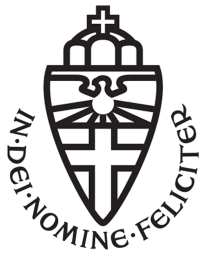
BACHELOR THESIS
COMPUTER SCIENCE



RADBOUD UNIVERSITY

Relevant biomarkers in the prediction of good and bad days for multiple sclerosis patients

Author
Tom Sandmann
s4330048

First supervisor & assessor
prof. dr. Tom Heskes
tomh@cs.ru.nl

Second assessor
dr. Saiden Abbas
saiden@orikami.nl

June 25, 2016

Abstract

Multiple sclerosis (MS) is a disease where the day quality of a patient can vary a lot. Therefore, it is hard to predict whether upcoming days will be 'good' or 'bad' days. By identifying relevant biomarkers for patients using data from wearables, we might be able to predict this day quality. These predictions can be used to get more personal advice on and insight in MS.

After we identified relevant biomarkers using literature research, we chose to research the following three biomarkers: Two-Minute Walk Test (2MWT), resting heart rate (RHR) and sleep duration. Using statistical analysis on each of these biomarkers, we investigated the correlation between the data belonging to this biomarker and the rating of the day after indicated by each of the participants. Although the period in which we gathered the data was reasonable, the amount of data was lower and of lesser quality than expected. This was especially the case for data that only could be gathered when participants entered specific information in the application that we used during our experiment, In the end, we did not find any evidence for a relation between each individual biomarker and the day rating for the next day. We did find that the process of gathering data from research subjects should preferably be fully automated (if possible). This is because participants tend to forget or ignore instructed tasks, although these tasks are essential for the data that will be used in research. Participant adherence was one of the major factors impacting the quality of our data and should be taken into account in future research.

Acknowledgements

I would like to thank my supervisor Tom Heskes for his time and support during my research. I would also like to thank Bram den Teuling for granting me the opportunity to write my thesis at Orikami and to work on this interesting topic. Finally I would also like to thank my parents for their never ending support.

Contents

Chapter 1

Introduction

1.1 Motivation and Research Questions

MULTIPLE SCLEROSIS (MS) is a long lasting disease that affects approximately 1 in 1000 people in the Netherlands. MS is a disease of the central nervous system, resulting in disrupted communication between the brain and other parts of the body. In MS, the insulating covers of nerve cells, called myelin, are damaged. The symptoms range from reduced vision to muscle weakness. Depression is also a common symptom of MS.

There are four different courses for MS [42]. The most common course is called relapsing-remitting MS (RRMS). RRMS is characterized by attacks, also called relapses, followed by periods of partial or complete recovery (remissions). These periods of attacks and recovery can, especially for MS patients with RRMS, result in good and bad days. Because there is no cure for MS, treatment tries to prevent and improve the recovering from these attacks.

Orikami is currently developing an app called DiaPro MS. Participants in the pilot experiment of this app will receive activity trackers. MS patients will wear these activity trackers for a period of two months. During this period, the activity tracker will capture biomarkers like heart rate, blood pressure, steps walked and the amount of sleep during the day. Together with the data gathered from filled in questionnaires, these biomarkers could be used to predict good and bad days for MS patients.

This thesis will focus on how to capture reliable biomarkers using wearables and try to overcome any difficulties faced during the pilot. It will provide more insight in the process of using wearables in research and how the captured data can be used for analysis. This is useful when the experiment will be done on a much larger scale.

The main question of this research is: **which biomarkers, captured by activity trackers, are possibly useful for the prediction of good and bad days for MS patients?** To answer this question, we need to answer the following subquestions:

- ⋆ What is the reliability and validity of biomarkers captured by activity trackers?
- ⋆ How are biomarkers measured?
- ⋆ How can biomarkers be used as an indicator of health?

1.2 Related Work

In this section, we will look at literature that is related to making predictions of the quality of upcoming days for people having MS. As we have seen in our previous section, MS is a complex disease with several processes playing role in the course of it. Biomarkers are used to get a measurement of the current state or condition of a person. This is done by measuring indicators. Traditionally, biomarkers are substances measured in body fluids, saliva and blood. Current research is continuously finding new relevant biomarkers for MS. Several papers exist that give an up to date overview of current biomarkers [10, 34]. These biomarkers can be very useful in distinguishing several subgroups of MS. New types of biomarkers like stress and neurodegeneration (the loss of neurons) are being explored. By exploring these biomarkers, researchers speed up the process of using these biomarkers in clinical practice. With the introduction of the activity trackers and smart watches, more biomarkers can be measured easily. Although research related to the reliability and validity of the measurements from these devices is still being conducted, the devices are currently being used in the research field.

By using measurements of heart rate, exercise capability (2MWT) and more we could give a patient more insight in his upcoming days. Prediction of quality of life in multiple sclerosis has been researched before. Health-related quality of life (HQOL) is used to determine an individual's well-being. This can be affected over time, depending on the disease or other conditions. For MS patients, HQOL is poor. In [8], several predictors (depression and self-reported fatigue) were considered simultaneously. In a group of 120 MS patients and in a control group of 44 people, HQOL was measured. It was found that MS patients reported lower HQOL compared to the control group. Depression and fatigue were the primary contributors to these results. This confirms the results of previous studies, where a strong relation was found between depression and HQOL in MS. A limitation of self reports is that the results are not objective: they depend on the mood of the patient. As we can see in previous studies, the prediction of upcoming days is something that remains unaddressed.

In [21], a web-based calculator was made for MS patients. This calculator was able to give estimates of the progression of the disease. By letting the user provide individual patient characteristics like disease course, number of attacks in the last two years and the age on which the first MS symptoms appeared, the calculator tried to find the best matching patients in a database. Using these matched patients, MS related prognoses were calculated for this specific individual having MS. One of these prognoses was the time it would take for the patient to transition from RRMS

to secondary progression MS (SPMS). These predictions where then compared to the predictions of 17 MS specialist neurologists. They were asked how long it would take for the presented MS patients to reach a value of 10 on the expanded disability status scale (EDSS) after their first MS symptoms. This means death due to MS. The predictive accuracy was measured using the Brier Score, which is a score function that measures the accuracy of probabilistic predictions. A score of 0 indicates perfect accuracy, while 0.5 indicates the same accuracy as chance. The Evidence-Based Decision Support Tool in Multiple Sclerosis (EBDiMS) was 100% consistent. Among the neurologists, there was a considerable inter-rate variability. Both the specialists and EBDiMS were in the Bier Score range of $0.1 - 0.2$, which indicates that the predictions are better than chance. For particular subgroups, EBDiMS did not do a better job than the specialists. The tool used data from a previous conducted longitudinal study. Although different biomarkers are commended and the period which will be predicted is longer, this approach shows similarities to what we want to achieve. Matching of patients against other patients might be a good idea for disease prediction when looking at longer periods of time, but might not be suitable when predicting over a shorter period (day or week). This is because over a shorter period of time, more variables will be influencing these results and patterns for each of the variables differ a lot. For this reason we decided to use biomarkers that can easily be measured using activity trackers. By eventually combining these biomarkers, we hope to get more insight in the day quality of MS patients. Before combining biomarkers, we first have to look at each of them individually. This gives us a feeling for the gathered data and hopefully new insights in using this data for predicting good and bad days for MS patients.

1.3 Outline Thesis

In this section, we will describe the outline of the rest of the thesis. In Chapter 2, we investigate the long term effects from wearing activity trackers extensively. This includes both the health related and the privacy related effects for the participants. In Chapter 3, we conduct a literature research on possibly relevant biomarkers for the prediction of good and bad days for MS patients. This is done by looking at how each biomarker is related to the health of an individual. We also look at how these biomarkers are currently being measured by activity trackers (if possible). This also includes the validity and reliability of the measurements done by these current activity trackers. In this same chapter, we also choose three relevant biomarkers that will be used for analysis in Chapter 5.

Chapter 4 describes how the experiment with the participants was set up, which devices were used in the process and how the data from these device were stored. In Chapter 5 we describe the analysis we performed on each of the selected biomarkers. Finally, Chapter 6 describes our findings in this thesis. We also reflect on everything we have done and what future research should take into account.

Chapter 2

Long Term Effects of Wearables

2.1 Privacy

ALTHOUGH we still have to find out if wearables will help people with MS to get more insight in the quality of upcoming days, there must be looked at the effects from wearing these devices extensively. One of these aspects is privacy. Because all of the information is stored on the device and in a database, data loss can have an enormous impact on the user. Companies providing these devices have a dominant position. What if the gathered data is sold for money? This can really effect patients in taking a health insurance. Not every patient wants their health insurance company to know in which stage of the disease they are currently residing. That data loss is a potential threat is shown in a research into the security of smartwatches by HP [30]. In total, ten popular smartwatches were investigated. During this research, it was found that some smartwatches were sending information to third parties. In 7 of the 10 smartwatches, information was send unencrypted to the server.

Another aspect is the influence of wearables on your choices during the day. Because almost everything is monitored, you cannot deny the patterns that can be made visible using the gathered data. For example, if you see that your calorie intake is sufficient for this day, you may decide to skip that snack you normally would have eaten. Of course, this is good for your health, but it puts some restrictions on your freedom, without you knowing. Next thing you know, you will receive a warning about your calorie intake. Similar things are already happening on Android devices. When users of Android devices put their volume too high, they will receive a warning saying that “listening at high volume for a long period may damage your hearing”. Although looking at all these small steps in technology individually makes them look harmless, combined they provide a stunning amount of information. It seems that people are not aware that this ever increasing amount of data is also continuously increasing the insight in customers and all of their patterns. Because people may not be aware of this, this increasing insight in their lives could be

unwanted and thus remains something that we have to watch out for.

2.2 Health

The intention of activity trackers is that they are worn 24/7. While wearing it, the device is constantly syncing data to your mobile phone. This makes use of Bluetooth, a wireless technology using the microwave frequency spectrum. Because people tend to sleep with their hands near their heads, this can cause a significant exposure of the brain during nighttime. When looking for research on the long term health effects of Bluetooth, it is remarkable that there are hardly any papers available. This might be due to the interests of the phone industry in the use of this technology. Looking into the specification of Bluetooth devices, we see that the maximum output for devices classified in the highest Bluetooth power class is still lower than the lowest powered mobile phones [26]. This implicitly suggests that the exposure to radiation emitted from devices using Bluetooth is save. However, some people question this implicit safety. Because more and more devices are making use of Bluetooth, people will be exposed to higher levels of radiation. This makes people question how safe the exposure to this increasing amount of radiation is: should we not reduce the amount of radiation when the levels of radiation we are exposed to are increasing? Although most smart watches and activity trackers are making use of Bluetooth and WiFi to function, there are also some brands that have a cellular chip in their devices. Wearing activity trackers with a cellular chip is like wearing a cellphone on your wrist. The W.H.O has a fact sheet about the relation between public health and mobile phones. On [74] we can read that there have been several longitudinal studies were the long-term risks from radiofrequency exposure were investigated. Although some studies are still ongoing, the results were not consistent. Because we cannot rule out the possibility of radiofrequency electromagnetic fields as being carcinogenic, they are are classified as possibly carcinogenic. This indicates that a causal association between radiofrequency electromagnetic fields and cancer is possible, but that confounding, bias or chance cannot be ruled with any certainty.

Chapter 3

Literature Research on Possibly Relevant Biomarkers

We first start with a literature research into the validity and reliability of activity trackers in Section 3.1. Afterwards, we will take a look at relevant biomarkers that could be useful in the prediction of good and bad days for MS patients (Section 3.2). This is also done by conducting a literature research. Based on these findings, we finally choose three biomarkers in Section 3.3 and further investigate how they could be used for analysis.

3.1 Reliable Data Collection

There is a whole range of different activity trackers on the market that can be used to capture health related data like distance walked or run, burned calories and sometimes even heart rate and quality of sleep. These trackers have been developed to increase the insight of an individual into their physical activities throughout the day. In order to use this data for research purposes, we must determine the validity and reliability of the data from these activity trackers and other related devices.

In [37], ten activity trackers available for consumers with the ability to measure step count were tested on healthy people. Testing was done both under laboratory conditions (using a treadmill) and under free-living conditions (on one working day between 9.00 and 16:30) using different groups. Besides the Omron, Moves app and the Nike+ Fuelband, most trackers showed a high reliability. This was analyzed using test-retest analysis[1] with Intraclass Correlation Coefficient (ICC)[2].

[1] Test-retest analysis is used to assess the consistency of a measurement from one time to another [69]. The consistency can be estimated by administering the same test to the same sample on two different moments.

[2] The Intraclass Correlation Coefficient (ICC) is used to quantify the degree of similarity between two or more values repeatedly measured on a continuous scale [36].

The validity of the trackers was measured by comparing each of the trackers using the most accurate test possible (gold standard) under laboratory and under free-living conditions.

Various studies have been published on the validity and reliability of different activity monitors [7, 13, 14, 19]. These devices are becoming widely accepted in the research field. In [18], the evidence for validity and reliability of two different activity trackers (Fitbit and Jawbone) was systematically reviewed. This was done for the measurement of sleep, energy expenditure, physical activity, steps and distance. The study showed that there was a higher validity of steps and physical activity. There was a lower validity of sleep and energy expenditure. Except for the measurement of physical activity, a high inter-device reliability was found. However, in 7 of the 22 studies, missing or lost data was reported. Some of the lost data was due to the validation criterion measure and not due to the trackers. Other lost data were attributable to research errors. It was also mentioned that one should expect data loss when using activity trackers, and that software updates of these trackers can influence measurements from these devices. These influences and the resulting risks should be taken into account during our data gathering and analysis.

In [40], consumer-based physical activity monitors were studied to examine the validation of the measurement of energy expenditure. This was done under semi-structured free-living conditions. Eight different activity monitors were worn during a 69-minute protocol, based on both free living and structured activities. Most of the consumer based monitors gave results similar to the Actigraph monitor, which is a research-grade monitor most commonly used in the field. This outcome is promising for the use of wearables in our experiment.

3.2 Possibly Relevant Biomarkers

In this section, we will discuss several biomarkers that will (or can) be used in the prediction of day quality for MS patients.

3.2.1 Resting Heart Rate

Resting heart rate is defined as the heart rate of a person that has not been subjected to any form of stimulation or exertion. For adults, the resting heart rate lies between 60 and 100 beats per minute. However, this also depends on the fitness of the person. Despite this variety, several studies have shown that a high heart rate is related with several risks [16]. Therefore it seems desirable to have a low resting heart rate. A lot of studies have addressed the importance of heart rate in healthy humans [9, 20, 54]. The resting heart rate can also be an interesting indicator for the prediction of day quality of MS patients. However, there is hardly any literature about the classification of heart rate in rest from heart rate data. Can we come up with an algorithm to extract this information?

