant at the entry and $time_{ij}$ is the time for the i^{th} participant on j^{th} training day that measurement were recorded. β_0 is the average intercept, β_1 is the age effect, β_2 is the gender effect, β_3 is the effect of the training, β_4 is average slope describing the linear age time effect, and β_5 is average slope describing the linear training time effect. $\boldsymbol{b_i} = (b_{0i}, b_{1i})$ is a vector of participant-specific random effects (for intercept and slope respectively). $\varepsilon_{ij} \sim N(0, \sigma^2 I_n)$ and $\boldsymbol{b_i} \sim N(0, \mathbf{D})$ are both assumed to be independent. The total number of steps per day for one participants are assumed to be independent to the total number of steps per day of the other.

It is also for interest to evaluate whether the total number of steps per day has increased or decreased in first two weeks and second two weeks of the training compared to the last weeks of training, means $1^{st} - 2^{nd}$ week of the training, and $3^{rd} - 4^{th}$ week of the training compared to 5^{th} week towards the last week. To achieve this, the model that include the weeks that the measurement was recorded as covariate was fitted. This model was presented in equation 2.

$$\begin{aligned} sqrtstep_{ij} =& (\beta_0 + b_{0i}) + \beta_1 * Age_i + \beta_2 * gder_i + \beta_3 * grp_i + \beta_4 * week1_i + \beta_5 * week2_i \\ &+ (\beta_6 * week1_i + \beta_7 * week2_i + \beta_8 * Age_i) * grp_i + \varepsilon_{ij} \end{aligned} \quad (2)$$

where $sqrtstep_{ij}$ is the square root of total number of steps per day for the i^{th} participant on j^{th} training occasion. β_4 is the average effect of the first 2 weeks periods for control group, β_5 is the average of the 3^{rd} to the 4^{th} weeks effect for control group, β_6 and β_7 are the average effects of the training effect for the first 2 weeks periods and 3^{rd}, and 4^{th} training weeks effect respectively. The reference is the 5^{th} week to the other weeks and the control was considered as the reference group. The model 2 was fitted under the same assumptions as model 1.
The same model as model 1 with the same assumptions and the same variable description was fitted for the total active time per day where the $sqrtstep_{ij}$ were replaced by $sqrtactivetime_{ij}$. $sqrtactivetime_{ij}$ is the square root of total active time per day for the i^{th} participant on j^{th} occasion.
The maximum likelihood was used for likelihood ratio test and restricted maximum likelihood was used for estimation to get unbiased estimates (Molenberghs and Verbeke, 2009). The mixture of chi square was used as the distribution for the likelihood ratio test for the random effect, since hypothesized values of random effects are on the boundary of the parameter space. The square root of total number of steps and total active time per day were used to transform the residual near to the normality. The training time was transformed to log scale because the original scale have a harder time converging on estimates.

2.2 Generalized Linear Mixed Model (GLMM)

The total number of steps per day was dichotomized based on the cut point of the total number of steps higher that 9500 considered as physically active and less than 9500 steps per day considered as physically inactive (Tudor-Locke and Bassett, 2004). A random effects model as an approach for handling participant specific parameters $\mathbf{b}_i$ in longitudinal data analysis was fitted. The model conditions on both the predictor variables and participant specific effects (Verbeke and Molenberghs, 2005). GLMM as an extension of the linear mixed models for non-gaussian responses was

used due to the interest lies in describing the evolution of each participants separately. The random effects $\mathbf{b}_i$ reflect how much the subject-specific profile deviates from the overall average profile and can be used to detect profiles of individuals evolving differently over time (Agresti, 2007).
The choice of the main effects and the random effects used, follow from the Linear Mixed Model in equation 1. Equation 3 presents random intercept model fitted.

$$logit[(\hat{\pi}_{ij})] = (\beta_0 + b_{0i}) + \beta_1 * Age_i + \beta_2 * gder_i + \beta_3 * grp_i + (\beta_4 * Age_i + \beta_5 * grp_i + \beta_6 + b_{1i}) time_{ij} \quad (3)$$

Where:
$\hat{\pi}_{ij}$ is the probability for i^{th} participants for being intensively active (total number of steps per day >9500) at training time period j.

2.3 Blocked Case Re-sampling Analysis

Later on, the investigation for the efficacy and utility of the linear mixed model, theoretically and empirically will be carried on for this data using, blocked case re-sampling to generate the replicated data. Blocked case re-sampling is used in several longitudinal studies with a small sample size and a reasonable number of repeated measurements for each participant (Hyunsu, 2015). Due to the dependence in current data, generating replicated data will be done by blocked case re-sampling process where the block taken as the participant measurements records (Hyunsu, 2015). The blocked case re-sampling should be carried out in such a way that the dependence structure should be captured.

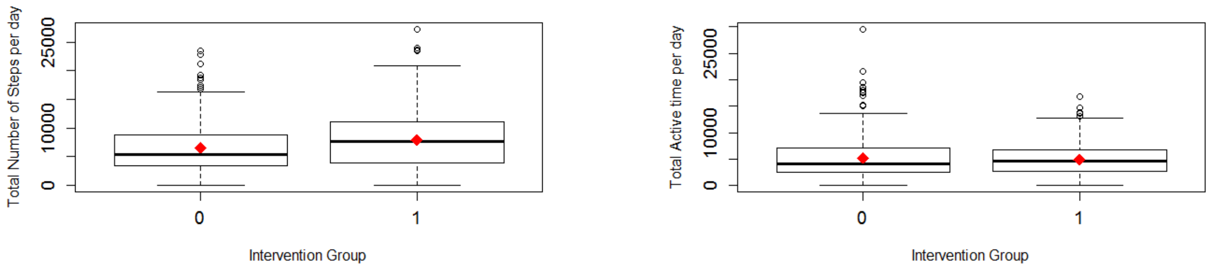
3 RESULTS

3.1 Exploratory Data Analysis (EDA)

For getting the insight on the data the exploratory data analysis was performed using simple descriptive statistics as well as graphical techniques for exploring the association between the outcomes and covariates of interest. This was intended to provide the data structure and the plausible implications that should be considered in modeling approach.

3.1.1 Boxplot of the Outcomes across the Training Groups

Figure 1 present the boxplot for the total number of steps per day and total active time per day across the training groups.



(a) (b)

Figure 1: *Boxplot of the total number of steps and total active time per day by the training group (a) and (b) the Red point indicate the mean in each group*

From Figure 1a and Figure 1b it can be observed that the outliers seem to be present on the total number of steps per day and the total active time per day respectively. The mean total number of steps per day and total active time for the intervention group are more or less the same as the control group.

