# The Six Spot Step Test for patients with Multiple Sclerosis

Ane Burheim Nermoen

Thesis submitted for the degree of
Master in Electrical Engineering, Informatics and Technology
60 credits

Department of Physics
Faculty of mathematics and natural sciences

UNIVERSITY OF OSLO

Spring 2021

# The Six Spot Step Test for patients with Multiple Sclerosis

Ane Burheim Nermoen

The Six Spot Step Test for patients with Multiple Sclerosis

http://www.duo.uio.no/

Printed: Reprosentralen, University of Oslo

## Acknowledgements

First, I would like to thank my supervisor Andreas Austeng for guidance and support with my thesis and my two co-supervisors, Pål Berg-Hansen and Stine Marit Moen, for helping me with information about Multiple Sclerosis. I would also like to thank Frédéric Meyer for helping me with the analysis and the rest of the AutoActive group for letting me write my thesis as part of their research project. Thank you to Pål Berg-Hansen, Thomas Dahl Klyve, and Henrik Negård for collecting the data I used in this thesis.
I want to thank my family and friends for their support and encouragement. Finally, I would like to thank the administration at the University of Oslo for doing what they could to keep the campus open as much as possible during the pandemic.

# Abstract

In later years wearable sensors have been used to determine if they can detect small nuances within walking mobility in people with neurological diseases, such as Multiple Sclerosis (MS). The Six Spot Step Test (SSST) is a clinical performance test used to assess gait speed, balance, and coordination. Within the test, the participant walks and kicks.

I used data from wearable sensors to investigate the data from patients with MS doing the SSST. I started with raw sensor data from a collection of many different tests and wanted to detect the SSST and the different kicks within the SSST. Finding the tests was easy, but identifying all kicks for all patients and controls turned out to be challenging. As a solution, I ended up using the raw data from the sensors with a video of the test being performed to find different time stamps within the tests to analyze.

I analyzed the time segments from the SSSTs for patients and controls to determine whether the time segments can be used to differentiate patients from healthy controls. I also wanted to see if one could detect progression after a rehabilitation stay and see if there is a learning effect from walking through the test one time. What I found is that there is a difference in the time used between the healthy controls and the patients with MS, but that this difference is small. I also found that one can see an improvement in the times used on the tests after a rehabilitation stay, but what this comes from, I could not say. Finally, I found that there is some learning effect from walking through the test one time.

# List of Figures

# List of Tables

# Contents

# Chapter 1

# Introduction

Multiple sclerosis (MS) is a cronical neurological disease that is often developed in young adults. This disease is one of the main reasons for early disability. The disease can affect the everyday mobility of the patients and end up giving them a strong handicap. The patients can improve their mobility with training and different aids. To find what can help them the most, their mobility should be assessed. Many different clinical performance tests assess the different body parts and mobility to help decide what training and aids can help. Unfortunately, their tests have proven less sensitive to pick up changes in the walking mobility for MS patients with little handicaps.

There has been an increase in the medicine domain to develop, test, and use wearable technologies, particularly when it comes to neurological diseases. One of the reasons for this is that the clinical performance tests one typically uses have difficulty detecting subtle gait dysfunction or progression. Wearable sensors have been used to examine gait balance control for people with concussions (Pitt et al., 2020), to find foot clearance for older people (Mariani et al., 2012), and to segment strides from free walking (Barth et al., 2015). These are just some of the research done with wearable sensors. When it comes to people with MS, it has been shown that with wearable sensors, one can detect mobility differences between people with MS and healthy controls where the traditional timed performance tests would not detect differences (Spain et al., 2012). Using wearable sensors, one hopes it will become easier to detect small and early changes in the mobility of people with MS and other neurological diseases.

One of the newer clinical performance tests that are used to assess people with MS is the Six Spot Step Test (Nieuwenhuis et al., 2006). The Six Spot Step Test (SSST) is a clinical performance test that assesses the patient's gait speed, mobility, balance, and coordination. The test was developed to get a quantitative test that is more sensitive than one of the most used clinical performance tests, the Timed 20-foot walk (T20FW), to assess gait mobility. Compared to many other walking tests that manly is about walking, the SSST includes kicks as well. This makes the test more complex, and one may pick up more differences between patients.

This MSc project is part of the main project, AutoActive: Tools and Methods for Autonomous Analysis of Human Activities from Wearable Device Sensor Data (*AutoActive*, [n.d.]). The project collaborates with researchers from NTNU, UiO, SINTEF, OUS, MS-Senteret Hakadal (MSSH), and Olympiatoppen. The main project is divided into different parts. This MSc project is part of a study where one wants to try sensors against clinical performance tests used today, focusing on walking abilities(gait function). One wants to see if the sensors give a more precise evaluation of this.

The purpose of this MSc project was to investigate the Six Spot Step Test to see if there were any differences in the ability to walk between patients with MS and control groups with the help of sensors used during the tests. The patients were tested during their first week at MSSH and the last week after a rehabilitation stay lasting from 2 to 4 weeks. The controls were only tested one time at MSSH.

The SSST can be divided into five sections with kicking and five with walking. I will be analyzing data from wearable sensors and the time used for the different sections within the SSSTs. There is possible to do other tests on these data, but this thesis is limited to the time used on the SSST and the analysis of these data. Some of the things I will be investigating from the data are the difference between the controls and the patients, the difference between the tests done in week 1 and at the end of the rehabilitation stay, and if there is a learning effect from walking through the SSST multiple times.

The thesis starts with an introduction to MS and the clinical performance tests used (Chapter 2). Then comes a theoretical chapter with an introduction into the statistical tools used in the analysis (Chapter 3), information about the data collection (Chapter 4), and how the data was processed (Chapter 5). The results are divided into five parts that handle different tests on the data (Chapter 6-10). Each of these chapters contains a discussion. The thesis ends with a final overall discussion and the conclusion (Chapter 11).

# Chapter 2

# Background

## 2.1 Multiple sclerosis

Multiple sclerosis (MS) is a chronic infectious disease that attacks the central nervous system, where the immune system attacks and "eats up" the isolation layer that surrounds the nerve threads in the brain and the spinal cord (Bø, [n.d.(a)]). Without the isolation layer, the nerve signals have problems traveling from the brain out to the rest of the body. The signals can be delayed, or in some cases, the signals do not manage to arrive.

The infection can appear anywhere in the central nervous system, so that the symptoms may vary from patient to patient. The symptoms a person with multiple sclerosis (pwMS) can get depend on the area where the infection is located and which nerves are attacked. Some of the symptoms can be paralysis, vision problems, balance problems, cognitive symptoms, and fatigue. The symptoms can develop to become a strong handicap. The core stability can be affected so that even a patient with a high function level and low age can still experience an increase in fall risk (Clausen, [n.d.]).

There are three different courses of the disease (Bø, [n.d.(a)]); relapsing-remitting multiple sclerosis (RRMS), secondary progressive multiple sclerosis (SPMS), and primary progressive multiple sclerosis (PPMS). There is no way of knowing how the disease will develop for the individual pwMS.

RRMS is the most common course of the disease, and 85-95% of the people with MS experience it. These patients will experience attacks. When a worsening of the symptoms appears and lasts for more than 24 hours, the symptom can be connected to the infection in the central nervous system. The symptoms one gets during an attack can last for some days up to months. When the attack is over, the symptoms from the attack will gradually improve. The attacks will lead to a slight deterioration where the pwMS will experience an increase in disability over several years.

The RRMS can, after years of gradual worsening from one year to another without attacks, become SPMS. There are fewer pwMS that experience this course due to the improvement of

the treatment given to the patients with RRMS.
PPMS is the course of the disease where the pwMS do not experience attacks but experience a gradual increase in their disability that is not connected to an attack from the beginning of the disease. This type of progression for the disease is the least common of the three.
We do not know why some people get MS. Today there is a belief that there is a correlation between genetics and the environment that makes people get MS. One has seen that the probability of getting MS can slightly increase if one has a close relative with MS. The environmental factors that may increase the likelihood of getting MS are smoking, vitamin D deficiency, and the Epstein-Barr virus (mono) (Bø, [n.d.(b)]). There are more females than males the get MS. The disease is most common to develop from age 20 to 50 but can appear at any age.
There has been registered an increase in the number of pwMS in Norway in later years. One does not know why there is an increase, but there is believed this increase has to do with improving the mapping methods used today. The increasing number of pwMS can also be because pwMS lives longer today because of improved treatment and the development of different medicines for slowing down the disease (*Statistikk om MS*, [n.d.]).
Some treatments can help to reduce the symptoms for the patients. Preventive treatments can help to reduce the risk of permanent dysfunction for pwMS with attacks, and some medicines can help reduce the risk of new attacks for people with RRMS (Bø, [n.d.(a)]). This can help to hinder or sink the progressive phase of the disease. For people with progressive MS, are there fewer medicines that can affect how the disease develops over time (Bø, [n.d.(c)]). There is essential that preventive treatments are followed up so they can be adjusted if needed.

## 2.2 Expanded Disability Statue Scale (EDSS)

The Expanded Disability Status Scale (EDSS) (Kurtzke, 1983) is a scale used to quantify the level of disability a pwMS has and helps with monitoring changes in the level of disability over time as the disease develops. It is a simple scale ranging from 0 to 10, with unit increments at 0.5. This scale covers from where a patient does not have any disability, 0, to where the patient has died due to the MS, 10. For the lower levels, 0 to 3.0, does the scale say that the patient had mild disabilities. From level 3.0 and upward, the patient, described by the scale, has moderate disabilities, which will worsen as the levels increase.
The different levels in the scale are based on the patients' results in the Functional system (FS). This is a system that contains eight different individual systems that cover different parts of the neural system. The eight individual systems are visual, bowel and bladder, brain stem, sensor, pyramidal, cerebellar, cerebral, and other (Kurtzke, 1983). During a neurological examination, the patient is given a score in all eight individual systems. The scoring in these eight systems is then used to place the patients in a level on the EDSS.
This scale is widely used as a tool in clinical trials to measure the progress in the disease and any outcomes after a clinical intervention (Meyer-Moock et al., 2014). Since the EDSS is so widely used, this makes it easier to compare different studies. Even though the scale is

widely used does not mean that the scale is perfect. The scale is shown to have low reliability and responsiveness, the levels are unevenly distributed, and the scale has a lack of precision (Blumhardt et al., 2004, p. 69).

## 2.3 Fampridin (Fampyra ®)

Fampridine is a medicine sometimes given to a pwMS with an EDSS between 4 and 7. The medication is taken to improve the gait function, but it has been shown that it may affect fatigue, function in the arms, the intestines, and bladder (Lunde et al., [n.d.]). It works on letting signals pass through the nerves more normally by helping to stop potassium from leaving the damaged nerve cells (*Fampridine (Fampyra)*, [n.d.]).

Before a patient can get the medicine, they have to take a walking test or a self-assessment that can be retaken 2 to 4 weeks later to see the effect of the medicine. After 2-4 weeks on the medicine, there must be a minimal improvement of 20% in the gait function to allow the treatment to continue (*Fampridin (Fampyra)*, [n.d.]; Lunde et al., [n.d.]). The renal function should also be checked before and after the treatment.

Some of the side effects of the medicine are urinal tract infections, dizziness, headache, back pain, difficulty sleeping, feeling sick, stomach upsets, balance disorder, and many more (Lunde et al., [n.d.]; *Fampridine (Fampyra)*, [n.d.]).

## 2.4 Clinical performance tests

One of the more common symptoms for a pwMS is the decrease in their walking abilities, with a change in their walking and balancing function (Clausen, [n.d.]). It is essential to evaluate the individual patients' abilities to walk and move and give them the best advice and guidance. A way to look at the walking abilities of a pwMS, and patients with other neurological diseases, is to do some clinical performance tests.

Many different clinical performance tests are executed with help from a physiotherapist to look at how the patient moves and walks. Some of them look at the upper body, and others look at the lower limbs. Some are only looking at the patients' balance, and some of the tests try to paint a more complex picture by looking at more than one aspect of the patients' mobility. When looking at the walking ability, there has been shown that there is no measure that is ideal (Bethouc et al., 2011). Because of this, one has to do different tests to get a complete picture of how the patient moves.

In the following subsections, five clinical performance tests are described. These tests look at walking speed, mobility, balance, and coordination. These five tests are the tests that have been chosen to be used in the AutoActive project. The participants in this study took the set of tests during one session as described later in the Data Collection (Chapter 4).

### 2.4.1 Timed 25-foot walking test (T25-FWT)

The times 25-foot walking test (T25FWT) is a clinical performance test used to look at the walking mobility and speed of the patients. The patient is told to walk 25 feet, or 7.62 meters, straightforward while the time is taken.
The T25FW is one of the components of the Multiple Sclerosis Functional Composite (MSFC) measure (Rudick et al., 2002) and is looked at as being a precise measure of disability for a pwMS (Blumhardt et al., 2004, p. 147). The test has been shown to be the test to measure the walking disability objective (Kieseier et al., 2012).
The thing that makes the test so widely used is that one only needs a stopwatch and a lain that measured 25 feet to do the test. That makes it possible to administer the test in a wide span of places.

### 2.4.2 6 Minutes Walking Test (6MWT)

The 6 Minutes Walking Test (6MWT) is a long clinical performance walking test that looks at the patients walking mobility, walking speed, and fatigue. The test is that the patient is to walk for 6 minutes. The results of the test are in how long the patient managed to walk during the 6 minutes. Even though one can look at the patients walking mobility and speed in this test, the main focus for this test is on the walking endurance of the patient (Bethouc et al., 2011).
The test has been shown to be reproducible and a measure that is reliable for a pwMS (Goldman et al., 2008). Compared to the T25FW test, where one only needs a 25 feet straight walking path, a long corridor or something similar where one can walk back and forth to do the test is needed. The gait speed from the long walking test, 6MW test, and the short walking test, T25FW test, has been shown to be closely correlated when looking at a pwMS compared to healthy peoples (Dalgas et al., 2012).

### 2.4.3 Timed Up and Go (TUG)

The Timed Up and Go (TUG) test is a clinical performance test that looks at the patients walking mobility, balance, and walking speed. The test starts with the patient sitting down in a chair with armrests. Then the patient is to stand up, walk 3 meters straightforward, turn, go back to the chair and sit down again. During the test, the instructor takes the time. The test is done three times, the first time to make the patient familiar with the test, and the two last to take the time of the test. The result of the test is the average test times of the two last tests (*TUG - The Timed "Up & Go"*, [n.d.]).
With the turning and transferring parts of the test, the test becomes more of a complex measure of the patient's mobility than the simple walking tests. There has been shown that the test is related to executive functions (Herman et al., 2011). This is most likely due to the complexity of the test. Also, with the variety of motions in the test, is the TUG test is proven to be a

valid measure of functional mobility (Sebastião et al., 2016).
If one wants to only look at a patient's walking ability, the TUG test is limited because the walking ability is only one of many parts that affect the time used to complete the test. But if one wants to look at more than just the walking ability, the TUG test can give good inside into the mobility-related functions for the patient (Bethouc et al., 2011).

### 2.4.4 Single Leg Stand (SLS)

Single Leg Stand (SLS) is a clinical performing test that only looks at the patients' static balance. The task is to stand on one leg while the time is taken.
The single leg stand test is a test that is part of the BESTest (Horak et al., 2009) and the mini-BESTest (Franchignoni et al., 2010). In the BESTest, it is item nr 11, under the Anticipatory Postural Adjustments system category. The mini-BESTest is a shorter version of the BESTest, and the SLS test is the third task in this test system. These two tests, BESTest, and mini-BeSTest have been shown to have good inter-rater and test-retest reliability (Hamrs et al., 2017).
This test is an easy test to implement. The only thing that is needed is a stopwatch and a good floor to stand on. Patients mildly affected with MS can experience a reduction in balance performance and balance confidence (Kanekar et al., 2013). This makes any test that looks at the balance of the patient an important test.

### 2.4.5 Six Spot Step Test (SSST)

The Six Spot Step Test (Nieuwenhuis et al., 2006) is a newer test that is a complex clinical performing walking test that combines walking and kicking. The combination makes the test good for testing walking abilities, speed, balance, and coordination.
The test is done in a 5× 1-meter marked-up rectangle course, where two cubes are placed on each of the long sides, and one cube in the middle of each of the short sides (Figure 2.1). The test starts at one of the short sides where the cube there is removed. Then the task is to walk from cube to cube as fast as possible, kicking each cube out of the course with either the left or the right foot. The instructor then takes the time the test taker uses.

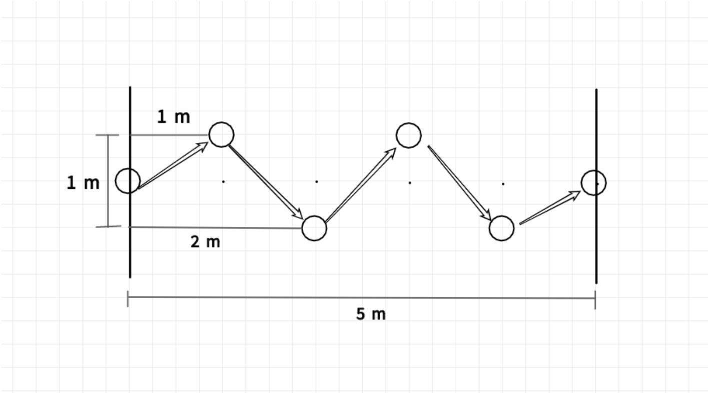


Figure 2.1: A diagram of the course for the six spot step test (SSST).

Because of the combination of movements in the test, the SSST has more variation in the time used compared to the T25FW test (Nieuwenhuis et al., 2006). Still, a high test-retest reliability has been reported (Bethouc et al., 2011).

# Chapter 3

# Theory

*In this chapter, short instructions into the statistical tests that are used later are presented. These statistical tests will be used later to gain information about the population I am looking at and aid in analyzing the data.*

## 3.1 Chi-square goodness-of-fit test of normality

There are many different statistical tests one can perform on a set of samples to gain information about the population represented. Many of the tests build on the assumption that the samples are normal distribution. To establish if this is the case, a common test to use is the Chi-square goodness-of-fit test (Devore et al., 2011, Ch 13.2).

If one has a set of samples and wants to determine if the underlying distribution is normal, one first estimates the $\mu$ and $\sigma$ to the samples and then does the chi-squared test to see if the samples are from a normal distribution.

Matlab has the function `chi2gof` that do the chi-squared goodness-of-fit test (*chi2gof: Chi-square goodness-of-fit test*, [n.d.]). One puts in a vector $x$ that contains the data samples. The function can return the test decision for the null hypothesis, $h$, the p-value of the hypothesis test, $p$, and a structure that contains the information about the test statistic. The null hypothesis is that the samples $x$ come from a normal distribution where the mean and variance are estimated from $x$. If the null hypothesis is true, the test decision $h = 0$. If the null hypothesis is rejected, the test decision $h = 1$. The significance level is set to 5% as default, but this can be changed.

## 3.2 Linear regression and R-squared

Suppose one has two variables, $x$ and $y$, and wants to find out the relationship between these two variables to gain information about one of them through knowing the other. In that case, one can do a regression analysis (Devore et al., 2011, Ch. 12). With regression analysis, one tries to fit a model to the data. The model can give information about one of the variables by

knowing the values of the other variable. With linear regression, one tries to fit a linear model to the data (Devore et al., 2011, p. 617-620).
The linear model equation that makes the dependent variable $Y$ related to the independent variable $x$ is

$$Y = \beta_0 + \beta_1 x + \varepsilon \tag{3.1}$$

where $\varepsilon$ is a random variable that is normally distributed with mean 0 and variance $\sigma^2$, $\beta_1$ the slope coefficient and $\beta_0$ the intercept coefficient. From the model equation (Eq 3.1) $n$ observed pairs are regarded as having been generated independently of each other (Devore et al., 2011, p. 617-618).
To see if the linear regression model has a good fit with the data, one can find the coefficient of determination (R-squared or $r^2$) (Devore et al., 2011, p. 633-635). This coefficient looks at the proportion of the observed $y$ variation that can be explained by the linear regression model,

$$r^2 = 1 - \frac{SSE}{SST} = \frac{SSR}{SST} \tag{3.2}$$

where SSE is the error sum of square (measuring of how much variation in $y$ that cannot be attributed to the linear regression model), SST is the total sum of square (measureing of the total amount of variation there is in the observed $y$ values), and SSR is the regression sum of squares (the total amount of variation that can be explained by the linear regression model). The smaller the $r^2$ is, the less do the linear regression model fit the data. $r^2_{adj}$ is the $r^2$ that have been adjusted to the number of predictor variables in the linear regression model (*Coefficient of Determination (R-Squared)*, [n.d.]) by multiplying $\frac{SSE}{SST}$ with $\frac{n-1}{n-p}$, where $n$ is the number of observations and $p$ is the number of regression coefficients.
In Matlab one can fit a linear regression model to a data set with the function `fitlm` (*fitlm: Fit linear regression model*, [n.d.]). The function works by putting a table or data set array containing two variables into the function, and the function returns the linear regression model. The last variable is used as the response variable as a default by the function. In the returned linear regression model, one can also find the number of observations, root mean squared error, error degrees of freedom, R-squared, adjusted R-squared, F-statistic vs. constant model, and the p-value.

## 3.3 Paired t-test

The paired t-test (Devore et al., 2011; *ttest: One-sample and paired-sample t-test*, [n.d.], p. 511-513) (*ttest: One-sample and paired-sample t-test*, [n.d.]) is a hypothesis test based on the `t` distribution looks at sets of paired data samples that are independent of each other. It is used when one has two observations and one wants to see if the difference between the two samples is a normal distribution with zero as a mean and unknown variance. If one has the sample set $X$ and the sample set $Y$, where $X$ is the first observations, and $Y$ is the second observations, and they are independent of each other, the expected difference is $\mu_D = \mu_X - \mu_Y$. To test if the two samples are independent, we can test the difference. The null hypothesis becomes

$$H_0 : \mu_D = \Delta_0, \tag{3.3}$$

where $\Delta_0$ is the null value (often $\Delta_0 = 0$). This test, when using the expected difference $\mu_D$, is the same as the one-sample t-test (Devore et al., 2011, p. 443-447) since the difference between $X$ and $Y$ constitute a normal random set of samples with mean $\mu_D$.
In Matlab can one use the function `ttest` (*ttest: One-sample and paired-sample t-test*, [n.d.]) to do both the one-sample and the paired-sample t-test. For the one-sample test, one sends in only one sample vector, and for a paired-sample test, one sends in two sample vectors. The function can return the test decision ($h$), the p-value ($p$), the confidence interval on the difference of the population means, and the information about the test statistic. If the test decision is $h = 1$, the test rejects the null hypothesis, and if the decision is $h = 0$, the null hypothesis is accepted. The null hypothesis is rejected at the 5% significant level as a default.

## 3.4 The two-sample t-test

The two-sample t-test (Devore et al., 2011, Ch 10.2) (*ttest2: Two-sample t-test*, [n.d.]) is a hypothesis test based on the $t$ distribution that compares two independent unpaired data sets. It tests if two sets are independent normal distributed with equal means and equal unknown variances. If we have a set of samples $X$ with mean $\mu_1$ and a set of samples $Y$ with mean $\mu_2$, and both are normally distributed and independent to each other. The null hypothesis is

$$H_0 : \mu_1 - \mu_2 = \Delta_0, \tag{3.4}$$

where $\Delta_0$ is the null value (often $\Delta_0 = 0$). Matlab has the function `ttest2` (*ttest2: Two-sample t-test*, [n.d.]) that do this test. Suppose one sends in two sample vectors $x$ and $y$ to the function. In that case, the function can return the test decision ($h$), the p-value ($p$), the confidence interval on the difference of the population means, and the information about the test statistic. If the test decision is $h = 1$, the test rejects the null hypothesis, and if the decision is $h = 0$, the null hypothesis is accepted. The null hypothesis is rejected at the 5% significant level as default.

## 3.5 P-value

P-values (Devore et al., 2011, Ch 9.4) is a probability calculated assuming that the null hypothesis is true. Generally speaking, the smaller the p-value, the more evidence is there against the null hypothesis in the sample data. The p-value is then used to decide if the null hypothesis is rejected or not, with the help of a selected significant level $\alpha$. If the p-value is smaller than the significant level, the null hypothesis is rejected. The significant level can be randomly selected,

but the standard are 5% ($\alpha = 0.05$) and 1% ($\alpha = 0.01$).
The p-value is not the probability that the null hypothesis is true. It is the probability of observing an equal or more extreme result than the actual observed result, given that the null hypothesis is true. An example of this is having a null hypothesis stating the data is normally distributed with zero mean and observing a low p-value ($p < \alpha$). The odds of an extreme observed outcome under the null hypothesis will be improbable.
In this project, attention will be paid to the p-value and if it is below or above the significant level that decides if the null hypothesis is rejected or not. I will also, in some situations, discuss how a change in significance level ($\alpha$) will affect if the null hypothesis is rejected or not.

# Chapter 4

# Data Collection

The data collection for this project was from five clinical performance walking tests that the participants did. The participants did the tests at MSSH (MS-senteret Hakadal) with the help of three different instructors, two physiotherapists, and one doctor. The raw data was a combination of video of the tests, sensor data of the tests, personal information about the participants, and observations from the tests by the instructors. The participants were a combination of a control group and a patient group. Some of the people in the patient group were tested twice, one time at the start of a rehabilitation stay and one time in the end. Not all the patients could do the second test because their rehabilitation stay was cut short due to the pandemic.

## 4.1 The equipment

To collect the sensor data, they used three Physilog®5 (*Physilog: Inertial Measurement Sensor (IMU)*, [n.d.]) IMUs (Internal Measurement Sensor) from GaitUp. Two of the IMUs were attached with an elastic band on the footrest. The last one was attached at the lower back with an elastic band. One of the IMUs on the footrest was configured as the master, and the two others were configured as slaves. Each of the IMUs has an accelerometer, a gyroscope, and a barometer inside. These recorded the movement of the feet and lower back.
To record the tests done by each of the participants, two GoPro Hero7 Black (*HERO7 BLACK*, [n.d.]) were used. One of them was put on a tripod and placed so that it did not film the participants' faces, and the other GoPro was placed on the participants' chest faced downwards to film their feet during the tests. The cameras were places such that the participants' faces did not show up in the video due to the participants' right to privacy. The cameras were used to film how the tests were performed to see any mistakes or pick up something one did not see during the testing.

## 4.2 The location

All the tests were done in the attic at MSSH. The room was measured to be 29 meters long with an even concrete floor. In the room, there were marked up a bunch of points to show where the different tests were to occur:

- one mark at each side of the room - for the 6MW test
- one mark where a chair is placed - for the TUG test
- one mark 3 meters in front of the chair - for the TUG test
- one mark 25 feet from the 3-meter mark - for the T25FW test
- the track for the SSST was marked up (Figure 4.1)
- marked up places the tripod were to be placed for the different tests

## 4.3 The individual tests

Before every test, the instructor taped three times on the master IMU on the left foot and waited then for 10 seconds before starting the tests. This was done so the video and the sensors data could be synchronized and that when looking at the raw data, one could easily find a specific test by searching for the three taps that come before the test in the data.

### 4.3.1 TUG (Timed up and go)

The patient started seated in the chair. The video camera was placed on the left side behind the patient. The test instructor taped three times on the left foot IMU while the camera is filming. After ten seconds, gave the test instructor the instruction - *walk at your normal speed to the marked point, turn, walk back and sit down.* The test was done two times, with a one-minute break between each test. Walking aids were allowed, and these would be placed in a preferred hand, in contact with the ground.

### 4.3.2 SLS (Single leg stand)

The patient walked to the 3-meter marked point from the TUG test. The chair was removed, and the video camera was placed one to two meters behind the patient. The video camera filmed the test instructor tapping three times at the left foot IMU and waited for ten seconds. Then the test instructor gave the instruction - *place your hands on your hip, eyes forward, lift the foot with the sole pointing backward. Stand as long as possible. The minute you have to place the foot down or touch the other leg, the test is finished.* The first test was done with the left foot, the second with the right foot, the third with the left foot, and the fourth with the right foot. There was a ten-second break between the first and second tests, the third and fourth tests, and a 30-second break between the second and third tests. If the patient needs aid to do the test, the test was stopped.

### 4.3.3 T25-FWT (Times 25-foot walking test)

The test started at the 3-meter mark, and the video camera was placed on the left side of the 25 feet mark. The test instructor taped three times on the left foot IMU and waited for ten seconds. The instructor gave the patient the instruction - *walk at your normal speed past the marker and do not stop until you have passed the mark. Walk back to the start position*. The time was taken from the start to the patient had passed the mark. Then there was a 30-second break before the second test was done. The instructor gave the patient the instruction - *walk as fast as you can pass the marker and not stop until you have passed the mark. Walk then back to the starting point at normal walking speed*. The instructor retook the time from the start to the patient had passed the mark.

### 4.3.4 SSST (Six spot step test)

The patient walked to one of the ends of the SSST track. The video camera was placed behind the patient on the left side, approximately one meter away. Then the instructor taped three times on the left foot IMU and waited for ten seconds. After this, the instructions were given - *you will walk to the other side of the track as fast as possible while kicking the cubes out from the circles with your dominant foot. After the last cube is kicked, then standstill on the place.* The test was repeated one more time with the same instructions. Then the test was to be done by using the non-domain foot to do the kicking. The new instructions were given - *walk to the other end of the track as fast as possible while kicking the cubes out from the circles with your non-dominant foot. After the last cube is kicked, then standstill on the place*.
If the right foot was used as a kicking foot, the cubes on the right were kicked with the outside of the foot, and the cubes on the left were kicked with the inside of the foot. If kicking with the left foot, the cubes on the right were kicked with the inside of the foot, and the cubes on the left were kicked with the outside of the foot.

Figure 4.1: The six-spot step test (SSST) track from the room where the tests were done.

### 4.3.5 6-MWT (6 minutes walking test)

The test started at one of the markers for the 6MWT. The video camera was placed at the same place as for the T25-FWT. The test instructor taped three times on the left foot IMU and waited for ten seconds. Then the instructor gave the instruction - *you are going to walk as far you can in six minutes, back and forth inside the beams. You decide the tempo yourself and can take breaks if you want to. You turn at the door, walk back, and remember that the goal is to walk as far as you can in six minutes. When the test is finished, stay in the place where you stopped.*

If something went wrong during one of the tests and had to be retaken, the test was retaken after the 6-MWT. When all the tests had been done, the video camera was stopped.