Currently, the technology of photoplethysmography (PPG) is being used to

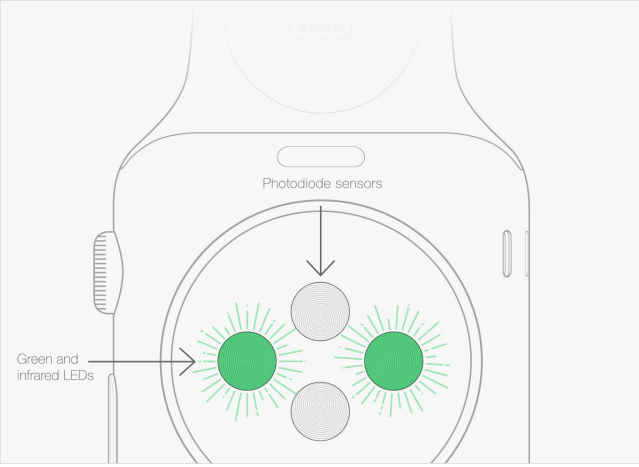


Figure 3.1: Illustration of the sensor on an Apple Watch

Source: `https://support.apple.com/en-us/HT204666`

develop small devices that can measure the heart rate [64]. This technology uses changes in blood volume in the tissue to determine the heart rate. The sensor contains a light source and a light detector. The light is emitted through the skin using the light source and the reflected light is measured using the light detector. Because blood is red, it reflects red and absorbs green light. By using a green light source, the change in the light intensity is therefore related to the blood perfusion of the tissue and can provide information on the pulse rate. When your heart beats, the blood flow in your wrist is greater, and so is the green light absorption. Between beats, the light absorption is less. By flashing the light many times per second, the number of heart beats per minute can be calculated. Fitbit's PurePulse technology is making use of the PPG technique, and so are the Basis Peak HR and the Microsoft Band II.

Many factors effect the performance of the heart rate sensor. If the amount of blood flowing through your skin is too low, the heart rate sensor will not get an accurate reading. Motion also plays a role: irregular movements decrease the accuracy of the sensor. Tattoos can also block the light emitted by the sensor.

3.2.2 Heart Rate Variability

Heart rate variability (HRV) is the variation in the time interval between heartbeats. It can be measured by measuring the time interval between consecutive heartbeats. To get a better understanding of this, we take a look at the ECG of a normal sinus rhythm which can be seen in Figure 3.2. In this figure, we also see the so-called QRS complex. In an ECG, the electrical activity of the heart is measured. The QRS complex is related to the depolarization of the left and right ventricles. By some complex processes (that are out of the scope of this thesis), this depolarization

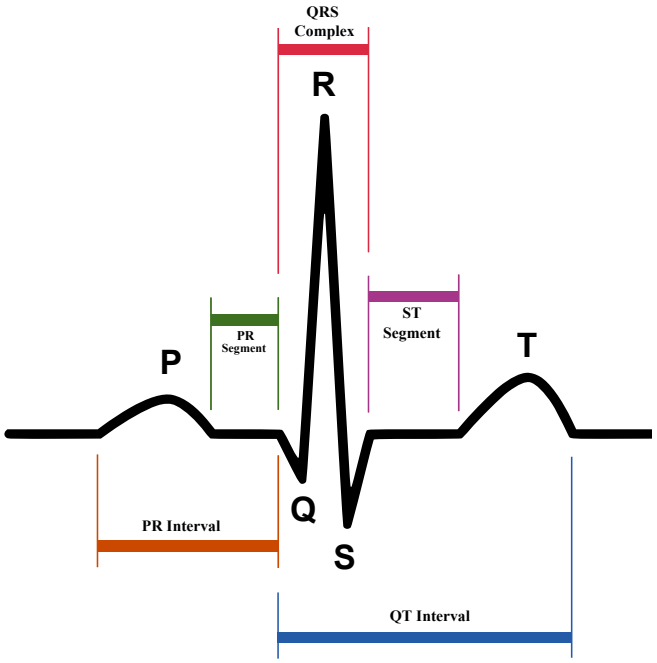


Figure 3.2: Schematic representation of a normal ECG.

Source: Wikipedia

eventually causes both ventricles to contract, resulting in a heartbeat. By measuring the interval between two consecutive QRS complexes in an ECG, one can determine the HRV. HRV has proven to be a reliable reflection of the autonomic nervous system, blood pressure and can also be related to gender, sex, age, sleep and diabetes [1]. The measurements of HRV can generally be grouped under time-domain and frequency domain methods [48]. In time domain methods, either the heart rate at any point in time or so called normal-to-normal (NN) intervals between consecutive QRS complexes are determined. NN is used to make clear that the beats are 'normal' beats. HRV is also referred to as RR variability, where R is the point on the QRS complex. Some examples of variables that can be calculated when time domain methods are used are the mean heart rate, shortest and longest NN interval and standard deviation.

In frequency domain methods, the heart rate variance is partitioned into underlying rhythms that occur at different frequencies. Fast Fourier Transform (FFT) is commonly used to find these underlying rhythms. When this is done, frequencies can be associated with different periodic rhythms. There are three different rhythms, each with their own frequency range:

- Very low frequency (VLF) (0.0033 - 0.04 Hz);
- Low frequency (LF) (0.04 - 0.15 Hz);

- High frequency (HF) (0.15 - 0.4 Hz).

Afterwards, the NN intervals in each frequency range are counted.

In [65], HRV as a biomarker for stress was investigated with a meta-analysis of studies. It was already found that individuals with a greater ability to regulate emotion had greater levels of resting HRV [3].

HRV could give more insight in the amount of stress a person is experiencing. If HRV is a potential biomarker for stress, it must be related to the perception of threat and safety, which are elements of stressors. That this relation exists was found in [28], where the effect of stress from computer work on HRV and blood pressure was investigated. Twelve female students participated in the study. They had to perform a computer task where random numbers had to be entered using the keyboard. The experiment consisted of three stages, in which the participants had to perform the described computer task. In two of these stages, the experimenter was unfriendly towards his colleagues and unsupportive to the participant. In these stages, the participants also had to perform a memory test. In the other stage, the experimenter was friendly and supportive, thus eliminating (almost) all stressors. In this stage no memory test was performed. Before and after the work, the participants filled in a questionnaire containing a scale to measure the experienced amount of stress. In stressful situations indicated by the participants, it was found that there was a reduction in the high-frequency component of HRV (calculated from ECG recordings). An increase in the low- to high-frequency ratio was also found in stressful situations. The stressors also led to an increase in blood pressure. This research gives a promising indication for the application of HRV as a biomarker for stress. In [49], a direct relation was found between the HRV spectrum and fatigue. In this study, neural network analysis was applied on HRV to assess the fatigue of drivers. Although the dataset was limited, they managed to get an accuracy of 90% in predicting whether a given subject was in an alert or in a fatigued state.

3.2.3 Two-Minute Walk Test

During the period of the pilot, MS patients that are participating in our experiment will be asked to do a two-minute walk test (2MWT) every day. Walking tests are generally used to monitor the effectiveness of a treatment for patients. The two-minute walk test is a shorter measure of walking performance than the six-minute walk test (6MWT). The 2MWT addresses some limitations of the 6MWT. The latter can be too fatiguing for older people and too time consuming in general. Especially the limitation of the 6MWT as being too fatiguing, that is addressed in the 2MWT, is relevant for MS patients [23]. Because nerve damage can leave muscles stiff or weak, MS patients can experience a reduced ability to move. Most of the MS patients also experience fatigue. Some studies have been done where the measurement properties of the 2MWT were examined [15] and the reliability and performance was evaluated [11]. The walking test forces the user to be subjected to physical exertion. Therefore, the test can be representative for factors like motivation and

fatigue which in turn could be related to day prediction. In [44], it was found that a worsening of the symptoms of MS may lead to a reduction of physical activities. This link could be made visible using the data from the walk test.

As mentioned before, participants will also perform the 2MWT walking test in our experiment. This will be done using the 'Mijn Kwik' app. This app will record the GPS coordinates as the participants perform the walking test. The walk test will be initiated as a notification on the phone of the patient. Only if the GPS accuracy is high enough, the walking test can be started. This will be explained in more detail in Section 5.1.

3.2.4 Stress

Stress is the response to a stressor. A stressor can be an external stimulus or a specific condition in the environment that causes stress. For humans, stress can be positive or negative. Although some forms of stress can have a positive effect on the performance, stress is mostly referred to in a negative way. Several studies have shown that stress affects MS. In [12], researchers investigated the relation of relapsing-remitting MS patients, that experienced attacks, with self-reported stressful events not related to MS. It was found that the experience of at least one stressful event in a period of four weeks would double the risk of an attack within the next week. That stress would play an important role in MS has been assumed for a long time.

Some wearables are able to measure the perspiration. This is done by measuring changes in the conductivity of the skin, also called electrodermal activity (EDA), which is related to your sweat production. The production of sweat is controlled by the sympathetic nervous system. EDA is related to arousal. Arousal is the activation state of the autonomic nervous system. This is related to the degree of mental awareness or consciousness. If the sympathetic nervous system, a branch of the autonomic nervous system, is highly aroused, more sweat will be released resulting in better conductance of the skin. It will also lead to an increased heart rate and blood pressure. In addition to motivating certain behaviors like the flight or fight response, arousal also plays an import role in emotion and performance. Several tasks require different levels of arousals for the best performance. For example, motivation dependent tasks require higher levels of arousal and concentration dependent tasks require lower levels of arousal for better performance.

In [35], the potential of EDA to measure arousal in a real life setting during exercise was researched. This was done using the Affectiva Q™ sensor that was developed in [51]. This sensor is designed to continuously measure EDA during everyday activities. While the measurement of heart rate can also be used to measure levels of arousal, heart rate is influenced by energy demand. Therefore, heart rate is only useful for the measurement of arousal in a resting state, in which energy demand is not influencing the heart rate. [35] illustrated that the measurement of arousal can be done in a high intensive real life setting, which is very promising for daily use of this sensor for MS patients.

3.2.5 Sitting, Walking and Lying Pattern

Most of the activity trackers nowadays are able to determine which activity you are doing right now. These activities include walking, running and walking. We can go into more detail by also keeping track of sitting, walking and lying patterns.

Fatigue is a common symptom for MS patients. Researchers have identified the characteristics of so called MS fatigue, that differs from normal fatigue [45]. These characteristics of fatigue include:

- Occurring on a daily basis;
- Coming easily and suddenly;
- Being generally more severe than normal fatigue.

We could use data from the wearables to get more insight in the pattern of sitting, walking and lying. This gives insight in the degree of functional ability and the level of activity during the day. This could be related to the feeling of a person. Several studies have researched the classification of activities. In [61], data from a triaxial accelerometer sensor was used for activity classification. The algorithm used for classification was able to classify running, walking and resting activities. According to the authors, the algorithm could be extended to also classify activities like standing and sitting. That this is possible was shown in [33], where a real-time movement classifier was created that could classify activities including falling, sitting, lying (including positions) and standing. The data being used was collected using a waist-mounted triaxial accelerometer. Because the classification was done in real time, the following constraints had to be satisfied:

- No knowledge of future events;
- Amount of buffered data is limited;
- Limited processing time: the device must be able to handle the continuous flow of incoming data.

When using this data for assessing the quality of upcoming days, the real-time constraints described before are not compulsory. Data can be send in batches to the database, from which the data can be fetched and used for data mining. In the described experiment, these real-time constraints led to some limitations of the developed system. However the experimental results showed that the system was able to classify most activities with a high accuracy.

3.2.6 Weather

That weather can influence your mood is something that many people have experienced. That this is no coincidence is confirmed by several studies. [29] related multiple mood and weather variables. In this study, it was found that weather was a major predictor for changes in mood. For example, a high humidity had a dominant influence on multiple mood variables:

- Decrease of concentration and potency;
- Increase of sleepiness and fatigue.

Because some weather data is freely available on the internet, it can be used easily for assessing day quality. During a meeting with MS patients organized by Orikami, it was found that some patients were more at ease during the summer, while others experienced the same during cold periods. Heat sensitivity of MS patients is something that has been subjected to extensive research. In [38], fatigue in MS was investigated. This was done by conducting structured interviews with 32 people having MS and 33 healthy adults. There were several questions about the characteristics of their fatigue, one of which was “Heat worsens it". 92% of the MS patients answered this question with ‘yes’. Increased body temperature also seems to influences MS patients. In [5], two cases were described in which heat exposure resulted in death for two patients having MS. In these cases, exposure to sun (and therefore heat) led to incapacitation of the patients, resulting in death. Although heat can enlighten the symptoms of MS, it can thus be very dangerous by causing sudden collapses of the thermoregulatory system.

3.2.7 Biorhythm

The theory of biorhythm claims that our life is influenced by rhythmic cycles. The theory is an example of pseudoscience, which means that it is presented as scientific but does not adhere to the scientific method. According to the biorhythm theory, three different cycles oscillating in a sine wave (Figure 3.3) influence three different aspects of our life [27]. The three cycles are:

- A 23 day cycle that influences **physical** aspects of behavior. The equation for this cycle is $\sin(2\pi t/23)$.
- A 28 cycle that influences **emotions**. The equation for this cycle is $\sin(2\pi t/28)$.
- A 33 day cycle that influences **intellectual** functions. The equation for this cycle is $\sin(2\pi t/33)$.

In each of the equations, t indicates the number of days since birth. According to modern biorhythm theory, these cycles start at the moment of birth and proceed to exist during the rest our lives. The durations of these periods do not change. The required knowledge to apply this theory gives us a great advantage. By only having knowledge of an individual’s date of birth, one is able to calculate the position on each cycle of a particular person, giving insight in the current state for each of the cycles. However, the theory of biorhythm remains controversial. No scientific evidence has been found that supports the theory. Some researchers even claim that it has no more predictive power than chance. It might be worth using biorhythm as a biomarker for day quality, despite its restricted expressive power.

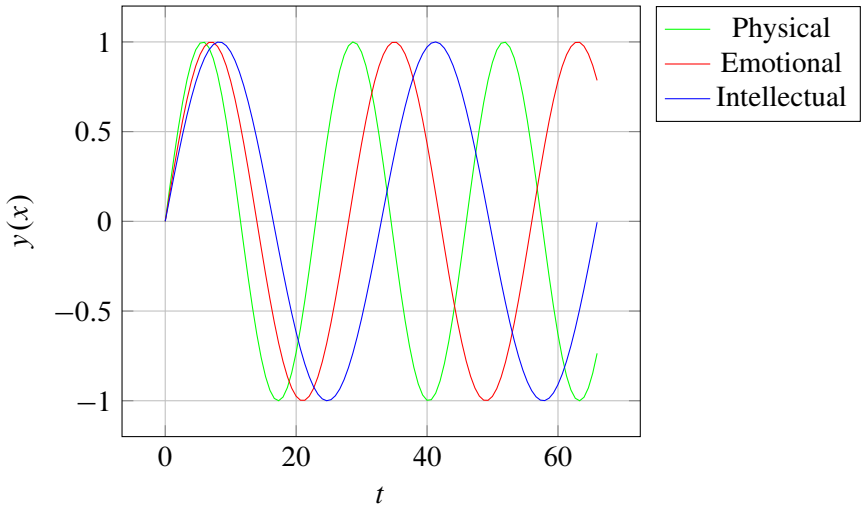


Figure 3.3: Biorhythm of the first 66 days after birth.