3.1.2 Individual Profile

The individual profiles is a scatter plot of the response connecting the repeated measurements for each participant over time. It shows whether there is a discernable pattern common to most subjects. These individual profiles can also provide some information on between-subject variability.

Figure 2 present individual profile with smoothed average curve for the total number of steps per day Figure 2a and Total active time per day Figure 2b respectively.

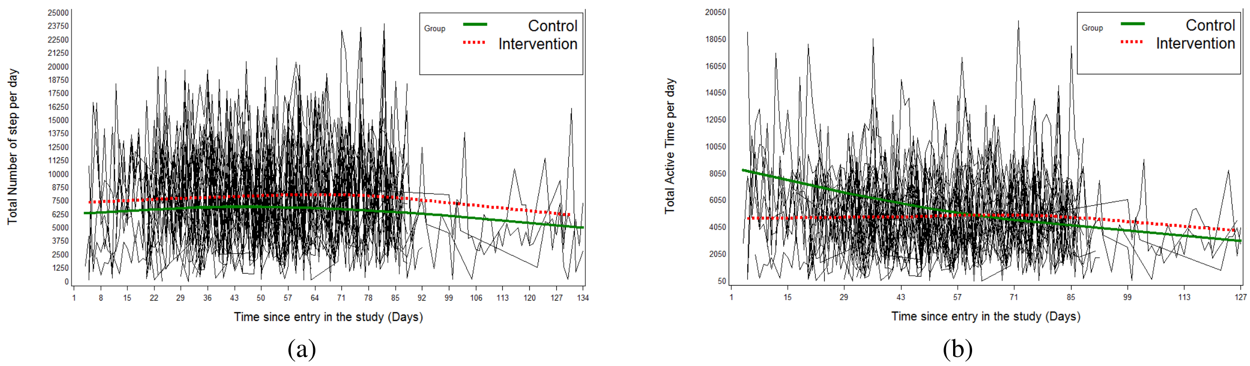
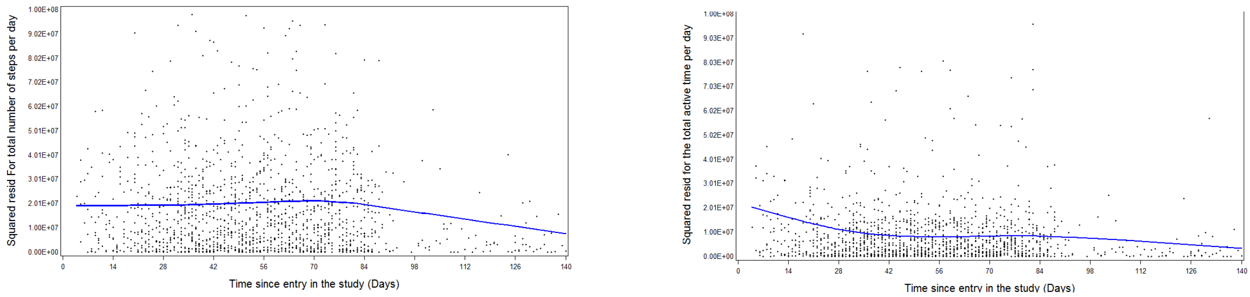


Figure 2: *Individual profile with smoothed average curve of total number of steps per day for control (solid line) and intervention group (dashed line) (a) and for total active time per day for control (solid line) and intervention group (dashed line) (b)*

From Figure 2, it can be observed that the evolution of total number of steps per day for the specific participant and total active time per day seemed to be different for participant to participant. The participants started with different level of physical activity and evolve differently. There seems to be a higher within-participant variability on the total number of steps per day. Figure 2a suggested that the average total number of steps per day for the participants in intervention group seemed to be higher compared to the control group and this average stayed high over time. Figure 2b suggested that the average total time of physical activity per day seems to be high in control group compared to the intervention group for early time, and at later time the average total active time per day seemed to be high in the intervention group compared to the control group. For total number of steps and total time in activity per day, the quadratic linear time effect can be assumed.

3.1.3 Variance Function

The covariance structure that is appropriate for the model is directly related to the component of variability which is dominant. After a candidate mean model is selected, fitting the model using ordinary least squares regression and examining the residuals may help to determine the appropriate covariance structure (Fitzmaurice *et al.*, 2004 and Molenberghs and Verbeke, 2009). For longitudinal data plotting the residuals by subject over time may indicate whether random intercepts and random slopes are needed (Molenberghs and Verbeke, 2009).



(a) (b)

Figure 3: *Smoothed average trend of squared ordinal least square residual for the total number of steps per day (a) and for total active time per day (b). The squared residual larger than* $1.00E8$ *are not shown*

Figure 3 present the smoothed average trend for squared ordinal least square residual to explore variance function over time for total number of steps and total active time per day. From Figure 3a and Figure 3b it can be observed that the variance function for the total number of steps and total active time per day seem to be less stationary. This suggest that the variability change over time and there are still some remaining systematic structure of variability that were not explained by the residual profile hence the inclusion of random effects for both outcomes should be reasonable to capture the remained unexplained variability.

Figure 6 in appendix present the smoothed average trend over the training period for the two outcomes by age quantiles. From Figure 6a it can be observed that at baseline, different age categories seems to be physically active in different way, it can be observed that the older participants at the baseline are more physically active per day compared to the younger participants but during the training periods the younger participants evolve so quickly while the slower evolution of physical activity for the older participants was observed.

3.2 Implication From EDA

In the current project, the information on participants in the study were collected for each participants across training periods. The distribution of the two outcomes was found to present some outliers. Since the repeated measurements within a participant are dependent, the data are correlated this was observed from the individual profile. The correlation among the measurement within participant violates the usual independence assumption. The mean structure of total number of steps per day over training time indicated that the evolution of the total number of steps per day for the participant in the intervention group is more or less the same as for the control group. The quadratic association between training time and physical activity per day was observed. We could assume that a line as well as quadratic curve are needed to characterize the mean response relation between training periods and the total number of steps per day. With regards to the between and within subjects variability, the individual profiles implied a need for random intercept and random slope. These findings imply that statistical models that takes into account the between and within patient variability seems to be more plausible.