## 4.4 The Test Form

During the testing, the instructors filled out a test form. This form contains:

- information about the participants (age, high, weight, gender, shoe size)
- which foot was the dominant foot
- if the participants use medicine and walking aids
- their EDSS (Expanded Disability Status Scale) and MSWS-12 (21-Item MS Walking Scale (Hobart et al., 2003)) results

- the instructor name
- the times that the instructor took from the tests
- time and place of the testing
- comments if anything went wrong during the tests

# Chapter 5

# Data Processing

The plan was to extract the raw data from the sensors that contained the information from the Six Spot Step Test (SSST) and then perform the needed analysis to process the data. The first challenge turned out to be the identification of the individual SSSTs for each participant. This turned out to be more complicated than anticipated. Considerable time was spent on automatically trying to identifying the different time segments before this strategy was abandon. The new plan became that I synchronized the sensor data with the video and went through the synchronized data to find different time stamps within each participants' tests.

## 5.1 Finding the data samples

The first thing I did was to identify where the SSST-tests were in the raw data, with the help of the taps done on the left foot IMU before the different tests. I pulled the data from the barometer on the left foot and found the raw data that observed the pressure change within the sensor. The barometer would not record any change during the test, but when the instructor tapped on the sensor, the barometer would experience a change in the pressure, and the change would be recorded. I used this data to identify the location on the time axis where all the tap groups (three taps together) were. The SSSTs were for most participants between tap groups five and six, so I located the fifth and sixth tap groups on the time axis. Inside these time points were the SSSTs.
Then I had to identify where the four individual SSSTs were inside these time parameters by finding the start and end of each of the SSSTs. The start of the individual tests was found by finding data from the foot sensors (gyroscope and accelerometer data). The participants went from standing still for a longer period to starting to walk. The end of the tests was when the last kick was finished. To find the ending, I had to identify where the kicks were in the test. The idea for finding the kicks was that to do a kick the opposite foot had to stand still. I started to look at the raw data to try to identify if this could be seen.

You can see an example on how the raw data looked like for the control 005 in the Figures 5.1 and 5.2, and from the patient 006 with en EDSS of 5.0 in the Figures 5.3 and 5.4. These examples are from the first SSST, where both of them were kicking with the right foot.

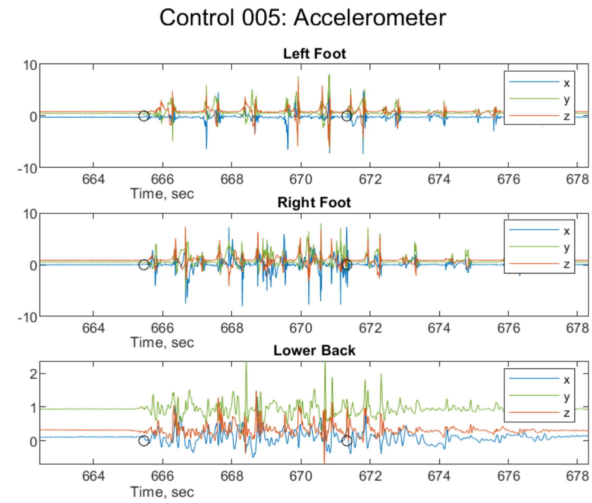


Figure 5.1: The three plots shows the raw data from the accelerometers in the left foot, the right foot and the lower back sensors for control 005.

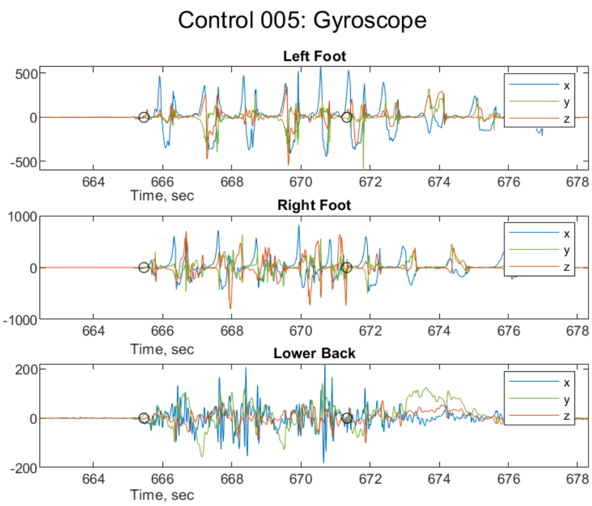


Figure 5.2: The three plots shows the raw data from the gyroscopes in the left foot, the right foot and the lower back sensors for control 005.

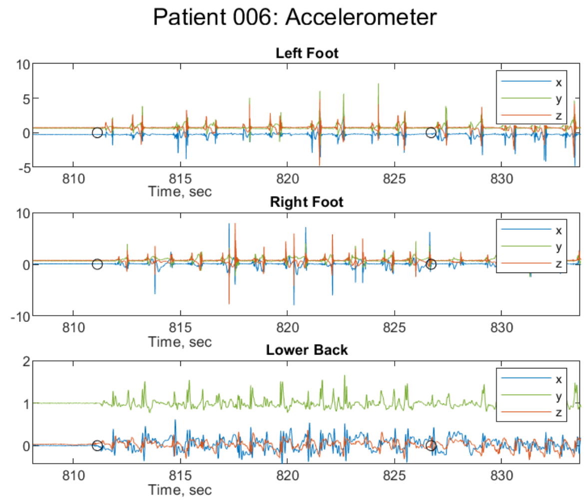


Figure 5.3: The three plots shows the raw data from the accelerometers in the left foot, the right foot and the lower back sensors for patient 006.

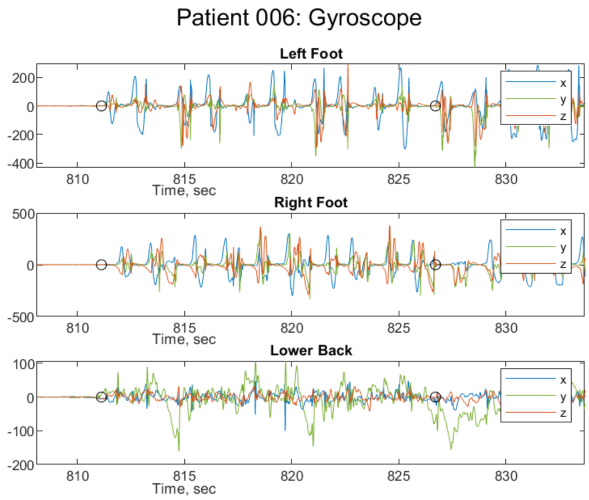


Figure 5.4: The three plots shows the raw data from the gyroscopes in the left foot, the right foot and the lower back sensors for patient 006.

From the data of control 005, one can observe differences between the left and the right foot for the accelerator and gyroscope data. For the right foot, it does not look like there is a pattern that is repeating itself. It looks like there is a repeating pattern from the left foot, but

the segments where the foot is standing still do not look like they change that much between each movement for the foot. The data from patient 006 do not look the same as that from control 005. Between the left and the right foot are there not much of an apparent difference as it was for the control. From the accelerometer (Figure 5.3), the data from the left and the right foot look the same, but the data from the right foot is slightly shifted. From the gyroscope data (Figure 5.4), there is a more notable difference between the left and the right foot, but there is no immediate pattern that emerges.
From comparing the data from the control and the patient, it became clear that it would be difficult to find a general algorithm for all the participants to find the kicks within the SSSTs. In total, I had data from 66 participants, all showing huge variations. After I looked at some of the videos from the tests, I discovered one reason why it would be challenging to find the kicks within the tests. From the videos, I was able to see that different participant kick very differently. Some of them stopped before the cubes and then kicks, while others started walking and kicked in the middle of their steps.
To fix the problem of finding when the participants were kicking the cubes in the test, it was suggested that adding a magnet inside the cubes that would react with a magnetometer would solve the problem. This could not be done since the IMUs that were used did not have a magnetometer. Another way of solving the problem could be to attach an accelerometer to each of the cubes that would record when the cubes were moved.
Since it became much more challenging to identify the kicks from the raw data, I had to use the videos to find the kicks. I started by synchronizing the raw data from the sensors with the video from the chest. The synchronizing was done with Activity Presenter (*Activity Presenter*, 2020-08-06). I found the beginning of the first SSST each of the participants did and used this as the synchronization point to synchronize the data. I could have used the taping on the left foot sensor, but the chest camera did not have a good view. The tripod video did have a good view of the taping but not a good view of each of the participants' kicks. The tool I used to synchronize the video with the data let me use only one video to be synchronized with the data.
With the synchronized video and data from the sensors, I went through frame by frame to find the different time stamps within the tests. Since the participants kicked in very different ways, the kicks were defined to start when the foot touched the cube for the first time and the end to be when the cube was kicked out of the circle, and the kicking foot was placed on the floor again. The time stamps ended up being when the patient started walking at the start of the tests, every beginning of a kick and every ending of a kick.
For every SSST, I ended up with eleven time stamps. The time stamps for each of the participants were noted in individual Excel files, and then I used Matlab to read the files, convert the time stamps into seconds, and change the times so that the first time stamp for each test was $t = 0$. From these times, I found the times from the hole test (time between the first and last time stamp), the times for the five kicks (time between touching the cube and placing the foot on the floor), and the times for the five walking parts of the test (time between placing the foot on the floor and touching the cube with the foot).

## 5.2 The participants

There were 46 patients with MS and 20 healthy controls that were tested in this project. Out of the 46 patients with MS were 32 of them retested at the end of a rehabilitation stay at MSSH, which lasted from 2 to 4 weeks. Out of the patients with MS, 13 of them were taking Fampridin to improve their walking. In Table 5.1 more details about the participants can be found.

| | | **Patients with MS** | **Control** |
|---|---|---|---|
| **Number of participants** | | 46 | 20 |
| **Male/Female** $[n(\%)]$ | | 17(37%)/29(63%) | 4(20%)/16(80%) |
| **Age, mean** $[years]$ | | 51 | 48 |
| **BMI, mean** | | 26.7 | 25.4 |
| **EDSS** | **median** | 3.04 | - |
| | **range** | $[1.0, 6.0]$ | - |
| **Fampridin** | | 13 | - |
| **Retested** | | 32 | - |
| **Used one/two canes to walk** | | 3 | - |

Table 5.1: The demographic of the participants.

Out of the 46 participants, I exclude eight; one control and seven patients with MS. The control was excluded because I had a hard time finding the time stamps from the video. The three patients who were using one or two canes to help with the walking were excluded since, with the cane in use, there was extra variability with them that I could not measure or see the effect on the raw data. The last four excluded patients were so because they did not complete all the four SSST in the first testing session. I ended up with 19 controls and 38 patients. Out of the 38 patients, 25 were retested, and 12 of the patients used Fampridin.

## 5.3 Dividing up the data

To investigate and compare that results the participants were sorted into groups based on there EDSS. This is one of the more cummon ways of looking at results when it comes to reasearch done on patients with MS. When I divided the patients into there EDSS I ended up with 9 patients groups and one control group, in total 10 groups (Table 5.2). Some of the EDSS groups ended up being only one patient.

| **EDSS** | 0.0 | 1.0 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 | 4.5 | 5.0 | 5.5 |
|---|---|---|---|---|---|---|---|---|---|---|
| **N** | 19 | 1 | 1 | 8 | 6 | 4 | 10 | 4 | 4 | 1 |

Table 5.2: The number of participants, N, with the EDSS levels.

To get fewer and bigger groups, I merged the EDSS groups, so I ended up with two groups representing the patients and one representing the controls (Table 5.3). I decided to divide the MS patients at EDSS level 4 because, from the scale, there is a shift in the description between levels 3.5 and 4. The patients with EDSS 3.5 are described as having no big problems with walking, while at EDSS 4, the patients can walk 500 meters without walking aids and rest.

| | | |
|---|---|---|
| **Group 1** | EDSS= 0 (controls) | N=19 |
| **Group 2** | $0 < \text{EDSS} < 4$ | N=20 |
| **Group 3** | $4 \leq \text{EDSS}$ | N=19 |

Table 5.3: How the participants were sorted into three groupings and how many participants, N, in each group.

# Chapter 6

# The test times

This chapter will analyze the time the participants use on the hole Six Spot Step Test (SSST) from the week 1 tests.
The participants will be divided into three groups based on their EDSS (Expanded Disability Statue Scale) (Section 5.3). The three groups will be:

- Group 1: the controls
- Group 2: the patients with an EDSS $< 4$
- Group 3: the patients with an EDSS $\geq 4$

During the analyze I looked at the mean test time for the participants based on

- the test times from all four tests
- the test times from the two tests where they kick with their dominant foot (D)
- the test times from the two tests where they kick with their non-dominant foot (ND)

I also looked separately at the test times from the four tests the participants did; the first test kicking with the dominant foot (D1), the second test kicking with the dominant foot (D2), the first test kicking with the non-dominant foot (ND1), and the second test kicking with the non-dominant foot (ND2).
I started by finding out if the test times for the participants in the three groups are from normal distributions. Then I looked into if the age of the participants has any effect on the mean test time they use for the SSSTs. I also saw if the use of the medicine Fampridin affects the test times. After this, I investigated if the mean test times between the three groups were different, if there was any difference in the test times between the dominant tests (D) and the non-dominant tests (ND), and look at the individual test times from the four tests (D1, D2, ND1, and ND2) the participants did.
I tested different null hypothesis $H_0$ with significant level $\alpha$:

- the mean test times for the patients with Fampridin and the patients without are both independent normal distributed, with equal mean and variance ($H_{0,1}$, with $\alpha = 0.05$).
- the mean D times for the patients with Fampiridin and the patients without are both independent normal distributed, with equal mean and variance ($H_{0,2}$, with $\alpha = 0.05$).
- the mean ND times for the patients with Fampiridin and the patients without are both independent normal distributed, with equal mean and variance ($H_{0,3}$, with $\alpha = 0.05$).
- the mean test times in Group $x$ and $y$ are both independent normal distributed, with equal mean and variance ($H_{0,4}$, with $\alpha = 0.05$).
- the difference between mean D times and mean ND times is normally distributed with mean$= 0$ ($H_{0,5}$, with $\alpha = 0.05$).
- mean D time in Group $x$ and $y$ are both independent normal distributed, with equal mean and variance ($H_{0,6}$, with $\alpha = 0.05$).
- mean ND time in Group $x$ and $y$ are both independent normal distributed, with equal mean and variance ($H_{0,7}$, with $\alpha = 0.05$).

## 6.1 Testing for normality

I started by testing if the test times within the three groups were normally distributed by doing the Chi-squared goodness-of-fit test. The test returns a result $h = 1$ if the samples are not normally distributed and $h = 0$ if they are. I did the test on the mean test times ($\mu$) for each of the participants and on the four test times (D1, D2, ND1, and ND2) for the participants in the three groups (Table 6.1).

| | **Group 1**, $[h]$ | **Group 2**, $[h]$ | **Group 3**, $[h]$ |
|---|---|---|---|
| **D1** | 0 | 0 | 0 |
| **D2** | 0 | 0 | 0 |
| **ND1** | 0 | 0 | 0 |
| **ND2** | 0 | 0 | 0 |
| $\mu$ | 0 | 0 | 0 |

Table 6.1: The result from the chi-squared goodness-of-fit test for the three groups, testing to see if the test times are normal distributed. The test were done on the four individual tests in the SSST and the mean test times. If $h = 0$ the samples are normal distributed, and if $h = 1$ are they not normal distributed.

## 6.2 The mean test time vs. age

To see if the age of the participants affected the time they used to complete the SSSTs, I first found the mean test time for the controls and plotted them against their age (Figure 6.1).

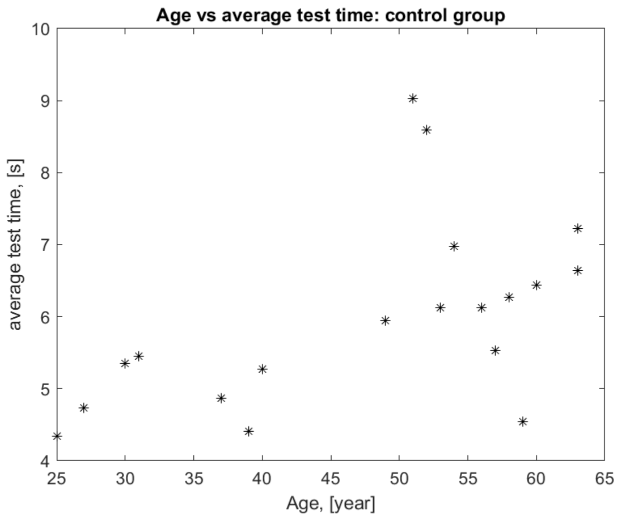


Figure 6.1: The mean test times for the control group plotted against there age in years.

Based on how the plot of the mean test times and the age for the controllers formed in Figure 6.1, I used the mean test times and the age of the controls to do a linear regression (Figure 6.2). I did the same for the patients (Figure 6.3) and for all of the participants (Figure 6.4).

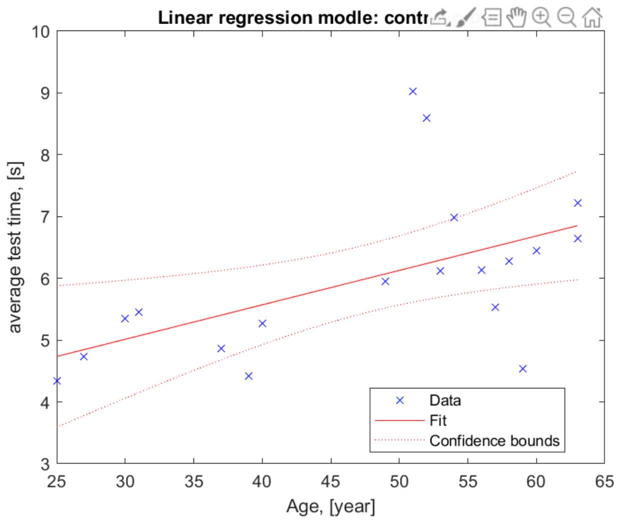


Figure 6.2: The linear regression of the mean test times and the age of the controllers.

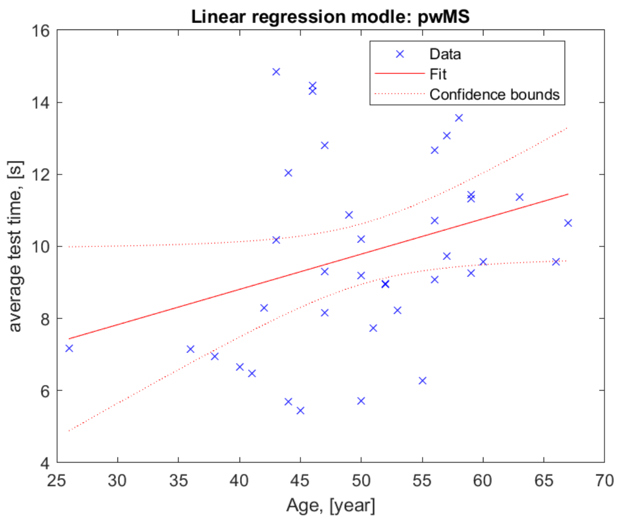


Figure 6.3: The linear regression of the mean test times and the age of the patients with MS.

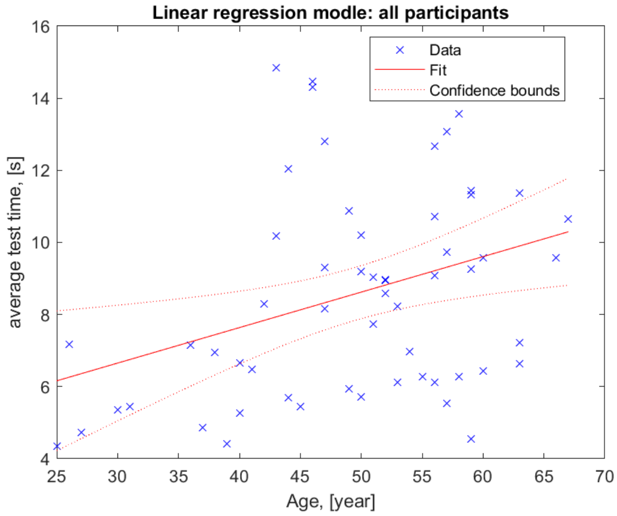


Figure 6.4: The linear regression of the mean test times and the age of all the participants.

From the linear regressions I found the adjusted r-squared ($r^2_{adj}$) corresponding to each of the linear regressions (Table 6.2).

| | $r^2_{adj}$ |
|---|---|
| **Control group** | 0.24968 |
| **Patient group** | 0.073353 |
| **All participants** | 0.098125 |

Table 6.2: The adjusted r-squared from the linear regressions done one the mean test times and the age for the controllers, the patients, and all the participants.

## 6.3 Fampridins effect on the test times

To see if the use of the medicine Fampridin affected the test times for the patients' SSSTs, I investigated the test times for the patients in Group 3. This is because the medicine is used by people with EDSS between 4 and 7. Out of the patients in Group 3, were 9 using the medicine, and 10 were not.
I started by looking at the mean test times for the patients (Figure 6.5) and the standard deviation of the test times for the patients (Figure 6.6). From these results, I found the mean and standard deviation of the mean test times for the patients using the medicine and the patients

that is not using it (Table 6.3).

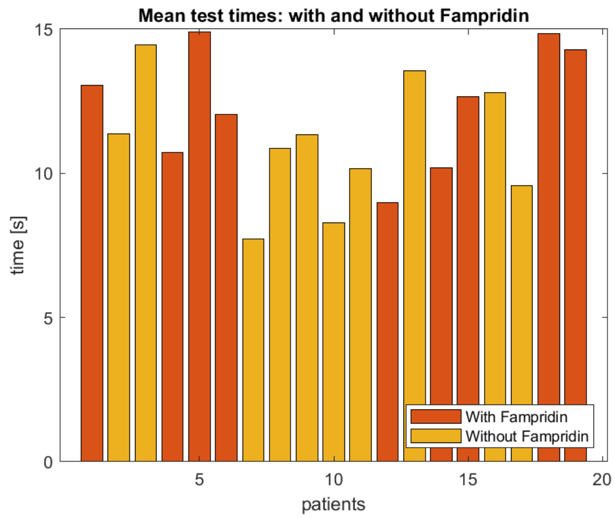


Figure 6.5: The mean test times for the participants in Group 3, color coordinated into two; use of Fampridin (With Fampridin) and not use of Fampridin (Without Fampridin)

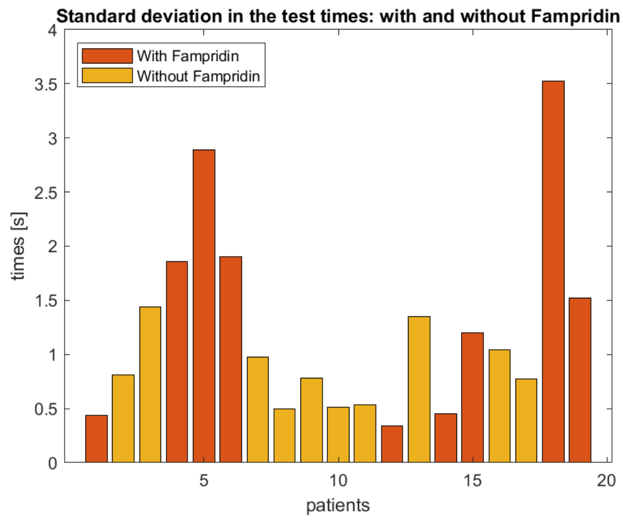


Figure 6.6: The standard deviation of the test times for the participants in Group 3, color coordinated into two; use of Fampridin (With Fampridin) and not use of Fampridin (Without Fampridin)

| | With Fampridin | Without Fampridin |
|---|---|---|
| $\mu[s]$ | 12.42 | 11.02 |
| $\sigma[s]$ | 2.12 | 2.18 |

Table 6.3: The mean ($\mu$) and standard deviation ($\sigma$) of the mean test times from the patient using Fampridin (With Fampridin) and the patient not using Fampridin (Without Fampridin).

After looking at the mean test times from all the tests, I found the mean times and the standard deviation for the D (Figures 6.7 and 6.8) and the ND (Figures 6.9 and 6.10). From these results, I found the mean and standers deviation from the D and the ND for the patients using the medicine and the patients that are not using it (Table 6.4)

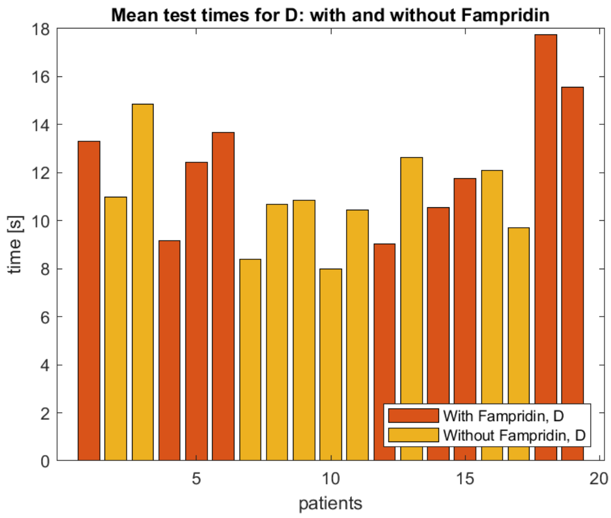


Figure 6.7: The mean test times from the dominant tests for the participants in Group 3, color coordinated into two; use of Fampridin (With Fampridin) and not use of Fampridin (Without Fampridin)

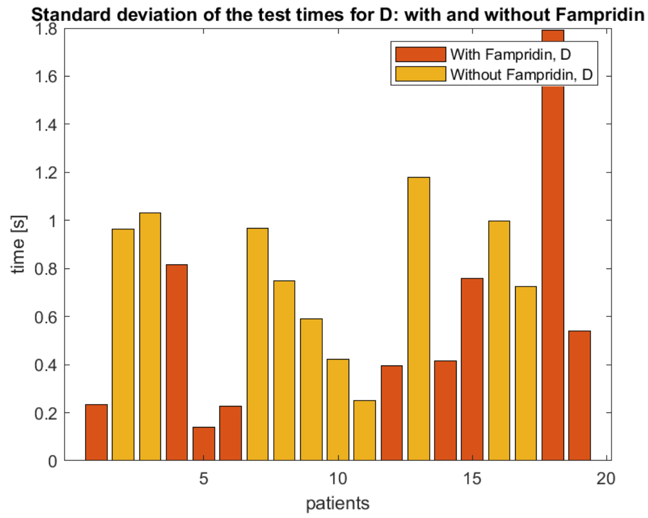


Figure 6.8: The standard deviation of the test times from the dominant tests for the participants in Group 3, color coordinated into two; use of Fampridin (With Fampridin) and not use of Fampridin (Without Fampridin)

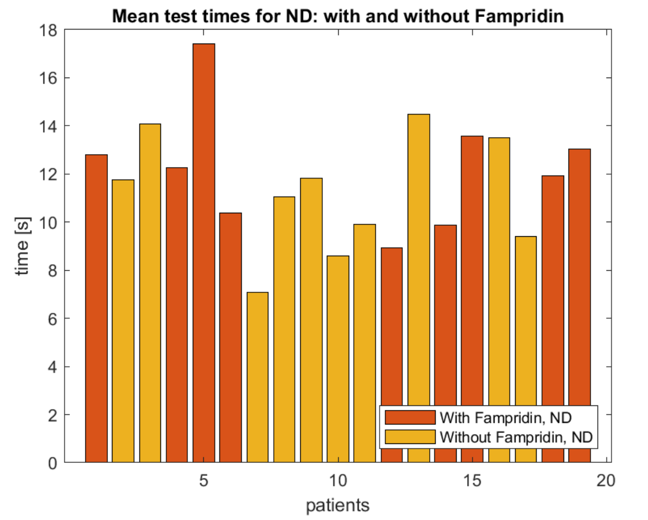


Figure 6.9: The mean test times from the non-dominant tests for the participants in Group 3, color coordinated into two; use of Fampridin (With Fampridin) and not use of Fampridin (Without Fampridin)

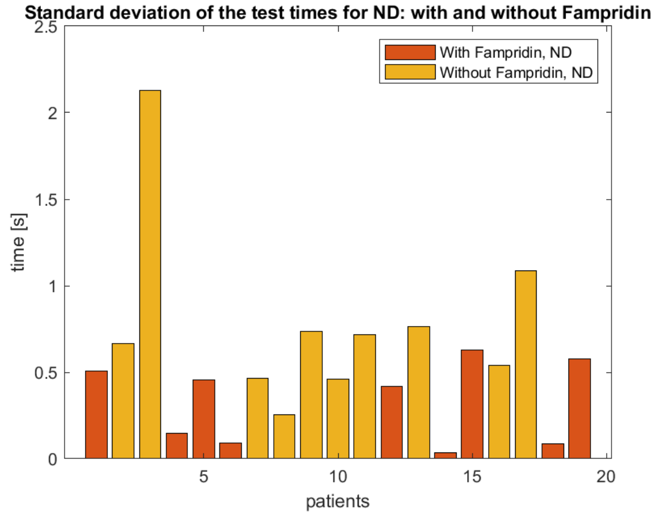


Figure 6.10: The standard deviation of the test times from the non-dominant tests for the participants in Group 3, color coordinated into two; use of Fampridin (With Fampridin) and not use of Fampridin (Without Fampridin)

| | **With Fampridin** | | **Without Fampridin** | |
|---|---|---|---|---|
| | **D** | **ND** | **D** | **ND** |
| $\mu[s]$ | 12.58 | 12.25 | 10.86 | 11.17 |
| $\sigma[s]$ | 2.89 | 2.49 | 2.02 | 2.44 |

Table 6.4: The mean ($\mu$) and standard deviation ($\sigma$) of the mean test times from the dominant (D) and the non-dominant (ND) tests, from the patient using Fampridin (With Fampridin) and the patient not using Fampridin (Without Fampridin).