3.2.8 Fatigue

Questionnaires in general can help to retrieve useful information from the user. The 'Mijn Kwik' app, which is also used for the 2MWT experiment, contains questionnaires as well. The corresponding questions can be seen in Figure 4.1 in Section 4.2. An example of such a question is the slider that can be used by a patient to indicate how energetic he is feeling right now. Such a question is just one indicator. There is no simple measure to determine if someone is fatigued, there are just different ways of estimating the level of fatigue. Below we will discuss several scales that could be used in questionnaires.

Visual Analogue Scales Visual Analogue Scales (VAS) can be used as a psychometric response scale in questionnaires. It consists of a straight line, where each end of the line contains an extreme statement. These statements are two opposite claims, for example the statements 'no pain' and 'the greatest pain I can imagine'. In the case where the patient must indicate how he is feeling today (Figure 4.1b in Section 4.2), this scale is also used. A problem with this method is that it is hard to make a distinction between all the possibilities. The choice made can only be partly explained by the underling property, making it hard to spot measurement errors. Due to its simplicity, VAS can easily be used. However, attention must be paid to its role. It should never be used alone, due to its biases [67]:

- Context Bias: a VAS score for a state depends on the number of better and worse states that are present at the same time. If a state is included with more states that are better, its value is depressed. This is in contrast with the effect that occurs when a state is included with more states being worse: its value is enhanced.

- End-aversion Bias: this bias refers to the reluctance of some respondents to use the extreme values of the scale, or to use the ends when using a continuous scale.

Samn-Perelli Seven Point Fatigue Scale The Samn-Perelli Seven Point Fatigue Scale (SPS) is a 7-point scale with scores ranging from 1 (“fully alert, wide awake”) to 7 (“completely exhausted, unable to function effectively”) [56], which has been validated and is being used widely.

Karolinska Sleepiness Scale The Karolinska Sleepiness Scale (KSS) can be used to evaluate sleepiness. It has been validated against EEG activity and is also widely used. KSS is a one-dimensional scale with scores ranging from 1 (“very alert”) to 9 (“very sleepy, great effort to keep awake”).

As we can see, there are a lot of verified techniques to get a measurement of fatigue. Which technique is the best in this setting could be determined by using these scales throughout the pilot and asking feedback from the participants.

3.2.9 Sleep

Sleep is a very important stage of the day in our lives. While sleeping, our body comes to rest. For humans (generally for mammals), sleep can be divided in two types: rapid eye movement (REM) and non-rapid eye movement (NREM). Each of these types have characteristic features.

REM During the REM phase in sleep, there is an increased brain activity compared to when being awake. The muscles in the body are completely relaxed and the eyes make horizontal and vertical movements. Only during this phase of sleep, people are able to dream. While sleeping, an adult remains on average 15% of his sleep in REM phase. Although it is not completely clear what the function of REM sleep is, there are several hypotheses about the functions of REM sleep. In general, sleep helps to save memories. Thus, REM might play a role in the processes that make sure that our memories are stored in our memory (memory consolidation), however REM is not a necessary condition for this process.

In [70], a totally different theory is proposed, which states that REM sleep evolved out of a defensive reflex called toxic immobility. This reflex results in total immobilization of an animal, which creates the suggestion that the animal is dead. This reflex is useful as a last line of defense when the animal is attacked by a predator. According to this theory, this reflex has similarities with REM sleep.

NREM In this stage of sleeping, there is hardly any eye movement. In contrary to REM, dreaming is rare in this stage of sleeping and there is no relaxation of the muscles. NREM can be divided in three stages [57]. While transitioning from the

first to the second stage, the movement of the eyes is reduced. In the last stage, it is more common to fall asleep than in the other stages.

As with the theories for REM, there also exist a lot of theories about the function of sleep:

- Increases the restoration of the body;
- Contributes to the processing of memories;
- Dreaming, that can occur during some sleep stages, can help processing experiences.

When researchers want to get insight in the sleep of one or several persons, sleep studies are conducted. A common sleep study is polysomnography (PSG), which is also the gold standard for evaluating a person's sleep. During this sleep study, the following will be measured [2]:

- Oxygen in blood;
- Eye movement;
- Heart rate;
- Brain waves;
- Activity of muscles;
- Lying position;
- Amount of air you breathe in and out.

Most of these measurements cannot be done with simple activity trackers, but they do provide detailed information about a night of sleep. This information is based on the movements in your sleep. At the end of the night, these movements are fed into algorithms that calculate sleep scores based on your relative amount of movement during the night. Although they can be a good indicator for your amount of sleep during the day, the quality for some other related measurements are not verified. This is especially the case for complex information like the time you spend in (N)REM during the night, for which it is unknown how these sleeping states are determined. Often a sleep score is assigned to a night of sleep. Some activity trackers show how this score is calculated, but it is hard to say how representative this score is. Sleep duration however is something that could be useful when looking at good or bad days.

3.2.10 Vitamin D

Vitamin D plays a very interesting role in the risks of relapses for MS patients. It is already known that vitamin D is an environmental factor that plays a role in the development of MS. In [55], the relation between vitamin D levels and the risk in relapsing-remitting MS of an exacerbation (a synonym for a relapse) was investigated in a longitudinal study with 73 patients having RRMS. It was found that higher levels of vitamin D were associated with a decrease in the risk of exacerbations. Vitamin D appears to suppress the autoimmune response of T cells attacking the myelin sheaths of the nerve cells and axons. When vitamin D was given in a high dose to MS patients, the activity of the T cells dropped significantly. However, some studies suggest that there is an opposite relation between vitamin D and a decreased risk of exacerbation in MS patients: low vitamin D levels are caused by less time spend outside due to increased disabilities for MS patients [71]. Vitamin D levels can be influenced by sunlight. The effectiveness of exposing the skin to sunlight, and thereby 'getting' vitamin D, depends on several factors:

- **Color of the skin**. People with a darker skin tone tend to produce less vitamin D when being exposed to sunlight.
- **Time of the day**. Ultraviolet B rays (UVB) play a role in the natural way of getting vitamin D. When the sun is shining on the earth at a large angle, these UVB rays are blocked. Thus during early and late parts of the day, exposing your skin to the sun will not help in the synthesis of vitamin D.
- **Where you live**. The closer you live to the equator, the easier it will be for your body to synthesize vitamin D from UVB rays.

It was already known that vitamin D had a protective effect at the start of MS, but new research shows that it can also effect the course of the disease. With the data from wearables, it might be possible to get an approximation of the time spent outdoors. Because time spent outdoors can influence vitamin D levels, the chance or increased risk of an exacerbation could be determined.

3.2.11 Neurogenesis

Neurogenesis is the name of the process responsible for growing new neurons. The hippocampus is a place in the brain of the human body where new neurons can be created. Neurogenesis plays an important role in learning, memory, mood and emotion [62]. What is also mentioned in this research, is that there is a clear relation between neurogenesis and depression. A reduced adult hippocampal neurogenesis (AHN) in different animal models, where antidepressants were able to restore this, is supporting this relation [17, 43]. It is also possible to influence the process of neurogenesis. In [63], an overview of different diets, and their influence on AHN for rats and mice was given. Diets consisting of high saturated fat, were found to reduce the production of new neurons. The same effect was found for the intake of

Effect on neurogenesis	
Increases	**Decrease**
Running	Aging
Sex	Sleep deprivation
Learning	Stress

Table 3.1: Events that increase or decrease neurogenesis [66].

ethanol, that is present in alcohol. Intermittent fasting and reducing your calorie intake were found to increase neurogenesis. This effect was also found for foods with a soft texture.

Despite the fact that a lot more factors influence depression and mood, in theory it would be possible to predict levels of depression based on someones diet. This would make it possible to prevent the effect of stress and even prevent depression. Because the evidence for the mediation of AHN on the effect of diets on mental health is limited, more studies are needed to confirm this relation. In Table 3.1, the effect on neurogenesis for some activities or events is given.

3.3 Selecting Relevant Biomarkers

As we have seen in Section 3.2, there is a large collection of biomarkers that could be related to good and bad days. In this section, we will identify which biomarkers could be useful. This will be done by looking at the use of biomarkers in other diseases for longer periods of time. We must also look at current technology being able to measure the biomarker.

3.3.1 Limiting the Set of Relevant Biomarkers

There are couple of biomarkers for which we decided they could not be used in our research. This applies to the following biomarkers: vitamin D, weather, neurogenesis, stress, 'sitting, walking and lying pattern' and HRV. We do actually see potential for each individual biomarker in future research. However, currently those biomarkers are not easy to measure, have not been investigated thoroughly, have not been used in practice or cannot be measured by consumer-available devices. All of these reasons made us choose a specific selection of biomarkers. Before looking into this selection, we will describe for each of these biomarkers listed above why we decided not use them in our research.

- For measuring vitamin D, we would need some sort of advanced sensor that could determine the amount of vitamin D in the blood. Although such devices exist, this would force our participants to manage multiple devices simultaneously. This would add unnecessary complexity. While writing this

thesis, there did not actually exist any consumer-available device that could measure vitamin D (as far as we know). All these reasons made us not to choose this biomarker.

- Neurogenesis is a new research topic that is currently being explored. As we have seen in Section 3.2.11, certain kind of activities and foods are influencing this process. Because the topic is relatively new, there is no data from which we can see how a certain type of food or activity influences neurogenesis. If this data was present however, this would require more user intervention: users would need to input their meals into the application. This seems undesirable. This observation of unwanted complexity ruled out neurogenesis as a biomarker.

- As seen in Section 3.2.4, stress and arousal are partly intertwined. We have seen that the measurement of sweat on the skin using a dedicated sensor can be used to measure levels of arousal. Some activity trackers are able to measure this sweat production on the skin, but not all of them. Because the functionality was not present on all the activity trackers we had available for our experiment, we chose not to include this biomarker.

- Activity classification is something that most activity trackers can do. However, the classification of positions (sitting / walking / lying pattern) is not a basic functionality for currently available activity trackers. Due to this currently lacking feature of activity trackers, we decided not to include this biomarker as well.

- Weather data is freely available, and given that most patients prefer certain temperatures this is a promising biomarker. Due to time restrictions, we decided not to include this biomarker in our research.

- Despite the restricted knowledge on HRV, it appears to be a reliable biomarker for responses to stress. While acute stress only influences HRV for a short term, chronic stress can even affect HRV during sleep. This was found in [25]. These stress related changes in HRV caused significant decreases in sleep maintenance, which is the ability to remain at sleep during night. Thus, the effects of some forms of stress on HRV can be seen for an extended period of time. Despite being a very useful biomarker, HRV cannot be measured using current available smartwatches. It requires advanced technology like ECG to get an accurate approximation of HRV. It is very likely that smartwatches in the future will have the ability to measure this indicator accurately. However, due to the current restriction of technology available for consumers, we will not select this biomarker.

3.3.2 Selected Biomarkers

Fatigue seems to be one of the most common symptom of MS. The results from the **2MWT** can be used as a reliable and valid measure of the physical function of patients. This is only possible when the test is standardized. A problem with the 2MWT is that there is a lack of reference values. This limits the interpretation of the results. In [60], this limitation was addressed by establishing a reference equation to predict the distance walked in the 2MWT. This was done for healthy participants. The equation was as follows:

$$\text{2MWT}_{\text{predicted}} = 252.583 - (1.165 \times \text{age}) + (19.987 \times \text{gender}^*)$$

$$\text{gender}^* = \begin{cases} 1 & \text{if male} \\ 0 & \text{if female} \end{cases}$$

The equation was found to be highly reproducible in healthy subjects. It might be possible to come up with an equation for MS patients that is capable of giving an estimation of the distance walked. This equation could have additional parameters, like the type of MS the patient has and the current severity of the disease.

Standardization of the 2MWT in the pilot is something that requires some more attention. Because the lack of a supervisor during this test, people might have different expectations on how to perform the walking test. There is documentation available in the 'Mijn Kwik' app about the test, but if people put this text in the same perspective as terms of use, it is not likely that this documentation is read thoroughly. Despite these uncertainties, the results of these tests seem promising as a potential biomarker. For MS, both the 6MWT and the 2MWT [22] were tested on relevance of habitual walking performance (HBW). Both the 2MWT and 6MWT were the best predictors among all the other tests for habitual walking performance, confirming its high potential.

Your **resting heart rate** (RHR) (Section 3.2.1) can be influenced by a lot of factors, including stress. When being exposed to stress for a long period, your body is constantly functioning in a high gear and therefore affecting your heart rate. Stress can also reduce your energy, sleep and make you feel cranky. Stress is also playing a big role in MS (Section 3.2.4). Although the use of resting heart rate from heart rate data from wearables in research is a new concept, it can prove useful. Combined with data from the questionnaire app, we can take a closer look at the relation between heart rate and the indicated quality of the next day.

There have been some follow up studies on the effect of elevated resting heart rate on the risk of all-cause mortality. In [31], this effect was researched on 2798 healthy middle-aged men in a 16-year follow up study. An inverse relation was found between resting heart rate (RHR) and physical fitness. An increased RHR was also related to mortality, independent of the physical fitness of a person

During **sleep** (Section 3.2.9), you lay the foundation for your next day. You will immediately notice the impact of lack of sleep from the previous night. Sleep also plays an important role in your mood. In [68], the relations between sleep and mood

were investigated. It was found that sleep was more related to subsequent well-being than prior well-being. Also an earlier onset of sleep was related to a better mood the next day. Earlier onset of sleep was also a better predictor of mood than sleep duration. This seems to be in conflict with the general opinion that duration of sleep has the most impact on your mood.

However, in [53] the effect of a short sleep duration on the intake of nutrients and mood was investigated. Undergraduate college students were used in this experiment. It was found that the total mood disturbance (TMD) score was linked to sleep duration and sleep quality. The Pittsburgh Sleep Quality Index (PSQI) was used to measure the quality and pattern of sleep. It also includes questions to estimate the sleep duration. Participants that had a sleep duration of at least 6 hours during the night had lower TDM scores ($p = 0.0091$) and therefore experienced a more positive mood state the next day. This suggests that a short sleep duration and a lower sleep quality have a negative impact on the mood state.

In [73], sleep as a biomarker for depression was investigated. Changes in a person's sleep pattern also seem to be important for diagnosing a subtype of depression, regardless of being positive (more sleep or an improvement of sleep quality) or negative (decrease in sleep or worsening of sleep quality). Sleep (and specifically sleep duration) can therefore be used as a biomarker to determine the increased risk for depression, which influences the quality of upcoming days.

Chapter 4

Description of Experiment

4.1 Setup

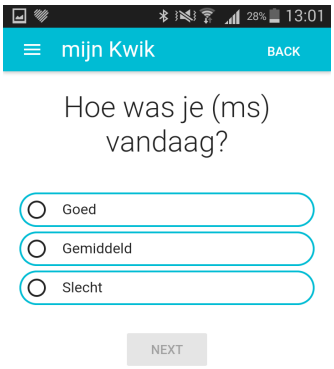
ORIKAMI is working on a project for MS patients, which includes the development of an app called 'Mijn Kwik'. In this project, daily data of MS patients will be gathered through wearables and a questionnaire app (Figure 4.1). All of this data is gathered with the intention to predict good and bad days. This study will be used to get more insight into the following aspects:

- Stability of the application;
- Ease of use of the wearables and app for MS patients;
- Problems with wearables during the experiment.