3.3 Linear Mixed Model for the total number of steps per day

The random effects were tested before fitting the model to see if it is necessary in the model with total number of steps and total active time per day outcome. Verbeke and Molenberghs, (2000) emphasized that the need for random effect can not be tested by using likelihood ratio test, due to the fact that the null hypothesis are on the boundary of the parameter space which implies that the likelihood ratio test are not asymptotically a chi-square distribution. Stram and Lee, (1994) showed that the asymptotic null distribution for the likelihood ratio test of the random effect hypothesis is a mixture of chi-square distribution.

Table 3 present the likelihood ratio test for the random effect using the mixture of chi-square distribution for the model of total number of steps per day outcome.

Random Effect	Parameter	-2ℓ	G^2	Distribution	P-value
No random effect	10	13996.41			
Random Int	11	13748.8	247.61	$\chi^2_{0,1}$	$< .0001$

Table 3: *Likelihood Ratio Test for Random Effects For Total Number of Steps per day*

From Table 3 it can be seen that the random intercept is more likely to be present in the model. The Hessian matrix was not positive definite when the random slope included in the model of total number of steps per day outcome hence we considered that there was a minor important of random slope in this case. Evidently the random intercept is sufficient to capture all the variability between participant in the model of total number of steps per day outcome.

Table 4 presents likelihood ratio test for covariance structure of the model for total number of steps per day outcome.

Variance Structure	Parameter	-2ℓ	G^2	P-value
Random Int and Simple Structure	11	13748.8		
Random Int and AR(1) Structure	12	13746.6	2.2	0.1293

Table 4: *Likelihood Ratio Test for variance structure of the model with total number of steps per day*

The simple covariance structure assume the independent measurements within-participant, where the within-subject error correlation is zero and the first-order autoregressive covariance structure takes into account a common trend in measurements recorded over time where the correlation between observations is a function of the number of time points apart (Littell *et al.*, 1996). In first-order autoregressive structure, the correlation between adjacent measurements records is ρ.
From Table 4 it can be observed that the first-order autoregressive (AR(1)) covariance structure seems to be likely compared to the simple covariance structure. Hence the model for only random intercept with AR(1) covariance structure was considered in this case. After selecting the covariance structure and the random component of the model it should be of interest to investigate the main effect that explain much on the variability of physical activity per day. This was done by using the likelihood ratio test for testing the main effect.
Table 13 in Appendix present the main effect selection using likelihood ratio test. Age, age*training group interaction, and quadratic log time*training group interaction were shown less contribution in explaining the variability of the total number of steps per day hence removed in the model. The time, quadratic time, training group and the log time*training group interaction were shown to have much contribution in explaining the variability of the total number of steps per day hence selected using likelihood ratio test.
Table 5 present the parameter estimates of the selected model for the total number of steps per day outcome.

Effect	Estimate(Standard Error)	P-value
intercept	19.217(19.779)	0.3317
time	35.510(9.875)	0.0004
grp_1	-5.792(12.539)	0.6446
$time^2$	-5.089(1.323)	0.0001
time*grp_1	2.046(2.785)	0.4628
σ_{b0i}	188.780(58.893)	
ρ	-0.032(0.028)	
σ^2	567.120(20.953)	

Table 5: *Parameter estimate of the model for the total number of steps per day*

From Table 5 it can be seen that there is no significant effect of the training group for increasing or decreasing the total number of steps per day at the baseline. The evolution of the total number of steps per day was found to be the same in the intervention group as for the control group over the training period. The log training time and quadratic log training time effects were statistically significant on the total number of steps per day. The higher the number of days in the training the more the number of steps per day. One day increase in log training time for the participant in control group, there is an increase of about 1261 total number of steps per day known that other factors are kept constant. For investigating whether the total number of steps per day was increased during the first two weeks of the training or the second two weeks or the fifth and followed weeks, the model that include the week that the observation were recorded was fitted. The parameter estimate for this model can be found in Table 6

Effect	Estimate(Standard Error)	P-value
Intercept	75.565(6.049)	<.0001
grp_1	6.357(6.896)	0.3648
$week_{1-2}$	4.341(3.213)	0.1773
$week_{3-4}$	2.662(3.249)	0.413
grp_1*week $_{1-2}$	-5.838(3.668)	0.1121
grp_1*week $_{3-4}$	-2.641(3.697)	0.4753
σ_{b0i}	188.780(58.893)	
ρ	-0.034(0.028)	
σ^2	562.650(20.953)	

Table 6: *Parameter estimate for the model that included weeks training effect for the total number of steps per day*

From Table 6 it can be observed that average effects of the training for the first 2 weeks training periods and 3^{rd}, and 4^{th} weeks effect were statistically the same as the one for the 5^{th} week to the other weeks. There is no evidence for the difference of total number of steps per day for intervention group compared to the control group for any week training periods.

3.4 Linear Mixed Model for total active time per day

Table 7 present the likelihood ratio test of the random effects for the model of the total active time per day.

Random Effect	Parameter	-2ℓ	G^2	Distribution	P-value
No random effect	10	13383.5			
Random Int	11	13101.4	282.1	$\chi^2_{0,1}$	$< .0001$
Random Int + Slope	13	13100.1	1.3	$\chi^2_{1,2}$	0.3881

Table 7: *Likelihood Ratio Test of random effects for total active time per day*

From Table 7 it can be observed that the random intercept are more likely to be present in the model. There is no evidence for the presence of the random slope in the model.
Table 8 present likelihood ratio test for covariance structure of the model for total active time per day.

Variance Structure	Parameter	-2ℓ	G^2	P-value
Random Int and Simple Structure	11	13136.1		
Random Int and AR(1) Structure	12	13127.1	9	0.0027

Table 8: *Likelihood Ratio Test for covariance structure of the model for total number of steps per day*

From Table 8 it can be seen that the model with random intercept and AR(1) covariance structure is fitting the data well compared to the model with simple structure hence the AR(1) structure were considered in the analysis for the model of total active time per day outcome.
Table 14 in Appendix present the main effects selection using likelihood ratio test. Age, age*training group interaction, gender, and gender*training group interaction had shown to have less contribution in explaining the variability of the total active time per day for the data considered in this thesis, hence removed in the model. The linear log time, training group and the linear log time*training

group and quadratic log time*training group interactions were found to contribute much in explaining the variability for the total active time per day hence selected using likelihood ratio test. The parameter estimates for the selected model of total active time per day was presented in the Table 9.