I did the two-sample t-test on the mean test times of all the test times, the D and the ND, between the patients that used the medicine and the patients that did not use the medicine to test the null hypothesis $H_{0,1}$, $H_{0,2}$ and $H_{0,3}$. The t-test returned the p-values ($p$) and the result on if the null hypothesis was rejected ($h = 1$) or not ($h = 0$) (Table 6.5).

| | $p$ | $h$ |
|---|---|---|
| **All tests**, $H_{0,1}$ | 0.16 | 0 |
| **Dominant tests**, $H_{0,2}$ | 0.15 | 0 |
| **Non-dominant tests**, $H_{0,3}$ | 0.35 | 0 |

Table 6.5: The results from the two-sample t-test between the patient using Fampridion the the patient the do not use it. The test was done on the mean test time from all the test, the dominant tests and the non-dominant test, to test the null hypothesis $H_{0,1}$, $H_{0,2}$ and $H_{0,3}$. The result from these tests were the p-value ($p$) and if the null hypothesis were rejected ($h = 1$) or not ($h = 0$)

## 6.4 The mean test times for each of the participants

I found the mean test times for each of the participants (Figure 6.11) and their standard deviation (Figure 6.12). From the mean test times for the patients, I found the mean and standard deviation of the test times for the three groups (Table 6.6).

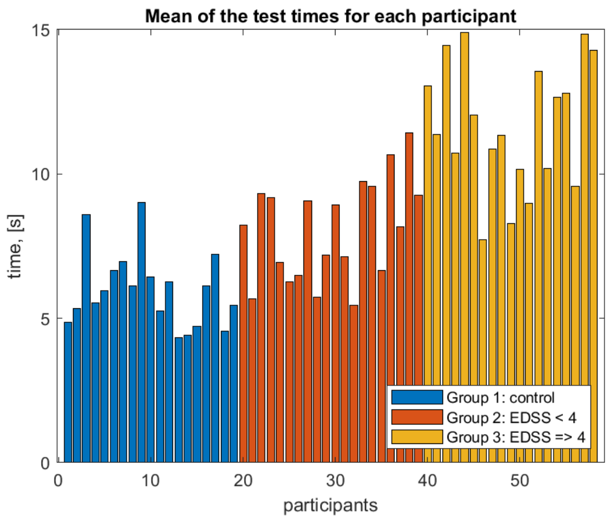


Figure 6.11: The mean test times for the participants, sorted in to the three groups and color coordinated.

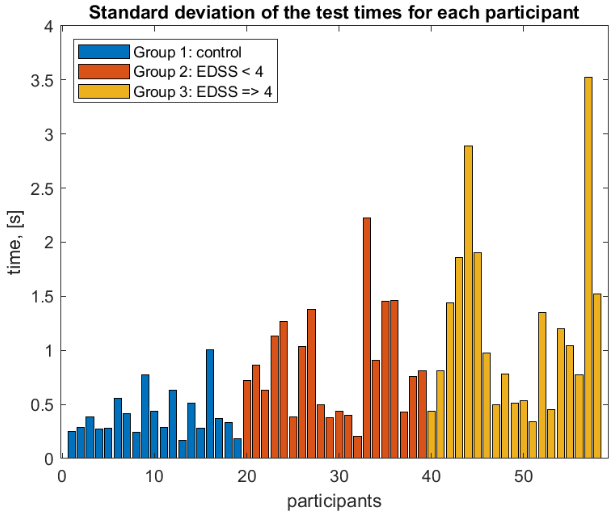


Figure 6.12: The standard deviation of the test times for the participants, sorted into the three groups and color coordinated.

| | $\mu$, $[s]$ | $\sigma$, $[s]$ |
|---|---|---|
| **Group 1** | 5.99 | 1.31 |
| **Group 2** | 8.06 | 1.74 |
| **Group 3** | 11.68 | 2.21 |

Table 6.6: The mean ($\mu$) and standard deviation ($\sigma$) of the mean test times for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

I then did a two-sample t-test to test the null hypothesis $H_{0,4}$. The t-test returned the p-values ($p$) and the decision result of if the $H_{0,4}$ were rejected ($h = 1$) or not ($h = 0$) (Table 6.7).

| | $p$ | $h$ |
|---|---|---|
| **Group 1, Group 2** | $p < 0.001$ | 1 |
| **Group 1, Group 3** | $p < 0.001$ | 1 |
| **Group 2, Group 3** | $p < 0.001$ | 1 |

Table 6.7: The p-values ($p$) and the decision result for rejecting or not the null hypothesis ($h$) of the two-sample t-test testing $H_{0,4}$.

### 6.4.1 Dominant vs non-dominant test times

After looking at all the tests together, I divided the four tests the participants did into two; the tests are done kicking with the dominant foot (D) and the non-dominant foot (ND).
I started by finding the mean and standard deviation of the test times for the D (Figures 6.13 and 6.15) and for the ND (Figures 6.14 and 6.16). The mean and standard deviation of the mean test times from the D and the ND for the three groups were also found (Table 6.8).

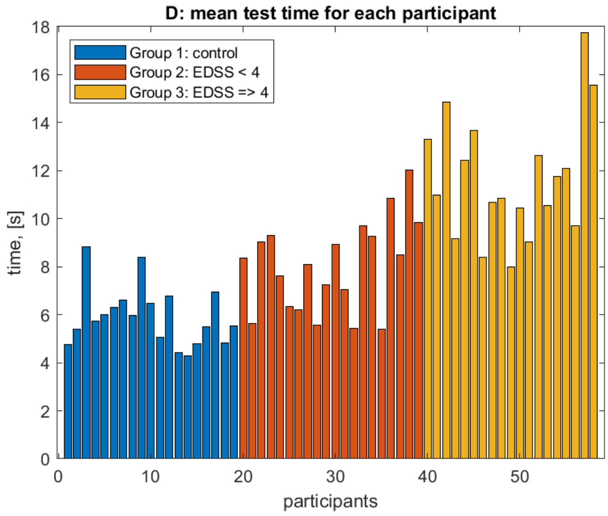


Figure 6.13: The mean test times of the two dominant (D) tests for the participants, sorted into the three groups and color coordinated.

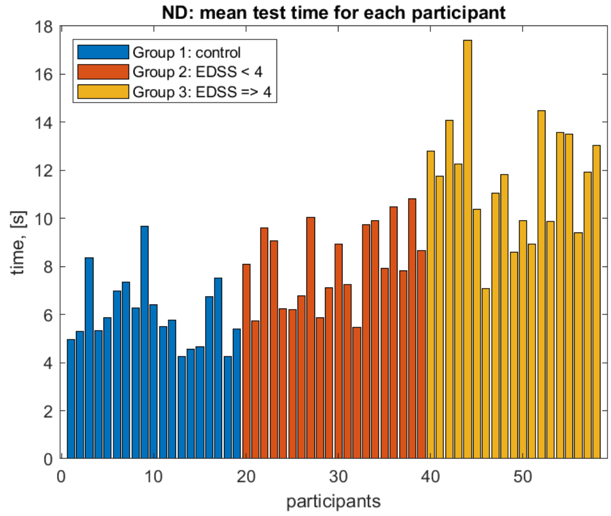


Figure 6.14: The mean test times of the two non-dominant (ND) tests for the participants, sorted into the three groups and color coordinated.

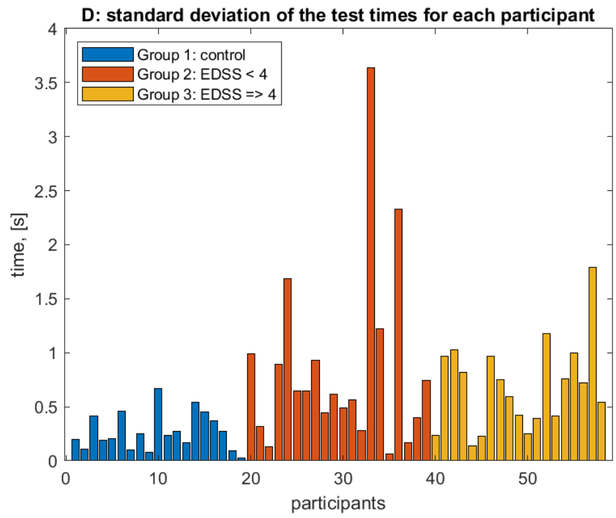


Figure 6.15: The standard deviation of the test times from the two dominant (D) tests for the participants, sorted into the three groups and color coordinated.

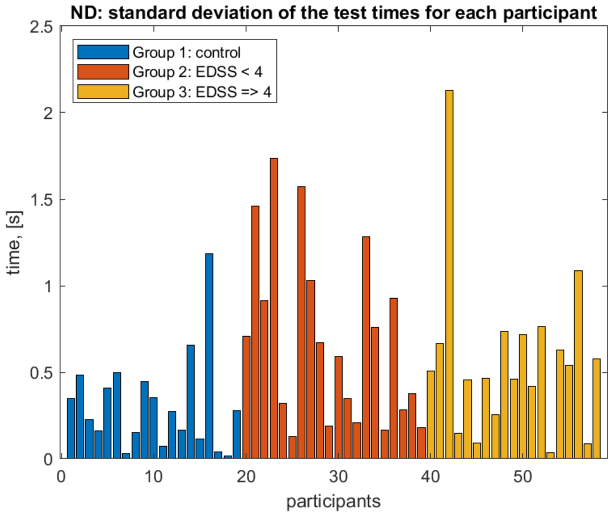


Figure 6.16: The standard deviation of the test times from the two non-dominant (ND) tests for the participants, sorted into the three groups and color coordinated.

| | **D**, $\mu \pm \sigma[s]$ | **ND**, $\mu \pm \sigma[s]$ |
|---|---|---|
| **Group 1** | $5.93 \pm 1.23$ | $6.06 \pm 1.44$ |
| **Group 2** | $8.02 \pm 1.90$ | $8.09 \pm 1.70$ |
| **Group 3** | $11.68 \pm 2.55$ | $11.68 \pm 2.46$ |

Table 6.8: The mean ($\mu$) and standard deviation ($\sigma$) from the dominant (D) and the non-dominant (ND) test, for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

After having found the mean and standard deviation for the D and the ND for the three groups, I found the mean and standard deviation for the three groups for the four test times (D1, D2, ND1, and ND2) (Table 6.9).

| | **D1**, $\mu \pm \sigma[s]$ | **D2**, $\mu \pm \sigma[s]$ | **ND1**, $\mu \pm \sigma[s]$ | **ND1**, $\mu \pm \sigma[s]$ |
|---|---|---|---|---|
| **Group 1** | $6.00\mu1.32$ | $5.86 \pm 1.18$ | $6.12 \pm 1.40$ | $5.99 \pm 1.54$ |
| **Group 2** | $8.62 \pm 2.22$ | $7.42 \pm 1.75$ | $8.33 \pm 1.89$ | $7.85 \pm 1.69$ |
| **Group 3** | $12.08 \pm 2.70$ | $11.28 \pm 2.47$ | $11.92 \pm 2.55$ | $11.44 \pm 2.45$ |

Table 6.9: The mean ($\mu$) and standard deviation ($\sigma$) for the four individual SSST (D1, D2, ND1 and ND2), for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

The difference in the mean test times between the D and the ND was then found, where I subtracted the mean test times of D from the mean test times of ND (Figure 6.17). If the result ended up being negative, the D mean test times are higher than ND.

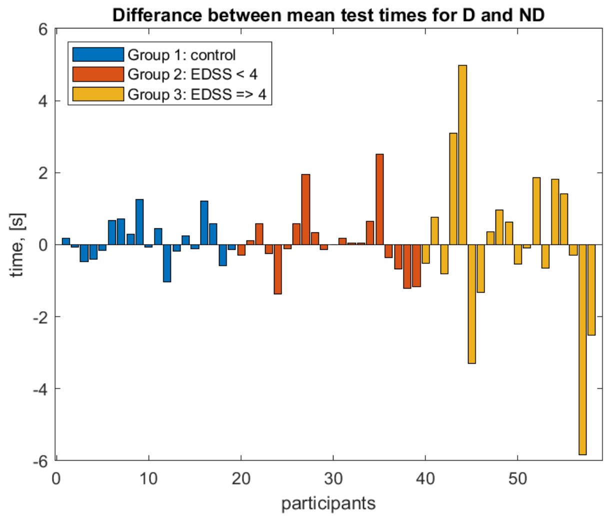


Figure 6.17: The difference in the mean test times between the non-dominant (ND) and the dominant (D) tests, sorted into the three groups and color coordinated.

I did a paired-sample t-test between the mean test times of the D and the ND (Table 6.10). I tested the null hypothesis $H_{0,5}$. The result from the test were the p-values ($p$) and a decision result of if the $H_{0,5}$ was rejected ($h = 1$) or not ($h = 0$).

| | D, ND | |
|---|---|---|
| | $p$ | $h$ |
| **Group 1** | 0.36 | 0 |
| **Group 2** | 0.74 | 0 |
| **Group 3** | 0.99 | 0 |

Table 6.10: The p-value ($p$) and the decision result for rejection or not the null hypothesis ($h$) of the paired-sample t-test testing $H_{0,5}$, for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

I then performed a two-sample t-test between the groups based on the mean test times from the D and the ND (Table 6.11) to test the null hypothesis $H_{0,6}$. The result from the test were the p-values ($p$) and a decision result of if the $H_{0,6}$ was rejected ($h = 1$) or not ($h = 0$).

| | D | | ND | |
|---|---|---|---|---|
| | $p$ | $h$ | $p$ | $h$ |
| **Group 1, Group 2** | $p < 0.001$ | 1 | $p < 0.001$ | 1 |
| **Group 1, Group 3** | $p < 0.001$ | 1 | $p < 0.001$ | 1 |
| **Group 2, Group 3** | $p < 0.001$ | 1 | $p < 0.001$ | 1 |

Table 6.11: The p-value ($p$) and the decision result for rejection or not the null hypothesis ($h$) of the two-sample t-test testing $H_{0,6}$.

## 6.5 Discussion

With the Chi-squares goodness-of-fit test, I found that the samples within the three groups, the mean of the test times and the test times from the four individual SSST, are normally distributed (Table 6.1).

I tried to find out if there were a correlation between the age and the mean test times for the patients, by performing three linear regressions (Figures 6.2, 6.3 and 6.4) on the controls, the patienst and all the participants. From these linear regressions, I found the adjusted r-squared (Table 6.2). The results show that there is no correlation between age and the mean test times. This can be because there may be young participants that are not that active and older participants that are very active. I have no data that informs me on how active the different participants are, so there is no way for me to look into if the activity of the participants correlate with the mean test time and age.

Then I looked into if the use of the medicine Fampridin affects the test times for the participants in Group 3. From the mean test times for the participants in Group 3 (Figure 6.5), it looks like there are no differences between the participants using the medicine and the participants not using the medicine. I found the mean and standard deviation of the mean test times for the participants using the medicine and for the participants that are not using it (Table 6.3), and it shows that the participants using the medicine has a higher mean than the participants not using it. The same were shown when I divided the test into the dominant tests and the non-dominant tests (Figures 6.7 and 6.9, and Table 6.4).
I did three two-sample t-tests to see if there were a big difference between the test times from the participants using Fampridin and the participants not using it by testing the null hypothesis $H_{0,1}$, $H_{0,2}$ and $H_{0,3}$ (Table 6.5). All three of the null hypothesis were accepted. This means that the test times from the participants using the medicine and the participants not using it are from a normal distribution with equal mean and variance.
In these analyses, I only had information on if the patient were on the medicine or not. I had no information on how long they had been on it and how bad their mobility was before starting on the medicine. For the participants on the medicine, their test times might have been worse if they had not been using Fampridin.

The result from Figure 6.11 and the Table 6.6 shows that the mean test times increases with the groups. There is an approximately even increase in the mean test time between the groups. The results from the two-sample t-test testing the null hypothesis $H_{0,4}$ (Table 6.7) showed that all the tests rejected the null hypothesis. This means that the three groups form three normal distributions with different mean and variance. If one looks at the mean $\pm$ the standard deviation of mean test times for the three groups (Table 6.6), there are big overlapping regions. Even though the three groups are from different normal distributions, with the big overlapping regions, it is hard to find a decision threshold that will give a good classification into the three groups based on the participants' mean test time.

After looking at the mean test times, I divided the tests into dominant (D) and non-dominant (ND) tests to see how the test times within the dominant and the non-dominant tests differ from each other (Figure 6.17). For both the dominant and the non-dominant tests, the mean and standard deviation for the three groups (Table 6.8) are approximately the same as they were for the mean test times for all the tests. The non-dominant tests have some higher mean test times for Group 1 and 2, but for Group 3, the mean is the same as for the dominant tests. I also looked at the four individual tests (D1, D2, ND1, and ND2) and found a decrease in the mean test times for the second dominant and non-dominant tests (Table 6.9). This will be investigated further later in the project (Chapter 9).

I did a paired-sample t-test between the dominant and the non-dominant tests to test the null hypothesis $H_{0,5}$ (Table 6.10). The results showed that the null hypothesis was accepted for all three groups. I also did two-sample t-tests to test the null hypothesis $H_{0,6}$ and $H_{0,6}$ (Table 6.11). The result showed that all the tests rejected the null hypothesis for both the dominant and the non-dominant tests. This means that for the mean test times from the dominant and the non-dominant tests, the three groups are still three different normal distributions with different mean and variance, but the difference in the test times between the dominant and the non-dominant tests are from normal distributions with mean$= 0$. The mean $\pm$ standard deviation of the dominant and non-dominant tests times for the three groups (Table 6.9) are still having a big overlapping region between the different groups. So there is not difference when it comes to classifying participants into the three groups based on the mean dominant/non-dominant test times, as it was with the mean test times.

To sum up, the test times to the participants are from different normal distributions based on the participants' EDSS, but this does not mean that one can classify the participants into the three groups based on their test times. With their mean $\pm$ standard deviations in the test times, there is too big of an overlapping region between the three groups. There is not a big difference between the test times from the dominant tests and the non-dominant tests.

# Chapter 7

# The kicking times

This chapter will analyze the time the participants use to kick in the Six Spot Step Test (SSST) from the week 1 tests. The kick is defined to start when the kicking foot makes contact with the cube and ends when the cube is out of the marked circle, and the kicking foot is touching the floor.
The participants will be divided into three groups based on their EDSS (Expanded Disability Statue Scale) (Section 5.3). The three groups will be:

- Group 1: the controls
- Group 2: the patients with an EDSS $< 4$
- Group 3: the patients with an EDSS $\geq 4$

During the analyze I looked at the mean kicking time for the participants based on

- all the kicking times from all four tests
- all the kicking times from the kicks done with the dominant foot (D)
- all the kicking times from the kicks done with the non-dominant foot (ND)
- all the kicking times from one test (D1, D2, ND1, ND2)
- both times for the first kick ($K1_D$), second kick ($K2_D$), third kick ($K3_D$), fourth kick ($K4_D$), or fifth kick ($K5_D$), done with the dominant foot
- both times for the first kick ($K1_{ND}$), second kick ($K2_{ND}$), third kick ($K3_{ND}$), fourth kick ($K4_{ND}$), or fifth kick ($K5_{ND}$), done with the non-dominant foot

I started by investigating if the mean kicking times between the three groups were different. Then I investigated the difference in the kicking times between kicking with the dominant foot (D) and the non-dominant foot (ND). I also looked at if the mean kicking times changed

between the four tests (D1, D2, ND1, ND2) and if the kicking times changed between which kick within the tests were different.
During these investigations, I tested different null hypothesis $H_0$ with significant level $\alpha$:

- the mean kicking times in Group $x$ and $y$ are both independent normal distributed, with equal mean and variance ($H_{0,1}$, with $\alpha = 0.05$).
- the difference between the mean D kicking times and the mean ND kicking times is normally distributed with mean$= 0$ ($H_{0,2}$, with $\alpha = 0.05$).
- the mean D kicking times in Group $x$ and $y$ are both independent normal distributed, with equal mean and variance ($H_{0,3}$, with $\alpha = 0.05$).
- the mean ND kicking times in Group $x$ and $y$ are both independent normal distributed, with equal mean and variance ($H_{0,4}$, with $\alpha = 0.05$).

## 7.1 The mean kicking time from all the tests

I started by looking at the mean kicking time from all the kicks the participants did during the four tests in week 1 (Figure 7.1) and the standard deviation of the kicking times (Figure 7.2). Then I found the mean and standard deviation of the kicking time for the three groups based on the mean kicking time to the participants in each group (Table 7.1).

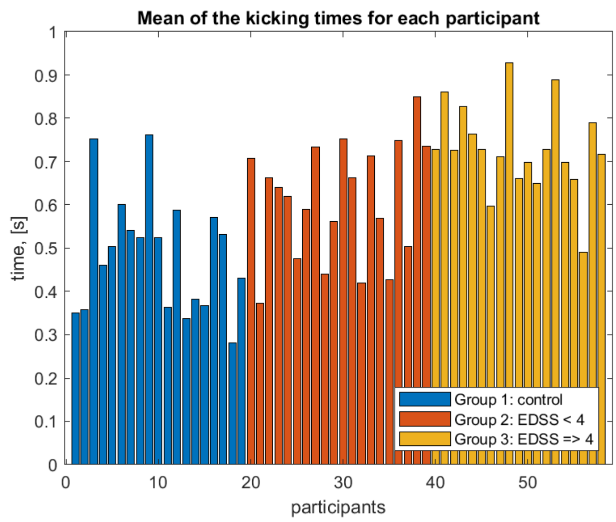


Figure 7.1: The mean kicking time from all the tests, for all the participants. The participants are sorted in the the three groups based on there EDSS, and color-coordinated.

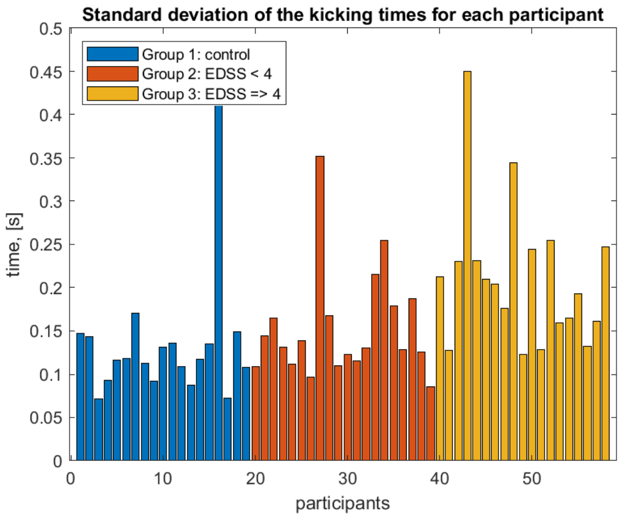


Figure 7.2: The standard deviation of the kicking times from all the tests for each of the participants. The participants are sorted into the three groups based on there EDSS, and color-coordinated.

| | $\mu$, $[s]$ | $\sigma$, $[s]$ |
|---|---|---|
| **Group 1** | 0.49 | 0.14 |
| **Group 2** | 0.61 | 0.13 |
| **Group 3** | 0.73 | 0.10 |

Table 7.1: The mean ($\mu$) and standard deviation ($\sigma$) of the mean kicking times for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

With the mean kicking time for each of the participants, I did a two-sample t-test between the three groups (Table 7.2) to test the null hypothesis $H_{0,1}$. The result from the test were the p-values ($p$) and a decision result of if the $H_{0,1}$ was rejected ($h = 1$) or not ($h = 0$).

| | $p$ | $h$ |
|---|---|---|
| **Group 1, Group 2** | 0.007 | 1 |
| **Group 1, Group 3** | $p < 0.001$ | 1 |
| **Group 2, Group 3** | 0.004 | 1 |

Table 7.2: The p-value ($p$) and the decision result for rejecting or not the null hypothesis ($h$) of the two-sample t-test testing $H_{0,1}$.

## 7.2 The mean kicking time from the dominant and non-dominant tests

After investigating the mean kicking time from all the kicks, I divided the kicks into two groups; the kicks done with the dominant foot (D) and the kicks done with the non-dominant foot (ND). From this I found the mean and standard deviation of the kicking times for each of the participants when using the D (Figures 7.3 and 7.5) and when using the ND (Figures 7.4 and 7.6).

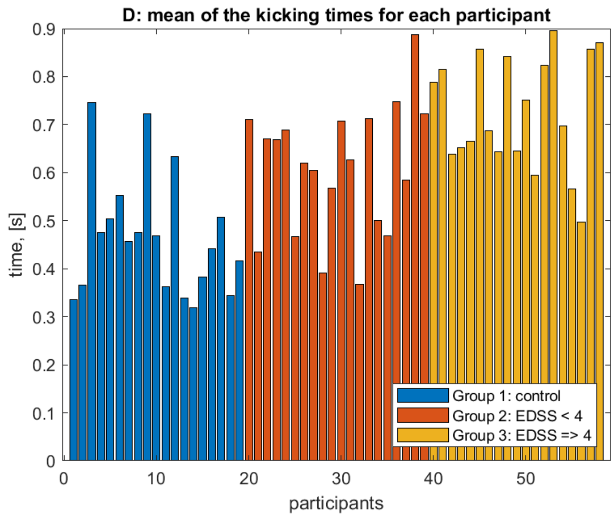


Figure 7.3: The mean kicking time from all the kicks done with the dominant foot (D), for all the participants. The participants are sorted into the three groups based on there EDSS, and color-coordinated.

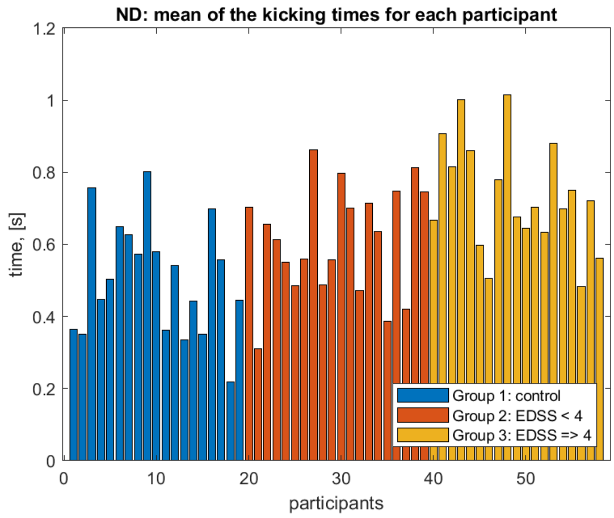


Figure 7.4: The mean kicking time from all the kicks done with the non-dominant foot (ND), for all the participants. The participants are sorted into the three groups based on there EDSS, and color-coordinated.

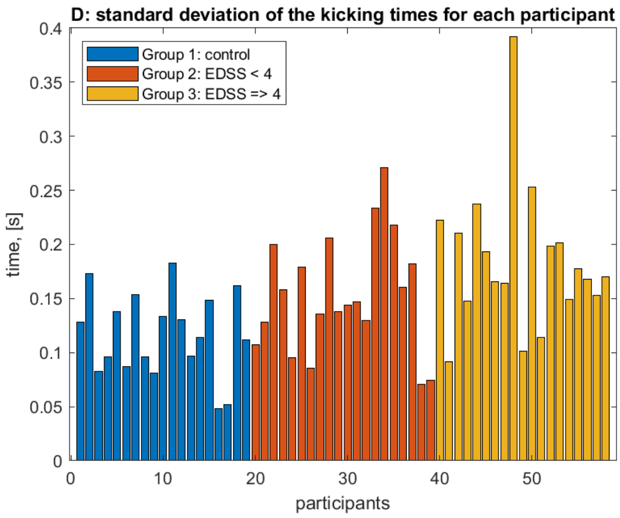


Figure 7.5: The standard deviation for the times from kicking with the dominant (D) foot, for each of the participants. The participants are sorted into the three groups based in there EDSS, and color-coordinated.

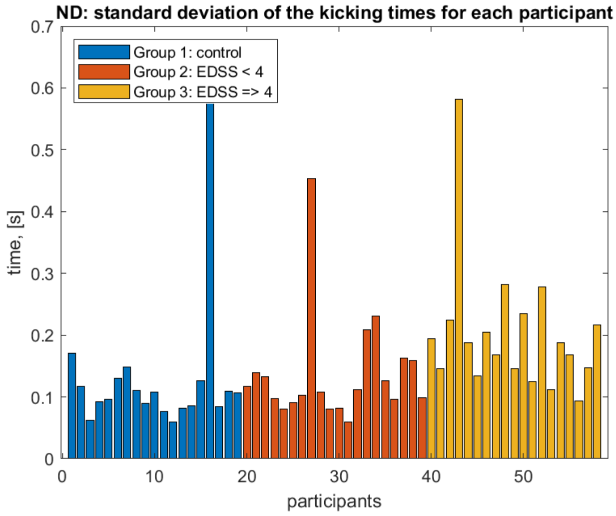


Figure 7.6: The standard deviation for the times from kicking with the non-dominant (ND) foot, for each of the participants. The participants are sorted into the three groups based in there EDSS, and color-coordinated.

From the mean of the kicking time for the kicks done with the D foot and the ND foot, I found the mean and standard deviation of the mean kicking time for the three groups (Table 7.3).

| | **D**, $\mu \pm \sigma[s]$ | **ND**, $\mu \pm \sigma[s]$ |
|---|---|---|
| **Group 1** | $0.47 \pm 0.17$ | $0.51 \pm 0.22$ |
| **Group 2** | $0.61 \pm 0.20$ | $0.61 \pm 0.21$ |
| **Group 3** | $0.73 \pm 0.22$ | $0.73 \pm 0.27$ |

Table 7.3: The mean ($\mu$) and standard deviation ($\sigma$) of the kicking time based on the kicks done with the dominant foot (D) and the non-dominant foot (ND), for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

The difference in the mean kick times between the D kicks and the ND kicks for each of the participants (Figure 7.7) was also found by subtracting the mean kicking time of D kicks from the mean kicking time of ND kicks. If this result became negative, it means that the mean kicking time of D kicks is higher than that of the ND kicks.

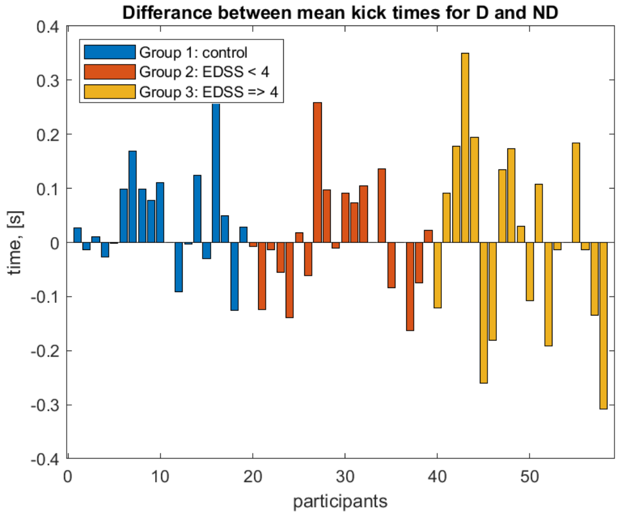


Figure 7.7: The difference in the mean kicking time between kicks done with the dominant (D) foot and the non-dominant (ND) foot, for each of the participants. The participants are divided into three groups and color-coordinated.