We explained our plans at a meeting for MS patients organized by Orikami. From all the people that attended the meeting, five patients volunteered to participate in the study. A second meeting was organized to explain all the necessary details to the participants. Each of the participants were given one of the following devices:

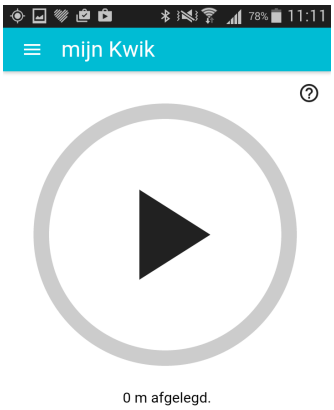
- Basis Peak;
- Fitbit Charge HR;
- Microsoft Wristband 2.

With our help, the participants were helped setting up the device and installing the corresponding mobile app on their phones. Also an user agreement along with a privacy contract were handed out. The latter had to be signed by each of the participants. In this privacy contract, it is also stated that all of the gathered data will only be used for research purposes.

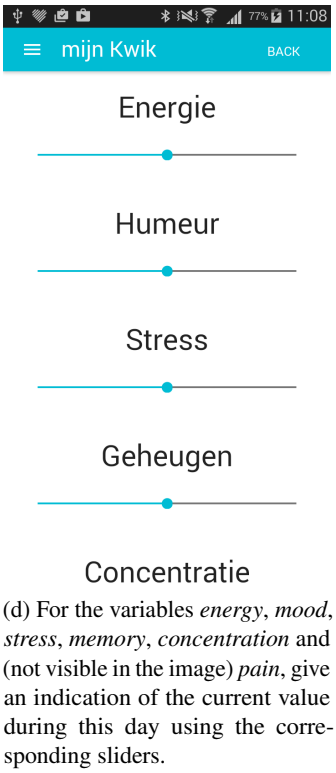


(a) Question about how your day was. It can be answered with *good*, *average* or *bad*.

(b) Question about your mood. The slider changes the happiness of the smiley, which represents your feeling today.



(c) The walking test.



(d) For the variables *energy*, *mood*, *stress*, *memory*, *concentration* and (not visible in the image) *pain*, give an indication of the current value during this day using the corresponding sliders.

Figure 4.1: Questions that must be answered and tasks that must be performed each day in the questionnaire app.

4.2 Data storage

Each of the activity trackers that are handed out to our participants will be linked to the Philips's "Connect to Healthy". Through this service, the captured data will be synced to a database that can be used to access all of the data from the participants. The data is stored in two databases, one containing the users, the measurement types and the actual measurements from the activity trackers. The other database contains the observations done through questions and the types of questions in the questionnaires.

Chapter 5

Analysis of Selected Biomarkers

In Section 3.3, we selected three relevant biomarkers: 2MWT, resting heart rate and sleep duration. In this chapter, we take a closer look at these selected biomarkers. For each of these selected biomarkers, we look at the corresponding data and combine this data with the questionnaire data for day ratings to see how useful the biomarker can be in predicting good and bad days. In particular, we want to relate the selected biomarkers to the 'How was your MS today?' question in the 'Mijn Kwik' application used by the participants. This specific question can be answered with 'good' (goed), 'average' (gemiddeld) or 'bad' (slecht). We will be looking at consecutive days, and how these biomarkers possibly play a role in the quality of upcoming days. Despite the limited dataset (consisting of data gathered over a period of a couple of weeks), we try to get a better understanding of these biomarkers and how they possibly relate to good and bad days. As noted before, we will be looking at the data of five participants, each of which is identifiable by a unique identity. For the patient that was wearing the Microsoft Band 2, the code to integrate the device's data with our database was not finished while writing this thesis. Therefore, the data from this participant was not used in any of the analyses of the biomarkers.

5.1 2MWT

We want to investigate if the distance walked during the experiments of the 2MWT could say anything about the quality of the next day. Before we can actually use the results of the 2MWT, we must research how accurate and reliable the GPS on mobile phones is. In Android, there is a method that can give us the accuracy of the retrieved location in meters [24]. However, this accuracy is only concerned with horizontal accuracy. Bearing, velocity or altitude are not included. The GPS is operational under all weather conditions [6] . Weather conditions do not influence the accuracy of the position [39]. Obstructions like roofs and walls do block GPS signals. The app should notify the user when the estimated error becomes too big.

Because the earth is not flat, there are actually no straight lines between two GPS coordinates. To address this problem, there exist several algorithms to calculate the distance between two GPS coordinates.

- The great-circle distance [72] assumes that the earth is a sphere and uses the spherical law of cosines formula.
- The haversine formula can also be used to give great-circle distances between two GPS coordinates. This formula is better for small distances, because the formula is not too sensitive for a change of input.
- Methods based on the great-circle distance are less accurate than Vincenty's formula, which assumes that the earth is an oblate spheroid. This formula can be accurate up to the millimeter.

The error in the great-circle distance can be up to 0.5%, which will result in an error of several centimeters in this case. Because the impact of this error is so low, we decided to use the great-circle distance instead.

To get more insight into the accuracy of the measured distance walked, two experiments were designed to test the inter device reliability and accuracy of calculated distance walked in the 2MWT using the 'Mijn Kwik' app. This is the same app that is used by the participants in our study.

5.1.1 Experiment 1

Method

In this experiment, we used the straight 100 meters on an all-weather running track as a comparison. The GPS coordinates, used for calculating the total distance walked, are registered at specific time intervals. These intervals depend on the GPS signal and the used chipset. In this case, the time between the registration of consecutive GPS signals was between 4000 and 6000 milliseconds. To calculate the distance walked, we first used linear interpolation to create a roughly smoothed line through the GPS coordinates. Afterwards, a Savgol (Savitzky-Golay) filter [52] was used to smooth this raw line. The coordinates returned by this Savgol filter where then used used to calculate the distance walked using the great-circle distance.

Results and Conclusion

The results of the experiment are shown in Table 5.2. The specification for each of the phones used in the experiment are shown in Table 5.1. During the experiment, one phone was not able to save the GPS coordinates. For the results from this phone, no smoothing was applied. It is unknown why this phone was not able to save these coordinates. Before using the method that gave us the results shown in Table 5.2 (this method will be discussed later), we used some other methods as well. As mentioned before, GPS coordinates are given an accuracy value. Because

Device	OS	GPS support
Samsung Galaxy S3 Neo GT-I9301I (1)	4.4.2	A-GPS, GLONASS
LG Leon (H320)	5.0.1	A-GPS
Samsung Galaxy S3 Neo GT-I9301I (2)	4.4.2	A-GPS, GLONASS

Table 5.1: Devices used in the experiment.

	Attempts								
Device	**#1**	**#2**	**#3**	**#4**	**#5**	**Avg. Error**	$\tilde{x}$	μ	σ
Samsung Galaxy S3 Neo GT-I9301I (1)	140m	89m	104m	95m	107m	13.4%	104m	107m	17.7m
LG Leon (H320)	96m	100m	85m	91m	33m	19%	91m	81m	24.52m
Samsung Galaxy S3 Neo GT-I9301I (2)	124m	127m	110m	273m	155m	57.8%	127m	157.8m	59.42m

Table 5.2: Calculated distance walking during each of the attemps. For each attempt the median ($\tilde{x}$) and the mean (μ) along with the standard deviation (σ) are shown.

some GPS coordinates can have an accuracy that is too low, at first we decided to use a threshold value of 50 meters. This means that all of the coordinates with an accuracy higher than 50 meter will be filtered. Along with this method, we applied the calculation of the walking speed between two consecutive GPS coordinates. If this walking speed was higher than 3.0 meters per second, the latter point would not be accepted as a valid one. The code can be seen in Listing A.3.

The previous mentioned techniques worked well for some GPS tracks, but failed in others. Therefore, we decided to test outlier detection using a robust covariance estimator. Scikit-learn [50] provides an object called Elliptic Envelope. This object can be used to fit an ellipse to the central data points, ignoring points outside the center. This gave us better results (Figure A.1.1), although in some cases the quality of the GPS track was too bad to make any prediction at all (Figure A.1.1d and Figure A.1.1e). As we can see in Table 5.2, the average error for each of the devices is high. Most of the readings are not accurate, and there are some extreme cases really differ from the actual distance walked. Inaccurate GPS signals were influencing our results. This was easily determined by looking at the unfiltered GPS coordinates. Even when standing still, the indicated distance walked would increase with tens of meters. In some extreme cases, this distance would increase with 50 meters during a couple of steps. The code that was used in the end can be seen in Listing A.1. The plots for the GPS tracks can be seen in Figure A.1.1.

From this experiment we can conclude that the accuracy and validity of GPS used by phones is something that might require some more investigation. Even with the two phones of the same model, the average error differs greatly. This is strange, because in each attempt, all phones were used at once to register the GPS coordinates which were eventually used to calculate the total distance walked. Because this experiment was only done with a couple of smartphones, these findings

Device	Attempts					Avg. Error	$\tilde{x}$	μ	σ
	#1	#2	#3	#4	#5				
Samsung Galaxy S3 Neo GT-I9301I	38m	41m	133m	111m	112m	30.6%	111m	87m	39.6m

Table 5.3: Calculated distance walked during each of the attemps.

are not very representative.

While using the Elliptic Envelope for fitting an ellipse to central data points worked fine when walking in a straight line, it does not work when an arbitrary route is walked which can also have other shapes than a straight line. When this is the case, other methods like One-class SVM [58] or clustering algorithms must be used.

5.1.2 Experiment 2

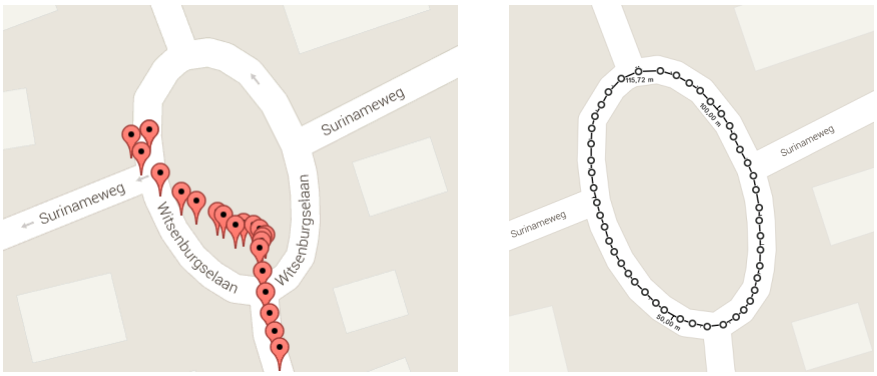
Method

As mentioned in the discussion of Experiment 1, the method applied for detecting outliers in that experiment was not suitable for non-straight tracks. Therefore, we decided to use a clustering algorithm to detect the outliers in these kind of GPS tracks. We choose to use the DBSCAN clustering algorithm. Density-based spatial clustering of applications with noise (DBSCAN) is a clustering algorithm based on density. Points that are closely packed together are grouped into clusters. Points that reside in regions with a low density are marked as outliers. Both of these classifications are based on the amount of and distance between neighbors. DBSCAN does not require specification of the number of clusters in the data beforehand. Another nice property of DBSCAN is that it has a notion of noise, which makes it robust to outliers. This makes it an excellent algorithm for doing the task of finding clusters in GPS data. For all the points in the clusters found with DBSCAN, the distance of the GPS coordinates were calculated and added. Notice that we do not calculate the distance within each cluster separately. We also calculate the distance between two consecutive points if both points were classified in different clusters.

The track we used in this experiment had a total distance of 115 meters and was circular shaped. For the calculation of the total distance walked, the same smoothing methods (linear interpolation and Savgol filter) were applied. As in Experiment 1, the same method for calculating the distance between two GPS-points (great-circle distance) was used. In this experiment, only one phone was used.

Results and Conclusion

The results for this experiment can be seen in Table 5.3. If we plot our GPS coordinates (Figure 5.1), we notice that the shape of a ellipse/circle is hardly visible. Sometimes the coordinates just show a straight line, like some internal system



(a) (b)

Figure 5.1: Plot of the GPS measurements of one 2MWT attempt (Figure 5.1a) compared to the original walked track (Figure 5.1b).

was used to predict future GPS coordinates when the quality decreased or when the signal was lost. The clustering algorithm (DBSCAN) we decided to use for detecting outliers is working fine for non-straight GPS tracks. The major problem is that some GPS measurements are not accurate. If this is the case, this will impact the accuracy of the calculation of the total distance walked. Because we do not know beforehand what the shape of a GPS track, belonging to one of our participants, is, we will also be using DBSCAN for the analysis of all of the experiments done by our participants.

5.1.3 General Discussion Experiments

For both Experiment 1 and Experiment 2, we decided to use linear interpolation and a Savgol filter to smooth our data. However, there exist a number of other techniques that could be useful as well. In [32], various smoothing techniques were applied to vehicle GPS. The performance of these algorithms were evaluated on minimizing the impact of random GPS errors in the estimation of speed, acceleration and distance. The algorithms were least square spline approximation, kernel based smoothing and a modified version of the Kalman filter. From all these algorithms, the latter was found the most accurate in distance estimation compared to the on-board monitor. Kalman filters [41] are also frequently used to smooth GPS data. We will not go into more details of this algorithm, because the math behind this filter is quite extensive. However, there seems to be confusion about post processing of GPS data in mobile phones. Some experts claim that current GPS chips extensively filter the GPS data, while others refute this statement. If post processing is done in the chips of mobile phones, we would need a second sensor, for example a velocity sensor, to actually improve the data using a Kalman filter. Kalman filtering is suitable when the input data are noisy. If it is used on GPS data with a small sample rate, the track will be much smoother. However, this will result in a loss of precision. Very noisy

points will be removed, so that the distance between some GPS points will increase, resulting in a small error in the estimation of the total distance walked.

5.1.4 Analysis

After the experiments into the results from the 2MWT, we take a look at the data from our participants so far. For the analysis, we used the same algorithm we applied to the data resulting from Experiment 2 (Section 5.1.2). This also includes the use of DBSCAN clustering on the GPS coordinates. Looking at data from our participants, we see huge differences between each of the participants. Some participants regularly did the walking tests while others did not. For the latter group, most of them have never done any experiment. We could make a fast conclusion that not having done any walking test makes it hard to say something about this person. But not having done a walking test for a couple of days or for a longer period of time could also mean a lot. If a person is too fatigued to be able to do the walking test, no walking tests results would be there. Therefore, no walking tests results could reflect a poor physical state. Of course, it is also possible that participants forgot to perform the instructed task for one or several days. This makes it more difficult to make claims about their physical capabilities. For the participants that did the walking test regularly, some experiments consisted of positions that would not produce anything useful using DBSCAN clustering. Looking at these experiments, we see that in the majority of the cases they contain two or even more duplicate coordinates: the start and the end positions. To be classified as a cluster, DBSCAN requires more points. Due to this lack of points, some experiments could not be used. In the early stages of the experiment also a bug was found in the app. This resulted in positions that did not contain any latitude or longitude information and making them useless. This reduced the number of usable experiments.