Effect	Estimate(Standard Error)	P-value
intercept	134.640(34.538)	0.0005
grp_1	-119.120(31.572)	0.0003
time	-23.599(13.319)	0.0770
$time^2$	1.779(1.818)	0.3281
grp_1*time	50.061(17.306)	0.0040
grp_1*$time^2$	-5.327(2.383)	0.0258
$var(b_{0i})$	123.970(38.832)	
ρ	0.087(0.028)	
σ^2	368.480(13.786)	

Table 9: *Parameter estimate of model for the total active time per day*

Table 9 present the parameter estimate for the model of total active time per day outcome. From Table 9 it can be seen that there was no significant effect of log training time and quadratic log training time at the baseline on total active time per day. The increase in log training time decrease the total active time per day for the participant in control group. However this decrement is not statistically significant. At the baseline the participants in the control group were more active compared to the participants in the intervention group which is in line to what observed for the smoothed average trend in Figure 2b. There is a significant difference between the intervention group compared to the control group. For one day increase in log training time, there is approximately 2510 seconds higher in physical activities for the participant in the intervention group compared to the control group known that other factors are kept fixed. There was a significant quadratic log time training trend difference between the two groups. Evidently, this suggest initially large evolution difference in intervention group compared to control group were observed, followed by a leveling off of this effect over the training period.

3.5 The Generalized Linear Mixed Model

The generalized linear mixed effect model which reflect how much the subject-specific profile probability deviates from the overall average probability profile over time was fitted to the dichotomized outcome for being physical active or physical inactive. The covariates selected in linear mixed model for the total number of steps per day were adopted in generalized linear mixed model. Table 10 present the parameter estimate and the standard error for GLMM.

Effect	Estimate(Standard Error)	P-value
Intercept	-3.949 (2.171)	0.0814
time	2.058(1.087)	0.0586
$time^2$	-0.346(0.145)	0.0175
grp_1	-1.376(1.338)	0.304
$time*grp_1$	0.436(0.326)	0.1808
$var(b_{0i})$	1.015(0.358)	

Table 10: *The parameter estimate for the model of binary outcomes*

From Table 10 it can be observed that there is no much evidence of the training group effect on the probability of being physically active. At the baseline the odds ratio of the participants in the intervention group is about 75% less compared to the participants in the control group known that other factors are kept constant and with zero random intercept. There is no significant effect of the linear log training time effect for the probability of being active but for the quadratic log training time effect there is significant effect. This suggest initially a large effect on probability of being physically active is present, followed by a leveling off of this probability over the training period. The evolution of the probability of being physically active is the same for two groups for the participants with zero random intercepts.

3.6 Model Assessment

3.6.1 Residual Assessment

The residual in linear mixed model is assumed to follow the normal distribution. Using the histogram and quantile normality probability plot from Figure 7 and Figure 8 in appendix this assumption was not satisfied for the total number of steps per day and total active time per day.

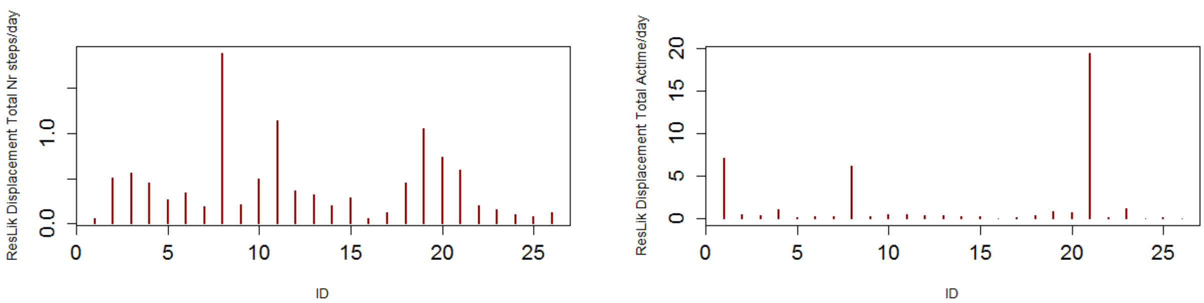
The alternative was to transform the response for getting the residual near to normal. The optimal transformation found was the square root of the total number of steps per day and total active time per day respectively.

3.6.2 Sensitivity Analysis

It is of interest to explore the model-data agreement. The sensitivity of the model is studied through measures that express its stability under perturbations, changes in the data or model components that should produce commensurate changes in the model output (Schabenberger, 2004). However Schabenberger, (2004) has discussed that the difficulty is to determine when the changes are substantive enough to warrant further investigation, possibly leading to a reformulation of the model or changes in the data, such as dropping outliers. In our case the influential analysis was used to determine which cases are influence and the manner in which they are important to the analysis.

Influence Analysis

Figure 4 present an overall influence statistics which measure the change in the restricted log likelihood being minimized for the total number of steps per day and total active time per day respectively. As the linear mixed model was fitted by restricted maximum likelihood (REML), an overall influence measure is the restricted likelihood distance referred to as restricted likelihood displacement which gives the amount by which the log-restricted likelihood of the full data changes if one were to evaluate it at the reduced-data estimates (Schabenberger, 2004).



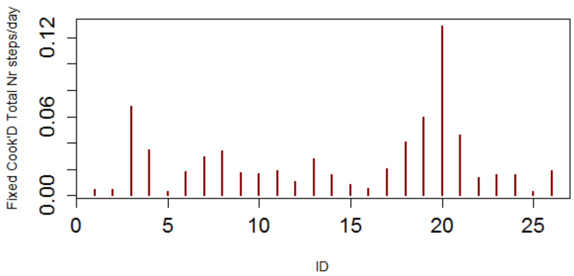
(a) (b)

Figure 4: *Overall influence statistics for measuring the change in restricted maximum likelihood for Total Number of Steps per day by leave-out subject (a) for Total Active Time per day (b).*

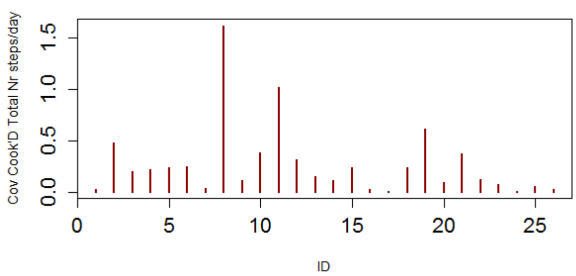
Figure 4 give the index plots of the overall influence measures on physical activities. Participant with ID 8, 11, and 19 stand out as the most overall influential compared to the others on the total number of steps per day outcome. The participant of ID 1, 8, and 21 were also seemed to be the overall influential on the total active time per day outcome.