To see if the mean kicking time were the same in both the dominant (D1 and D2) and non-dominant (ND1 and ND2) tests I found the mean and standard deviation of the mean kicking time from each tests for the three groups (Table 7.4).

| | **D1**, $\mu \pm \sigma[s]$ | **D2**, $\mu \pm \sigma[s]$ | **ND1**, $\mu \pm \sigma[s]$ | **ND1**, $\mu \pm \sigma[s]$ |
|---|---|---|---|---|
| **Group 1** | $0.47 \pm 0.17$ | $0.47 \pm 0.17$ | $0.51 \pm 0.28$ | $0.51 \pm 0.18$ |
| **Group 2** | $0.63 \pm 0.21$ | $0.59 \pm 0.19$ | $0.63 \pm 0.23$ | $0.60 \pm 0.20$ |
| **Group 3** | $0.73 \pm 0.21$ | $0.72 \pm 0.23$ | $0.74 \pm 0.30$ | $0.72 \pm 0.22$ |

Table 7.4: The mean ($\mu$) and standard deviation ($\sigma$) for the kicking done in the four individual tests (D1, D2, ND1, and ND2) for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

After looking at the mean kicking time for the different tests and the groups, I did a paired-sample t-test to test the null hypothesis $H_{0,2}$ (Table 7.5). The result from the test were the p-values ($p$) and a decision result of if the $H_{0,2}$ was rejected ($h = 1$) or not ($h = 0$).

| | D, ND, $H_{0,2}$ | |
|---|---|---|
| | $p$ | $h$ |
| **Group 1** | 0.07 | 0 |
| **Group 2** | 0.88 | 0 |
| **Group 3** | 0.89 | 0 |

Table 7.5: The p-value ($p$) and the decision result for rejection or not the null hypothesis ($h$) of the paired-sample t-test testing $H_{0,2}$, for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

I also did a two-sample t-test (Table 7.6) to test the null hypothesis $H_{0,3}$ and $H_{0,4}$. The result from the test were the p-values ($p$) and a decision result of if the $H_{0,3}/H_{0,4}$ was rejected ($h = 1$) or not ($h = 0$).

| | **D**, $H_{0,3}$ | | **ND**, $H_{0,4}$ | |
|---|---|---|---|---|
| | $p$ | $h$ | $p$ | $h$ |
| **Group 1, Group 2** | 0.002 | 1 | 0.04 | 1 |
| **Group 1, Group 3** | $p < 0.001$ | 1 | $p < 0.001$ | 1 |
| **Group 2, Group 3** | 0.006 | 1 | 0.018 | 1 |

Table 7.6: The p-value ($p$) and the decision result for rejection or not the null hypothesis ($h$) of the two-sample t-test testing $H_{0,3}$ and $H_{0,4}$.

## 7.3 The mean kicking times for the five individual kicks within the tests

There were five kicks within each of the tests ($K1$, $K2$, $K3$, $K4$, and $K5$). I wanted to investigate if the participants' time for each kick differed between the kicks and between the three groups of participants. So I found the mean kicking times for each kicking within the tests based on the participants' kicking times. This I did for the kicks done with the dominant foot (Table 7.7) and for the kicks done with the non-dominant foot (Table 7.8).

| | Group 1 $\mu \pm \sigma, [s]$ | Group 2 $\mu \pm \sigma, [s]$ | Group 3 $\mu \pm \sigma, [s]$ |
|---|---|---|---|
| $K1_D$ | $0.55 \pm 0.13$ | $0.72 \pm 0.15$ | $0.82 \pm 0.19$ |
| $K2_D$ | $0.40 \pm 0.20$ | $0.54 \pm 0.23$ | $0.69 \pm 0.25$ |
| $K3_D$ | $0.52 \pm 0.11$ | $0.66 \pm 0.16$ | $0.75 \pm 0.22$ |
| $K4_D$ | $0.37 \pm 0.17$ | $0.51 \pm 0.24$ | $0.65 \pm 0.19$ |
| $K5_D$ | $0.49 \pm 0.17$ | $0.62 \pm 0.16$ | $0.71 \pm 0.21$ |

Table 7.7: The mean ($\mu$) and the standard deviation ($\sigma$) of the kicking time for each of the five kicks done with the dominant foot, separated into the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

| | Group 1 $\mu \pm \sigma, [s]$ | Group 2 $\mu \pm \sigma, [s]$ | Group 3 $\mu \pm \sigma, [s]$ |
|---|---|---|---|
| $K1_{ND}$ | $0.50 \pm 0.18$ | $0.60 \pm 0.20$ | $0.74 \pm 0.24$ |
| $K2_{ND}$ | $0.56 \pm 0.16$ | $0.66 \pm 0.18$ | $0.76 \pm 0.23$ |
| $K3_{ND}$ | $0.43 \pm 0.19$ | $0.58 \pm 0.20$ | $0.69 \pm 0.24$ |
| $K4_{ND}$ | $0.49 \pm 0.14$ | $0.60 \pm 0.16$ | $0.81 \pm 0.34$ |
| $K5_{ND}$ | $0.55 \pm 0.39$ | $0.62 \pm 0.30$ | $0.66 \pm 0.23$ |

Table 7.8: The mean ($\mu$) and the standard deviation ($\sigma$) of the kicking time for each of the five kicks done with the non-dominant foot, separated into the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

## 7.4 Discussion

There is a variation within the groups when looking at the mean kicking times of all the kicks done by the participants (Figure 7.1). I found from the mean and standard deviation of the mean kicking times within the three groups (Table 7.1) that the mean kicking times increase with the groups. All three of the groups have approximately the same standard deviation.
Looking at the standard deviation in the kicking time for each of the participants (Figure 7.2), there are a few participants in all three groups that have a higher standard deviation than the rest of the participants. This is most likely because the participants have missed one kick and therefore have to repeat it to complete the kick.
The results from the two-sample t-test testing the null hypothesis $H_{0,1}$ (Table 7.2) show that all the tests reject the $H_{0,1}$. This means that the three groups are normal distributed with different

mean and variance. Even though the mean kicking times for the different groups are from different normal distributions, with the mean ± the standard deviation, there are big overlapping regions in the mean kicking times between the three groups. This overlap is too big to set a decision threshold to classify participants into the three groups based on the mean kicking times.

After I had looked at the mean kick times for all the kicks, I wanted to see if there were any differences between the kicks done with the dominant foot and the kicks done with the non-dominant foot. From seeing the mean kicking time for the different feet (Figures 7.3 and 7.4), both feet have the same pattern when it comes to looking at the difference between the groups and how the mean kicking time varies within the groups. There is a small difference between the kicks done with the dominant foot and the non-dominant foot. This is more clearly shown when looking at the mean and standard deviation of the mean kicking times from the dominant and the non-dominant foot for the three groups (Table 7.3). It shows that the mean kicking time increases with the groups, for both feet, and so does the standard deviation. The kicks done with the non-dominant foot have a higher standard deviation than the kicks done with the dominant foot, and Group 1 also has a higher mean kicking time.
For the standard deviation in the kicking time for the individual participants (Figures 7.5 and 7.6), there are three participants with higher standard deviation than the others when kicking with the non-dominant foot. This shows that the high standard deviation in the mean kicking time for some of the participants (Figure 7.2) is most likely from one or more of the kicks done with the non-dominant foot.
When looking at the difference in the mean kicking time between the dominant and non-dominant foot for each of the participants (Figures 7.7), there is no immediate pattern that emerges. Some participants use less time to kick with the dominant foot, and some participants use less time to kick with the non-dominant foot. From the mean and standard deviation for the three groups of the mean kicking times from dominant kicks and non-dominant kicks (Table 7.3) are there shown that there is no difference in the kicking time between kicking with the dominant and non-dominant foot for the participants in Group 2 and 3. For Group 1, a small increase in the mean kicking time between kicks is done with the dominant foot and the non-dominant foot.
I also looked at how the mean kicking time was for the four individual kicks in the tests for the three groups (Table 7.4). The same trend seen when looking at the other mean kicking times shows up here where the mean kicking time increases with the groups. For Groups 2 and 3, the mean kicking time decreases between the first and second tests done with both the dominant and non-dominant feet as the kicking foot. The same happens with the standard deviation. This is investigated further later in the project (Chapter 9).

From the paired-sample t-tests testing the null hypothesis $H_{0,2}$ (Table 7.5), I found that all the tests accepted the null hypothesis $H_{0,2}$. This means that the differences between D kicks and ND kicks are small and from a normal distribution where the mean is zero.

When it comes to the two-sample t-test I did to test the null hypothesis $H_{0,3}$ and $H_{0,4}$ (Table 7.6), the result shows that all the tests rejected the null hypothesis. This means that the three groups' dominant and non-dominant kicking times are from normal distributions with different mean and variance. The tests of the $H_{0,4}$ between Group 1 and 2, and between Group 2 and 3, would have been accepted if I had sat the significant level to be 0.01. By changing the significance level, I would have had stricter requirements to reject the null hypothesis.

Lastly, I wanted to see how the different kicks within the tests compared to each other for the kicks done with the dominant foot (Table 7.7) and the non-dominant foot (Table 7.8). The mean kicking time between $K1_D$ and $K2_D$, and between $K3_D$ and $K4_D$, decreases for the three groups. The mean time between $K2_D$ and $K3_D$, and between $K4_D$ and $K5_D$, increases for all the groups. This makes sense since $K2_D$ and $K4_D$ are done in the same way, and $K1_D$ and $K3_D$ in the same way for the individual participants.
For the kicks done with the non-dominant foot, the pattern is reversed, where the mean kicking time between $K1_{ND}$ and $K2_{ND}$ and between $K3_{ND}$ and $K4_{ND}$ increases. For the time between $K2_{ND}$ and $K3_{ND}$, the mean kick time decreases for all the groups. For Group 1 and 2 are the mean kick time increase between $K4_{ND}$ and $K5_{ND}$, but for Group 3 are the mean kick time decrease.

To sum up, the mean kicking time between kicks done with the dominant foot and the non-dominant foot has a minimal and not significant difference. There is also a difference in the mean kicking time between the three groups, where Group 1 has the lower mean kicking time and Group 3 has the higher mean kicking time. The mean $\pm$ standard deviation of the kicking times for the three groups has a big overlapping region, making it hard to classify a participant into one of the groups based on the mean kicking time.

# Chapter 8

# The walking times

This chapter will analyze the time the participants use to walk in the Six Spot Step Test (SSST) from the week 1 tests. The walking parts are defined to start when the kicking foot touches the floor after a kick and end when the kicking foot makes contact with the cube. The first walking part in the tests starts when the test starts.
The participants will be divided into three groups based on their EDSS (Expanded Disability Statue Scale) (Section 5.3). The three groups will be:

- Group 1: the controls
- Group 2: the patients with an EDSS $< 4$
- Group 3: the patients with an EDSS $\geq 4$

During the analysis, I looked at the mean time of the walking parts for the participants based on

- the times of all the walking parts from all four tests
- the times of the walking parts from the two tests where they kick with their dominant foot (D)
- the times of the walking parts from the two tests where they kick with their non-dominant foot (ND)
- the times of the walking parts from one test (D1, D2, ND1, ND2)
- both the times of the first walking part ($W1_D$), second walking part ($W2_D$), third walking part ($W3_D$), fourth walking part ($W4_D$), or fifth walking part ($W5_D$), from the two tests where they kick with there dominant foot

- both the times of the first walking part ($W1_{ND}$), second walking part ($W2_{ND}$), third walking part ($W3_{ND}$), fourth walking part ($W4_{ND}$), or fifth walking part ($W5_{ND}$), from the two tests where they kick with there non-dominant foot

I started by investigating if the mean time for the walking parts between the three groups were different. Then I investigated the differences in the times between the walking parts in the dominant tests (D) and the non-dominant tests (ND). I also looked at if the mean times for the walking parts changed between the four tests (D1, D2, ND1, ND2) and if the times for the different walking parts within the tests were different from each other.
During these investigations, I tested different null hypothesis $H_0$ with significant level $\alpha$:

- the mean times of the walking parts in Group $x$ and $y$ are both independent normal distributed, with equal mean and variance ($H_{0,1}$, with $\alpha = 0.05$).
- the difference between the mean time of the waling parts in D and ND is normally distributed with mean$= 0$ ($H_{0,2}$, with $\alpha = 0.05$).
- the mean time of the walking part from D in Group $x$ and $y$ are both independent normal distributed, with equal mean and variance ($H_{0,3}$, with $\alpha = 0.05$).
- the mean time of the walking part from ND in Group $x$ and $y$ are both independent normal distributed, with equal mean and variance ($H_{0,4}$, with $\alpha = 0.05$).

## 8.1 The mean time for the walking part from all the tests

The first thing I did was to find the mean time of the walking parts for all of the participants, based on all the walking parts in all four tests (Figure 8.1). I also found the standard deviation of the time for the walking parts to each of the participants (Figure 8.2).

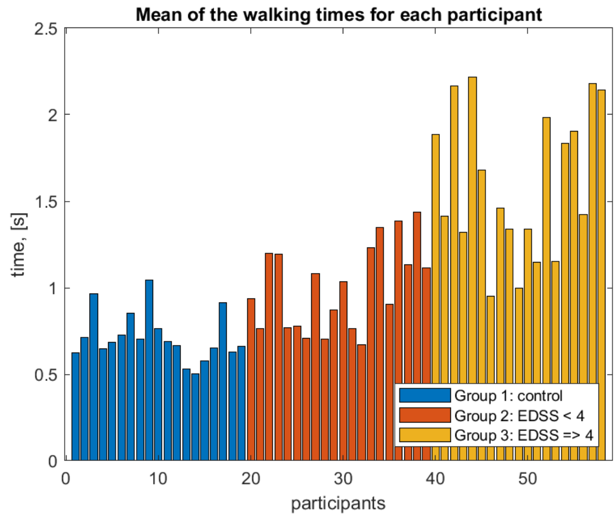


Figure 8.1: The mean time for the walking part from all the tests for all the participants. The participants are sorted into the three groups based on their EDSS and color-coordinated.

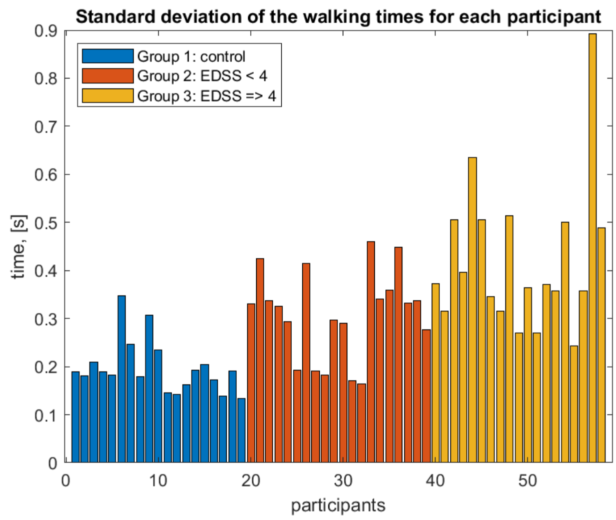


Figure 8.2: The standard deviation of the time for the walking part from all the tests for all the participants. The participants are sorted into the three groups based on there EDSS, and color-coordinated.

From the mean time of the walking parts for the participants I found the mean time for the three groups, and their standard deviation (Table 8.1).

| | $\mu$, $[s]$ | $\sigma$, $[s]$ |
|---|---|---|
| **Group 1** | 0.71 | 0.14 |
| **Group 2** | 1.00 | 0.25 |
| **Group 3** | 1.61 | 0.42 |

Table 8.1: The mean ($\mu$) and standard deviation ($\sigma$) of the mean time for the walking parts for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

To see if there were any connection between the mean time for the walking parts between the participants in the three groups, I did a two-sample t-test. The t-test tested the null hypothesis $H_{0,1}$ (Table 8.2). The result from the test were the p-values ($p$) and a decision whether the $H_{0,1}$ was rejected ($h = 1$) or not ($h = 0$).

| | $p$ | $h$ |
|---|---|---|
| **Group 1, Group 2** | $p < 0.001$ | 1 |
| **Group 1, Group 3** | $p < 0.001$ | 1 |
| **Group 2, Group 3** | $p < 0.001$ | 1 |

Table 8.2: The p-value ($p$) and the decision result for rejecting or not the null hypothesis ($h$) of the two-sample t-test testing $H_{0,1}$.

## 8.2 The mean time for the walking parts from the dominant and non-dominant tests

After looking at the mean time for the walking part based on all the tests, I divided the tests into two; the dominant tests (D) and the non-dominant tests (ND). Then I found the mean and standard deviation of the time for the walking parts, for each of the participants separated into D tests (Figures 8.3 and 8.5) and ND tests (Figures 8.4 and 8.6).

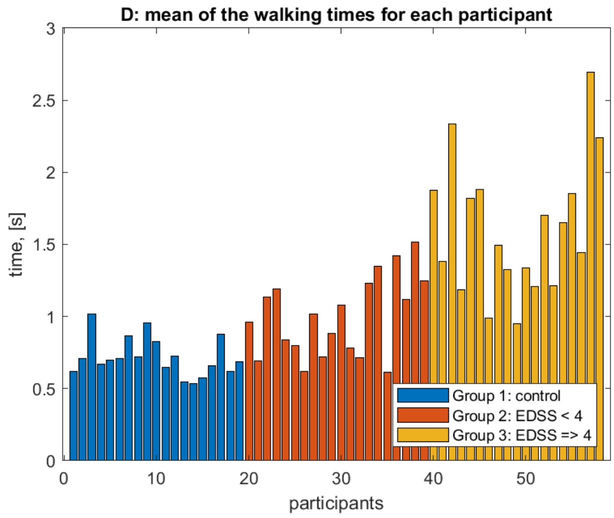


Figure 8.3: The mean time of the walking parts from the two dominant (D) tests, for all the participants. The participants are sorted into the three groups based on there EDSS, and coloc-coordinated.

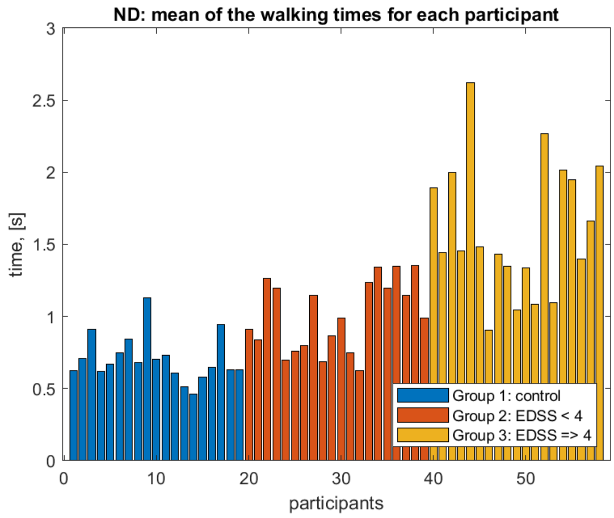


Figure 8.4: The mean time of the walking parts from the two non-dominant (ND) tests, for all the participants. The participants are sorted into the three groups based on there EDSS, and coloc-coordinated.

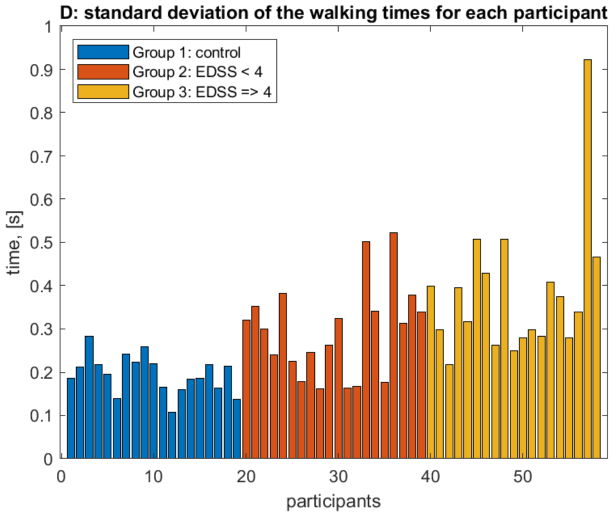


Figure 8.5: The standard deviation for the times for the walking parts done in the two dominant (D) tests, for each of the participants. The participants are sorted into the three groups based on there EDSS, and color-coordinated.

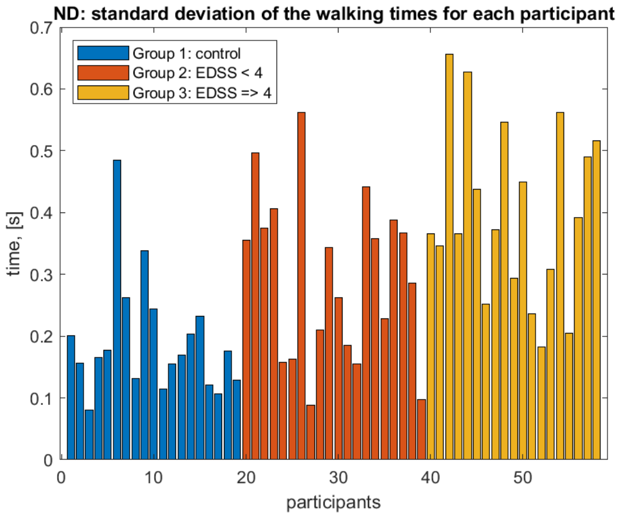


Figure 8.6: The standard deviation for the times for the walking parts done in the two non-dominant (ND) tests, for each of the participants. The participants are sorted into the three groups based on there EDSS, and color-coordinated.

Then I found the mean and standard deviation of the time for the walking parts, for the three groups, based on the mean time of the walking parts from the D and the ND test (Table 8.3). I also found the difference in the mean time of the walking parts between the D and the ND tests for all the participants (Figure 8.7) by subtracting the mean time of the walking parts in D from the mean time of the walling parts in ND. If the difference becomes negative, the mean time for the walking parts in D is higher than in ND.

| | **D**, $\mu \pm \sigma[s]$ | **ND**, $\mu \pm \sigma[s]$ |
|---|---|---|
| **Group 1** | $0.72 \pm 0.23$ | $0.71 \pm 0.25$ |
| **Group 2** | $1.00 \pm 0.40$ | $1.01 \pm 0.39$ |
| **Group 3** | $1.61 \pm 0.60$ | $1.61 \pm 0.60$ |

Table 8.3: The mean ($\mu$) and standard deviation ($\sigma$) of the time of the walking parts based on the walking parts from the dominant (D) and non-dominant (ND) tests, from the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

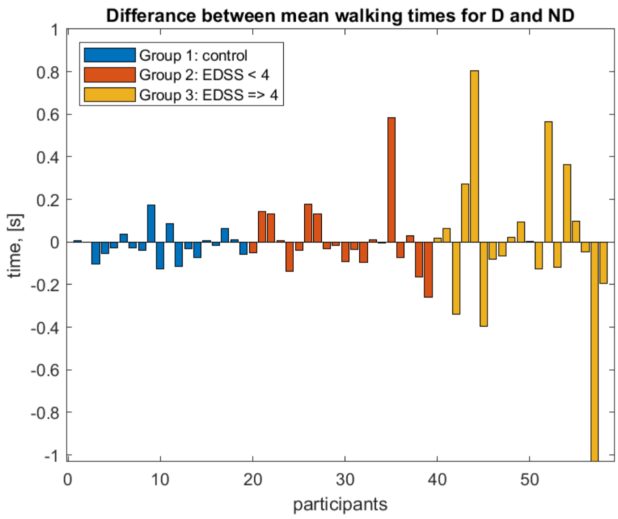


Figure 8.7: The difference in the mean time of the walking parts between the parts from the dominant (D) tests and the non-dominant (ND) tests, for each of the participants. The participants are sorted into the three groups based on there EDSS, and color-coordinated.

I wanted to see if the mean time for the walking parts changed between the four tests, so I found the mean and standard deviation of the time for the walking parts in the four tests (D1, D2, ND1, and ND2) for the three groups (Table 8.4).

| | **D1**, $\mu \pm \sigma[s]$ | **D2**, $\mu \pm \sigma[s]$ | **ND1**, $\mu \pm \sigma[s]$ | **ND2**, $\mu \pm \sigma[s]$ |
|---|---|---|---|---|
| **Group 1** | $0.73 \pm 0.24$ | $0.71 \pm 0.22$ | $0.72 \pm 0.23$ | $0.69 \pm 0.28$ |
| **Group 2** | $1.09 \pm 0.43$ | $0.90 \pm 0.34$ | $1.04 \pm 0.40$ | $0.98 \pm 0.38$ |
| **Group 3** | $1.69 \pm 0.63$ | $1.53 \pm 0.57$ | $1.64 \pm 0.64$ | $1.57 \pm 0.57$ |

Table 8.4: The mean ($\mu$) and standard deviation ($\sigma$) time of the walking parts from the four individual tests (D1, D2, ND1, and ND2) for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

I used a paired-sample t-test to test the null hypothesis $H_{0,2}$ (Table 8.5). The result from the test were the p-values ($p$) and a decision result of if the $H_{0,2}$ was rejected ($h = 1$) or not ($h = 0$).

| | **D, ND**, $H_{0,2}$ | |
|---|---|---|
| | $p$ | $h$ |
| **Group 1** | 0.40 | 0 |
| **Group 2** | 0.79 | 0 |
| **Group 3** | 0.96 | 0 |

Table 8.5: The p-value ($p$) and the decision result for rejection or not the null hypothesis ($h$) of the paired-sample t-test testing $H_{0,2}$, for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

After having investigated how the difference in the time of the walking parts was between the D and the ND tests, I wanted to test the null hypothesis $H_{0,3}$ and $H_{0,4}$ with the use of the two-sample t-test (Table 8.6). The result from the test were the p-values ($p$) and a decision whether the $H_{0,3}/H_{0,4}$ was rejected ($h = 1$) or not ($h = 0$).

| | D | | ND | |
|---|---|---|---|---|
| | $p$ | $h$ | $p$ | $h$ |
| **Group 1, Group 2** | $p < 0.001$ | 1 | $p < 0.001$ | 1 |
| **Group 1, Group 3** | $p < 0.001$ | 1 | $p < 0.001$ | 1 |
| **Group 2, Group 3** | $p < 0.001$ | 1 | $p < 0.001$ | 1 |

Table 8.6: The p-value ($p$) and the decision result for rejection or not the null hypothesis ($h$) of the two-sample t-test testing $H_{0,3}$ and $H_{0,4}$.

## 8.3 The mean walking times for the five individual walking parts within the tests

In each of the tests, there are five walking parts before/in between each kick ($W1$, $W2$, $W3$, $W4$, and $W5$). I wanted to see if the participants' time on the five walking parts differed from each other. To do this, I found the mean and standard deviation time for the five walking parts, based on the walking part from the D test (Table 8.7) or the ND test (Table 8.8).

| | **Group 1** $\mu \pm \sigma, [s]$ | **Group 2** $\mu \pm \sigma, [s]$ | **Group 3** $\mu \pm \sigma, [s]$ |
|---|---|---|---|
| $W1_D$ | $1.02 \pm 0.16$ | $1.26 \pm 0.35$ | $1.65 \pm 0.36$ |
| $W2_D$ | $0.63 \pm 0.19$ | $0.94 \pm 0.38$ | $1.55 \pm 0.55$ |
| $W3_D$ | $0.73 \pm 0.20$ | $1.07 \pm 0.44$ | $1.73 \pm 0.78$ |
| $W4_D$ | $0.61 \pm 0.17$ | $0.93 \pm 0.42$ | $1.73 \pm 0.56$ |
| $W5_D$ | $0.61 \pm 0.13$ | $0.80 \pm 0.25$ | $1.38 \pm 0.63$ |

Table 8.7: The mean ($\mu$) and standard deviation ($\sigma$) time of the walking parts for each of the five parts done within the two dominant tests, separated into the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

| | Group 1 $\mu \pm \sigma, [s]$ | Group 2 $\mu \pm \sigma, [s]$ | Group 3 $\mu \pm \sigma, [s]$ |
|---|---|---|---|
| $W1_{ND}$ | $0.93 \pm 0.21$ | $1.17 \pm 0.26$ | $1.71 \pm 0.38$ |
| $W2_{ND}$ | $0.69 \pm 0.20$ | $1.10 \pm 0.47$ | $1.62 \pm 0.57$ |
| $W3_{ND}$ | $0.61 \pm 0.22$ | $0.97 \pm 0.40$ | $1.72 \pm 0.68$ |
| $W4_{ND}$ | $0.69 \pm 0.20$ | $1.00 \pm 0.42$ | $1.74 \pm 0.72$ |
| $W5_{ND}$ | $0.61 \pm 0.30$ | $0.79 \pm 0.27$ | $1.24 \pm 0.48$ |

Table 8.8: The mean ($\mu$) and standard deviation ($\sigma$) time of the walking parts for each of the five parts done within the two non-dominant tests, separated into the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

## 8.4 Discussion

From looking at the plot that shows all the participants' mean time of the walking parts (Figure 8.1), one observes a small increase in the mean time from Group 1 to Group 3. One observes less variance between the participants in Group 1 than in the two others. This is also shown with the mean and standard deviation of the mean time for the walking parts for the three groups (Table 8.1). The mean of the mean walking time increases with the groups, and so those the standard deviation.
When looking at the standard deviation of the times from the walking parts for every participant (Figure 8.2), there are some of the participants in each group that has a higher standard deviation than the rest of the participants within the groups. In Group 3, particularly, one participant has a much higher standard deviation than the other.
The results from the two-sample t-test (Table 8.2) testing the null hypothesis $H_{0,1}$, show that all the tests reject the null hypothesis. This means that the mean times of the walking parts in the three groups are independent normally distributed with different mean and variance. With this knowledge and looking at the mean $\pm$ the standard deviation of the mean walking times for the three groups (Table 8.1), the overlapping region between each group is too big. Consequently, it is hard to find a suitable decision threshold to classify participants based on the mean time for the walking parts. However, the overlap between the three groups is smaller here than for the test times and the kicking times.