To get more insight in all the data, we created plots which are shown in Figure 5.2. Each letter of the plot corresponds with the participants in Table A.3. Only for the person that did not do a single walking test (Table A.3e) no plot is shown. For both participant (a) and participant (b), some results of the walking tests (200+ meters walked) seem to be inaccurate. Given that an average healthy person walks 5 kilometers per hour, it is possible to walk a distance of roughly $5/3.6 \cdot 120 \approx 167$ meters. Because MS patients often find it difficult to walk, it is unlikely that such distances would be walked in the performed walking tests. It is also possible that some people really wanted to challenge themselves and therefore performed really well. In [47], factors for lower walking speed for persons having MS were analyzed. By using only a track of 10 meters, it was found that 85% of the patients showed a lower walking speed. This makes the statement of highly motivated participants less likely. To validate the hypothesis of inaccurate GPS tracks causing a high distance walked, we would need more participants for comparison.

Before actually looking at the data to inspect the relation between day ratings and measurements, we have to find out how both the distance walked and the day rating are related with each other. For this we need to use some form of statistical

test. Before actually choosing a statistical test, we need to know what the level of measurement for all of our involved variables is. For the ratings we have three different categories: 'good', 'average' and 'bad'. The level of measurement for the variable 'day rating' clearly is ordinal. The level of measurement for the variable 'distance walked' is ratio (this also same for the measurements of heart rate and sleep duration). There are a lot of statistical tests available to research the association between two variables. Using [46], we find that both the Spearman's rank correlation coefficient ρ and the Kendall's rank correlation coefficient τ are applicable. The values of Kendall's τ are usually smaller than the values Spearman's ρ. Because Spearman's ρ is based on deviations, it is more sensitivity to errors compared to Kendall's τ. Because it is very likely that our data is contaminated with inaccurate data, we decided to choose Kendall's τ to investigate the association between the two variables. Kendall's rank correlation is a correlation coefficient based on the ranks of the data instead on the data itself. The results of the application of the algorithm can be seen in Table 5.4. In this table, we also see a so called p-value. When the Kendall's rank correlation coefficient is used to determine whether two variables are related or not, the null hypothesis of independence of the two used variables has an expected value of zero. According to the documentation of the used implementation of Kendall τ from SciPy, this p-value is "the two-sided p-value for a hypothesis test whose null hypothesis is an absence of association" [59]. The two-sided (or two-tailed) p-value allows to test the statistical significance in both directions. This means that you are testing for the possibility of a relation in both directions. As we can seen in Table 5.4, the correlation coefficients for participant (b) and (c) are not shown. This is because the algorithm for calculating the correlation coefficient returned NaN (Not a Number) for both of the corresponding measurements. Because we investigate the relation between the measurement and the rating of the day after, we drop our first rating and our last measurement. This makes sure that for each measurement, the rating of the next day is used to see if there is a relation between the two of them. If we have only two measurements and corresponding ratings, the method described above will leave us only with one measurement and one rating. This is not enough to calculate the Kendall's correlation coefficient. Therefore, the entries for these two figures resulted in NaN values.

Participant	τ	p
(a)	0.40	0.414935
(d)	0.0	1.0

Table 5.4: Expressing the rank correlation between distance walked and day rating.

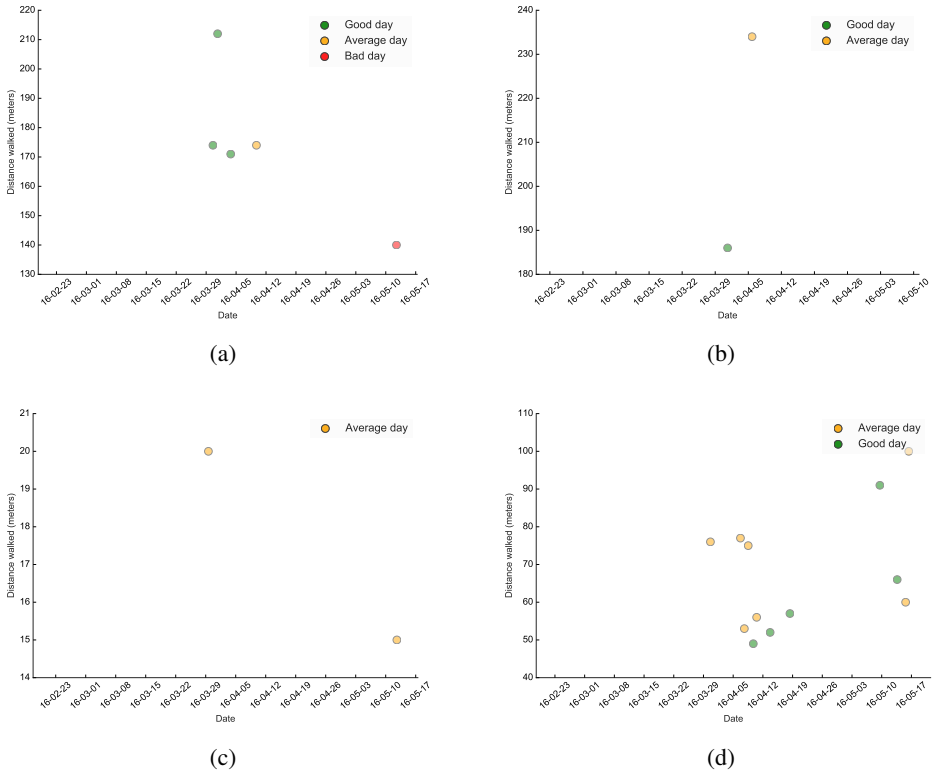


Figure 5.2: Plots of the walking durations of the participants. Only dates that both had a rating and a measurement of distance walked are shown. As can be seen in the legend of each of the figures, each point is colored according to the rating given by the participant for that specific day.

To interpret the rest of the coefficients, we must look at the outcome of the correlation coefficient.

- A positive correlation coefficient means that the ranks of both the variables are increasing.
- A negative correlation coefficient means that if the rank of one variable increases, the rank of the other variable decreases.
- If both variables are independent, a coefficient with a value of zero would be expected.

For the other participants ((a) and (d)), the amount of samples is better. For all of the coefficients, the corresponding p-value is never lower than 0.05. Therefore, there is no sign of statistical significance for each of the coefficients. This lack of statistical significance makes it useless to try to interpret the correlation coefficients. In the end, it seems like the amount of experiments for most of the participants is just too low to get meaningful results using statistical analysis.

5.2 Resting Heart Rate

Every minute, the heart rate of a participant is measured by the activity tracker and eventually stored in the database. In Table A.2, we see the heart rates of our participants. The corresponding plots can be seen in Figure 5.3. As one can notice, not all the dates visible in each of the boxplots in Figure A.3.1 are in these plots. This is because all participants did not give a rating for some days. Although we do not know for sure why this is the case, this is probably because they forgot to do so. We also excluded these days without ratings in each of the tables in Table A.2. In contrary to our first investigated biomarker (2MWT), we can see that a lot more measurements are present in Figure 5.3. Because each device automatically measures the heart rate, there is no user intervention required which definitely has a positive effect on the amount of data.

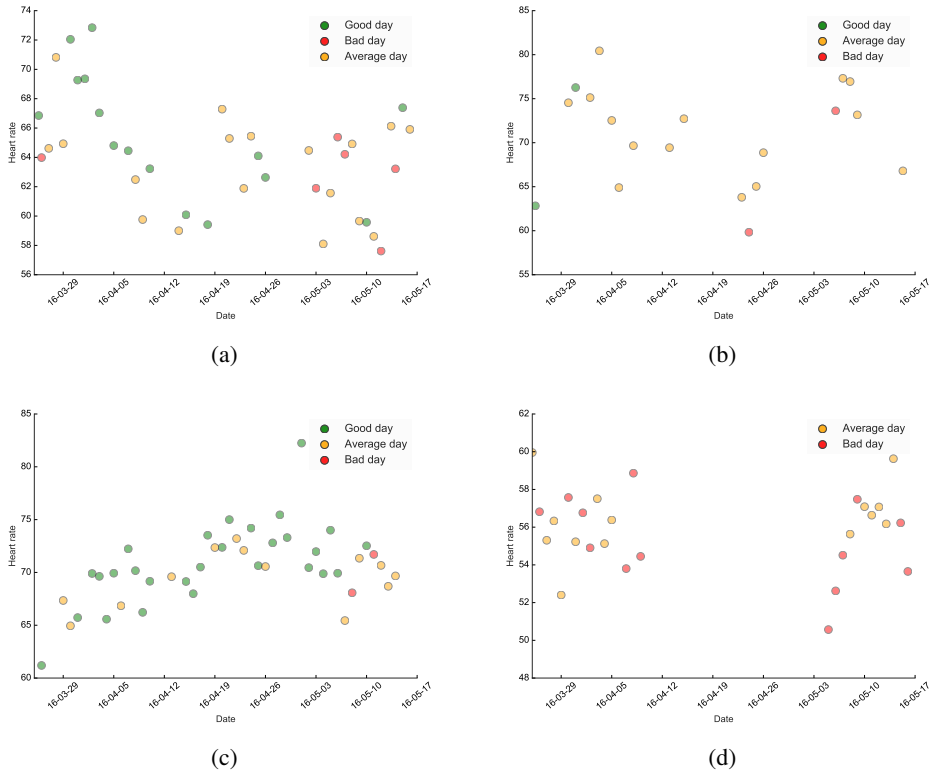
If we take a look at the boxplots in Figure A.3.1, we see that the majority of the fliers (outliers) are in the high regions of the boxplots. In the lower parts, we rarely see fliers. This observation gave us the insight that the average of the lowest 5% of the measurements is probably a good and stable estimation of the resting heart rate of the participant for a single day. Therefore, the plots show the average of the lowest 5% of the heart rates measurements and not the average of all the measurements for that day. For each participant, Kendall's τ correlation coefficient along with the p-values can be seen in Table 5.5. As we have seen in Section 3.2.1, the resting heart rate for adult lies between 60 and 100 beats, which is also suggested by the figures in Figure 5.3. In Section 3.2.1, we have also seen that a low resting heart rate is more desirable than a high one. Looking at the plot from participant (b), we see exactly the opposite of what we would have expected: for most of the lower resting heart rates, the day ratings tend to go worse. The day on which the

participant had his lowest resting heart rate was rated as ‘bad’. For participant (a), such a relation is not visible as the data seems to be spread randomly.

In Table 5.5, we see that the correlation coefficient for participant (c) is close to zero. The coefficients for participants (b) and (d) tend to go negative, suggesting that if the rank of the average of the lowest 5% of the heart rate measurements decreases, the rank of the rating increases (which we also noticed by ourselves). None of the p-values is near 0.05. This indicates that the effect observed in the data is likely, even assuming that the null hypothesis of no effect is true. For this particular biomarker, the null hypothesis would state that there is no relation between the average of the lowest 5% of the heart rates and the rating of the day after.

Participant	τ	p
(a)	0.128873	0.254708
(b)	−0.176887	0.305310
(c)	0.021340	0.842163
(d)	−0.174944	0.210129

Table 5.5: Expressing the rank correlation between the average of the lowest 5% of the heart rate measurements and the day rating.



(a) (b) (c) (d)

Figure 5.3: Plots of the average from the lowest 5% of the heart rates measurement of the participants for each day. Only dates that both had a rating and measurements of heart rates are shown. As can be seen in the legend of each of the figures, each point is colored according to the rating given by the participant for that day.

5.3 Sleep Duration

In this section we take a look at the sleep duration. As we can see in Table A.1, there are a lot of day ratings without a measurement of sleep duration and vice versa. Another thing is that there can be multiple entries that have the same date but a different sleep duration. This is only the case for the devices that resulted in the data shown in Table A.1a and Table A.1b. These 'splits' occur when a person's sleep is interrupted. A sleep interruption is detected by the device, which then saves the new measurement as a separate one. Some participants also took one or more naps during the day, resulting in more entries with the same data. For dates without a sleep measurement, the format is Y-M-D. For dates with one or multiple measurements, the date is formatted as Y-M-D H-M-S. This latter date format indicates the time the person stopped sleeping and awakened. Due to this formatting, we can distinguish if an interruption belongs to the previous or the current night. For the used devices that resulted in the measurements in Table A.1c and Table A.1d, the sleep durations are never split. The API returned the total sleep duration of that day. The dates in these measurements only show the day the measurement belongs to and do not indicate the time at which the participant fell asleep or awakened. For analysis, splits belonging to the same day were summed and represented the total sleep duration for that day. Looking at Table A.1, it is strange that a lot of values are missing. Especially in Table A.1a, we see a lot of measurements where the time of awakening is around 5 in the morning. We would expect that this is due to an interruption and that the person would resume his sleep. This is however not confirmed by the data, because in most of these cases there is no additional entry for that specific date. It could also be possible that this particular participant decided to take of his wristband during the night because it was becoming too uncomfortable. Another explanation could be that a couple of hours before awakening, people reside in a sleep stage from which it is easier to awake and also involves more movement. Due to this increased amount of movement, the device could classify your sleep state as 'not sleeping', although you are. For dates with no measurements, it might be the case that the participant decided not to wear the activity tracker at all.

Participant	τ	p
(a)	0.426694	0.042305
(b)	0.188608	0.447772
(c)	0.070367	0.558409
(d)	0.051105	0.686289

Table 5.6: Expressing the rank correlation between sleep duration and day rating.

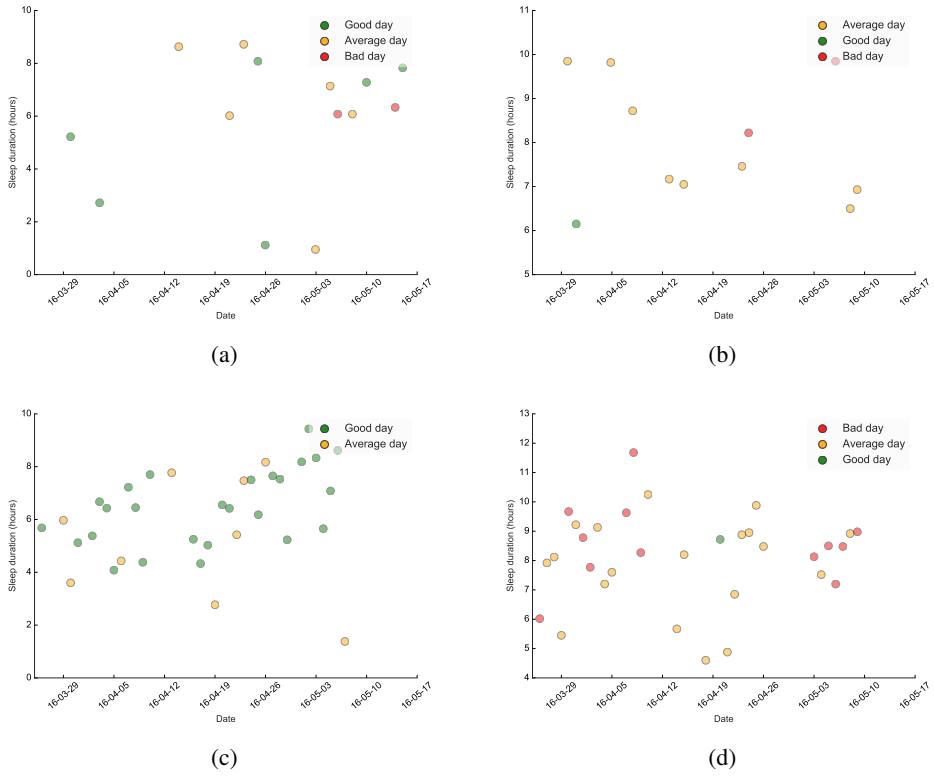


Figure 5.4: Plots of the sleep durations of the participants. Only dates that both had a rating and a measurement of the sleep duration are shown. As can be seen in the legend of each of the figures, each point is colored according to the rating given by the participant for that day.