Influence on Fixed Parameters

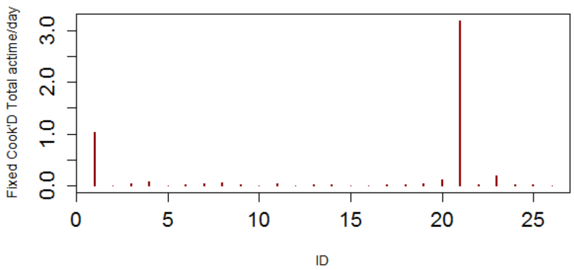
In investigating the influential case on fixed parameters, the more focus is in assessing the influence of the ith participant on the obtained parameter estimates, more generally, the combination of linearly independent of the parameters. Using restricted maximum likelihood estimates by omitting the ith observation, the REML estimate of $\hat{\beta}$ is $\hat{\beta_{(i)}}$. For assessing the influential observation, the generalized Cook's statistic as a measure of the influence of the ith observation on $\hat{\beta}$ by leave-out observation was used. For more details about generalized cook's distance we referred to Zewotir and Galpin, (2005) and Schabenberger, (2004). Figure 5 present index plot of the influential analysis resulted on the fixed and covariance parameters in leave-out participants analysis.



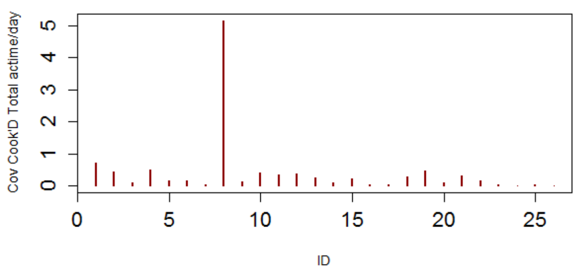
(a)



(b)



(c)



(d)

Figure 5: *Generalized Cooks's distance for measuring the change in fixed parameter for total number of steps per day by leave-out subject (a) and variance-covariance for total number of steps per day (b) Fixed parameter for total active time per day (c) and variance-covariance for total active time per day (d).*

From Figure 5a it can be observed that participant of ID 20 is the most influential observation on the fixed effects, and participant with ID 3, 19, and 21 regarded as somewhat less influential on the total number of steps per day outcome. Form Figure 5b, for the total number of steps per day covariance, the participant with ID 8 was observed to be the most influential and participant of ID 11 and 19 are seemed to be less influential. From Figure 5c it can be observed that the most influential on total active time per day for fixed effect parameters is participant of ID 21 and participants of ID 1 seems to be less influential. For covariance parameter of total active time per day, the only most influential case is the participant with ID 9. For the overall analyses, participant with ID 8 and 20 seems to have a large influence for the total number of steps per day. For the total active time per day, the participant with ID 21 have a large influence. For dealing with the observed influential cases on the restricted log likelihood function, the fixed parameters and covariance parameters the influential cases were left out in the data and then the same model was refitted on the reduced data. Overall, for the total number of steps per day, the influences were not found to have a potential effect on fixed and the covariance parameters. For the total active time per day the same process was done but for this outcome the participant of ID 21 was found to have a potential influence on the model parameters due to the model fitted on the reduced data, the quadratic log time and the quadratic log time*training group interaction were not found to be significant hence further investigation are needed to infer the findings on the fitted model. The parameter estimate for the model fitted on the reduced data can be found in Table 15 in Appendix

3.6.3 Blocked Case Re-sampling Study

The blocked case re-sampling study was conducted based upon 1000 replicates of data within each block. The original model were fitted for each sample and parameter estimate were estimated, and then look at the collection of values as the estimate of the overall sampling distribution of the real estimator. The underlying idea is to compare the results obtained from model fitted on the real data against the possible results if we re-label the data points, then see how extreme the results from the real data are, when compared against this alternatives arrangements of the data. We did it by re-sampling 1000 datasets (simple random sampling with replacement within subject repeated measurements) from the original sample. Typically for case re-sampling, we wanted to estimate the unbiased estimates and the standard error for the parameter. This case re-sampling was a non-parametric method, since we are not making specific assumptions about the distribution from which the data arise.

The case re-sampling results from 1000 data were analyzed separately and combined together to have common value based on the large dataset for the two outcomes. The parameter estimate were presented in Table 11 for the total number of steps per day and Table 12 for total active time per day. For the total number of steps per day outcome, it can be observed that the estimated parameters for the original data and the estimates obtained in case re-sampled data are more or less the same this suggest the stability of the results. For the total active time per day outcome there is slight difference to the estimated parameter for the original data compared to the case re-sampled data. The quadratic time and the quadratic log time-training group interaction which were significant in the original data were found to be not significant in the replicated data. This was not surprise because these parameters were also found to be non significant in leave out influential analysis. Due to the original parameter estimate were based on the small sample size and one participant can have an effect on the parameter estimates. These show how certain target parameter estimates are improved by upgrading the sample size in the re-sampling random samples within participant measurements.

Effect	Estimate(Standard Error)
intercept	18.816(24.855)
grp_1	-6.184(22.329)
time	35.497(12.613)*
$time^2$	-5.062(1.690)*
time*grp_1	2.055(3.529)
σ_{b0i}	187.505(60.794)
ρ	-0.034(0.046)
σ^2	567.555(26.492)

*:P-value < 0.05

Table 11: *Parameter estimate obtained for the combination of the blocked case re-sampled parameters for the total number of steps per day*

Effect	Estimate(Standard Error)
intercept	135.627(34.021)*
grp_1	-118.506(39.497)*
time	-24.123(16.474)
$time^2$	1.848(2.362)
time*grp_1	49.244(20.239)*
$time^2$*grp_1	-5.238(2.628)
σ_{b0i}	124.707(40.518)
ρ	0.091(0.042)
σ^2	368.542(17.399)

*:P-value < 0.05

Table 12: *Parameter estimate obtained for the combination of the blocked case re-sampled parameters for the total active time per day*

4 DISCUSSION AND CONCLUSION

This study aimed to evaluate change in total number of steps and total active time per day of the MS participants over the training period. In this study, 42 participants with mild MS symptoms received personal training program to be performed at home. Before and after this training program, 42 MS patients performed tests to measure the walking capacity, physical fitness and cognitive functions, as well as brain imaging scans to determine the effect of physical training. This thesis focused on the data records taken during the training periods by the withings device. Withings devices were able to monitor and track fitness-related metrics such as total number of steps per minute, and the total time that the user was physically active per day. The outcomes of interest thus were the total number of steps per day which ranged from 24 to 27228 steps per day and total active time per day ranging from 60 to 29520 seconds per day. The records for the participants in the study were collected repeatedly across different training periods, this was induced correlation among the repeated measurements within participant. There was a total of 1495 measurements records from 26 participants. The number of measurements per participants ranged between 7 and 96 with only one participant having just 7 measurements records. The linear average trend of the total number of steps per day over training period suggested equal evolution for the intervention and control group. There seemed to be a quadratic relation between time and total physical activity per day. The linear as well as quadratic trend were needed to characterize the mean response relation between training periods and the total physical activity per day. The individual profiles curve suggested the higher between and within participant variability.