When I divided the walking parts into those in the dominant tests (D) and those in the non-dominant tests (ND), there was not a big difference between the mean walking times (Figures 8.3 and 8.4). The mean and standard deviation of the mean walking time for the different groups from the different tests (Table 8.3) showed the same. The difference between the dominant and the non-dominant tests ended up being less than 0.5%.

From the standard deviation in the walking time for each of the participants (Figures 8.5 and 8.6), the same participant in Group 3 that had a high standard deviation has a high standard deviation in the dominant tests. This means that whatever went wrong within the tests for the participant happened in one of the dominant tests.
When doing the paired-sample t-test (Table 8.5) testing the null hypothesis $H_{0,2}$, all the tests accept the null hypothesis. I also did some two-sample t-tests (Table 8.6) to test the null hypothesis $H_{0,3}$ and $H_{0,4}$. None of these tests accepted the null hypothesis. I can say that the participants' times in the groups are from different normal distributions with different mean and variance, but the difference between dominant and non-dominant tests are from normal distributions with mean= 0.
In the mean and standard deviation of the walking time for the three groups from the four individual tests (Table 8.4), some changes were observed. Between D1 and D2, and between ND1 and ND2, all the groups experience a decrease in the mean waling time. There is a difference in how prominent this decrease is for each of the groups. This is looked closer on later in the project (Chapter 9).
The three groups' mean time for the walking parts from the dominant and the non-dominant tests are different normal distribution. However, with their mean $\pm$ the standard deviation of the mean walking times (Table 8.3), there is still an overlapping region between the three groups, making it hard to classify any participants by their mean time for the walking parts.

Lastly, I wanted to see if there were any differences in the walking times between the five walking parts within the tests (Tables 8.7 and 8.8). Between $W1_D$ and $W2_D$ all the groups had a decrease in the mean time, while between $W2_D$ and $W3_D$, all the groups had an increase in the time. In the non-dominant (ND) tests, the same observation was not found. Group 1 and 2 experience the same pattern when it comes to decrease and increase in the mean walking time, and both of the groups have their highest mean walking time in $W1$. Group 3, on the other hand, does not experience the same pattern.

To sum up, there is a change in the mean walking time between the three groups. This was the same when I divided the walking parts into those done in the dominant (D) tests and non-dominant (ND) tests. The difference between the groups is not big enough to classify the participants into one of the three groups based on their walking time. When looking into the differences between the different walking parts, Group 1 and 2 have similar patterns, while Group 3 is the odd one.

# Chapter 9

# The learning effect in the test

In this chapter, I will be looking into the learning effect of the Six Spot Step Test (SSST) by investigating the differences between the four SSST each of the participants did from the week 1 tests and the retests.
It is essential that the four tests from each of the participants I am looking at are done in the same order. The participants are supposed to go through the test four times. But if something goes wrong during a test, that test will be retaken at the end of the test session. Because of this, I only use the participants that do not have to retake any of the tests.
The participants will be divided into three groups based on their EDSS (Expanded Disability Statue Scale) (Section 5.3). The three groups will be:

- Group 1: the controls
- Group 2: the patients with an EDSS $< 4$
- Group 3: the patients with an EDSS $\geq 4$

During the analysis, I looked at different times for the participants:

- the test times from the first and second test where they are kicking with their dominant foot, from both the week 1 ($D1_{W1}$ and $D2_{W1}$) tests and the retests ($D1_{RT}$ and $D2_{RT}$)
- the test times from the first and second test where they are kicking with their non-dominant foot, from both the week 1 ($ND1_{W1}$ and $ND2_{W1}$) tests and the retests ($ND1_{RT}$ and $ND2_{RT}$)
- the mean kicking times from the first ($K_{D1}$) and second ($K_{D2}$) test where they are kicking with their dominant foot
- the mean kicking times from the first ($K_{ND1}$) and second ($K_{ND2}$) test where they are kicking with their non-dominant foot

- the kicking time for the five kicks within the first $(K1_{D1}, \cdots, K5_{D1})$ and second $(K1_{D2}, \cdots, K5_{D2})$ tests where they kick with there dominant foot
- the kicking time for the five kicks within the first $(K1_{ND1}, \cdots, K5_{ND1})$ and second $(K1_{ND2}, \cdots, K5_{ND2})$ tests where they kick with there non-dominant foot
- the mean times of the walking parts from the first and second tests where they are kicking with their dominant foot, from both the week 1 ($W_{D1,W1}$ and $W_{D2,W1}$) tests and the retests ($W_{D1,RT}$ and $W_{D2,RT}$)
- the mean times of the walking parts from the first and second tests where they are kicking with their non-dominant foot, from both the week 1 ($W_{ND1,W1}$ and $W_{ND2,W1}$) tests and the retests ($W_{ND1,RT}$ and $W_{ND2,RT}$)
- the times of the five walking part within the first and second tests where they kick with their dominant foot, from both the week 1 ($W1_{D1,W1}, \cdots, W5_{D1,W1}$ and $W1_{D2,W1}, \cdots, W5_{D2,W1}$) tests and the retests ($W1_{D1,RT}, \cdots, W5_{D1,RT}$ and $W1_{D2,RT}, \cdots, W5_{D2,RT}$)
- the times of the five walking part within the first and second tests where they kick with their non-dominant foot, from both the week 1 ($W1_{ND1,W1}, \cdots, W5_{ND1,W1}$ and $W1_{ND2,W1}, \cdots, W5_{ND2,W1}$) tests and the retests ($W1_{ND1,RT}, \cdots, W5_{ND1,RT}$ and $W1_{ND2,RT}, \cdots, W5_{ND2,RT}$)

I started by looking into the test from week 1. First, I investigated the difference in the mean test times between the first and second tests for the dominant and the non-dominant tests. Then, I investigated the difference in the kicking times and the times from the walking parts. After looking at the tests from week 1, I looked at if the difference found in week 1 had changed in the retests.
During the investigations I tested different null hypothesis $H_0$ with significant level $\alpha$:

- the differences between the $D1_{W1}$ and $D2_{W1}$ are normal distributed where the mean$= 0$ ($H_{0,1}$, with $\alpha = 0.05$)
- the differences between the $ND1_{W1}$ and $ND2_{W1}$ are normal distributed where the mean$= 0$ ($H_{0,2}$, with $\alpha = 0.05$)
- the differences between the $K_{D1,W1}$ and $K_{D2,W1}$ are normal distributed where the mean$= 0$ ($H_{0,3}$, with $\alpha = 0.05$)
- the differences between the $K_{ND1,W1}$ and $K_{ND2,W1}$ are normal distributed where the mean$= 0$ ($H_{0,4}$, with $\alpha = 0.05$)
- the difference between the $Ki_{D1,W1}$ and $Ki_{D2,W1}$, $i = 1, 2, 3, 4, 5$, are normal distributed where the mean$= 0$ ($H_{0,5}$, with $\alpha = 0.05$)

- the difference between the $Ki_{ND1,W1}$ and $Ki_{ND2,W1}$, $i = 1, 2, 3, 4, 5$, are normal distributed where the mean= 0 ($H_{0,6}$, with $\alpha = 0.05$)
- the differences between the $W_{D1,W1}$ and $W_{D2,W1}$ are normal distributed where the mean= 0 ($H_{0,7}$, with $\alpha = 0.05$)
- the differences between the $W_{ND1,W1}$ and $W_{ND2,W1}$ are normal distributed where the mean= 0 ($H_{0,8}$, with $\alpha = 0.05$)
- the difference between the $Wi_{D1,W1}$ and $Wi_{D2,W1}$, $i = 1, 2, 3, 4, 5$, are normal distributed where the mean= 0 ($H_{0,9}$, with $\alpha = 0.05$)
- the difference between the $Wi_{ND1,W1}$ and $Wi_{ND2,W1}$, $i = 1, 2, 3, 4, 5$, are normal distributed where the mean= 0 ($H_{0,10}$, with $\alpha = 0.05$)
- the differences between the $D1_{RT}$ and $D2_{RT}$ are normal distributed where the mean= 0 ($H_{0,11}$, with $\alpha = 0.05$)
- the differences between the $ND1_{RT}$ and $ND2_{RT}$ are normal distributed where the mean= 0 ($H_{0,12}$, with $\alpha = 0.05$)
- the differences between the $W_{D1,RT}$ and $W_{D2,RT}$ are normal distributed where the mean= 0 ($H_{0,13}$, with $\alpha = 0.05$)
- the differences between the $W_{ND1,RT}$ and $W_{ND2,RT}$ are normal distributed where the mean= 0 ($H_{0,14}$, with $\alpha = 0.05$)
- the difference between the $Wi_{D1,RT}$ and $Wi_{D2,rt}$, $i = 1, 2, 3, 4, 5$, are normal distributed where the mean= 0 ($H_{0,15}$, with $\alpha = 0.05$)
- the difference between the $Wi_{ND1,RT}$ and $Wi_{ND2,RT}$, $i = 1, 2, 3, 4, 5$, are normal distributed where the mean= 0 ($H_{0,16}$, with $\alpha = 0.05$)

## 9.1 Testing week 1

Improvements in the test spent doing an SSST may be caused by learning, i.e. the total test time may decrease because the participant has practiced the task before. In searching for such an effect, I started looking at the four SSSTs done in week 1. First, I will start with the total test time. Then I will break the tests up into segments and see if there is a change in the time spent for kicking or walking. An underlying assumption is that the participants have not done an SSST before. We are pretty sure this is a valid assumption.

### 9.1.1 Test times

The first thing I started with was to see how the test times for each of the participants changed between the first and second time they walked through the test kicking with the same foot (changes in time between $D1_{W1}$ and $D2_{W1}$, and $ND1_{W1}$ and $ND2_{W1}$).
I found the difference in the test times between $D1_{W1}$ and $D2_{W1}$, and between $ND1_{W1}$ and $ND2_{W1}$ for all the participants (Figures 9.1 and 9.2) by subtracting $D2_{W1}$ and $ND_{W1}$ from $D1_{W1}$ and $ND1_{W1}$ respectively. From the difference for each of the participants I found the mean and standard deviation of the difference for the three groups (Table 9.1).

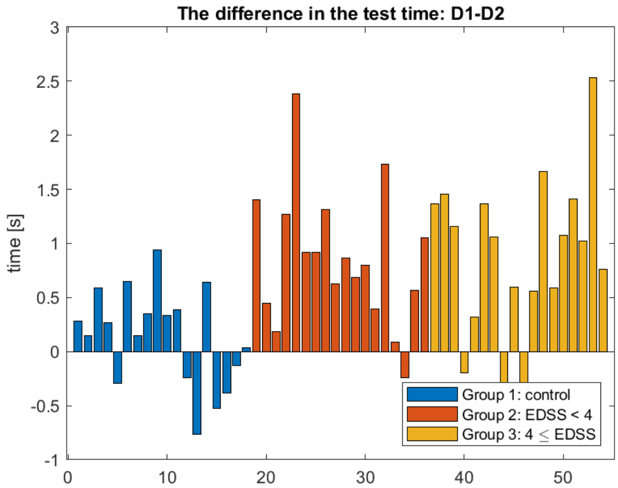


Figure 9.1: The time difference between the first and second dominant test times ($D1_{W1}$ and $D2_{W1}$) for every participants. The participants are sorted into the three groups based on there EDSS, and color-coordinated.

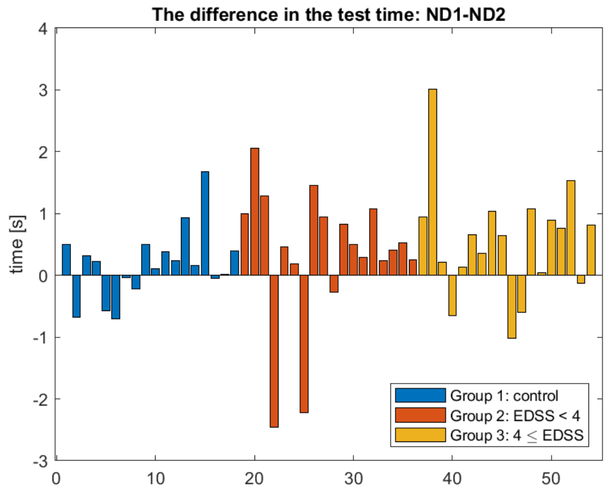


Figure 9.2: The time difference between the first and second non-dominant test times ($ND1_{W1}$ and $ND2_{W1}$) for every participants. The participants are sorted into the three groups based on there EDSS, and color-coordinated.

| | **Dominant tests** | | **Non-dominant tests** | |
|---|---|---|---|---|
| | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] |
| **Group 1** | 0.14 | 0.45 | 0.18 | 0.57 |
| **Group 2** | 0.86 | 0.63 | 0.37 | 1.13 |
| **Group 3** | 0.86 | 0.80 | 0.54 | 0.91 |

Table 9.1: The mean ($\mu$) and standard deviation ($\sigma$) of the difference in the test times between the first and second tests; $D1_{W1}$ and $D2_{W1}$ (Dominant tests) and $ND1_{W1}$ and $ND2_{W1}$ (Non-dominant tests). This for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

Then I did a paired-sample t-test on the test times to test the null hypothesis $H_{0,1}$ and $H_{0,2}$ (Table 9.2) for the three groups. The t-test returned the p-values ($p$) and the decision result of if the $H_{0,1}/H_{0,2}$ were rejected ($h = 1$) or not ($h = 0$).

| | | $p$ | $h$ |
|---|---|---|---|
| **Group 1** | D1, D2 | 0.22 | 0 |
| | ND1, ND2 | 0.21 | 0 |
| **Group 2** | D1, D2 | $p < 0.001$ | 1 |
| | ND1, ND2 | 0.19 | 0 |
| **Group 3** | D1, D2 | $p < 0.001$ | 1 |
| | ND1, ND2 | 0.02 | 1 |

Table 9.2: The p-values ($p$) and the decision result for rejecting or not the null hypothesis ($h$) of the paired sample t-tests testing $H_{0,1}$ and $H_{0,2}$. This was done for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

### 9.1.2 Kick times

After looking at the test times, I wanted to investigate the difference in the kicking times between the first and second tests using the dominant foot and non-dominant foot. I started by finding the mean kicking time within each of the tests for the participants ($K_{D1,W1}$, $K_{D2,W1}$, $K_{ND1,W1}$, and $K_{ND2,W1}$). Then I found the difference in the mean kicking time between $K_{D1,W1}$ and $K_{D2,W1}$ (Figure 9.3), and between $K_{ND1,W1}$ and $K_{ND2,W1}$ (Figure 9.4). The difference was found by subtracting the $K_{D2,W1}$ and $K_{ND2,W1}$ from the $K_{D1,W1}$ and $K_{ND1,W1}$ respective. With the difference in kicking times for the participants, I found the mean and standard deviation of the difference for each of the groups (Table 9.3).

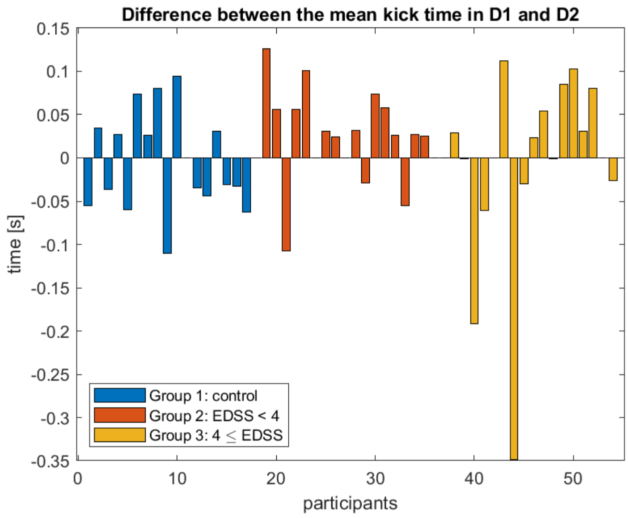


Figure 9.3: The difference between the mean kicking times from the first and second dominant tests ($K_{D1,W1}$ and $K_{D2,W1}$) for all the participants. Each of the participants is represented by a bar, and they are sorted into their respective groups by colors.

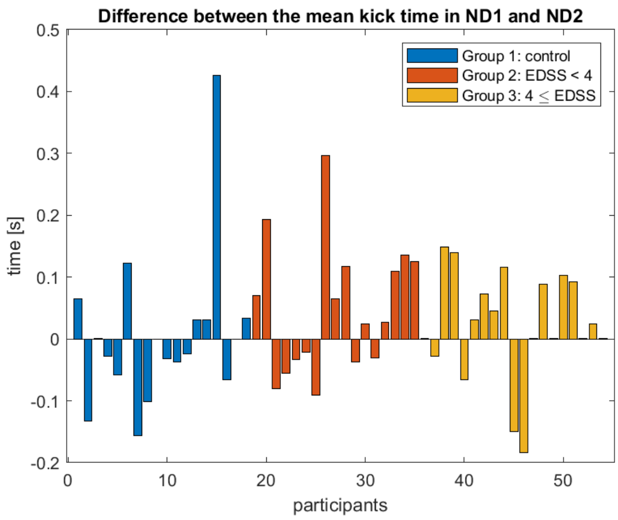


Figure 9.4: The difference between the mean kicking times from the first and second non-dominant tests ($K_{ND1,W1}$ and $K_{ND2,W1}$) for all the participants. Each of the participants is represented by a bar, and they are sorted into their respective groups by colors.

| | Dominant tests | | Non-dominant tests | |
|---|---|---|---|---|
| | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] |
| **Group 1** | -0.01 | 0.06 | $0.42 \times 10^{-2}$ | 0.13 |
| **Group 2** | 0.03 | 0.05 | 0.05 | 0.10 |
| **Group 3** | -0.01 | 0.11 | 0.02 | 0.09 |

Table 9.3: The mean ($\mu$) and the standard deviation ($\sigma$) of the difference in the kick times between the mean kicks done in the first and second tests; $K_{D1,W1}$ and $K_{D2,W1}$ (Dominant tests) and $K_{ND1,W1}$ and $K_{ND2,W1}$ (Non-dominant tests). This for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

I also wanted to investigate the individual kicks in the tests. I found the time change in kicks 1, 2, 3, 4, and 5 between the first and second dominant tests and between the first and second non-dominant tests for each of the participants. From this, I found the mean ($\mu$) and standard deviation ($\sigma$) of the difference in the time for the three groups (Table 9.4).

| | Kick | Dominant tests $\mu$, [s] | Dominant tests $\sigma$, [s] | Non-dominant tests $\mu$, [s] | Non-dominant tests $\sigma$, [s] |
|---|---|---|---|---|---|
| **Group 1** | 1 | -0.02 | 0.11 | -0.03 | 0.12 |
| | 2 | 0.03 | 0.11 | 0.01 | 0.14 |
| | 3 | 0.01 | 0.10 | -0.05 | 0.10 |
| | 4 | $0.01 \times 10^{-2}$ | 0.12 | $0.27 \times 10^{-2}$ | 0.11 |
| | 5 | -0.04 | 0.17 | 0.10 | 0.52 |
| **Group 2** | 1 | -0.01 | 0.13 | $-0.05 \times 10^{-2}$ | 0.17 |
| | 2 | 0.02 | 0.13 | 0.07 | 0.19 |
| | 3 | 0.05 | 0.15 | 0.02 | 0.14 |
| | 4 | 0.02 | 0.17 | 0.02 | 0.16 |
| | 5 | 0.05 | 0.10 | 0.12 | 0.40 |
| **Group 3** | 1 | 0.03 | 0.19 | $-0.44 \times 10^{-2}$ | 0.31 |
| | 2 | -0.09 | 0.35 | 0.13 | 0.26 |
| | 3 | -0.09 | 0.20 | -0.09 | 0.20 |
| | 4 | -0.05 | 0.20 | 0.10 | 0.40 |
| | 5 | 0.08 | 0.20 | -0.02 | 0.21 |

Table 9.4: The mean ($\mu$) and standard deviation ($\sigma$) of the difference in the kick time for the individual kicks, between the first and second time dominant test, and non-dominant test, for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

Then I did a paired-sample t-test on the mean kicking times and the five individual kicks to test the null hypothesis $H_{0,3}$, $H_{0,4}$, $H_{0,5}$ and $H_{0,6}$ (Table 9.5) for the three groups. The t-test returned the p-values ($p$) and the decision result of if the null hypothesis were rejected ($h = 1$) or not ($h = 0$).

| | $i$ | Group 1 $p$ | Group 1 $h$ | Group 2 $p$ | Group 2 $h$ | Group 3 $p$ | Group 3 $h$ |
|---|---|---|---|---|---|---|---|
| $Ki_{D1,W1}$, $Ki_{D2,W1}$ | 1 | 0.43 | 0 | 0.82 | 0 | 0.53 | 0 |
| | 2 | 0.34 | 0 | 0.58 | 0 | 0.31 | 0 |
| | 3 | 0.77 | 0 | 0.23 | 0 | 0.85 | 0 |
| | 4 | 0.998 | 0 | 0.70 | 0 | 0.28 | 0 |
| | 5 | 0.33 | 0 | 0.04 | 1 | 0.10 | 0 |
| $K_{D1,W1}$, $K_{D2,W1}$ | | 0.68 | 0 | 0.07 | 0 | 0.77 | 0 |
| $Ki_{ND1,W1}$, $Ki_{ND2,W1}$ | 1 | 0.27 | 0 | 0.99 | 0 | 0.95 | 0 |
| | 2 | 0.87 | 0 | 0.16 | 0 | 0.05 | 0 |
| | 3 | 0.04 | 1 | 0.64 | 0 | 0.09 | 0 |
| | 4 | 0.92 | 0 | 0.53 | 0 | 0.29 | 0 |
| | 5 | 0.45 | 0 | 0.21 | 0 | 0.64 | 0 |
| $K_{ND1,W1}$, $K_{ND2,W1}$ | | 0.89 | 0 | 0.08 | 0 | 0.28 | 0 |

Table 9.5: The p-values ($p$) and the decision result for rejecting or not the null hypothesis ($h$) of the paired sample t-tests testing $H_{0,3}$, $H_{0,4}$, $H_{0,5}$ and $H_{0,6}$. This was done for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

### 9.1.3 Walk times

I then investigated the walking times between the first and second dominant teats ($W_{D1,W1}$ and $W_{D2,W1}$), and the non-dominant tests ($W_{ND1,W1}$ and $W_{ND2,W1}$). The first thing I did was to find the mean walking times in the four tests ($W_{D1,W1}$, $W_{D2,W1}$, $W_{ND1,W1}$, and $W_{ND2,W1}$) for each of the participants, and then I found the time difference between $W_{D1,W1}$ and $W_{D2,W1}$ (Figure 9.5) and between $W_{ND1,W1}$ and $W_{ND2,W1}$ (Figure 9.6). The difference was found by subtracting the $W_{D2,W1}$ and $W_{ND2,W1}$ from the $W_{D1,W1}$ and $W_{ND1,W1}$ respective. With the difference in kicking times for the participants, I found the mean and standard deviation of the difference for each of the groups (Table 9.6).

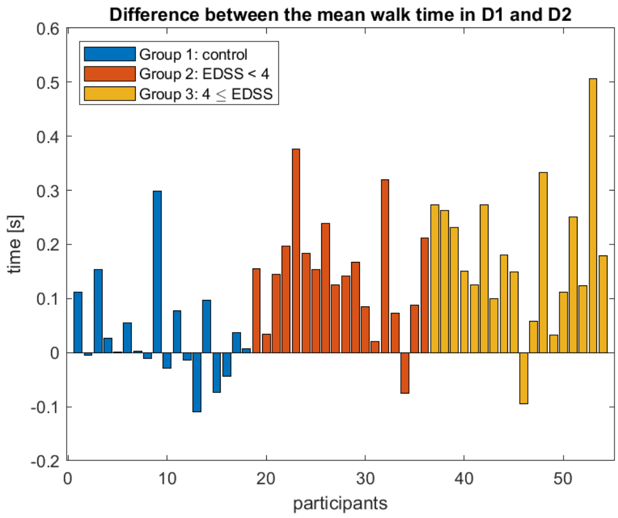


Figure 9.5: The difference between the mean walking times from the first and second dominant tests ($W_{D1,W1}$ and $W_{D2,W1}$) for all the participants. Each of the participants is represented by a bar, and they are sorted into their respective groups by colors.

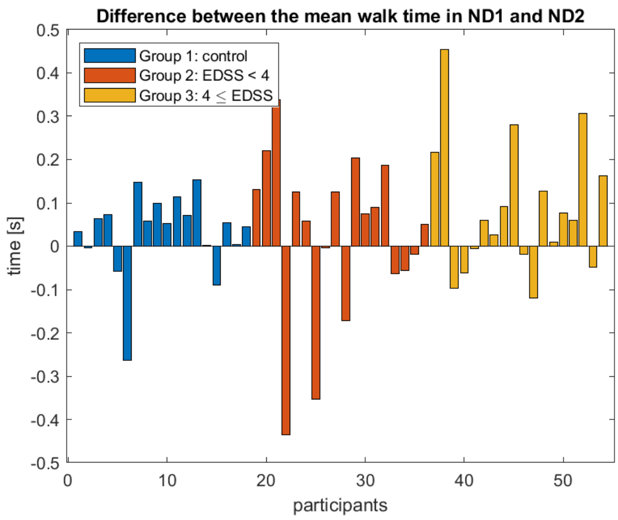


Figure 9.6: The difference between the mean walking times from the first and second non-dominant tests ($W_{ND1,W1}$ and $W_{ND2,W1}$) for all the participants. Each of the participants is represented by a bar, and they are sorted into their respective groups by colors.

| | Dominant tests | | Non-dominant tests | |
|---|---|---|---|---|
| | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] |
| **Group 1** | 0.03 | 0.09 | 0.03 | 0.10 |
| **Group 2** | 0.15 | 0.11 | 0.03 | 0.20 |
| **Group 3** | 0.18 | 0.13 | 0.08 | 0.15 |

Table 9.6: The mean ($\mu$) and the standard deviation ($\sigma$) of the difference in the walking part times between the mean walking parts done in the first and second tests; $W_{D1,W1}$ and $W_{D2,W1}$ (Dominant tests) and $W_{ND1,W1}$ and $W_{ND2,W1}$ (Non-dominant tests). This for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

Then I found the difference in time for the individual walking parts between the first and second dominant tests and the first and second non-dominant tests. From these results, I found the mean and standard deviation of the difference in the individual walking parts between the tests (Table 9.7).

| | | Dominant tests | | Non-dominant tests | |
|---|---|---|---|---|---|
| | **Walk** | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] |
| **Group 1** | 1 | 0.02 | 0.13 | 0.02 | 0.18 |
| | 2 | 0.04 | 0.11 | 0.07 | 0.15 |
| | 3 | 0.05 | 0.24 | 0.01 | 0.08 |
| | 4 | 0.05 | 0.13 | 0.13 | 0.13 |
| | 5 | 0.01 | 0.18 | -0.08 | 0.38 |
| **Group 2** | 1 | 0.09 | 0.23 | 0.10 | 0.22 |
| | 2 | 0.27 | 0.29 | 0.09 | 0.48 |
| | 3 | 0.15 | 0.29 | -0.15 | 0.59 |
| | 4 | 0.15 | 0.16 | 0.11 | 0.26 |
| | 5 | 0.07 | 0.27 | -0.01 | 0.20 |
| **Group 3** | 1 | 0.10 | 0.18 | 0.17 | 0.27 |
| | 2 | 0.07 | 0.54 | 0.04 | 0.53 |
| | 3 | 0.21 | 0.50 | -0.05 | 0.56 |
| | 4 | 0.34 | 0.48 | 0.31 | 0.41 |
| | 5 | 0.19 | 0.37 | -0.05 | 0.23 |

Table 9.7: The mean ($\mu$) and the standard deviation ($\sigma$) of the difference in the walking time for the individual walking parts, between the first and second dominant test, and the non-dominant test, for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

After finding the difference in the walking times, I did some paired-sample t-tests on the mean time for the walking parts and on the five individual walking parts times to test the null hypothesis $H_{0,7}$, $H_{0,8}$, $H_{0,9}$ and $H_{0,10}$ (Table 9.8) for the three groups. The t-test returned the p-values ($p$) and the decision result of if the null hypothesis were rejected ($h = 1$) or not ($h = 0$).

| | | Group 1 | | Group 2 | | Group 3 | |
|---|---|---|---|---|---|---|---|
| | $i$ | $p$ | $h$ | $p$ | $h$ | $p$ | $h$ |
| $Wi_{D1,W1}$, $Wi_{D2,W1}$ | 1 | 0.56 | 0 | 0.12 | 0 | 0.03 | 1 |
| | 2 | 0.18 | 0 | 0.001 | 1 | 0.60 | 0 |
| | 3 | 0.41 | 0 | 0.04 | 1 | 0.10 | 0 |
| | 4 | 0.11 | 0 | $p < 0.001$ | 1 | 0.009 | 1 |
| | 5 | 0.86 | 0 | 0.28 | 0 | 0.05 | 1 |
| $W_{D1,W1}$, $W_{D2,W1}$ | | 0.16 | 0 | $p < 0.001$ | 1 | $p < 0.001$ | 1 |
| $Wi_{ND1,W1}$, $Wi_{ND2,W1}$ | 1 | 0.64 | 0 | 0.07 | 0 | 0.016 | 1 |
| | 2 | 0.08 | 0 | 0.41 | 0 | 0.77 | 0 |
| | 3 | 0.59 | 0 | 0.30 | 0 | 0.71 | 0 |
| | 4 | $p < 0.001$ | 1 | 0.09 | 0 | 0.005 | 1 |
| | 5 | 0.42 | 0 | 0.84 | 0 | 0.37 | 0 |
| $W_{ND1,W1}$, $W_{ND2,W1}$ | | 0.19 | 0 | 0.55 | 0 | 0.03 | 1 |

Table 9.8: The p-values ($p$) and the decision result for rejecting or not the null hypothesis ($h$) of the paired sample t-tests testing $H_{0,7}$, $H_{0,8}$, $H_{0,9}$ and $H_{0,10}$. This was done for the three groups; Group 1 the controls, Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

## 9.2 Retest

I wanted to investigate if the difference that was found between the first and second dominant tests, and non-dominant tests, in the SSSTs from week 1 can be found in the retests. For this, I only looked at the participants that were retested after a rehabilitation stay. None of the controls were retested, so I will only be looking at the participants in Groups 2 and 3.
From the analysis done above, I found no significant changes in the kicking times between the first and second tests, dominant and non-dominant. For that reason, I will only be investigating the test times and the walking times when investigating if there are some changes in the retest.