In Figure 5.4, we see the plots for each of the participant. The corresponding coefficients are shown in Table 5.6. Looking at these coefficients, we see that all of them are smaller than 0.5 and have a positive sign. In fact, most of them have a value that is close to zero. For participant (c) and (d), the coefficient is very small. Although a bit bigger, the same applies for the value of the coefficient for participant (b). The coefficient for participant (a) is the biggest. Because most values of the coefficients are low, they suggest that the measurement and the rating of the next day are statistically independent. Only for participant (a), the p-value is significant. For all the other entries, there are no significant p-values. In some of the figures, we see some measurements that have a total sleep duration of less than 4 hours. Although this is possible, it is not very likely. We could add an extra survey to specifically ask for the quality of the night, so that we can verify our measurements against these answers (triangulation). This could also include a question where the participant can give an indication of the total sleep he or she had during the night. Using these extra questions, we could determine more easily if the measurements are reliable or not.

Chapter 6

Conclusion and Discussion

In this thesis, we have looked at relevant biomarkers in the prediction of good and bad days for multiple sclerosis (MS) patients. Because activity trackers were used, the validity and reliability was reviewed in Section 3.1. In this section it was found that activity trackers are widely used in the research field and that the measurement of most of the health related variables are an accurate representation of the true values. In Section 3.2, we have conducted a literature review for relevant biomarkers. This was done by looking at how biomarkers can be measured and how they are currently being used or could be used to get more insight in the health state of (MS) patients. In Section 5.1, two experiments were designed and conducted to test the reliability of GPS coordinates that were used to calculate the distance walked in the 2MWT. In both experiments it was found that the quality of the GPS tracks would differ a lot. There were tracks that would give a good approximation of the total distance walked but other tracks were of insufficient quality to get a good estimation. From the result of the final experiment, it was decided that applying clustering on the GPS points belonging to a GPS track would detect the outliers. These outliers are excluded in the algorithm for calculating the total distance walked. After reviewing existing usage of biomarkers in Section 3.3, we chose the following biomarkers for investigating their relation with good and bad day analysis: 2MWT, resting heart rate (RHR) and sleep duration. For the results of the 2MWT, the final algorithm which makes use of clustering was used to calculate the distance walked in the analysis of the data from the participants. For each individual biomarker, we investigated its relation with the day rating of the next day by using statistical analysis. Although we found some relations that confirmed or refuted some of our hypotheses, their was too much diversity to conclude anything. The restricted dataset and number of participants contributed to this. We also found that some data seemed to be missing and that participants were forgetting to rate days using the 'Mijn Kwik' app. Especially data, that required the participant to perform specific tasks in order to become available, was missing. To minimize the risk of missing data due to participant in future research, most of the

measurement should be done automatically if possible.

Although we focused on a small selection of biomarkers to investigate their potential as a biomarker in the prediction of good and bad days for MS patients, we do not exclude the use of others. As an example, another biomarker was 'found' during a hackathon. In the weekend of 21-22 May 2016, this MS hackathon was organized in Amsterdam, the Netherlands[1]. In this hackathon, a variety of people worked together to gain new insights into MS for researchers, doctors and patients themselves. A team from Orikami also participated in this hackathon. Currently there is not an objective scale to measure fatigue. People that have to indicate themselves how fatigued they are, seem to have problems with this: they are not able to correctly make an estimation of their current fatigue. Relatives however are often able to see when people are fatigued, although these people cannot see this by themselves. In current literature, the relation between eye movement and fatigue has already been investigated [4]. Based on these findings, we came up with the idea to measure fatigue based on recordings from the eye. By letting the user focus on a screen on which circles appear randomly, we can determine how fast the eye responds on these events. Using these measurement, we would like to get an objective measurement of the fatigue level from MS patients. The idea was also appreciated by specialists and MS patients and we even managed to win the 2nd place. In the future, this new technique should be investigated even more to get a more accurate measurement on fatigue for MS patients.

Because this thesis can be seen as an exploratory study, the lack of any relation between the investigated biomarkers and the day ratings was in line with our expectations. In upcoming research, investigating the relation between biomarkers and the quality of upcoming days could be done using more sophisticated methods. This should also include analysis in which multiple biomarkers are considered at the same time.

[1] `http://www.mshackathon.nl`

Bibliography

[1] U Rajendra Acharya, K Paul Joseph, Natarajan Kannathal, Choo Min Lim, and Jasjit S Suri. Heart rate variability: a review. *Medical and biological engineering and computing*, 44(12):1031–1051, 2006.

[2] A.D.A.M. Medical Encyclopedia. Polysomnography. `https://www.nlm.nih.gov/medlineplus/ency/article/003932.htm`. Accessed: June 7, 2016.

[3] Bradley M Appelhans and Linda J Luecken. Heart rate variability as an index of regulated emotional responding. *Review of general psychology*, 10(3):229, 2006.

[4] Zainal Arief, Djoko Purwanto, Dadet Pramadihanto, Tetsuo Sato, and Kotaro Minato. Relation between eye movement and fatigue: Classification of morning and afternoon measurement based on fuzzy rule. In *International Conference on Instrumentation, Communication, Information Technology, and Biomedical Engineering 2009*. Institute of Electrical & Electronics Engineers (IEEE), nov 2009.

[5] Simon P Avis and William EM Pryse-Phillips. Sudden death in multiple sclerosis associated with sun exposure: a report of two cases. *Canadian Journal of Neurological Sciences/Journal Canadien des Sciences Neurologiques*, 22(04):305–307, 1995.

[6] Yoaz Bar-Sever, Bruce Haines, Willy Bertiger, Shailen Desai, and Sien Wu. Geodetic reference antenna in space (grasp)—a mission to enhance space-based geodesy. In *COSPAR colloquium: scientific and fundamental aspects of the Galileo program, Padua*, 2009.

[7] David R Bassett Jr, Ann V Rowlands, and Stewart G Trost. Calibration and validation of wearable monitors. *Medicine and science in sports and exercise*, 44(1 Suppl 1):S32, 2012.

[8] Ralph HB Benedict, Elizabeth Wahlig, Rohit Bakshi, Inna Fishman, Frederick Munschauer, Robert Zivadinov, and Bianca Weinstock-Guttman. Predicting quality of life in multiple sclerosis: accounting for physical disability, fatigue, cognition, mood disorder, personality, and behavior change. *Journal of the neurological sciences*, 231(1):29–34, 2005.

[9] Athanase Benetos, Annie Rudnichi, Frédérique Thomas, Michel Safar, and Louis Guize. Influence of heart rate on mortality in a french population role of age, gender, and blood pressure. *Hypertension*, 33(1):44–52, 1999.

[10] Bibiana Bielekova and Roland Martin. Development of biomarkers in multiple sclerosis. *Brain*, 127(7):1463–1478, 2004.

[11] Richard W Bohannon, Ying-Chih Wang, and Richard C Gershon. Two-minute walk test performance by adults 18 to 85 years: Normative values, reliability, and responsiveness. *Archives of physical medicine and rehabilitation*, 96(3):472–477, 2015.

[12] D Buljevac, WCJ Hop, W Reedeker, ACJW Janssens, FGA Van Der Meché, PA Van Doorn, and RQ Hintzen. Self reported stressful life events and exacerbations in multiple sclerosis: prospective study. *Bmj*, 327(7416):646, 2003.

[13] Nancy F Butte, Ulf Ekelund, and Klaas R Westerterp. Assessing physical activity using wearable monitors: measures of physical activity. *Med Sci Sports Exerc*, 44(1 Suppl 1):S5–12, 2012.

[14] Kong Y Chen, Kathleen F Janz, Weimo Zhu, and Robert J Brychta. Re-defining the roles of sensors in objective physical activity monitoring. *Medicine and science in sports and exercise*, 44(1 Suppl 1):S13, 2012.

[15] DM Connelly, BK Thomas, SJ Cliffe, WM Perry, and RE Smith. Clinical utility of the 2-minute walk test for older adults living in long-term care. *Physiotherapy Canada*, 61(2):78–87, 2009.

[16] Roberto Sa Cunha, Bruno Pannier, Athanase Benetos, Jean-Philippe Siché, Gérard M London, Jean Michel Mallion, and Michel E Safar. Association between high heart rate and high arterial rigidity in normotensive and hypertensive subjects. *Journal of hypertension*, 15(12):1423–1430, 1997.

[17] Alex Dranovsky and René Hen. Hippocampal neurogenesis: regulation by stress and antidepressants. *Biological psychiatry*, 59(12):1136–1143, 2006.

[18] Kelly R Evenson, Michelle M Goto, and Robert D Furberg. Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1):1–22, 2015.

[19] Patty Freedson, Heather R Bowles, Richard Troiano, and William Haskell. Assessment of physical activity using wearable monitors: recommendations for monitor calibration and use in the field. *Medicine and science in sports and exercise*, 44(1 Suppl 1):S1, 2012.

[20] Yoshihisa Fujiura, Hisashi Adachi, Makoto Tsuruta, David R Jacobs, Yuji Hirai, and Tsutomu Imaizumi. Heart rate and mortality in a japanese general population:: An 18-year follow-up study. *Journal of clinical epidemiology*, 54(5):495–500, 2001.

[21] Ian Galea, Christian Lederer, Anneke Neuhaus, Paolo A Muraro, Antonia Scalfari, Nils Koch-Henriksen, Christoph Heesen, Sasha Koepke, P Stellmann, Holger Albrecht, et al. A web-based tool for personalized prediction of long-term disease course in patients with multiple sclerosis. *European Journal of Neurology*, 20(7):1107–1109, 2013.

[22] Domien Gijbels, Geert Alders, Elke Van Hoof, Caroline Charlier, Machteld Roelants, Tom Broekmans, Bert Op't Eijnde, and Peter Feys. Predicting habitual walking performance in multiple sclerosis: relevance of capacity and self-report measures. *Multiple Sclerosis*, 2010.

[23] Domien Gijbels, BO Eijnde, and Peter Feys. Comparison of the 2-and 6-minute walk test in multiple sclerosis. *Multiple Sclerosis Journal*, 17(10):1269–1272, 2011.

[24] Google and Open Handset Alliance. Location. `http://developer.android.com/reference/android/location/Location.html`. Accessed: June 7, 2016.

[25] Martica Hall, Raymond Vasko, Daniel Buysse, Hernando Ombao, Qingxia Chen, J David Cashmere, David Kupfer, and Julian F Thayer. Acute stress affects heart rate variability during sleep. *Psychosomatic medicine*, 66(1):56–62, 2004.

[26] M Hietanen and T Alanko. Occupational exposure related to radiofrequency fields from wireless communication systems. In *Proceedings. Union Radio-Scientifique International*, 2005.

[27] Terence M Hines. Comprehensive review of biorhythm theory. *Psychological reports*, 83(1):19–64, 1998.

[28] Nis Hjortskov, Dag Rissén, Anne Katrine Blangsted, Nils Fallentin, Ulf Lundberg, and Karen Søgaard. The effect of mental stress on heart rate variability and blood pressure during computer work. *European journal of applied physiology*, 92(1-2):84–89, 2004.

[29] Edgar Howarth and Michael S Hoffman. A multidimensional approach to the relationship between mood and weather. *British Journal of Psychology*, 75(1):15–23, 1984.

[30] HP Fortify. Internet of Things Security Study: Smartwatches. Technical report, HP, 2015.

[31] Magnus Thorsten Jensen, Poul Suadicani, Hans Ole Hein, and Finn Gyntelberg. Elevated resting heart rate, physical fitness and all-cause mortality: a 16-year follow-up in the copenhagen male study. *Heart*, 99(12):882–887, 2013.

[32] Jungwook Jun. *Smoothing methods designed to minimize the impact of GPS random error on travel distance, speed, and*

acceleration profile estimates. PhD thesis, Department of Civil Engineering, Clemson University, 2005.

[33] Dean M Karantonis, Michael R Narayanan, Merryn Mathie, Nigel H Lovell, and Branko G Celler. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *Information Technology in Biomedicine, IEEE Transactions on*, 10(1):156–167, 2006.

[34] Serafeim Katsavos and Maria Anagnostouli. Biomarkers in multiple sclerosis: an up-to-date overview. *Multiple sclerosis international*, 2013, 2013.

[35] Melanie Knufinke. The measurement of arousal by the means of electrodermal activity during an actually performed balance beam routine and observational learning of the same routine. Master's thesis, University of Twente, 2012.

[36] Gary G Koch. Intraclass correlation coefficient. *Encyclopedia of statistical sciences*, 1982.

[37] Thea JM Kooiman, Manon L Dontje, Siska R Sprenger, Wim P Krijnen, Cees P van der Schans, and Martijn de Groot. Reliability and validity of ten consumer activity trackers. *BMC sports science, medicine and rehabilitation*, 7(1):1, 2015.

[38] Lauren B Krupp, Luis A Alvarez, Nicholas G LaRocca, and Labe C Scheinberg. Fatigue in multiple sclerosis. *Archives of neurology*, 45(4):435–437, 1988.

[39] LandAirSea. Do Weather Conditions Affect GPS Accuracy? `https://www.landairsea.com/gps-tracking-news/do-weather-conditions-affect-gps-accuracy`. Accessed: June 7, 2016.

[40] Jung-Min Lee. Validity of consumer-based physical activity monitors and calibration of smartphone for prediction of physical activity energy expenditure. Master's thesis, Iowa State University, 2013.

[41] Larry J Levy. The kalman filter: navigation's integration workhorse. *GPS World*, 8(9):65–71, 1997.

[42] Fred D Lublin, Stephen C Reingold, Jeffrey A Cohen, Gary R Cutter, Per Soelberg Sørensen, Alan J Thompson, Jerry S Wolinsky, Laura J Balcer, Brenda Banwell, Frederik Barkhof, et al. Defining the clinical course of multiple sclerosis the 2013 revisions. *Neurology*, 83(3):278–286, 2014.

[43] Jessica E Malberg and Ronald S Duman. Cell proliferation in adult hippocampus is decreased by inescapable stress: reversal by fluoxetine treatment. *Neuropsychopharmacology*, 28(9):1562–1571, 2003.

[44] RW Motl, PA Arnett, MM Smith, FH Barwick, B Ahlstrom, and EJ Stover. Worsening of symptoms is associated with lower physical activity levels in individuals with multiple sclerosis. *Multiple sclerosis*, 14(1):140–142, 2008.