Linear mixed model approach was used in the analysis as the method that allows a wide variety of correlation patterns to be explicitly modeled (Fitzmaurice *et al.*, 2004). Due to the high between and within participant variability, a random intercept and slope were was suggested in the model for the total number of steps per day and the total time of physical activity per day. The random intercept for the linear mixed models for both total number of steps and total time of physical activities per day was evidently present. The first autoregressive covariance structure was plausible for both models of total number of steps and total active time per day.

By using likelihood ratio test, Age, age*training group interaction, and quadratic log time*training interaction were shown to explain only a little part of the variability of the total number of steps per day. The log training time, quadratic log training time, training group and the log time*training group interaction were found to explain more on the variability of the total number of steps per day outcome and for the total time in physical activities per day added with quadratic log time*training interaction. There was no statistical evidence for showing an effect of the training group on the total number of steps per day at baseline. Interestingly the evolution of the total number of steps per day was found to be the same in the intervention group as in control group over the 12 weeks training period. The time and quadratic time was found to have a statistical significant effect on the total number of steps per day. It has shown that an increase in training periods increases the total number of steps per day. One day increase in log training period there was an increase of about 1261 total number of steps per day for the participant in the control known that other factors are kept constant. It was of interest to see if the total number of steps per day was high during the first 2 weeks of the training or the second 2 weeks or the fifth and followed weeks. It has observed that average effects of the training for the first 2 weeks periods and 3^{rd}, and the 4^{th} weeks effect were statistically the same as the one for the 5^{th} week to the followed weeks. There is no evidence for the difference in intervention group compared to control group for any week training periods. For the total active time per day outcome there were no significant effect of log training time, quadratic log training time on total active time per day at the baseline. The increase in log training time decrease the total active time per day for the participant in control group however this decrement was not statistically significant. There was a significant difference in total active time per day between the intervention group compared to the control group. The total active time per day for the participant in intervention group is about 2500 seconds higher compared to control group for one day increased in log training periods known that other factors are kept fixed. Quadratically there was a significant difference between the two groups.

The generalized linear mixed effect model which was fitted to the binary outcome for being physical intensively active or physical inactive. The participant training group had the significant effect on the probability of being physically active at the baseline. At the baseline the log odds ratio of the participants in intervention group is about 75% less compared to the participants in control group with zero random intercept. There is no significant effect of the linear log training time effect for the probability of being physically active, the same for the quadratic log training time effect. The evolution of the probability of being physically active is the same for two groups for the participants with zero random intercepts.

The influence analysis was performed to check how much the fitted mixed models were sensitive to peculiar observations for one participant that can have an unusually large influence on the results of the analysis.

Participant with ID 8,11, and 19 had the most overall influential on the restricted likelihood compared to the others on the total number of steps per day outcome. The participants of ID 1,8, and 21 ware also standed out to be the overall influential on the total active time per day outcome.

Overall influential analyses has shown that the participant with ID 8 and 20 seems to have a large influence for the total number of steps per day. For the total active time per day outcome, the participant with ID 21 had a large influence.

For dealing with the observed influential cases, the influence participant has left out in the data and the same model was refitted on the reduced data. Overall, for the total number of steps per day, the influences was not found to have a potential effect on fixed and the covariance parameters because the conclusion did not change. For the total active time per day the same process was done but for this case, the participant of ID 21 was found to have a potential influence on the model parameters due to for the model fitted on the reduced data, the quadratic log time and the quadratic log time*training interaction were not found to be significant see in Table 15 in Appendix.

For comparing the results obtained from linear mixed model fitted on the real data against the possible results if we re-label the data points, then see how extreme the results form the real data are, when compared against this alternatives arrangements of the data. The blocked case re-sampling study was conducted based upon 1000 replicates of data within each block. The original model were fitted for each sample and parameter estimate were estimated for each replicate then combined together to have common value based on the large dataset for the two outcomes.

For the total number of steps per day outcome, the estimated parameters for the original data and the estimates obtained in case re-sampled data were more or less the same this suggested the stability of the results. For the total active time per day outcome there were slight difference to the estimated parameter for the original data compared to the case re-sample data. The quadratic time and the quadratic log time-training group interaction which were significant in the original data were found to be not significant in the replicated data. This was not surprise because these parameters were also found to be non significant in leave out influential analysis. Due to the original parameter estimate were based on the small sample size and one participant can have an effect on the parameter estimates. So we based on the model fitted on the reduced data due to for being in line with the one based on larger sample size. These results led to the conclusion that there were no evidence for the difference between the intervention and control group for the evolution of the total number of steps per day across the training time periods. The participants in intervention group have shown to be daily active more than the participants in control group during the 12 weeks of the training.

4.1 Recommendation for future study

In the current study the data are slightly imbalanced and the sample size was rather small. We recommend for the future studies focusing on MS patients to gather considerable preliminary evidence on related participants before conducting the study in order to define the sample size needed for determining the critical effect. Thus, testing of a null hypothesis might be particularly challenging when confronted with limited sample size. Due to this small sample size the power and the effectiveness of the study may be diminished. However these limitations can be tackled by calculating the accurate sample size base on predefined level of power. Also we recommend for the future studies to control for factors that can confound the relation between intervention and physical activity patterns and to assure balance between groups in order to make a valid and accurate comparison.

References

1. Agresti, A. (2007). *Categorical Data Analysis*. Second Edition. New Jersey: Wiley Series in Probability and Statistics.

2. Bennett, E.J. (2013). Predictors of Physical Activity Among Persons With Multiple Sclerosis. All Graduate Theses and Dissertations. Paper 1983. Available at [http://digitalcommons.usu.edu/etd/1983] [Cited on 8^{th} August 2015]

3. Carton, H., Vlietinck, R., Debruyne, J., De Keyser, J., Truyen, L., D'Hooghe, M.D., Medaer, R., Sadovnick, A.D., Loss, R., Yee, I.M.L. (1997). Risks of multiple sclerosis in relatives of patients in Flanders, Belgium. *Journal of Neurology, Neurosurgery and Psychiatry*, **62**, 329-333.