### 9.2.1 Test times

I started by looking at the test times for the participants from week 1 and the retest. First I found the difference in the test times between $D1_{W1}$ and $D2_{W1}$, and $D1_{RT}$ and $D2_{RT}$ (Figure 9.7). I did the same for the non-dominant tests, $ND1_{W1}$ and $ND2_{W1}$, and $ND1_{RT}$ and $ND2_{RT}$ (Figure 9.8). Based on these results, I found the mean and standard deviation of the difference in the test times between the first and second dominant/non-dominant tests from week 1 and the retest for the two groups (Table 9.9).

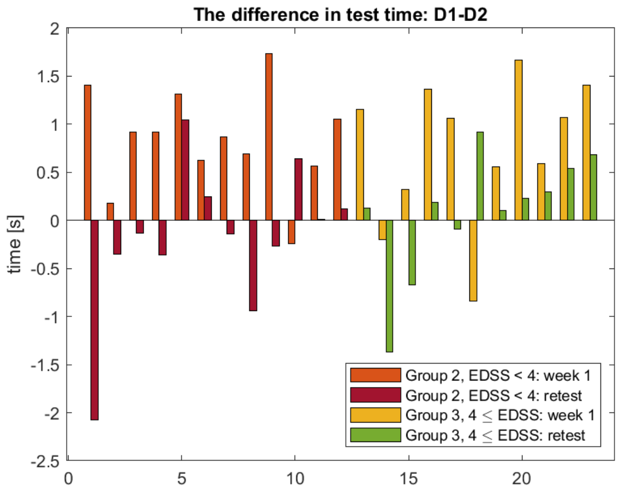


Figure 9.7: The time difference between the first and second dominant test times ($D1_{W1}$ and $D2_{W1}$, and $D1_{RT}$ and $D2_{RT}$) for every participants. The participants are sorted into the three groups based on there EDSS, and color-coordinated.

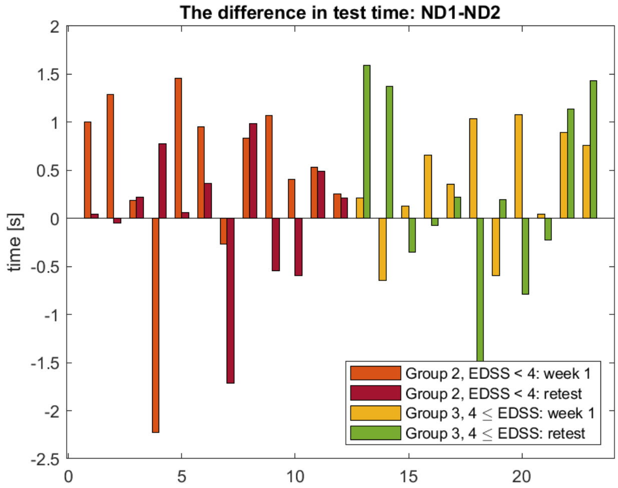


Figure 9.8: The time difference between the first and second non-dominant test times ($ND1_{W1}$ and $ND2_{W1}$, and $ND1_{RT}$ and $ND2_{RT}$) for every participants. The participants are sorted into the three groups based on there EDSS, and color-coordinated.

| | **Dominant tests** | | | | **Non-dominant tests** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Week 1** | | **Retest** | | **Week 1** | | **Retest** | |
| | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] |
| **Group 2** | 0.84 | 0.54 | -0.18 | 0.78 | 0.46 | 0.98 | 0.02 | 0.72 |
| **Group 3** | 0.74 | 0.75 | 0.09 | 0.64 | 0.36 | 0.60 | 0.24 | 1.07 |

Table 9.9: The mean ($\mu$) and standard deviation ($\sigma$) of the difference in the test times between the first and second tests from week 1 testing and the retests; $D1_{W1}$, $D2_{W1}$, $D1_{RT}$, and $D2_{RT}$ (Dominant tests) and $ND1_{W1}$, $ND2_{W1}$, $ND1_{RT}$, and $ND2_{RT}$ (Non-dominant tests). This for two of the groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

I preformed some paired-sample t-tests to test the null hypothesis $H_{0,11}$, $H_{0,12}$, $H_{0,1}$ and $H_{0,2}$ (Table 9.10) for the two groups. The t-test returned the p-values ($p$) and the decision result of if the null hypothesis were rejected ($h = 1$) or not ($h = 0$).

| | | $p$ | | $h$ | |
|---|---|---|---|---|---|
| | | **Week 1** | **Retest** | **Week 1** | **Retest** |
| **Group 2** | D1, D2 | $p < 0.001$ | 0.44 | 1 | 0 |
| | ND1, ND2 | 0.14 | 0.92 | 0 | 0 |
| **Group 3** | D1, D2 | 0.009 | 0.66 | 1 | 0 |
| | ND1, ND2 | 0.08 | 0.47 | 0 | 0 |

Table 9.10: The p-value ($p$) and the decision result for rejecting or not the null hypothesis ($h$) of the paired sample t-test testing the null hypothesis $H_{0,11}$, $H_{0,12}$, $H_{0,1}$ and $H_{0,2}$. This was done for the two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

### 9.2.2 Walking times

After looking at the difference in the test time from week 1 and the retest, I wanted to investigate any changes in the walking times differences from week 1 and the retest.
First, I found the mean waking time for each participant from the eight tests, four from week 1 and four from the retest. Then I found the difference in the mean walking time between the first and second dominant tests (D1 and D2) for both week 1 and the retest (Figure 9.9), and the difference in the mean walking time between the first and second non-dominant tests (ND1 and ND2) from both week 1 and the retest (Figure 9.10). From these results were the mean and standard deviation of the difference in the walk time for the two groups, from week 1 and the retest, found (Table 9.11).

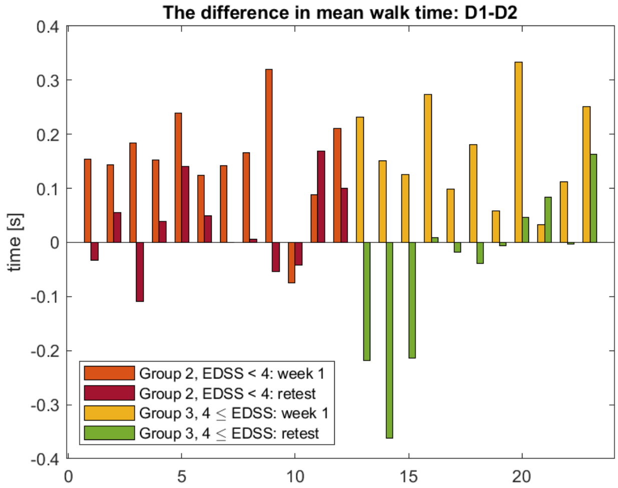


Figure 9.9: The difference between the mean walking times from the first and second dominant tests from week 1 ($W_{D1,W1}$ and $W_{D2,W1}$) and the retest ($W_{D1,RT}$ and $W_{D2,RT}$), for all the participants. Each of the participants is represented by a bar, and they are sorted into their respective groups by colors.

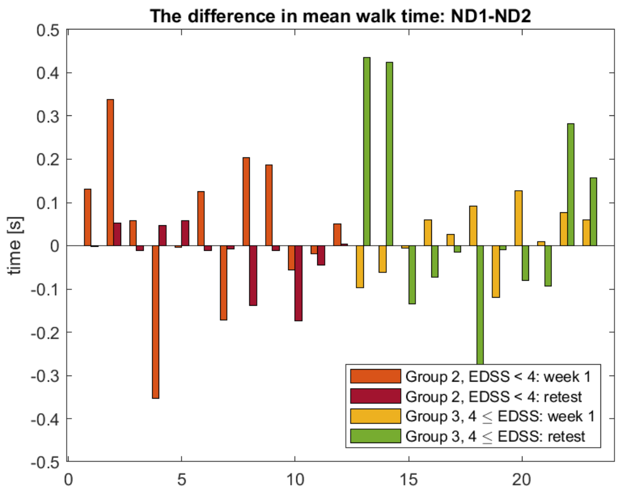


Figure 9.10: The difference between the mean walking times from the first and second non-dominant tests from week 1 ($W_{ND1,W1}$ and $W_{ND2,W1}$) and the retest ($W_{ND1,RT}$ and $W_{ND2,RT}$), for all the participants. Each of the participants is represented by a bar, and they are sorted into their respective groups by colors.

| | Dominant tests | | | | Non-dominant tests | | | |
|---|---|---|---|---|---|---|---|---|
| | Week 1 | | Retest | | Week 1 | | Retest | |
| | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] |
| **Group 2** | 0.15 | 0.09 | 0.03 | 0.08 | 0.04 | 0.18 | -0.02 | 0.07 |
| **Group 3** | 0.17 | 0.10 | -0.05 | 0.15 | 0.02 | 0.08 | 0.04 | 0.26 |

Table 9.11: The mean ($\mu$) and the standard deviation ($\sigma$) of the difference in the walking part times, from week 1 and the retest, between the mean walking parts done in the first and second tests; $W_{D1,W1}$, $W_{D2,W1}$, $W_{D1,RT}$, and $W_{D2,RT}$ (Dominant tests) and $W_{ND1,W1}$, $W_{ND2,W1}$, $W_{ND1,RT}$, and $W_{ND2,RT}$ (Non-dominant tests). This for two of the groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

I also found the difference in the times for the individual walking parts between the first and second dominant/non-dominant test for week 1 and the retest. From these difference were the mean and standard deviation of the difference for the five individual walking parts, from week 1 and the retest, found (Table 9.12).

| | | Dominant tests | | | | Non-dominant tests | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Week 1 | | Retest | | Week 1 | | Retest | |
| | **Walk** | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] |
| **Group 2** | 1 | 0.10 | 0.27 | 0.03 | 0.18 | 0.07 | 0.14 | -0.04 | 0.08 |
| | 2 | 0.19 | 0.14 | -0.05 | 0.19 | 0.21 | 0.48 | -0.0003 | 0.25 |
| | 3 | 0.16 | 0.33 | 0.09 | 0.24 | -0.21 | 0.69 | -0.10 | 0.17 |
| | 4 | 0.18 | 0.19 | 0.06 | 0.21 | 0.11 | 0.22 | 0.01 | 0.17 |
| | 5 | 0.15 | 0.30 | 0.01 | 0.15 | 0.02 | 0.17 | 0.03 | 0.22 |
| **Group 3** | 1 | 0.17 | 0.18 | 0.09 | 0.14 | 0.15 | 0.25 | 0.07 | 0.30 |
| | 2 | -0.0002 | 0.68 | -0.01 | 0.13 | -0.09 | 0.40 | -0.22 | 0.51 |
| | 3 | 0.01 | 0.16 | -0.09 | 0.28 | -0.09 | 0.50 | 0.04 | 0.44 |
| | 4 | 0.42 | 0.55 | -0.18 | 0.32 | 0.22 | 0.45 | 0.44 | 0.59 |
| | 5 | 0.25 | 0.41 | -0.07 | 0.55 | -0.11 | 0.25 | -0.11 | 0.78 |

Table 9.12: The mean ($\mu$) and the standard deviation ($\sigma$) of the difference in the walking time for the individual walking parts from week 1 and the retest, between the first and second dominant test, and the non-dominant test, for two of the groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

With the walking times for the five walking parts, and the mean walking times, I did some paired-sample t-test to test the null hypothesis $H_{0,7}$, $H_{0,8}$, $H_{0,9}$, $H_{0,10}$, $H_{0,13}$, $H_{0,14}$, $H_{0,15}$ and $H_{0,16}$ (Table 9.13) for the two groups. The t-test returned the p-values ($p$) and the decision result of if the null hypothesis were rejected ($h = 1$) or not ($h = 0$).

| | | **Group 2** | | | | **Group 3** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | **Week 1** | | **Retest** | | **Week 1** | | **Retest** | |
| | $i$ | $p$ | $h$ | $p$ | $h$ | $p$ | $h$ | $p$ | $h$ |
| $Wi_{D1}$, $Wi_{D2}$ | 1 | 0.22 | 0 | 0.60 | 0 | 0.01 | 1 | 0.06 | 0 |
| | 2 | $p < 0.001$ | 1 | 0.38 | 0 | 0.99 | 0 | 0.74 | 0 |
| | 3 | 0.13 | 0 | 0.20 | 0 | 0.93 | 0 | 0.34 | 0 |
| | 4 | 0.007 | 1 | 0.36 | 0 | 0.03 | 1 | 0.10 | 0 |
| | 5 | 0.12 | 0 | 0.91 | 0 | 0.07 | 0 | 0.70 | 0 |
| $W_{D1}$, $W_{D2}$ | | $p < 0.001$ | 1 | 0.28 | 0 | $p < 0.001$ | 1 | 0.30 | 0 |
| $Wi_{ND1}$, $Wi_{ND2}$ | 1 | 0.12 | 0 | 0.16 | 0 | 0.08 | 0 | 0.48 | 0 |
| | 2 | 0.15 | 0 | 0.97 | 0 | 0.49 | 0 | 0.18 | 0 |
| | 3 | 0.32 | 0 | 0.07 | 0 | 0.57 | 0 | 0.77 | 0 |
| | 4 | 0.12 | 0 | 0.89 | 0 | 0.14 | 0 | 0.03 | 1 |
| | 5 | 0.66 | 0 | 0.65 | 0 | 0.16 | 0 | 0.65 | 0 |
| $W_{ND1}$, $W_{ND2}$ | | 0.46 | 0 | 0.35 | 0 | 0.54 | 0 | 0.58 | 0 |

Table 9.13: The p-value ($p$) and the decision result for rejecting or not the null hypothesis ($h$) of the paired sample t-test testing the null hypothesis $H_{0,7}$, $H_{0,8}$, $H_{0,9}$, $H_{0,10}$, $H_{0,13}$, $H_{0,14}$, $H_{0,15}$ and $H_{0,16}$. This was done for the two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

## 9.3 Discussion

Looking at the difference in the test times between the first and second dominant tests from week 1 (Figure 9.1), most of the participants in Groups 2 and 3 are decreasing their test times in the second walkthrough. For the participants in Group 1, approximately a third of them increase the test times in the second walkthrough. When looking at the difference between the first and second non-dominant tests from week 1 (Figure 9.2), there are approximately the same amount of participants as it was in the dominant tests for each group that increases their test times in the second walkthrough. From the mean difference for the different groups (Table 9.1), the difference decrease between the dominant and the non-dominant tests for Groups 2 and 3. Group 1 increases their mean difference in the non-dominant tests.
The results from the paired-sample t-tests of the null hypothesis $H_{0,1}$ and $H_{0,2}$ (Table 9.2)

shows that $H_{0,1}$ were rejected for both Group 2 and 3. For the null hypothesis, $H_{0,2}$ only Group 3 rejected it, but if I had set the significant level to be $\alpha = 0.01$, they would have accepted the null hypothesis. With changing the significant level, the requirement for rejecting the null hypothesis would have been stricter.
The difference in the mean test times between the first and second walkthrough is not from a normal distribution with mean$= 0$ for Group 2 and 3 in the dominant tests, and for the non-dominant tests only the difference in Group 2 is from a normal distribution with mean$= 0$.

When looking at the difference in the kicking times between the first and second dominant tests (Figure 9.3) and non-dominant tests (Figure 9.4) from week 1, approximately half of the participants in Group 1 increases there mean kicking time in the second walkthrough for both the dominant and non-dominant tests. For Group 2, the participants increase their mean kicking times in the second dominant test. In the non-dominant tests, approximately a third of the participants increase their mean kicking time. In Group 3, for both the dominant and non-dominant tests, a third of the participants increased their mean kicking times in the second walkthrough.
The mean difference in the kicking times for the three groups (Table 9.3) show that the difference between the first and second tests for both the dominant and non-dominant tests are small. For the dominant tests, the difference is negative for Groups 1 and 3. This means that, on average, the mean kicking times in these groups increase in the second walkthrough. I also found the mean difference for the five individual kicks between the first and second dominant/non-dominant tests (Table 9.4). In these results, some kicks have small differences, and some have higher differences. The size of the difference does not depend on kick, dominant and non-dominant, or group.
I did paired-sample t-tests to test out the null hypothesis $H_{0,3}$, $H_{0,4}$, $H_{0,5}$ and $H_{0,6}$ (Table 9.5). There were only two rejected tests, for Group 2 kick number 5 and Group 1 kick number 3. For both of these kicks, if I had used the significant level $\alpha = 0.01$, these tests would have accepted the null hypothesis. With changing the significant level, the requirement for rejecting the null hypothesis would have been stricter. The results of the paired-sample t-tests tell me the difference in the kicking times between the first and second walkthrough are very small, close to zero and from a normal distribution with mean$= 0$.

For the difference in times from the walking parts between the first and second dominant tests from week 1 (Figure 9.5), all but one of the participants in Group 2 and 3 decrease their mean times in the second walkthrough. In Group 1, approximately half of the participants decreased the mean walking time in the second walkthrough. When looking at the difference in the times for the non-dominant tests in week 1 (Figure 9.6), fewer participants in Group 1 increase their mean time in the second walkthrough compared to that from the dominant tests. For Groups 2 and 3, there is an increase in the participants that increase their mean walk time in the second walkthrough compared to the dominant tests. The same can be seen in the mean

difference of the walk time difference for the participants in the groups (Table 9.6). The mean difference for Group 2 and 3 decreases between the dominant and the non-dominant tests, but for Group 1, the difference stays the same.
When looking at the difference for the individual walking parts (Table 9.7), there are differences between the walking parts, the groups, and if one is looking at the dominant or the non-dominant tests. Walking parts 3 and 5 have large changes between the dominant and non-dominant tests. Group 2 has larger changes between the dominant and non-dominant tests for walking parts 2 and 3. Group 1, the dominant and non-dominant tests differ significantly for walking parts 4 and 5.
From the paired-sample t-tests testing the null hypothesis $H_{0,7}$, $H_{0,8}$, $H_{0,9}$ and $H_{0,10}$ (Table 9.8) 12 tests reject the null hypothesis. Both Group 2 and 3 reject the null hypothesis $H_{0,7}$. Group 3 is the only group that rejected the null hypothesis $H_{0,8}$, but if the significant level had been $\alpha = 0.01$, the null hypothesis would have been accepted. When it comes to the null hypothesis $H_{0,9}$ for Group 2 and 3, reject some of the tests. Group 2 rejects 3 of the 5 tests, where one of the tests would have been accepted with a significant level $\alpha = 0.01$. Group 3 also rejects 3 of 5 tests, where two of the tests would have been accepted with a significant level $\alpha = 0.01$. For the null hypothesis $H_{0,10}$, are Group 3 rejecting 2 of the 5 tests, and one of them would have been accepted with a significant level $\alpha = 0.01$. Group 1 rejects one of the tests on the $H_{0,10}$.
By changing the significant level, the requirement for rejecting the null hypothesis would have been stricter. This would not change much, as most of the null hypothesis for the dominant tests to Groups 2 and 3 were rejected, which means that the difference in the walking parts between the first and second walkthrough for Groups 2 and 3 are not so small and so the differences is not from a normal distribution with mean$= 0$.

After looking at the difference between the first and second walkthrough of the dominant and the non-dominant tests in week 1, I looked at the difference in the retest and compared them to the results from week 1. I started by looking at the difference in the test times from the dominant tests (Figure 9.7) and the non-dominant tests (Figure 9.8). For the dominant tests, most of the participants decrease their test time in the second walkthrough in week 1, but only approximately half of them do the same in the retest. In the non-dominant tests, there was almost equal when it comes to how many participants decreased their test time in the second walkthrough for the week 1 test and the retests.
Looking at the mean and standard deviation of the difference in the test times for the dominant and non-dominant tests (Table 9.9) from week 1 and retest, there is a big decrease in the difference between week 1 and retests. The decrease between week 1 and retest is the biggest when looking at the dominant tests. The smallest change is between week 1 and retest in the non-dominant tests for Group 3.
The results from the paired-sample t-test of the null hypothesis $H_{0,1}$, $H_{0,2}$, $H_{0,11}$ and $H_{0,12}$ (Table 9.10) shows that both the groups rejects the $H_{0,1}$, and the rest of the null hypothesis are accepted. This means that the difference in the test time between the first and second walk-

through is not from a normal distribution with mean= 0 in week 1 but is from that distribution in the retests.

I looked at the walking times from both week 1 and the retest. The difference between the first and second dominant tests (Figure 9.9) shows that all the participants in Group 3 decrease their mean walking time in the second dominant test from week 1, but approximately half decrease their mean walking time in the retests. For Group 2, one participant increases the mean walking time in week 1, and around a third of the participants increase their mean walking time in the retest.
When it comes to the difference between the first and second non-dominant tests (Figure 9.10), the mean walking time increases from the first to the second walkthrough for approximately half of the participants in each group, for week 1 and the retest. For the mean difference of the mean waling times (Table 9.11), both the groups have a higher decrease between week 1 and retest for the dominant tests compared to the non-dominant tests. Some of the same things can be observed from the mean difference for the five individual walking parts (Table 9.12), but it varies very between the two groups, the walking parts, and between the dominant and the non-dominant tests.
From the paired-sample t-tests testing the different null hypotheses (Table 9.13) there was only one test when looking at the non-dominant tests that were rejected. This was for walking part 4 in Group 3 from the retests. The rest of the week 1 dominant tests were rejected. Both groups rejected the null hypothesis $H_{0,7}$. For Group 2, there were two tests on the week 1 dominant tests that were rejected, and for Group 3 were also two tests on the week 1 dominant tests that were rejected. The two rejected tests for Group 3 would have been accepted if the significant level was changed to $\alpha = 0.01$. With changing the significant level, the requirement for rejecting the null hypothesis would have been stricter. With all the retest tests accepted, the difference between the first and second walkthrough is from a normal distribution with mean= 0. The few rejected tests in week 1 means that the difference for these tests was not from a normal distribution with mean= 0. This means that the differences found bewteen the first and second walkthrough in week 1 that made the tests reject the null hypotesis, has been reduced in the retests so the null hypothesis were accepted.

There seems to be a change in the times between the first and second walkthrough, and this change in time is greatest for the dominant tests and decreases when looking at the non-dominant tests. From the analysis, it looks like the change in the times most likely are from the walking parts. When looking at the different results from the paired-same t-tests, it indicates a difference between the first and second dominant tests that is not there when looking at the same tests from the retests and looking at the non-dominant tests in both week 1 and retests. These observations point to that there is some learning effect from walking through the test one time.

# Chapter 10

# Effect in a rehabilitation stay

This chapter will investigate if there is a change in the participants' times from the Six Spot Step Test (SSST) between the week 1 tests and the retest.
There were only 32 patients that were retested (25 of them I used in this analysis) and non of the controls. Therefore I only looked at the retested patients and their SSSTs from both the week 1 testing and the retests. The patients will be divided into two groups based on their EDSS (Expanded Disability Statue Scale) (Section 5.3). The two groups will be:

- Group 2: the patients with an EDSS $< 4$
- Group 3: the patients with an EDSS $\geq 4$

From the analysis I did on the learning effects in the SSSTs, I found a change between the first and second time the participants walked through kicking with their dominant foot and a smaller change between the first and second time the participants walked through kicking with their non-dominant foot. I also found that this change between tests was not there when looking at the retesting of the SSSTs. Due to this, I have decided only to analyze the second time the participants walk through the SSST kicking with the dominant and non-dominant foot from the week 1 testing and then retesting. During the analyze I will be looking at different times from the participants:

- the mean time they used based on the second SSST kicking with their dominant and non-dominant foot in the week 1 testing ($W1$)
- the mean time they used from both the second SSST kicking with their dominant foot and non-dominant foot in the retest ($RT$)
- the time they used for the second SSST kicking with their dominant foot ($D_{W1}$) and the non-dominant foot ($ND_{W1}$) in the week 1 testing
- the time they used for the second SSST kicking with their dominant foot ($D_{RT}$) and the non-dominant foot ($ND_{RT}$) in the retest

- the mean kicking time they used from both the second SSST kicking with their dominant foot and their non-dominant foot in the week 1 testing ($K_{W1}$)
- the mean kicking time they used from both the second SSST kicking with their dominant foot and non-dominant foot in the retest ($K_{RT}$)
- the mean kicking time they used from the second SSST kicking with their dominant/non-dominant foot in the week 1 testing ($K_{D,W1}/K_{ND,W1}$)
- the mean kicking time they used from the second SSST kicking with their dominant/non-dominant foot in the retest ($K_{D,RT}/K_{ND,RT}$)
- the mean time for the walking part they used in both the second SSST kicking with their dominant foot and their non-dominant foot in the week 1 testing ($W_{W1}$)
- the mean time for the walking part they used in both the second SSST kicking with their dominant foot and non-dominant foot in the retest ($W_{RT}$)
- the mean time for the walking part they used in the second SSST kicking with their dominant/non-dominant foot in the week 1 testing ($W_{D,W1}/W_{ND,W1}$)
- the mean time for the walking part they used in the second SSST kicking with their dominant/non-dominant foot in the retest ($W_{D,RT}/W_{ND,RT}$)

I started by investigating differences in the mean test times between the week 1 tests and the retests and seeing if there were any differences when looking separately at the dominant and non-dominant tests. Then I looked at the mean kicking times from the week 1 tests and the retests and if there were changes in the difference between the two testings when kicking with the dominant and non-dominant foot. I also looked into the mean times for the walking parts from the week 1 tests and the retests to determine if there were changes in the times.
Within the analysis I did, I also tested out different null hypothesis $H_0$ with significant level $\alpha$:

- the difference between the mean test times from the week 1 tests and the retests is normally distributed with mean= 0 ($H_{0,1}$, with $\alpha = 0.05$)
- mean time from $D_{W1}/ND_{W1}$ and $D_{RT}/ND_{RT}$ are both independent normal distributed, with equal mean and variance ($H_{0,2}$, with $\alpha = 0.05$)
- the difference between the mean kicking times from the week 1 tests and the retests is normally distributed with mean= 0 ($H_{0,3}$, with $\alpha = 0.05$)
- mean time from $K_{D,W1}/K_{ND,W1}$ and $K_{D,RT}/K_{ND,RT}$ are both independent normal distributed, with equal mean and variance ($H_{0,4}$, with $\alpha = 0.05$)

- the difference between the mean times of the walking parts from the week 1 tests and the retests is normally distributed with mean= 0 ($H_{0,5}$, with $\alpha = 0.05$)
- mean time from $W_{D,W1}/W_{ND,W1}$ and $W_{D,RT}/W_{ND,RT}$ are both independent normal distributed, with equal mean and variance ($H_{0,6}$, with $\alpha = 0.05$)
- mean test time from the week 1 tests for the controls and the participants in Group 3 are both independent normal distributed, with equal mean and variance ($H_{0,7}$, with $\alpha = 0.05$)
- mean test time for the controls in the week 1 tests and the participants in Group 3 from the retests are both independent normal distributed, with equal mean and variance ($H_{0,8}$, with $\alpha = 0.05$)
- mean kicking time from the week 1 tests for the controls and the participants in Group 3 are both independent normal distributed, with equal mean and variance ($H_{0,9}$, with $\alpha = 0.05$)
- mean kicking time for the controls in the week 1 tests and the participants in Group 3 from the retests are both independent normal distributed, with equal mean and variance ($H_{0,10}$, with $\alpha = 0.05$)

## 10.1 Test times

The first thing I did was investigate any changes in the test times between the week 1 tests (W1) and the retests (RT) after the rehabilitation stay. I started by investigating the mean test times for every participant (Figure 10.1) and found the difference between the W1 times and RT times for every participant (Figure 10.2) by subtracting the RT times from the W1 times.

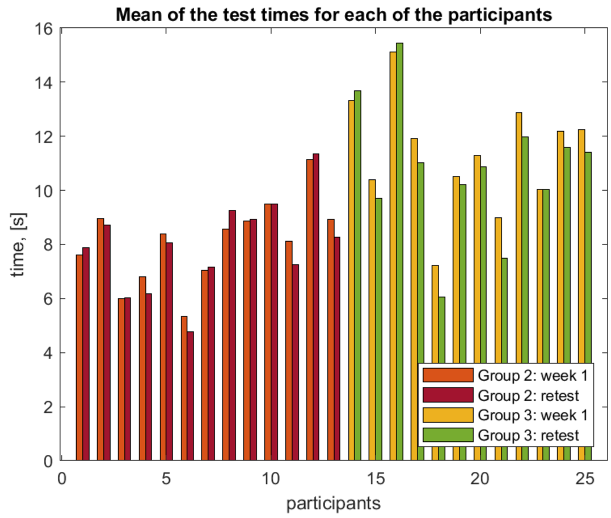


Figure 10.1: The mean test time for each of the participants for week 1 (W1) testing and the retests (RT), based on the two test times from each of the weeks. The participants are sorted and color coordinated into two groups based on there EDSS.

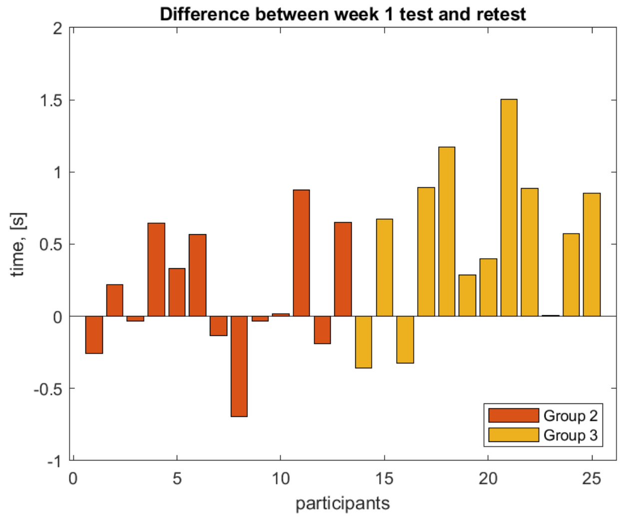


Figure 10.2: The difference in the mean test times between the tests from week 1 (W1) and the retest (RT) for each of the participants. The participants are sorted and color coordinated into two groups based on there EDSS.