[45] National Multiple Sclerosis Society. MS Fatigue. `http://www.nationalmssociety.org/Symptoms-Diagnosis/MS-Symptoms/Fatigue`. Accessed: June 7, 2016.

[46] Barun K Nayak, Avijit Hazra, et al. How to choose the right statistical test? *Indian journal of ophthalmology*, 59(2):85, 2011.

[47] Leandro Alberto Calazans Nogueira, Luciano Teixeira dos Santos, Pollyane Galinari Sabino, Regina Maria Papais Alvarenga, and Luiz Claudio Santos Thuler. Factors for lower walking speed in persons with multiple sclerosis. *Multiple sclerosis international*, 2013, 2013.

[48] Task Force of the European Society of Cardiology et al. Heart rate variability standards of measurement, physiological interpretation, and clinical use. *Eur Heart J*, 17:354–381, 1996.

[49] M Patel, SKL Lal, Diarmuid Kavanagh, and Peter Rossiter. Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert systems with Applications*, 38(6):7235–7242, 2011.

[50] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, 12:2825–2830, 2011.

[51] Ming-Zher Poh, Nicholas C Swenson, and Rosalind W Picard. A wearable sensor for

unobtrusive, long-term assessment of electrodermal activity. *Biomedical Engineering, IEEE Transactions on*, 57(5):1243–1252, 2010.

[52] William H Press and Saul A Teukolsky. Savitzky-golay smoothing filters. *Computers in Physics*, 4(6):669–672, 1990.

[53] Allison Rehagen, Stephanie Harris, and Lynn Cialdella-Kam. Short sleep duration: Effects on nutrient intake and mood state. *The FASEB Journal*, 29(1 Supplement):900–14, 2015.

[54] A Reunanen, J Karjalainen, P Ristola, M Heliövaara, P Knekt, and A Aromaa. Heart rate and mortality. *Journal of internal medicine*, 247(2):231–239, 2000.

[55] Tessel Runia. *Multiple Sclerosis-Predicting The Next Attack*. PhD thesis, Erasmus Universiteit Rotterdam, 2015.

[56] Sherwood W Samn and Layne P Perelli. Estimating aircrew fatigue: a technique with application to airlift operations. Technical report, DTIC Document, 1982.

[57] Hartmut Schulz. Rethinking sleep analysis. *J Clin Sleep Med*, 4(2):99–103, 2008.

[58] scikit-learn. Novelty and Outlier Detection. `http://scikit-learn.org/stable/modules/outlier_detection.html`. Accessed: June 7, 2016.

[59] SciPy. scipy.stats.kendalltau. `http://docs.scipy.org/doc/scipy-0.17.0/reference/generated/scipy.stats.kendalltau.html`. Accessed: June 11, 2016.

[60] Jessyca PR Selman, Anderson A de Camargo, Jenifer Santos, Fernanda C Lanza, and Simone Dal Corso. Reference equation for the 2-minute walk test in adults and the elderly. *Respiratory care*, 59(4):525–530, 2014.

[61] Annapuma Sharma, Amit Purwar, Young-Dong Lee, Young-Sook Lee, and Wan-Young Chung. Frequency based classification of activities using accelerometer data. In *Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on*, pages 150–153. IEEE, 2008.

[62] Natalie Smith, Sophie Miquel-Kergoat, and Sandrine Thuret. The impact of mastication on cognition: Evidence for intervention and the role of adult hippocampal neurogenesis. *Nutrition and Aging*, 3(Preprint):1–10, 2016.

[63] Doris Stangl and Sandrine Thuret. Impact of diet on adult hippocampal neurogenesis. *Genes & nutrition*, 4(4):271–282, 2009.

[64] Toshiyo Tamura, Yuka Maeda, Masaki Sekine, and Masaki Yoshida. Wearable photoplethysmographic sensors—past and present. *Electronics*, 3(2):282–302, 2014.

[65] Julian F Thayer, Fredrik Åhs, Mats Fredrikson, John J Sollers, and Tor D Wager. A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. *Neuroscience & Biobehavioral Reviews*, 36(2):747–756, 2012.

[66] Sandrine Thuret. You can grow new brain cells. here's how. `http://www.ted.com/talks/sandrine_thuret_you_can_grow_new_brain_cells_here_s_how`. Accessed: June 6, 2016.

[67] George W Torrance, David Feeny, and William Furlong. Visual analog scales do they have a role in the measurement of preferences for health states? *Medical Decision Making*, 21(4):329–334, 2001.

[68] Peter Totterdell, Shirley Reynolds, Brian Parkinson, and Rob B Briner. Associations of sleep with everyday mood, minor symptoms and social interaction experience. *Sleep: Journal of Sleep Research & Sleep Medicine*, 1994.

[69] William MK Trochim. Types of reliability, 2006.

[70] Ioannis Tsoukalas. The origin of rem sleep: A hypothesis. *Dreaming*, 22(4):253, 2012.

[71] Ingrid AF van der Mei, A-L Ponsonby, Terry Dwyer, Leigh Blizzard, Bruce V Taylor, Trevor Kilpatrick, Helmut Butzkueven, and AJ McMichael. Vitamin d levels in people with multiple sclerosis and community controls in tasmania, australia. *Journal of neurology*, 254(5):581–590, 2007.

[72] Eric W Weisstein. Great circle. `http://mathworld.wolfram.com/GreatCircle.html`, 2002. Accessed: June 24, 2016.

[73] Adam Wichniak, Aleksandra Wierzbicka, and Wojciech Jernajczyk. Sleep as a

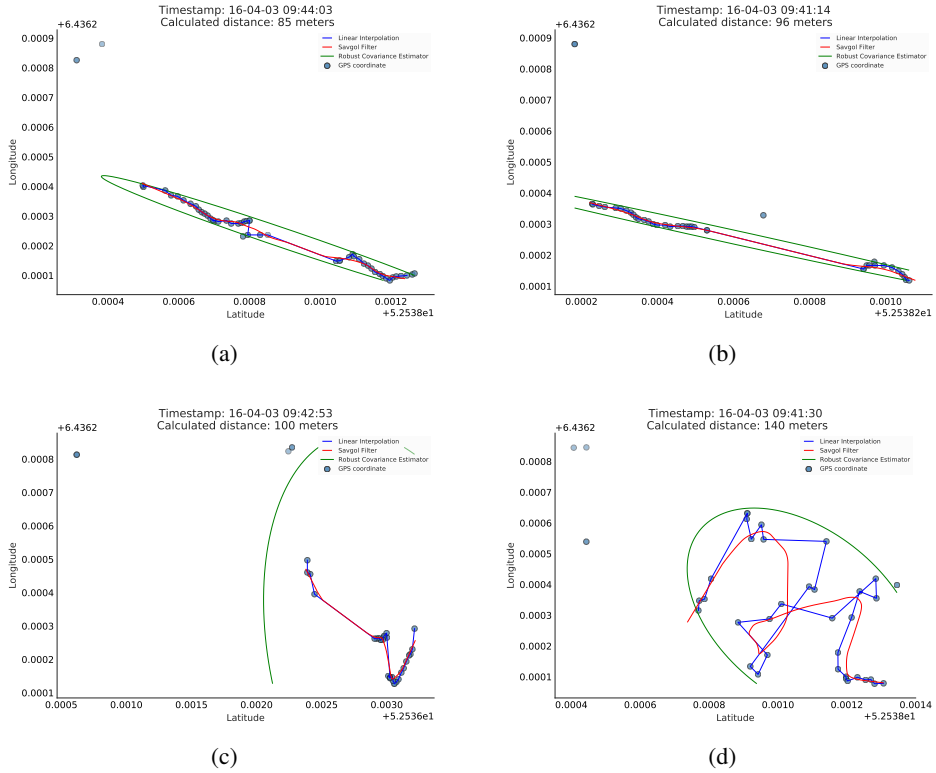
biomarker for depression. *International Review of Psychiatry*, 25(5):632–645, 2013.

[74] World Health Organization (WHO). Electromagnetic fields and public health: mobile phones. `http://www.who.int/mediacentre/factsheets/fs193/en/`. Accessed: June 6, 2016.

Appendix A

Appendix

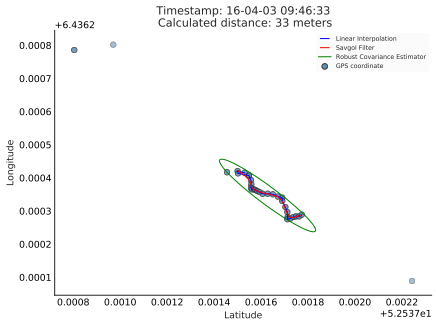
A.1 GPS Plots Straight Path



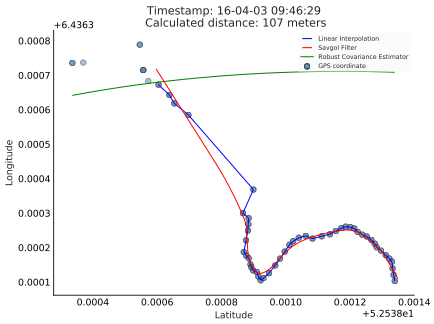
(a) (b)

(c) (d)

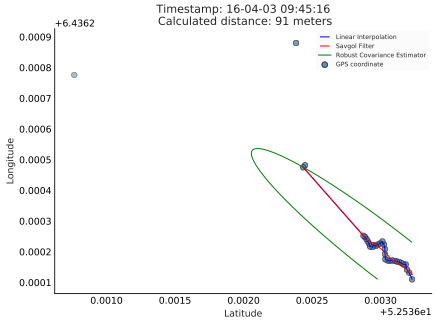
Figure A.1.1: Plots of the GPS coordinates of the walking test before and after smoothing. First, outlier detection is done using Elliptic Envelope. Afterwards, inliners are smoothed in two stages: 'Lineair Interpolation' is the first smoothing step and 'Savgol Filter' the last.



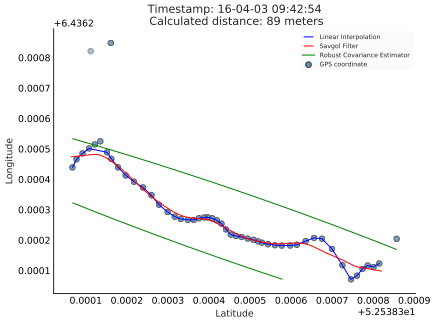
(e)



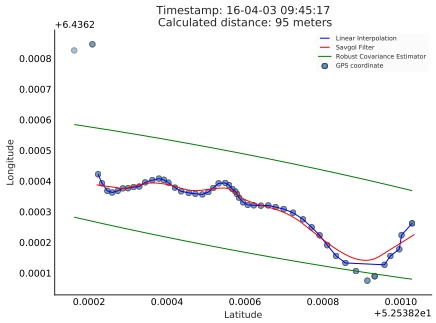
(f)



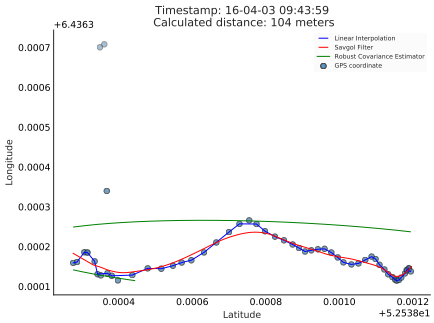
(g)



(h)



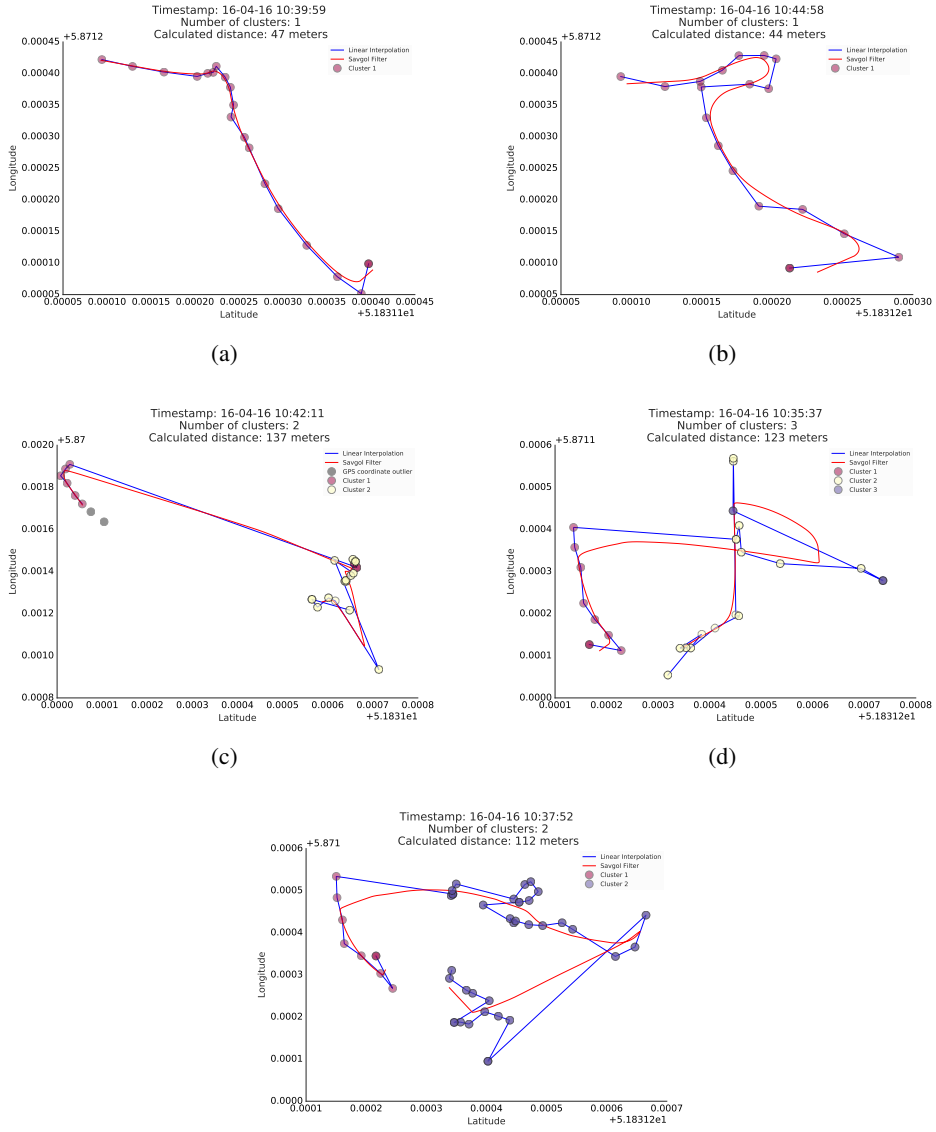
(i)



(j)