4. Colley , R.C., Janssen ,I. and Tremblay, M.S. (2012). Daily Step Target to Measure Adherence to Physical Activity Guidelines in Children. *Medecine & Science in Sport & Exercise*, **44**(5), 977–982.

5. Fitzmaurice, G., Davidian, M., Molenberghs, G., and Verbeke, G. (2009). *Longitudinal Data Analysis*. New York: Chapman & Hall.

6. Hyunsu, J.(2015). Moving Block Bootstrap for Analyzing Longitudinal Data. *Commun Stat Theory Methods*, **44**(6), 1130–1142.

7. Janet, E.J. and Johnston, R.B. (2008). *Multiple Sclerosis: Current Status and Strategies for the Future.* Institute of Medecine, National Academy Press. Washington, D.C.

8. Koch-Henriksen, N. and Sørensen, P.S. (2010). The changing demographic pattern of multiple sclerosis epidemiology. *The Lancet Neurology*, **9**, 520–532.

9. Littell, R.C., Milliken, G.A., Stroup, W.W., and Wolfinger, R.D. (1996). SAS® System for Mixed Models, Cary, NC: SAS Institute Inc.

10. Medical Disability Advisor. (2012). Multiple Sclerosis. Reed Group, Ltd. in the United States. Available at: [http://www.mdguidelines.com/multiple-sclerosis] [Cited on 4^{th} July 2015]

11. Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.

12. National Multiple Sclerosis. (2009). Physical Activity for People with Multiple Sclerosis. An Introduction to MS for Health and Wellness Professionals. 2nd Edition

13. Pugliatti, M., Rosati, G., Carton, H., Riise, T., Drulovic, J., Vécsei, L. and Milanov, I. (2006). The epidemiology of multiple sclerosis in Europe. *European Journal of Neurology*, **13**, 700–722.

14. Puyau, M.R., Adolph, A.L., Vohra, F.A., Zakeri, I., and Butte, N.F. (2004). Prediction of Activity Energy Expenditure Using Accelerometers in Children. *Medecine & Science in Sport & Exercise*, **36**(9), 1625–1631.

15. Schabenberger, O. (2004). Mixed Model Influence Diagnostics. Cary, NC: SAS Institute Inc.

16. Stram, D.O., Lee, J.W. (1994). Variance Components Testing in the Longitudinal Mixed Effects Model. *Biometrics*, **50**(4), 1171-1177.

17. Tudor-Locke, C., Bassett, D.R. (2004). How many steps/day are enough? Preliminary pedometer indices for public health. *Medicine and Science in Sports Medicine*, **34**, 1-8.

18. Van Ooteghem, P., De Hooghe, M.B., Vlietnick, R. and Carton, H. (1994). Prevalence of multiple sclerosis in Flanders, Belgium. *Neuroepidemiology*, **13**(9), 220–225.

19. Verbeke, G., and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New-York: Springer.

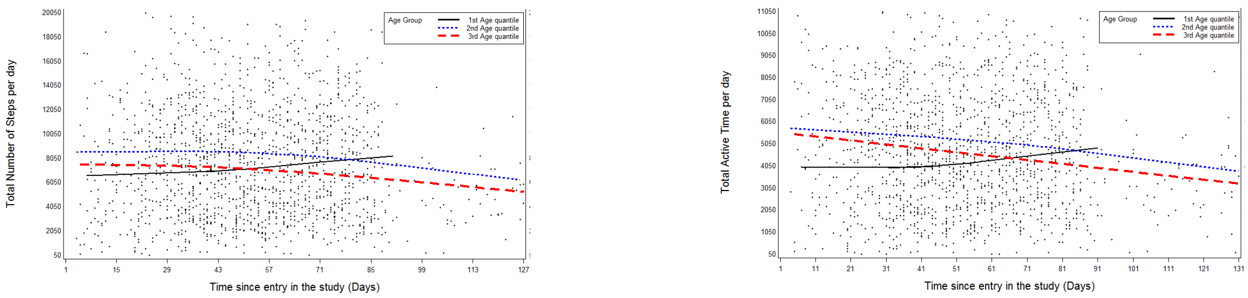
20. Withings Health Institute. (2014). Impact of Physical Activity on Sleep. Available at: [`http://www2.withings.com/eu/en/health-institute/`] [Cited on: 7^{th} July 2015].

21. World Health Organization. (2008). Atlas multiple sclerosis resources in the world Geneva: WHO Press.

22. Zewotir, T., and Galpin, J. S. (2005). Influence Diagnostics for Linear Mixed Models. *Journal of Data Science*, **3**, 153-177.

Appendix

Evolution of the responses by age quantiles



(a) (b)

Figure 6: *Smoothed average trend by age quantiles for total number of steps per day (a) and for total active time per day (b)*

Main effect selectio for the model of the total number of steps per day

Model	# Param	-2ℓ	G^2	P-value
Full Model	11	13786.0		
Time,age,grp,gend,time^2,gend*grp,time*grp,time^2*grp	10	13786.6	0.6	0.4385
time,age,grp,gend,gend*grp,time^2,time*grp	9	13786.1	0.5	0.4795
time,age,grp,gender,time^2,time*grp	8	13788.1	2	0.1572
time,grp,age,time^2,time*grp	7	13789.4	1.3	0.2542
time,grp,time^2,time*grp	6	13792.0	2.6	0.1068
time,grp,time*grp	5	13806.8	14.8	0.0001

Table 13: *Likelihood Ratio Test for fixed effects for total number of steps per day*

Main effect selection for the model of the total active time per day

Model	# Parameter	-2ℓ	G^2	Pvalue
Full Model	11	13127.1		
time,age,grp,time^2,gend,time*grp,time^2*grp,age*grp	10	13128.6	1.5	0.2206
time,age,gend,grp,time^2,time*grp,time^2*grp	9	13128.7	0.1	0.7518
time,age,grp,time^2,time*grp,time^2*grp	8	13129.4	1.3	0.5220
time,grp,time^2,time*grp,time^2*grp	7	13131.7	2.3	0.1293
time,grp,time^2,time*grp	6	13136.7	5	0.0253