From the mean test times for W1 and RT, and from the difference between W1 times and

RT times, I found the mean and standard deviation for the two groups (Table 10.1).

| | Week 1 | | Retest | | Difference | |
|---|---|---|---|---|---|---|
| | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] |
| **Group 2** | 8.10 | 1.55 | 7.95 | 1.72 | 0.15 | 0.45 |
| **Group 3** | 11.34 | 2.10 | 10.80 | 2.49 | 0.55 | 0.57 |

Table 10.1: The mean ($\mu$) and standard deviation ($\sigma$) of the mean test time from week 1 and the retest, and of the difference in time between mean test time from week 1 (W1) and retests (RT). This for the two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

I then did a paired-sample t-test on the mean test times to test the null hypothesis $H_{0,1}$ for both the groups (Table 10.2). The t-test returned the p-values ($p$) and the decision result of if the $H_{0,1}$ were rejected ($h = 1$) or not ($h = 0$).

| | Group 2 | | Group 3 | |
|---|---|---|---|---|
| | $p$ | $h$ | $p$ | $h$ |
| **Week 1, Retest** | 0.24 | 0 | 0.007 | 1 |

Table 10.2: The p-values ($p$) and the decision result for rejecting or not the null hypothesis ($h$) of the paired-sample t-tests testing $H_{0,1}$. This was done both of the groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

### 10.1.1 Dominant and non-dominant tests

After looking at the mean test times, I wanted to see if there were any differences in the times used between the tests done using the dominant foot to kick and the non-dominant foot. I started by finding the test times for the dominant foots from week 1 ($D_{W1}$) and the retest ($D_{RT}$) (Figure 10.3), and the test times for the non-dominant foot from week 1 ($ND_{W1}$) and the retest ($ND_{RT}$) (Figure 10.5). Then I found the difference between the test times from week 1 ($D_{W1}$ and $ND_{W1}$) and the retest ($D_{RT}$ and $ND_{RT}$) for each of the participants (Figures 10.4 and 10.6). This I did by subtracting the times from the retests ($D_{RT}$ and $ND_{RT}$) from the times from week 1 ($D_{W1}$ and $ND_{W1}$).

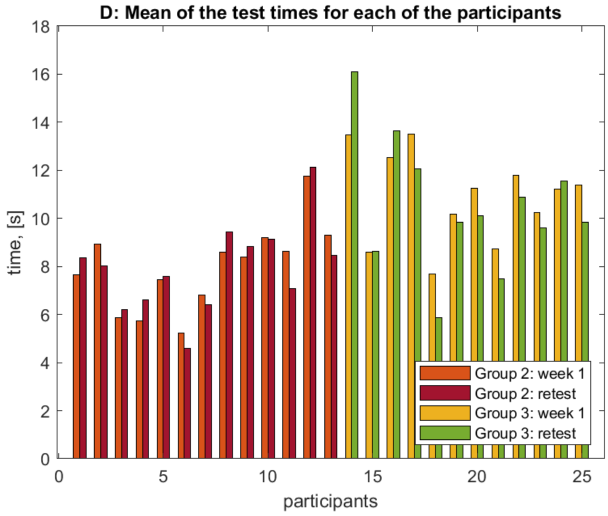


Figure 10.3: The dominant test (D) time for each of the participants from week 1 and the retest. The participants are sorted and color coordinated into two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

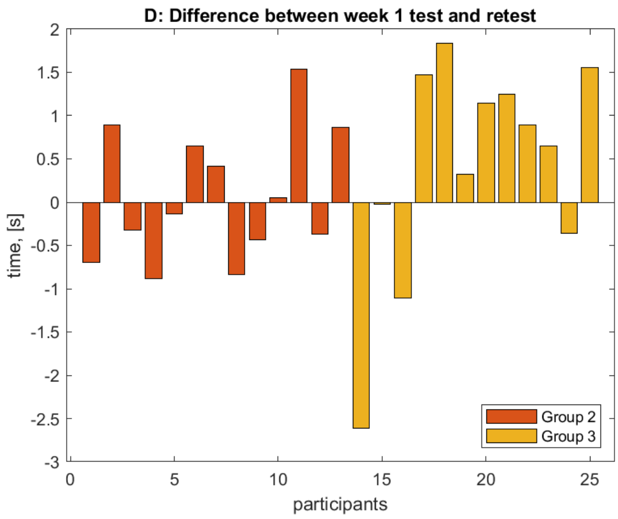


Figure 10.4: The difference in the dominant test (D) times between the tests from week 1 and the retest for each of the participants. The participants are sorted and color coordinated into two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

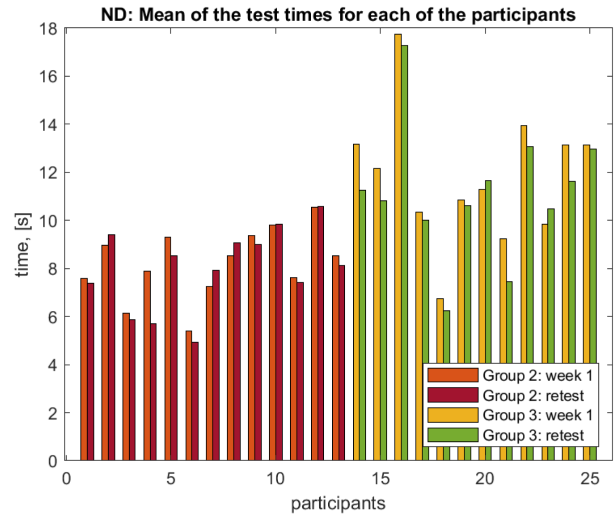


Figure 10.5: The non-dominant test (ND) time for each of the participants for week 1 and the retest. The participants are sorted and color coordinated into two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

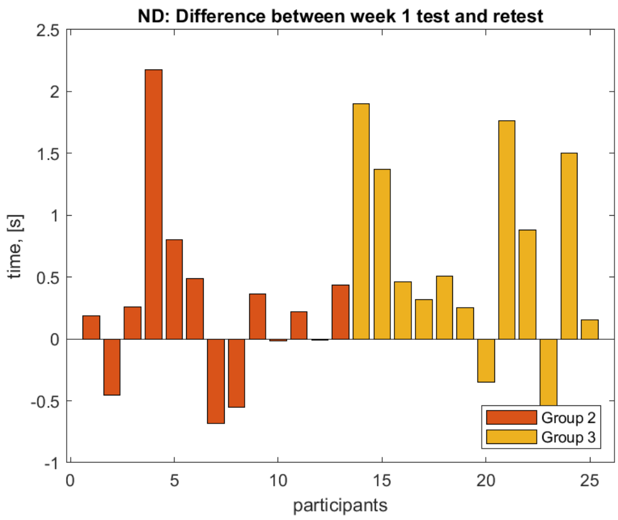


Figure 10.6: The difference in the non-dominant test (ND) times between the tests from week 1 and the retest for each of the participants. The participants are sorted and color coordinated into two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

From the mean test times for $D_{W1}$, $D_{RT}$, $ND_{W1}$ and $ND_{RT}$ I found the mean ($\mu$) and standard deviation ($\sigma$) for the two groups (Table 10.3). I also found the mean ($\mu$) and standard deviation ($\sigma$) for the differences between the times from week 1 (W1) and the retest (RT) for the two groups (Table 10.4).

| | $\mathbf{D}_{W1}$ | | $\mathbf{D}_{RT}$ | | $\mathbf{ND}_{W1}$ | | $\mathbf{ND}_{RT}$ | |
|---|---|---|---|---|---|---|---|---|
| | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] |
| **Group 2** | 7.97 | 1.78 | 7.91 | 1.86 | 8.23 | 1.46 | 7.98 | 1.69 |
| **Group 3** | 10.89 | 1.87 | 10.47 | 2.71 | 11.80 | 2.77 | 11.12 | 2.78 |

Table 10.3: The mean ($\mu$) and standard deviation ($\sigma$) of the dominant test times ($D_{W1}$ and $D_{RT}$) and the non-dominant test times ($ND_{W1}$ and $ND_{RT}$) from the week 1 testing and the retesting, for the two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

| | **Dominant tests** | | **Non-dominant tests** | |
|---|---|---|---|---|
| | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] |
| **Group 2** | 0.06 | 0.76 | 0.25 | 0.72 |
| **Group 3** | 0.42 | 1.29 | 0.68 | 0.82 |

Table 10.4: The mean ($\mu$) and standard deviation ($\sigma$) of the difference in time between $D_{W1}$ and $D_{RT}$ (Dominant tests), and $ND_{W1}$ and $ND_{RT}$ (Non-dominant tests), for the two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

I then did the paired-sample t-test to test the null hypothesis $H_{0,2}$ (Table 10.5). The t-test returned the p-values ($p$) and the decision result of if the $H_{0,2}$ were rejected ($h = 1$) or not ($h = 0$).

| | **Group 2** | | **Group 3** | |
|---|---|---|---|---|
| | $p$ | $h$ | $p$ | $h$ |
| $\mathbf{D}_{W1}$, $\mathbf{D}_{RT}$ | 0.79 | 0 | 0.28 | 0 |
| $\mathbf{ND}_{W1}$, $\mathbf{ND}_{RT}$ | 0.24 | 0 | 0.015 | 1 |

Table 10.5: The p-value ($p$) and the decision result for rejecting or not the null hypothesis ($h$) of the paired-sampled t-test $H_{0,2}$. This is done for the two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

## 10.2 Kicking times

After looking at the times used for the hole test, I wanted to investigate the changes when the patients kicked during the test after a rehabilitation stay. First, I started by looking at the mean kicking time from all the SSST the patients did in week 1 ($K_{W1}$) and the retests ($K_{RT}$) (Figure 10.7) and the difference between the mean times from $K_{W1}$ and $K_{RT}$ (Figure 10.8).

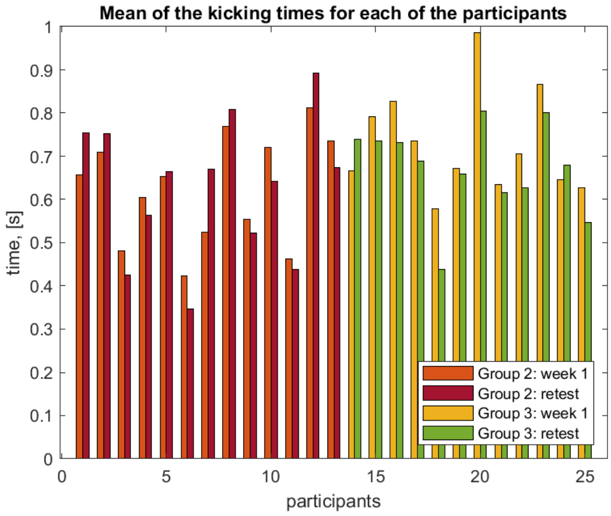


Figure 10.7: The mean kicking times for the participants from the week 1 tests ($K_{W1}$) and the retests ($K_{RT}$). The participants are sorted and color-coordinated into two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

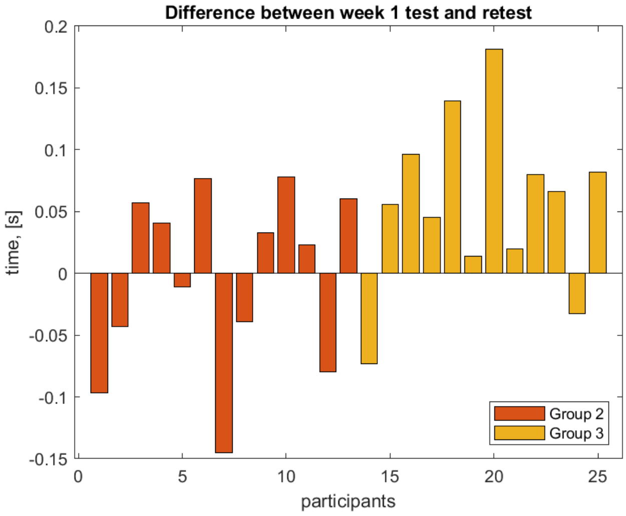


Figure 10.8: The difference in the mean kicking times between $K_{W1}$ times and $K_{RT}$ times for each of the participants. The participants are sorted and color coordinated into two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

I found the mean ($\mu$) and standard deviation ($\sigma$) of the mean kicking times $K_{W1}$ and $K_{RT}$, and the difference in the mean kicking times between $K_{W1}$ and $K_{RT}$ (Table 10.6). The difference in the mean kicking times was found by subtracting the $K_{RT}$ times from the $K_{W1}$ times.

| | **Week 1** | | **Retest** | | **Difference** | |
|---|---|---|---|---|---|---|
| | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] |
| **Group 2** | 0.62 | 0.13 | 0.63 | 0.16 | $-0.36\times10^{-2}$ | 0.07 |
| **Group 3** | 0.73 | 0.12 | 0.67 | 0.11 | 0.06 | 0.07 |

Table 10.6: The mean ($\mu$) and standard deviation ($\sigma$) of the mean kicking times from the week 1 testing and the retest, and the mean ($\mu$) and standard deviation ($\sigma$) of the difference in the mean kicking times. This was found for the two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

Then I did a paired-sample t-test to test the null hypothesis $H_{0,3}$ (Table 10.7) for both of the groups. The t-test returned the p-values ($p$) and the decision result of if the $H_{0,3}$ were rejected ($h = 1$) or not ($h = 0$).

| | Group 2 | | Group 3 | |
|---|---|---|---|---|
| | $p$ | $h$ | $p$ | $h$ |
| **Week 1, Retest** | 0.86 | 0 | 0.018 | 1 |

Table 10.7: The p-value ($p$) and the decision result for rejecting or not the null hypothesis ($h$) of the paired-sampled t-test $H_{0,3}$. This is done for the two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

### 10.2.1 Dominant and non-dominant tests

I wanted to see if there were differences in the kicking time when looking only at the kicks done with the dominant foot $D$ and the non-dominant foot (ND). The first thing I did were to find the mean kicking time done with the dominant foot from week 1 ($K_{D,W1}$) and the retest ($K_{D,RT}$) (Figure 10.9), and the mean kicking time done with the non-dominant foot from week 1 ($K_{ND,W1}$) and the retest ($K_{ND,RT}$) (Figure 10.11). Then I found the difference in the mean kicking time between the kicking done in week 1 ($K_{D,W1}$ and $K_{ND,W1}$) and the retest ($K_{D,RT}$ and $K_{ND,RT}$) (Figures 10.10 and 10.12).

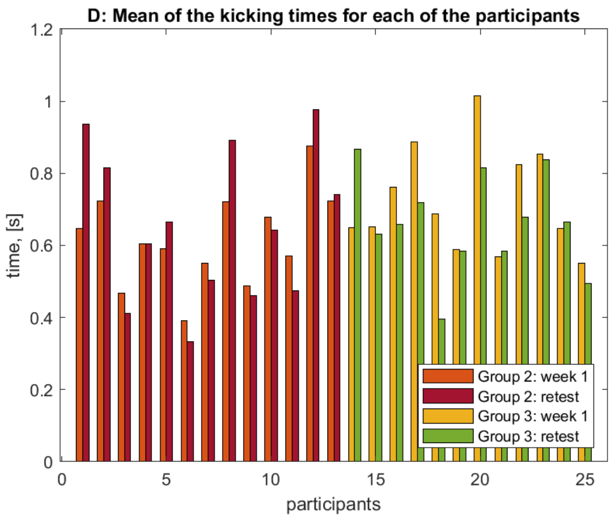


Figure 10.9: The mean kicking time done with the dominant foot for each of the participants, from week 1 and the retest. The participants are sorted and color-coordinated into two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

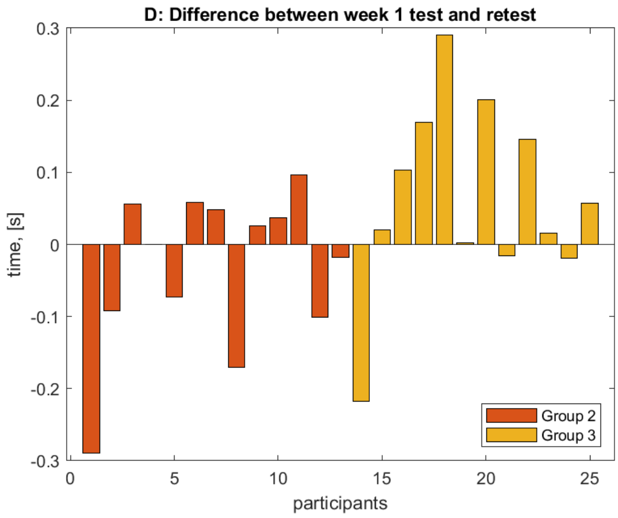


Figure 10.10: The difference in the mean kicking time done with the dominant foot between week 1 tests and the retest, for each of the participants. The participants are sorted and color-coordinated into two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

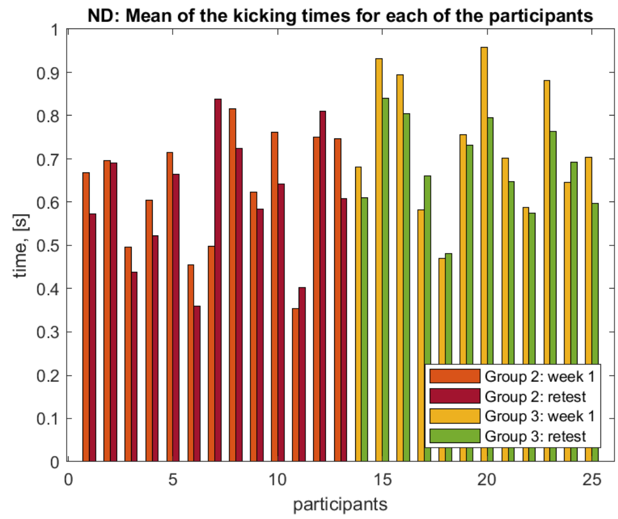


Figure 10.11: The mean kicking time done with the non-dominant foot for each of the participants from week 1 and the retest. The participants are sorted and color coordinated into two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

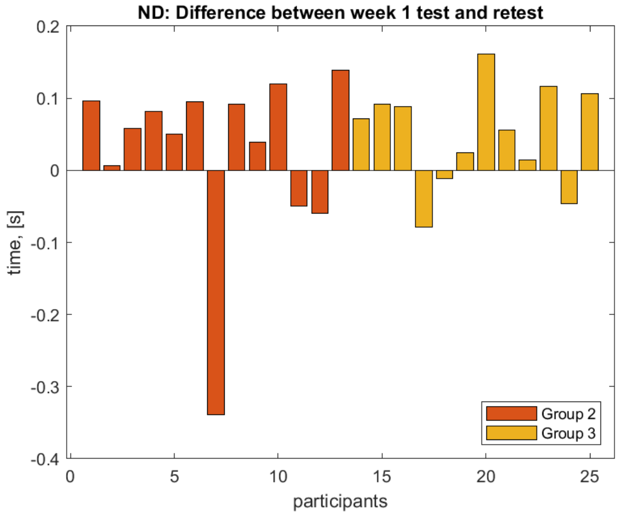


Figure 10.12: The difference in the mean kicking time done with the non-dominant foot between week 1 tests and the retest, for each of the participants. The participants are sorted and color coordinated into two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

The mean ($\mu$) and standard deviation ($\sigma$) of the mean kicking time done with the dominant and the non-dominant foot from week 1 ($K_{D,W1}$ and $K_{ND,W1}$) and the retest ($K_{D,RT}$ and $K_{ND,RT}$) were found (Table 10.8). I also found the mean ($\mu$) and standard deviation ($\sigma$) of the difference in the kicking time between the week 1 test ($K_{D,W1}$ and $K_{ND,W1}$) and the retest ($K_{D,RT}$ and $K_{ND,RT}$) (Table 10.9).

| | $K_{D,W1}$ | | $K_{D,RT}$ | | $K_{ND,WT}$ | | $K_{ND,RT}$ | |
|---|---|---|---|---|---|---|---|---|
| | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] |
| **Group 2** | 0.62 | 0.13 | 0.65 | 0.21 | 0.63 | 0.14 | 0.60 | 0.15 |
| **Group 3** | 0.72 | 0.15 | 0.66 | 0.14 | 0.73 | 0.16 | 0.68 | 0.11 |

Table 10.8: The mean ($\mu$) and standard deviation ($\sigma$) of the mean kicking times done with the dominant ($k_{D,W1}$ and $K_{D,RT}$) and the non-dominant foot ($K_{ND,W1}$ and $K_{ND,RT}$), for the two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

| | Dominant tests | | Non-dominant tests | |
|---|---|---|---|---|
| | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] |
| **Group 2** | -0.03 | 0.11 | 0.03 | 0.13 |
| **Group 3** | 0.06 | 0.13 | 0.05 | 0.07 |

Table 10.9: The mean ($\mu$) and standard deviation ($\sigma$) of the difference between the mean kicking times $K_{D,W1}$ and $K_{D,RT}$ (Dominant tests), and $K_{ND,W1}$ and $K_{ND,RT}$ (Non-dominant tests), for the two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

Having found the difference, I did paired-sample t-test to test the null hypothesis $H_{0,4}$ (Table 10.10) for both the groups. The t-test returned the p-values ($p$) and the decision result of if the $H_{0,4}$ were rejected ($h = 1$) or not ($h = 0$).

| | Group 2 | | Group 3 | |
|---|---|---|---|---|
| | $p$ | $h$ | $p$ | $h$ |
| $K_{D,W1}$, $K_{D,RT}$ | 0.31 | 0 | 0.13 | 0 |
| $K_{ND,W1}$, $K_{ND,RT}$ | 0.48 | 0 | 0.03 | 1 |

Table 10.10: The p-value ($p$) and the decision result for rejecting or not the null hypothesis ($h$) of the two-sampled t-test testing $H_{0,4}$. This is done for the two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

## 10.3 Walking times

After looking at the kicking part of the SSST, I wanted to investigate the possible changes in the walking parts of the test after a rehabilitation stay. First, I found the mean times of the walking parts for each of the patients from week 1 ($W_{W1}$) and the retest ($W_{RT}$) (Figure 10.13) and the difference in the mean time for the walking parts between the $W_{W1}$ and the $W_{RT}$ (Figure 10.14).

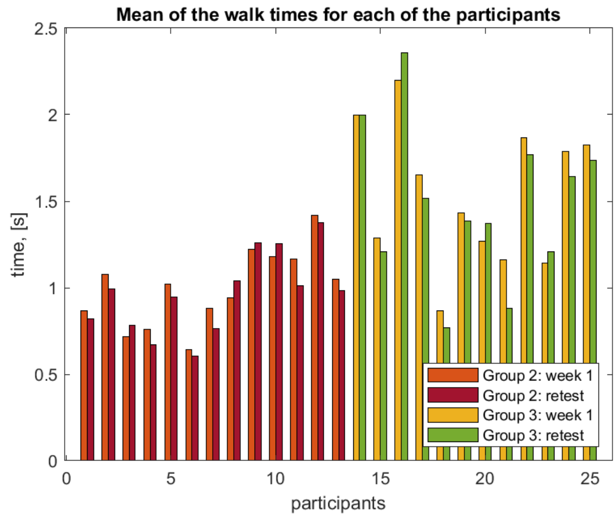


Figure 10.13: The mean times of the walking parts for the participants from the week 1 tests ($W_{W1}$) and the retests ($W_{RT}$). The participants are sorted and color-coordinated into two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

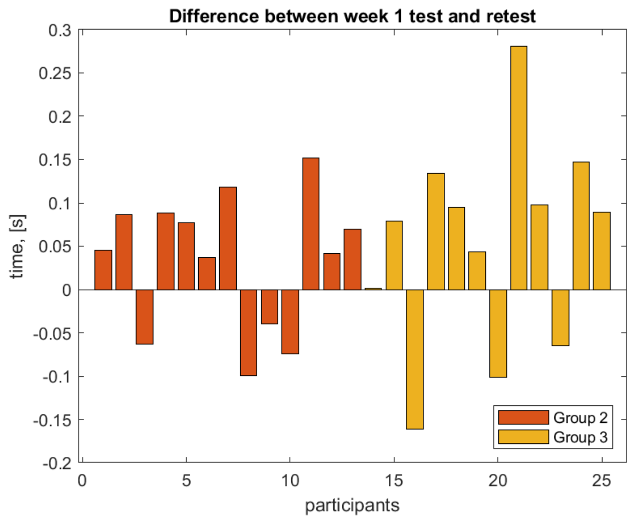


Figure 10.14: The difference in the mean time of the walking parts between the tests from $W_{W1}$ and the $W_{RT}$ for each of the participants. The participants are sorted and color coordinated into two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

The mean ($\mu$) and standard deviation ($\sigma$) of the mean times for the walking parts from

$W_{W1}$ and the $W_{RT}$, and the difference between the mean times from $W_{W1}$ and the $W_{RT}$ were found (Table 10.11) for the two groups.

| | **Week 1** | | **Retest** | | **Difference** | |
|---|---|---|---|---|---|---|
| | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] |
| **Group 2** | 1.00 | 0.22 | 0.96 | 0.24 | 0.03 | 0.08 |
| **Group 3** | 1.54 | 0.41 | 1.49 | 0.45 | 0.05 | 0.12 |

Table 10.11: The mean ($\mu$) and standard deviation ($\sigma$) of the mean times for the walking parts from the week 1 testing and the retests, and the mean ($\mu$) and standard deviation ($\sigma$) of the difference in the mean times for the walking parts. This was found for the two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

I then did a paired-sample t-test to test the null hypothesis $H_{0,5}$ (Table 10.12) for both of the groups. The t-test returned the p-values ($p$) and the decision result of it the $H_{0,5}$ were rejected ($h = 1$) or not ($h = 0$).

| | **Group 2** | | **Group 3** | |
|---|---|---|---|---|
| | $p$ | $h$ | $p$ | $h$ |
| **Week 1, Retest** | 0.15 | 0 | 0.15 | 0 |

Table 10.12: The p-value ($p$) and the decision result for rejecting or not the null hypothesis ($h$) of the paired-sample t-test testing $H_{0,5}$. This is done for the two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

### 10.3.1 Dominant and non-dominant tests

I wanted to divide the walking parts into the parts done in the dominant test (D) and the non-dominant test (ND). I started by finding the mean time for the waling times from the dominant test in week 1 ($W_{D,W1}$) and the retest ($W_{D,RT}$) (Figure 10.15), and from the non-dominant test in week 1 ($W_{ND,W1}$) and the retest ($W_{ND,RT}$) (Figure 10.17). Then I found the difference between the times from week 1 ($W_{D,W1}$ and $W_{ND,W1}$) and the retest ($W_{D,RT}$ and $W_{ND,RT}$) (Figures 10.16 and 10.18).