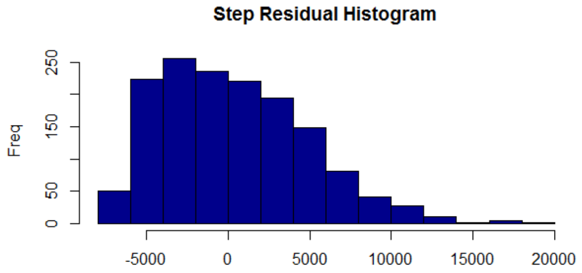
Table 14: *Likelihood Ratio Test for fixed effects for total active time per day*

Parameter estimate for total active time per day obtained leaving out the influential ID

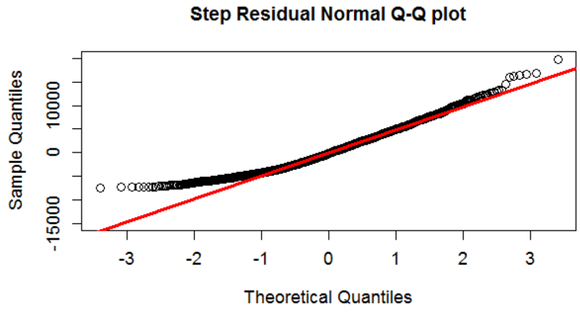
Effect	Estimate	Standard Error	p-value
Intercept	130.63	26.792	<.0001
time	-22.399	14.211	0.1196
$time^2$	1.760	1.916	0.3604
grp_1	-106.76	37.882	0.0087
$time*grp_1$	44.672	20.344	0.0328
$time^2*grp_1$	-4.815	2.754	0.0849
σ_{b0i}	126.391	40.355	
ρ	0.0417	0.047	
σ^2	359.461	16.758	

Table 15: *Parameter estimate for total active time per day obtained using the reduced data after leaving out the influential participant with ID 21*

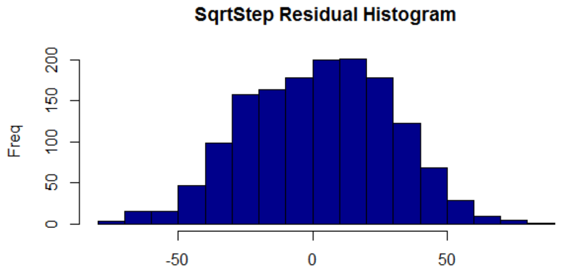
Residual Analysis for the total number of steps per day



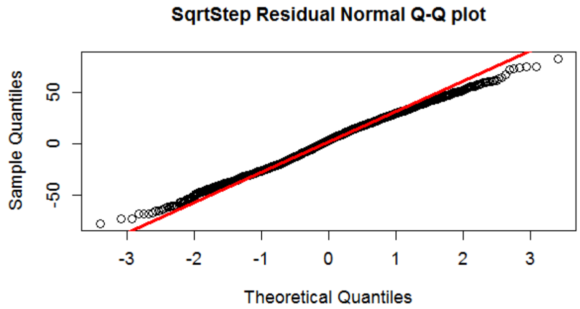
(a)



(b)



(c)



(d)

Figure 7: *The histogram and normal probability plot for the residual of the total number of steps per day and its square root transformation*

Residual Analysis for total active time per day

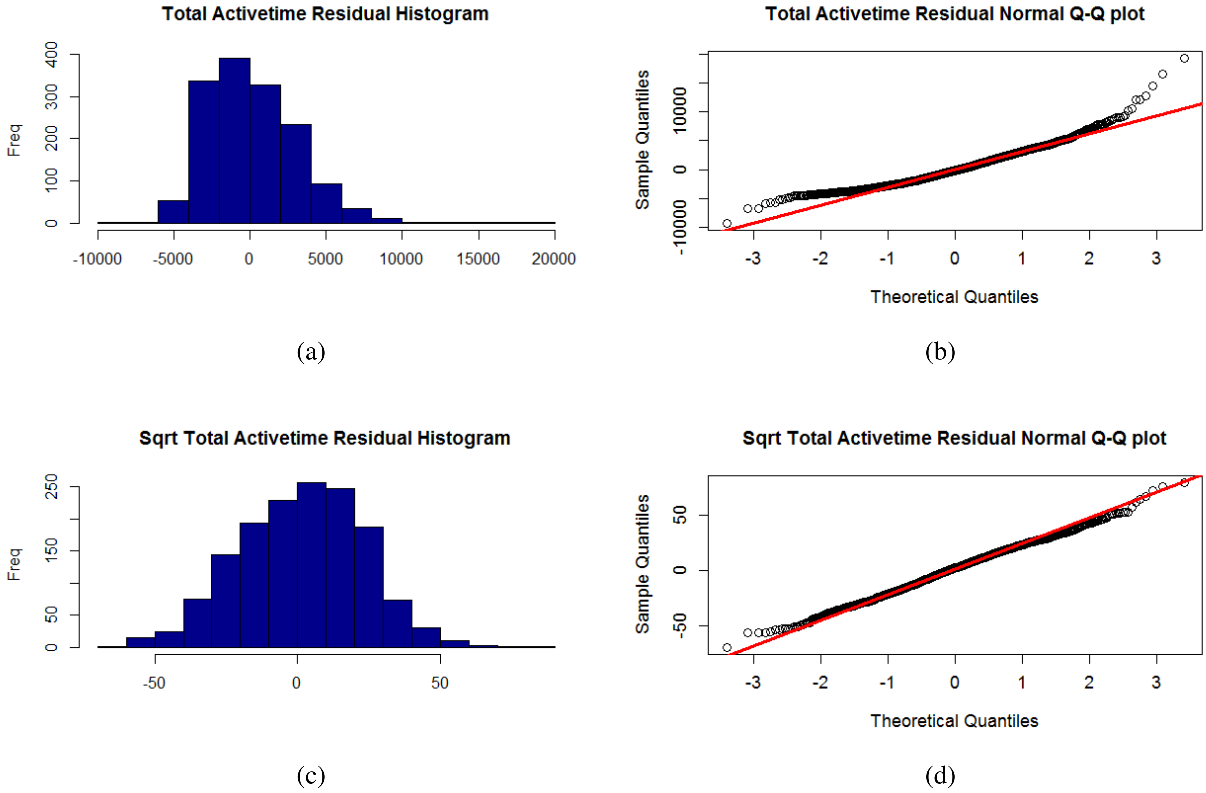


Figure 8: *The histogram and normal probability plot for the residual of the total active time per day and its square root transformation*

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Patterns in physical activity during a 5km start-to-run program for persons with Multiple Sclerosis

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,

Dusabimana, Alfred

Datum: **1/09/2015**