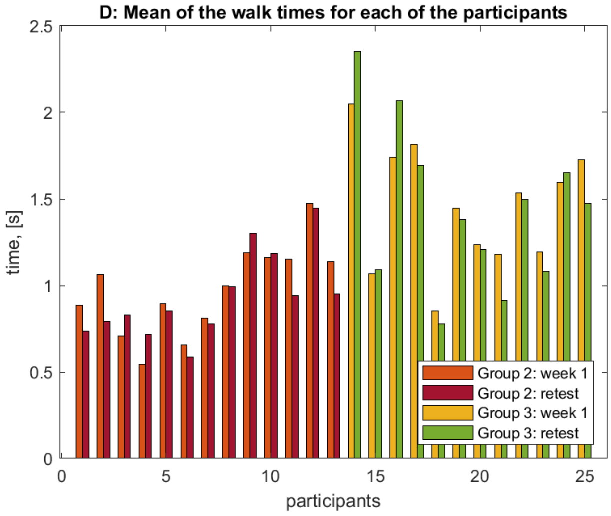


Figure 10.15: The mean times of the walking times from the dominant tests for each of the participants for week 1 and the retest. The participants are sorted and color coordinated into two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

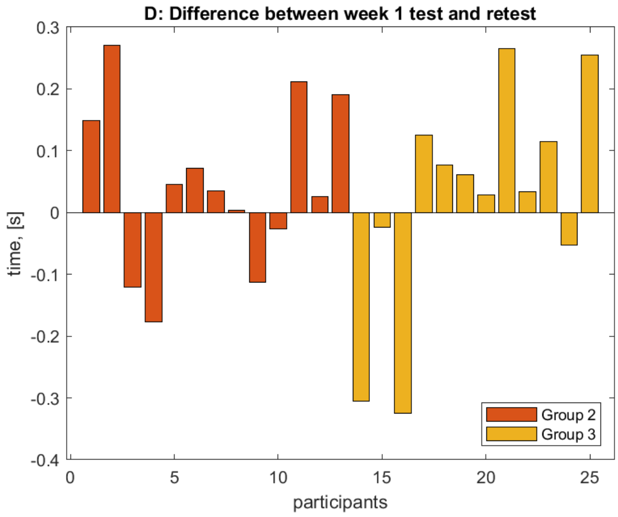


Figure 10.16: The difference in the mean time of the walking parts from the dominant tests between week 1 tests and the retest for each of the participants. The participants are sorted and color-coordinated into two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

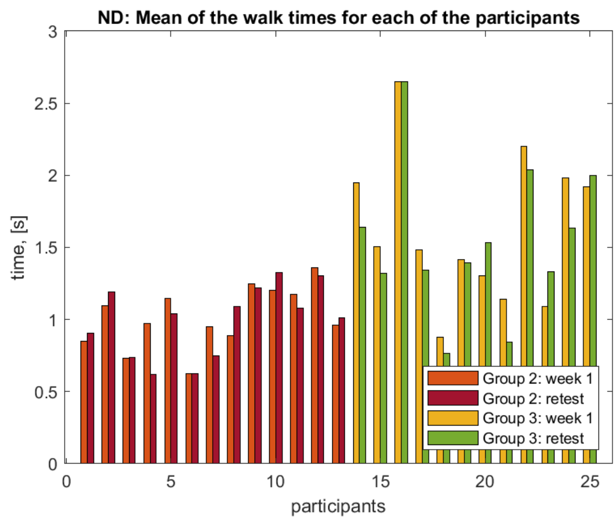


Figure 10.17: The mean times of the walking times from the non-dominant tests for each of the participants for week 1 and the retest. The participants are sorted and color coordinated into two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

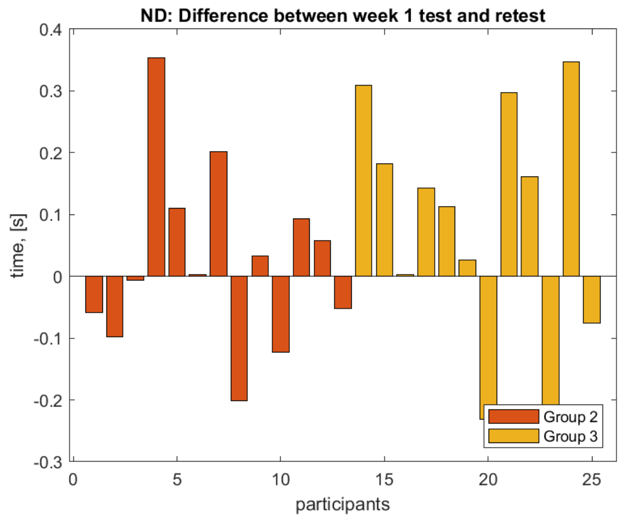


Figure 10.18: The difference in the mean time of the walking parts from the non-dominant tests between week 1 tests and the retest for each of the participants. The participants are sorted and color-coordinated into two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

I then found the mean ($\mu$) and standard deviation ($\sigma$) of the mean time from the waling parts for the dominant tests and the non-dominant tests in the week 1 tests ($W_{D,W1}$ and $W_{ND,W1}$) and the retests ($W_{D,RT}$ and $W_{ND,RT}$) (Table 10.13). The mean ($\mu$) and standard deviation ($\sigma$) of the difference between the times of the walking parts from week 1 ($W_{D,W1}$ and $W_{ND,W1}$) and the retest ($W_{D,RT}$ and $W_{ND,RT}$) were also found (Table 10.14).

| | $W_{D,W1}$ | | $W_{D,RT}$ | | $W_{ND,W1}$ | | $W_{ND,RT}$ | |
|---|---|---|---|---|---|---|---|---|
| | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] |
| **Group 2** | 0.98 | 0.26 | 0.93 | 0.25 | 1.02 | 0.21 | 0.99 | 0.25 |
| **Group 3** | 1.45 | 0.35 | 1.43 | 0.46 | 1.63 | 0.52 | 1.54 | 0.52 |

Table 10.13: The mean ($\mu$) and standard deviation ($\sigma$) of the mean times for the walk parts from the dominant tests ($W_{D,W1}$ and $W_{D,RT}$) and the non-dominant tests ($W_{ND,W1}$ and $W_{ND,RT}$), for the two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

| | **Dominant tests** | | **Non-dominant tests** | |
|---|---|---|---|---|
| | $\mu$, [s] | $\sigma$, [s] | $\mu$, [s] | $\sigma$, [s] |
| **Group 2** | 0.04 | 0.14 | 0.02 | 0.15 |
| **Group 3** | 0.02 | 0.18 | 0.09 | 0.20 |

Table 10.14: The mean ($\mu$) and standard deviation ($\sigma$) of the difference between the mean times for the walking parts $W_{D,W1}$ and $W_{D,RT}$ (Dominant tests), and $W_{ND,W1}$ and $W_{ND,RT}$ (Non-dominant tests), for the two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

I did a paired-sample t-test to test the null hypothesis $H_{0,6}$ (Table 10.15) for both the groups. The t-test returned the p-values ($p$) and the decision result of it the $H_{0,6}$ were rejected ($h = 1$) or not ($h = 0$).

| | Group 2 | | Group 3 | |
|---|---|---|---|---|
| | $p$ | $h$ | $p$ | $h$ |
| $W_{D,W1}$, $W_{D,RT}$ | 0.27 | 0 | 0.70 | 0 |
| $W_{ND,W1}$, $W_{ND,RT}$ | 0.56 | 0 | 0.16 | 0 |

Table 10.15: The p-value ($p$) and the decision result for rejecting or not the null hypothesis ($h$) of the two-sampled t-test testing $H_{0,6}$. This is done for the two groups; Group 2 patients with EDSS $< 4$, and Group 3 patients with EDSS $\geq 4$.

## 10.4 Control group vs the retest times

After investigating if there were any significant changes in the SSST times after a rehabilitation stays, I wanted to see if the changes in the times for the participants in Group 3 from the retests have affected the results from the two-sample t-tests Group 3 and the controls.
I did the two-sample t-test on the mean test times of the controls and Group 3, both of them from week 1 and when the Group 3 times were from the retests (Table 10.16) to test the null hypothesis $H_{0,7}$ and $H_{0,8}$. I also did the same two-sample t-tests but on the mean kicking times for the controls and the Group 3 participants (Table 10.17) to test the null hypothesis $H_{0,9}$ and $H_{0,10}$. The t-tests returned the p-values ($p$) and the decision result of it the $H_{0,7}$, $H_{0,8}$, $H_{0,9}$, and $H_{0,10}$ were rejected ($h = 1$) or not ($h = 0$).

| | Week $1_{control}$, Week $1_{Group3}$ | | | Week $1_{control}$, Retest$_{group3}$ | | |
|---|---|---|---|---|---|---|
| | All | D | ND | All | D | ND |
| $p$ | $p < 0.001$ | $p < 0.001$ | $p < 0.001$ | $p < 0.001$ | $p < 0.001$ | $p < 0.001$ |
| $h$ | 1 | 1 | 1 | 1 | 1 | 1 |

Table 10.16: The p-values ($p$) and the decision result for rejecting or not the null hypothesis ($h$) of the two-sample t-test testing $H_{0,7}$, and $H_{0,8}$. This is the testing of the test times between the control and group 3 for the difference in the week 1 tests and the difference in the week 1 and retest tests.

| | Week $1_{control}$, Week $1_{Group3}$ | | | Week $1_{control}$, Retest$_{group3}$ | | |
|---|---|---|---|---|---|---|
| | **All** | **D** | **ND** | **All** | **D** | **ND** |
| $p$ | 0.025 | $p < 0.001$ | $p < 0.001$ | 0.055 | $p < 0.001$ | 0.002 |
| $h$ | 1 | 1 | 1 | 0 | 1 | 1 |

Table 10.17: The p-values ($p$) and the decision result for rejecting or not the null hypothesis ($h$) of the two-sample t-test testing $H_{0,9}$, and $H_{0,10}$. This is the testing of the kicking times between the control and group 3 for the difference in the week 1 tests and the difference in the week 1 and retest tests.

## 10.5 Discussion

When looking at the mean test times for the patients and the difference in the time between the tests from week 1 (W1) and the retest (RT) (Figure 10.2), there are almost half of the patients in Group 2 that increase their mean test times in the retest. In Group 3, are there only two participants that experience the same. When looking at the mean test times in both groups (Table 10.1), a similar result can be observed. The mean for the two groups decreased for the retests. The decrease in the mean test time for Group 2 is smaller than what it is for Group 3. When doing the paired-sample t-test testing the null hypothesis $H_{0,1}$ (Table 10.2) only Group 3 rejected the null hypothesis. This means that the difference between week 1 and the retes is not a normal distribution with mean= 0. If I had placed the significant level to be $\alpha = 0.01$, Group 3 would accept the null hypotheses, as Group 2 did. By changing the significant level, I would have made it stricter to reject the null hypothesis.

For the dominant tests (Figure 10.4), around half of the participants in Group 2 increase their mean test time between week 1 and the retest. In Group 3, almost all participants reduce their test time between week 1 and the retest tests. Looking at the mean test times for the two groups and the mean difference (Tables 10.3 and 10.4), Group 2 has a very small decrease in their test time. For Group 3, the decreases in their test time by around 4%. Looks at the result from the paired-sample t-test testing the null hypothesis $H_{0,2}$ on the dominant tests (Table 10.5), both groups accept the null hypothesis. This means that the difference is from a normal distribution with mean= 0.
When looking at the non-dominant tests, the participants in the two groups have, for the most part, decreased their mean teat time from week 1 to the retest (Figure 10.6). For Group 2, around a third increases their test times from week 1 to the retest, while for Group 3, only two participants experience the same. Looking at the mean test times and the mean difference for the groups (Tables 10.3 and 10.4), there is a larger difference between the week 1 tests and the retests for both the groups compared to the dominant tests. The decrease in the test

times for Group 2 was higher than I found for the dominant tests. For Group 3, the decrease was almost 8%. The result from the paired-sample t-test of the null hypothesis $H_{0,2}$ on the non-dominant tests (Table 10.5) showed that Group 3 rejected the null hypothesis while Group 2 accepted it. This means that the difference between week 1 and the retest for Group 2 is from a normal distribution with mean= 0, but for Group 3, this is not the case. If I had placed the significant level to be $\alpha = 0.01$, Group 3 would also have accepted the null hypothesis. By changing the significant level, I would have made it stricter to reject the null hypothesis.

After looking at the test times, I investigated the mean kicking times between the tests and how they changed between week 1 and the retest. The mean kicking time from all the kicks decreases from week 1 to the retest for all but two of the participants in Group 3 (Figure 10.8). For the participants in Group 2, around half increased their kicking time in the retests. When looking at the mean kicking time and the mean difference in times for the two groups (Table 10.6), the differences between week 1 tests and the retests are small for both of the groups. Group 2 is the difference negative, which meant that, on average, the participants increased their kicking time in the retests. For both of the groups, the change in the mean kicking time is lower than 0.5%. The results from the paired t-test of the null hypothesis $H_{0,3}$ (Table 10.7) show that Group 3 rejects the null hypothesis and Group 2 accepts it. This means that for Group 2, the difference between week 1 and the retest is from a normal distribution with mean= 0. This is not the case for Group 3. If the significant level had been $\alpha = 0.01$, Group 3 would also have accepted the null hypothesis. By changing the significant level, I would have made it stricter to reject the null hypothesis.

When I separated the kicks into two, kicking with the dominant and the non-dominant foot, for kicks done with the dominant foot, around half of the participants in Group 2 increase their kicking time during the retests (Figures 10.10). For the participants in Group 3, fewer participants increased their kicking times. The mean difference in the kicking times, from kicking with the dominant foot, is negative for Group 2 (Table 10.9). This means that, on average, the participants in Group 2 increase their kicking times in the retest. For both groups, the mean increase/decrease in the kicking times is lower than 0.5%.
For the kick done with the non-dominant foot, most of the participants in both groups decrease their mean kicking time from week 1 to the retest (Figure 10.12). The mean difference, however, has not changed that much from the mean difference when kicking with the dominant foot (Table 10.9). The decrease in the kicking time from week 1 to the retest for the non-dominant kicks smaller than 0.5%. For Group 2, the average difference in kicking time decreases for the non-dominant foot. For the dominant foot, it increased.
The paired-sample t-test results of the null hypothesis $H_{0,4}$ (Table 10.10) shows that the null hypothesis is only rejected for non-dominant kicks for Group 3. The rest of the tests were accepted. This means that most of the differences between week 1 and the retests are from a normal distribution with mean= 0 (accepted null hypothesis).

The patients' mean time for walking in the tests changed from week 1 to the retest (Figure 10.14). Approximately a third of the groups' participants increased their mean time for the walking parts in the retest. On average, the difference in the mean time for the walking parts for both of the groups was minimal (Table 10.11). The decrease in the mean time for the walking parts for both the groups was smaller than 0.5%. From the paired-sample t-test to test the null hypothesis $H_{0,5}$, both the groups accepted the null hypothesis, which means that the difference between week 1 and the retest is from a normal distribution with mean$= 0$.

When I divided the walking parts into two groups based on whether they were from a dominant test or a non-dominant test, approximately a third of the participants in both of the groups increased their walking times for both the dominant and the non-dominant tests (Figures 10.15 and 10.17). On average, the mean time for the walking parts decreased in the retest for both the groups and for both dominant and non-dominant tests (Table 10.14). However, the decrease is again very small, where the participant, on average, decreases their mean time for the walking parts by less than 0.5%.
From the paired-sample t-test I did on the null hypothesis $H_{0,6}$ (Table 10.15), the result was that all the tests accepted the null hypothesis. Accepting the null hypothesis means that the difference between week 1 and the retest is from a normal distribution with mean$= 0$.

I did some two-sample t-tests to see how the test times and kicking times for Group 3 compared to the control group (Tables 10.16 and 10.17). I tested the null hypothesis $H_{0,7}$, $H_{0,8}$, $H_{0,9}$ and $H_{0,10}$. All of the tests I did, except one, ended up rejecting the null hypothesis. The groups are from normal distributions with different mean and variance with the rejection of the null hypothesis. The one test that accepted the null hypothesis was when I tested the mean kicking time between the control and retest times from Group 3. If I had changed the significant level to be $\alpha = 0.01$, the test on the mean kicking time between the controls and the Group 3 both from week 1 would also be accepted. By changing the significant level, I would have made it stricter to reject the null hypothesis.

I have no information about the patients' fitness levels except for their EDSS, and I do not know what they did during their rehabilitation stay. Because of this, I can not say if the improvement in the test times comes from specific training or if the patients have just gotten more used to moving the body. The tests were done at different times of the day. The time of the day the tests were done could affect the test times because of fatigue. One of the symptoms of MS is fatigue, and after a long day, the patients could be tired, which could affect the results of the tests.

To sum up, there is a difference in the test times between the week 1 tests and the retests. For the most part, this improvement can be seen in the participants from Group 3. The difference in the test times most likely comes from a change in the kicking times for the participants. Since I have no information on what the participants did during the rehabilitation stay, can I not say anything about why the participants improved their times.

# Chapter 11

# Discussion and Conclution

First, I tried to use the raw data from the sensors to find the different segments within each of the Six Spot Step Tests (SSSTs). I was not able to use the raw data to find these segments because I could not identify the kicks within the tests. Identifying the kicks turned out to be very difficult because different participants kicked in different ways. This became clear when I looked at the video of the tests. One resolution to this could have been to place sensors on the cubes to record the cubes' movements. Suppose we had used IMUs that, in addition to having gyroscopes, accelerometers, and barometers within, also had a magnetometer, and we had added a magnet inside or on the cubes. In that case, this could also help to find the kicks from the raw data much easier.

From the analysis of the total test times of the SSSTs, I found that the three groups I divided the participants into made three normal distributions with different mean and variance. However, the regions described by the groups mean $\pm$ standard deviation overlap. This makes it hard to use the mean test times to classify the participants into one of the three groups. When I looked at the difference between the dominant and the non-dominant tests, I could not find any significant difference. The mean test times from the dominant and non-dominant tests for the three groups are normally distributed with different means and variances. They also have overlapping regions given by the mean $\pm$ standard deviation.

When I looked into the time used to kick in the SSSTs, I found that the mean kicking times for the three groups formed three normal distributions with different mean and variance. The mean $\pm$ standard deviation for the three groups gave overlapping regions between them. I then looked into any differences between the time used to kick with the dominant foot and the non-dominant foot. Here as well, was there no noticeable difference in the kicking times.

Then I looked into the time used for the walking parts in the tests to see if there were any differences between the three groups. Here as well, I found that the walking times from the three groups were normal distributed with different mean and variance. There mean $\pm$ standard deviation created overlapped regions between them, but not as big of a region as when I looked into the kicking time and the time for the whole test.

After looking into the times from the tests, I looked into if there were any differences between

the first and second time the participants walked through the SSSTs kicking with the dominant foot and with the non-dominant foot. I found a change between the first and second walkthroughs for both the dominant and the non-dominant tests from the mean test times. This change was more notable for the participants in Groups 2 and 3 than it was for the participants in Group 1. The difference was larger for the dominant tests than the non-dominant tests. For the kicking times, there was not much of a change between the first and second walkthroughs. For the walking parts, I found the same trend as I found for the test times. The participants in Groups 2 and 3 were the ones that experienced the most change between the first and second walkthrough. There was a more noticeable difference when it came to the dominant test than for non-dominant tests. When I compared the changes between the tests done in week 1 and the retests, I found that for the test times, the change between the first and second walkthrough from the dominant and non-dominant tests in week 1 had reduced in the retests. Only the dominant tests reduced the difference between the first and second walkthrough from week 1 to the retests for the times from the walking parts.
I ended the analysis by looking into the changes the participants experienced between the week 1 tests and the retests after a rehabilitation stay. Group 2 and 3 participants experienced a decrease between week 1 and retest for the mean test times, but the changes were more prominent for Group 3. There was a greater improvement of the mean test times in the non-dominant tests for both groups than in the dominant tests. There was an improvement between week 1 tests and the retests when it came to the mean kicking times. The change in the mean kicking times was more prominent for Group 3 compared to Group 2. When I divided the kicking into kicking with the dominant foot and kicking with the non-dominant foot, the changes were approximately the same for Group 3. Group 2 had the dominant kicking improved, but the mean non-dominant kicking times increased in the retests. There were some changes between week 1 and the retests in the times for the walking parts, but these changes were minimal. In the end, I investigate if the improvement in Group 3 had made their times from the tests closer to the controls times from week 1. For the mean kicking times, I found that the participants in the control group from week 1 and Group 3 from retests ended up being normally distributed with the same mean and variance.
During the analysis in this thesis, I tested multiple null hypotheses with the significant level $\alpha = 0.05$. For some of the null hypotheses, the p-value results of the tests ended up being below 0.05 but above 0.01. This means that this null hypothesis was rejected with the statistic level I sat but would have been accepted if I had used $\alpha = 0.01$. Even though the decisions to reject or accept these null hypotheses would change with the change of the statistic level, would the overall findings not change. The test times for the three groups' participants are normally distributed with different mean and variance. The mean $\pm$ standard deviation has too much of an overlapping region between these three groups. This makes it hard to use the times the participants use for the tests to classify which group they belong to. The times for the walking parts had a smaller overlapping region between the three groups than the test times and the kicking times. It could have been interesting to take a closer look into the walking part and

if there is some significance with the time leading up to the kicks. The Six Spot Step Test is complex, and my results have shown that for both Group 2 and 3, there is a decrease in the times they use for the test the second time they walk through the tests. This decrease in the time spent is much smaller in the retests. There had been interesting to see if the control groups would experience the same change in the difference of the first and second walkthrough between week 1 and some weeks later. One could use that the patients with MS had more improvement between the first and second walkthrough than the controls to distinguish the MS patients from the controls, or one could let all the participants walk through the test one time before starting to take the time. From the retesting times, I found that there is an improvement in the test and kicking times. Since I have no idea of what the different participants did during their rehabilitation stay and if they were used to moving their bodies or not, I can not say if this improvement is from rehabilitation and training or from getting used to moving their bodies.

In this thesis, I have investigated if there is possible to use the time a person with MS uses on the Six Spot Step Test (SSST) to say something about the EDSS level the person has. From my results, this is not the case because the different groups mean ± standard deviation of the test times have too big of an overlapping region. I have also investigated if it is possible to observe an effect of a rehabilitation stay from the SSST and if there is a learning effect from walking through the SSST one or more times. My results showed that there is possible to see an effect from the rehabilitation stay, for the most part for the patients with EDSS $> 4$. They also showed that there is some learning effect by walking through the SSST, but this decreases by a lot after the first walkthrough.
Using the time used on the SSST alone is not a good enough measure of the patients EDSS levels, but combining the times used on the test with a physiotherapists' observation could give more of a complete outcome. If one had used the raw data from the sensors, one could have found more details in the walking pattern for the participants, which could have helped find differences between people with MS and healthy controls. To say if this is the case, one has to do further investigations into this.

# Bibliography

*Activity Presenter*, 2020-08-06 [online] [visited on 2021-04-06]. Available from: https://www.microsoft.com/en-us/p/activity-presenter/9n01v94ljlx7?cid=msft_web_chart%5C&activetab=pivot:overviewtab.

AutoActive, [n.d.]. *AutoActive: tools and methods for autonomous analysis of human activities from wearable device sensor data* [online]. SINTEF [visited on 2021-05-21]. Available from: https://www.sintef.no/projectweb/autoactive/.

Barth, Jens; Oberndorfer, Cäcilia; Pasluosta, Cristian; Schülein, Samuel; Gassner, Heiko; Reinfelder, Samuel; Kugler, Patrick; Schuldhaus, Dominik; Winkler, Jürgen; Klucken, Jochen; Eskofier, Björn M, 2015. Stride segmentation during free walk movements using multi-dimensional subsequence dynamic time warping on inertial sensor data. *Sensors (Basel, Switzerland)*. Vol. 15, no. 3, pp. 6419–6440. ISSN 1424-8220.

Bethouc, Francois; Bennett, Susan, 2011. Evaluating Walking in Patients with Multiple Sclerosis. *International jurnal of MS Care*. Vol. 13, no. 1, pp. 4–14. Available from DOI: 10.7224/1537-2073-13.1.4.

Blumhardt, Lance; Lin, Xia, 2004. *Dictionary of Multiple Sclerosis*. New York, USA: CRC Press LLC. ISBN 9781853178665.

Bø, Lars, [n.d.(a)]. *Norsk MS-veileder: 2.1 Generelt om MS* [online]. Nasjonal kompetansetjenest for MS [visited on 2021-05-06]. Available from: https://msveileder.no/artikkel/1/generelt-om-ms.

Bø, Lars, [n.d.(b)]. *Norsk MS-veileder: 2.2 Årsaken til MS* [online]. Nasjonal kompetansetjenest for MS [visited on 2021-05-06]. Available from: https://msveileder.no/artikkel/6/arsaken-til-ms.

Bø, Lars, [n.d.(c)]. *Norsk MS-veileder: 4.1 Generelst om forebyggende behandling* [online]. Nasjonal kompetansetjenest for MS [visited on 2021-05-06]. Available from: https://msveileder.no/artikkel/32/generelt-om-forebyggende-behandling.

*chi2gof: Chi-square goodness-of-fit test*, [n.d.] [online]. The MathWorks, Inc [visited on 2021-05-06]. Available from: https://www.mathworks.com/help/stats/chi2gof.html.

Clausen, Camilla, [n.d.]. *Norsk MS-veileder: 3.1 Gange- og balanseproblemer* [online]. Nasjonal kompetansetjenest for MS [visited on 2021-05-06]. Available from: `https://msveileder.no/artikkel/78/gange-og-balanseproblemer`.

*Coefficient of Determination (R-Squared)*, [n.d.] [online]. The MathWorks, Inc [visited on 2021-05-06]. Available from: `https://www.mathworks.com/help/stats/coefficient-of-determination-r-squared.html`.

Dalgas, Ulrik; Severinsen, Kaare; Overgaard, Kristian, 2012. Relations Between 6 Minute Walking Distance and 10 Meter Walking Speed in Patients With Multiple Sclerosis and Stroke. *Archives of Physical Medicine and Rehabilitation*. Vol. 93, no. 7, pp. 1167–1172. ISSN 0003-9993. Available from DOI: `https://doi.org/10.1016/j.apmr.2012.02.026`.

Devore, Jay L; Berk, Kenneth N, 2011. *Modern Mathematical Statistics with Applications*. New York, NY: Springer New York. Springer texts in statistics. ISBN 9781461403906.

*Fampridin (Fampyra)*, [n.d.] [online]. Nye Metoder [visited on 2021-05-07]. Available from: `https://nyemetoder.no/metoder/fampridin-fampyra`.

*Fampridine (Fampyra)*, [n.d.] [online]. Multiple Sclerosis Trust [visited on 2021-05-07]. Available from: `https://mstrust.org.uk/a-z/fampridine-fampyra`.

*fitlm: Fit linear regression model*, [n.d.] [online]. The MathWorks, Inc [visited on 2021-05-06]. Available from: `https://www.mathworks.com/help/stats/fitlm.html`.

Franchignoni, Franco; Horak, Fay; Godi, Marco; Nardone, Antonio; Giordano, Andrea, 2010. Using psychometric techniques to improve the Balance Evaluation System's Test: the mini-BESTest. *J Rehabil Med*. Vol. 42, no. 4, pp. 323–331. Available from DOI: `10.2340/16501977-0537`.

Goldman, Myla D; Marrie, Ruth Ann; Cohen, Jeffrey A, 2008. Evaluation of the six-minute walk in multiple sclerosis subjects and healthy controls. *Multiple Sclerosis Journal*. Vol. 14, no. 3, pp. 383–390. Available from DOI: `10.1177/1352458507082607`. PMID: 17942508.

Hamrs, Charlotta; Botolfsen, Pernille; Tangen, Gro Gujord; Helbostad, Jorunn L., 2017. Interrater and test-retest reliability and validity of the Norwegion version of the BESTest and mini-BESTest in people with increased risk of falling. *BMC Geriatrics*. Vol. 17, no. 1, p. 92. Available from DOI: `10.1186/s12877-017-0480-x`.

Herman, Talia; Giladi, Nir; Hausdorff, Jeffery M., 2011. Properties of the 'Timed Up and Go' Test: More than Meets the Eye. *Gerontology*. Vol. 57, pp. 203–210. Available from DOI: `10.1159/000314963`.

*HERO7 BLACK*, [n.d.] [online]. GoPro [visited on 2021-05-08]. Available from: `https://gopro.com/en/us/shop/cameras/hero7-black/CHDHX-701-master.html`.

Hobart, J.C; Riazi, A; Lamping, D.L; Fitzpatrick, R; Thompson, A.J, 2003. Measuring the impact of MS on walking ability: The 12-item MS Walking Scale (MSWS-12). *Neurology*. Vol. 60, no. 1, pp. 31–36. ISSN 0028-3878.

Horak, Fay B; Wrisley, Diane M; Frank, James, 2009. The Balance Evaluation Systems Test (BESTest) to Differentiate Balance Deficits. *Physical Therapy*. Vol. 89, no. 5, pp. 484–498. ISSN 0031-9023. Available from DOI: 10.2522/ptj.20080071.

Kanekar, Neeta; Aruin, Alexander S., 2013. The Role of Clinical and Instrumented Outcome Measures in Balance Control of Individuals with Multiple Sclerosis. *Multiple sclerosis international*. Vol. 2013, no. 190162. ISSN 2013. Available from DOI: 10.1155/2013/190162.

Kieseier, Bernd C.; Pozzilli, Carlo, 2012. Assessing walking disability in multiple sclerosis. *Multiple Sclerosis Journal*. Vol. 18, no. 7, pp. 914–924. Available from DOI: 10.1177/1352458512444498. PMID: 22740603.

Kurtzke, John F., 1983. Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology*. Vol. 33, no. 11, pp. 1444–1452. ISSN 0028-3878. Available from DOI: 10.1212/WNL.33.11.1444.

Lunde, Hanne Marie Bøe; Torkildsen, Øyvind Fredvik Grytten; Ljøstad, Unn; Mygland, Åse, [n.d.]. *Fampridin - Fampyra* [online]. NEL Nevrologi: Norsk Helseinformatikk AS [visited on 2021-05-07]. Available from: https://nevrologi.legehandboka.no/handboken/sykdommer/demyeliniserende-sykdommer/behandlinger/fampridin---fampyra/#fagmedarbeider

Mariani, Benoit; Rochat, Stephane; Büla, Christophe J.; Aminian, Kamiar, 2012. Heel and Toe Clearance Estimation for Gait Analysis Using Wireless Inertial Sensors. *IEEE Transactions on Biomedical Engineering*. Vol. 59, no. 11, pp. 3162–3168. Available from DOI: 10.1109/TBME.2012.2216263.

Meyer-Moock, Sandra; Feng, You-Shan; Maeurer, Mathias; Dippel, Franz-Werner; Kohlmann, Thomas, 2014. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. *BMC neurology*. Vol. 14, no. 1, pp. 58–58. ISSN 1471-2377.

Nieuwenhuis, M M; Tongeren, H Van; Sørensen, P S; Ravnborg, M, 2006. The Six Spot Step Test: a new measurement for walking ability in multiple sclerosis. *Multiple Sclerosis Journal*. Vol. 12, no. 4, pp. 495–500. Available from DOI: 10.1191/1352458506ms1293oa. PMID: 16900764.

*Physilog: Inertial Measurement Sensor (IMU)*, [n.d.] [online]. GaitUp [visited on 2021-05-08]. Available from: https://research.gaitup.com/physilog/.

Pitt, Will; Chen, Szu-Hua; Chou, Li-Shan, 2020. Using IMU-based kinematic markers to monitor dual-task gait balance control recovery in acutely concussed individuals. *Clinical Biomechanics*. Vol. 80, p. 105145. ISSN 0268-0033. Available from DOI: https://doi.org/10.1016/j.clinbiomech.2020.105145.

Rudick, R. A.; Cutter, G.; Reingold, S., 2002. The Multiple Sclerosis Functional Composite: a new clinical outcome measure for multiple sclerosis trials. *Multiple Sclerosis Journal*. Vol. 8, no. 5, pp. 359–365. Available from DOI: 10.1191/1352458502ms845oa. PMID: 12356200.

Sebastião, Emerson; Sandroff, Brian M.; Learmonth, Yvonne C.; Motl, Robert W., 2016. Validity of the Timed Up and Go Test as a Measure of Functional Mobility in Persons With Multiple Sclerosis. *Archives of Physical Medicine and Rehabilitation*. Vol. 97, no. 7, pp. 1072–1077. ISSN 0003-9993. Available from DOI: `https://doi.org/10.1016/j.apmr.2015.12.031`.

Spain, R.I.; St. George, R.J.; Salarian, A.; Mancini, M.; Wagner, J.M.; Horak, F.B.; Bourdette, D., 2012. Body-worn motion sensors detect balance and gait deficits in people with multiple sclerosis who have normal walking speed. *Gait & Posture*. Vol. 35, no. 4, pp. 573–578. ISSN 0966-6362. Available from DOI: `https://doi.org/10.1016/j.gaitpost.2011.11.026`.

*Statistikk om MS*, [n.d.] [online]. Multippel Sklerose Forbundet [visited on 2020-05-30]. Available from: `https://www.ms.no/fakta-om-ms/hva-er-ms/statistikk-om-ms`.

*ttest: One-sample and paired-sample t-test*, [n.d.] [online]. The MathWorks, Inc [visited on 2021-05-09]. Available from: `https://www.mathworks.com/help/stats/ttest.html`.

*ttest2: Two-sample t-test*, [n.d.] [online]. The MathWorks, Inc [visited on 2021-05-09]. Available from: `https://www.mathworks.com/help/stats/ttest2.html`.

*TUG - The Timed "Up & Go"*, [n.d.] [online]. Norsk Helseinformatikk (NHI) [visited on 2021-05-03]. Available from: `https://nhi.no/skjema-og-kalkulatorer/skjema/geriatripleie/timed-up-and-go-tug/`.