Figure A.1.1

A.2 GPS Plots Circular Path



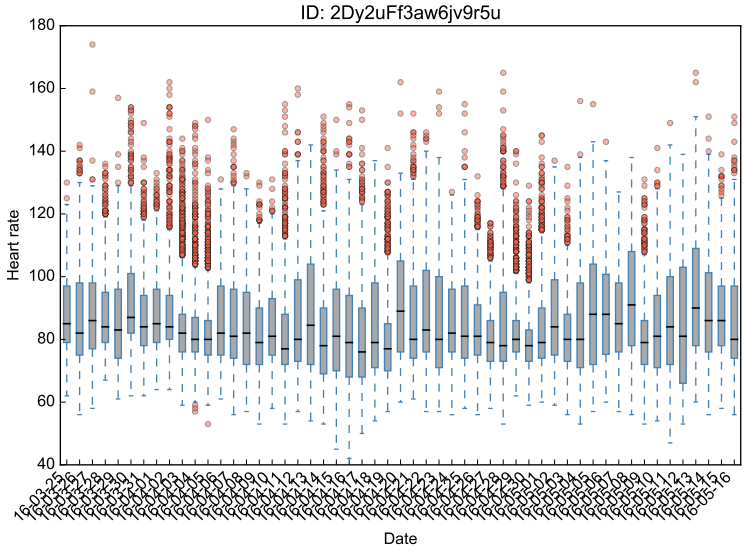
(a) (b)

(c) (d)

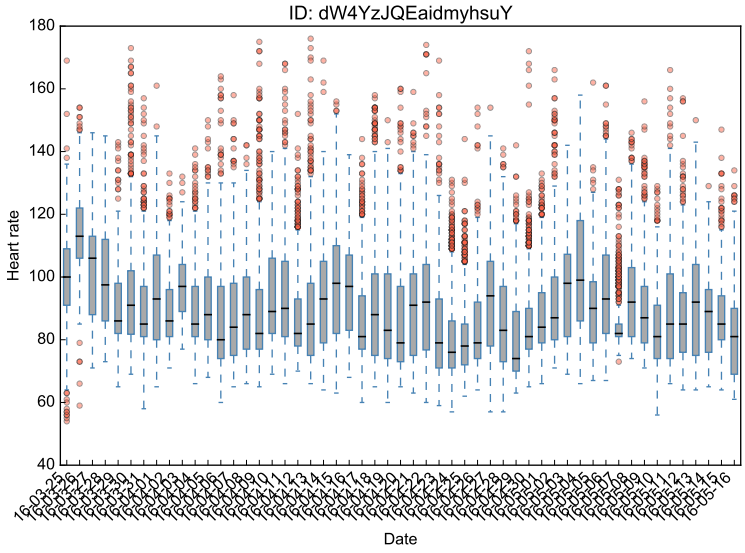
(e)

Figure A.2.1: Plots of the GPS coordinates of the walking test using a circular track. First, clustering is used to filter out the outliers. Afterwards, inliners are smoothed in two stages: ‘Lineair Interpolation’ is the first smoothing step and ‘Savgol Filter’ the last. Black points are points classified as noise.

A.3 Heart Rate Boxplots

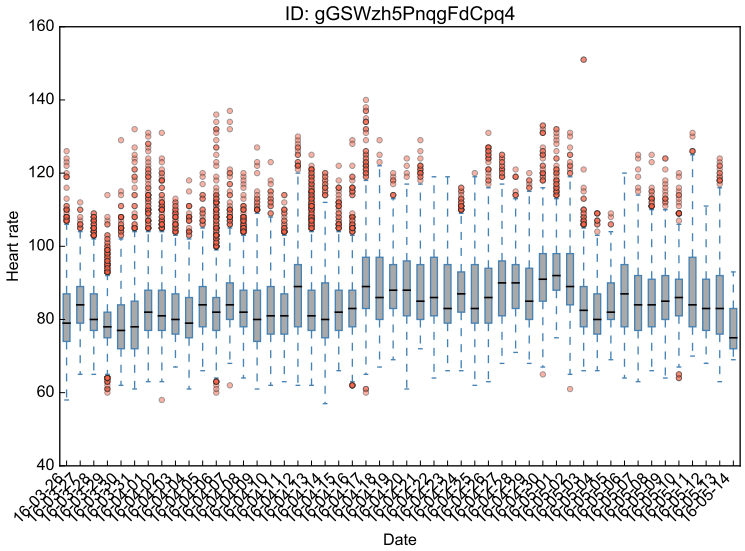


(a)

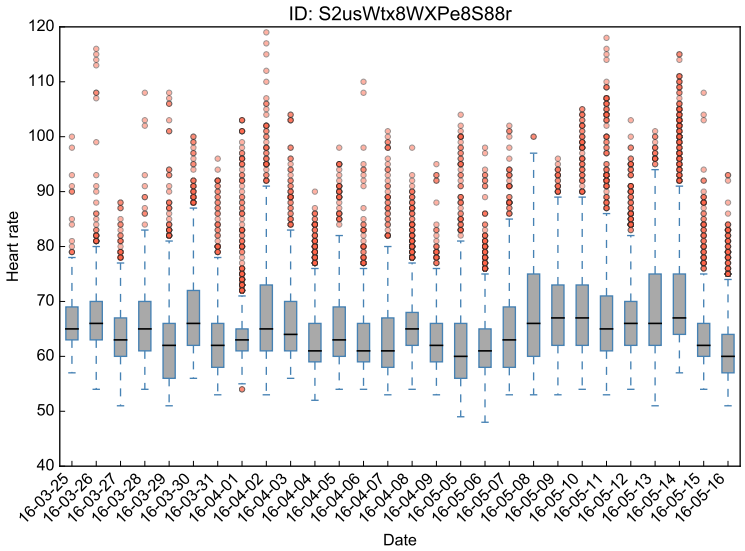


(b)

Figure A.3.1: Boxplots of the heart rates for each day.



(c)



(d)

Figure A.3.1

A.4 Elliptic Outlier Detection

```
def PlotEllipticOutliers(rawPositions, ID, timestmp):
    """
    This method is used for outlier detection on a straight track.
    Uses the EllipticEnvelope object from Scikit-learn
    :param rawPositions:    The raw GPS positions
    :param ID:              The ID of the user
    :param timestmp:        The timestamp
    """
    folder = "out/"
    plotDir = folder + "plots/Walking Test"

    lats = []
    longs = []
    timestamps = []
    for pos in rawPositions:
        lat = pos["latitude"]
        long = pos["longitude"]
        timestamp = pos["timestamp"]
        lats.append(lat)
        longs.append(long)
        timestamps.append(timestamp)

    # writeToFile(ID, "output", lats, longs)
    classifiers = {
        "Robust Covariance Estimator": EllipticEnvelope(contamination=0.08,
        ↪  support_fraction=0.8)
    }
    alpha = 0.5
    linewidth = 0.15

    classifierName = "Robust Covariance Estimator"
    classifier = classifiers[classifierName]
    X = zip(lats, longs)
    xx1, yy1 = np.meshgrid(np.linspace(min(lats), max(lats), 1000),
    ↪  np.linspace(min(longs), max(longs), 1000))
    classifier.fit(X)
    Z1 = classifier.decision_function(np.c_[xx1.ravel(), yy1.ravel()])
    Z1 = Z1.reshape(xx1.shape)
    CS = plt.contour(xx1, yy1, Z1, levels=[0], linewidths=1, colors="g")
    CS.collections[0].set_label(classifierName)
    plt.scatter(lats, longs, s=50, alpha=alpha, linewidth=linewidth,
    ↪  edgecolor=almost_black, color="steelblue",
                label="GPS coordinate")
    plt.autoscale(enable=True, axis="both")
    filteredPositions = []

    for lat, long, timestamp in zip(lats, longs, timestamps):
        z = classifier.decision_function(np.c_[lat, long])
        if z > 0:  # Check if coordinate (lat, long) is within the ellipse
            filteredPositions.append((lat, long, timestamp))

    y = zip(*filteredPositions)[0]  # the latitudes
```

```
50      x = zip(*filteredPositions)[1]  # the longitudes
51      t = zip(*filteredPositions)[2]  # the timestamps
52
53      x2, y2, newx2, newy2 = smooth(y, x, t)
54      plt.plot(y2, x2, label="Linear Interpolation")
55      totalDistance = calcDistanceWalked(newy2, newx2)
56
57      plt.plot(newy2, newx2, label="Savgol Filter", color="r")
58      plt.title("Timestamp: %s\n Calculated distance: %i meters" % (timestmp,
        ↪  totalDistance))
59      plt.xlabel("Latitude")
60      plt.ylabel("Longitude")
61      fancyPlot()
62      writeToPdf(ID, plotDir)
```

Listing A.1: Code used for detecting outliers in a straight path.

A.5 Clustering Outlier Detection

```
def clustering(lats, longs, timestamps, ID, timestmp, multiPDF=False):
    """
    Clusters the GPS coordinates using DBSCAN
    :param timestmp:                The timestamp
    :param ID:                      The ID
    :param timestamps:              The timestamps of the GPS coordinates
    :param lats:                    The latitudes
    :param longs:                   The longitudes
    :return:                        The rounded distance
    """
    folder = "out/"
    plotDir = folder + "plots/Walking Test Analysis"

    R = 6371  # Radius of the earth in km
    cartesianX = []
    cartesianY = []
    cartesianZ = []

    for lat, long in zip(lats, longs):
        # Convert to cartesian coordinates
        x = R * cos(lat) * cos(long)
        y = R * cos(lat) * sin(long)
        z = R * sin(lat)
        cartesianX.append(x)
        cartesianY.append(y)
        cartesianZ.append(z)

    combined = np.vstack((cartesianX, cartesianY, cartesianZ)).T
    (core_samples, labels) = dbscan(combined, eps=0.5)
    grouped = zip(labels, core_samples)
    nonGroupedPositions = []

    for (label, core_sample) in grouped:
        if label != -1:
            lat = lats[core_sample]
            long = longs[core_sample]
            stamp = timestamps[core_sample]
            nonGroupedPositions.append((lat, long, stamp))

    if len(nonGroupedPositions) > 0:
        y = zip(*nonGroupedPositions)[0]  # the latitudes
        x = zip(*nonGroupedPositions)[1]  # the longitudes
        t = zip(*nonGroupedPositions)[2]  # the timestamps
        x2, y2, newx2, newy2 = smooth(y, x, t)

        plt.plot(y2, x2, label="Linear Interpolation")
        plt.plot(newy2, newx2, label="Savgol Filter", color="r")
        distance = calcDistanceWalked(newy2, newx2)
        grouped = sorted(grouped, key=itemgetter(0))

        clusters = {}
        labels = []
```

```
53          for key, group in groupby(grouped, key=itemgetter(0)):
54              # group the clusters based on their label
55              labels.append(key)
56              clusters[key] = [el[1] for el in group]
57
58          noise = False
59          colors = plt.get_cmap("Spectral")(np.linspace(0, 1, len(clusters)))
60          for label in labels:
61              indices = clusters[label]
62              latitudes = []
63              longitudes = []
64              size = 10
65              alpha = 0.5
66              lineWidth = 0.15
67              for i in indices:
68                  latitudes.append(lats[i])
69                  longitudes.append(longs[i])
70              if label == -1:
71                  # outliers are identified with a label of -1
72                  plt.plot(latitudes, longitudes, "o",
                    ↪  markerfacecolor=almost_black,
                    ↪  markeredgecolor=almost_black,
73                          markersize=size, alpha=alpha, linewidth=lineWidth,
                            ↪  label="Outlier")
74                  noise = True
75              else:
76                  plt.plot(latitudes, longitudes, "o",
                    ↪  markerfacecolor=colors[label],
                    ↪  markeredgecolor=almost_black,
77                          markersize=size, alpha=alpha, linewidth=lineWidth,
                            ↪  label="Cluster %i" % (label + 1))
78
79          plt.title("Timestamp: %s\n Number of clusters: %i\n Calculated
            ↪  distance: %i meters" % (
80              timestmp, (len(clusters) - 1) if noise else len(clusters),
                ↪  round(distance)))
81          plt.xlabel("Latitude")
82          plt.ylabel("Longitude")
83          fancyPlot()
84          writeToPdf(ID, plotDir)
85          return True, distance
86      else:
87          # DBSCAN gave back an empty array, therefore we cannot perform any
            ↪  smoothing or distance calculation
88          return False, 0
```

Listing A.2: Code used for detecting outliers in a non-straight path.

A.6 Calculation of Walking Speed

```
def filterPositions(positions):
filteredPositions = []
threshold = 50
previousPos = None
for pos in positions:
  if pos["accuracy"] <= threshold and not pos["heading"] is None:
      if previousPos is None:
          previousPos = pos
          filteredPositions.append(pos)
      else:
          timeDifference = (pos["timestamp"] - previousPos["timestamp"]) /
          ↪  1000.0
          previousCoordinate = (previousPos["latitude"],
          ↪  previousPos["longitude"])
          currentCoordinate = (pos["latitude"], pos["longitude"])
          distanceWalked = great_circle(previousCoordinate,
          ↪  currentCoordinate).meters
          meterPerSecond = distanceWalked / timeDifference if timeDifference
          ↪  > 0 else 0
          if 0 < meterPerSecond <= 3:
              filteredPositions.append(pos)
              previousPos = pos
return filteredPositions
```

Listing A.3: Code used for calculating the walking speed between GPS coordinates.

A.7 Tables

A.7.1 Sleep

(a)

Date	Rating	Duration (h)
16-03-25	Goed	-
16-03-26	Goed	-
16-03-27	Goed	-
16-03-28	Goed	-
16-03-29	Goed	-
16-03-30 05:02:28	Goed	5.22
16-03-31	Goed	-
16-04-01	Goed	-
16-04-02	Goed	-
16-04-03 03:00:36	Goed	2.72
16-04-04 04:24:41	-	3.97
16-04-05	Goed	-
16-04-06 04:50:04	-	6.23
16-04-07	Goed	-
16-04-08	Goed	-
16-04-09	Goed	-
16-04-10	Goed	-
16-04-13 05:04:22	-	4.95
16-04-14 08:00:00	Gemiddeld	8.63
16-04-15	Goed	-
16-04-16 07:52:35	-	9.13
16-04-17 10:34:52	-	1.77
16-04-18	Goed	-
16-04-19 02:59:08	-	3.48
16-04-19 08:04:08	-	4.82
16-04-20	Goed	-
16-04-21 06:23:58	Gemiddeld	6.02
16-04-23 08:11:00	Gemiddeld	8.72
16-04-24	Goed	-
16-04-25 06:38:34	Goed	8.08
16-04-26 00:05:21	Goed	1.12
16-04-28 06:02:50	-	7.47
16-05-02 22:59:47	Gemiddeld	0.95
16-05-03	Goed	-
16-05-04	Goed	-
16-05-05 00:25:54	Gemiddeld	1.62
16-05-05 06:43:49	Gemiddeld	5.52
16-05-06 02:46:46	Slecht	2.9
16-05-06 06:29:46	Slecht	3.18
16-05-07	Goed	-
16-05-08 06:48:16	Gemiddeld	6.08
16-05-09	Goed	-
16-05-10 07:01:25	Goed	7.28
16-05-11	Goed	-
16-05-12	Goed	-
16-05-13	Goed	-
16-05-14 02:47:40	Slecht	3.68
16-05-14 04:24:02	Slecht	1.1
16-05-14 08:11:33	Slecht	1.55
16-05-15 07:48:47	Goed	7.83
16-05-16	Goed	-

(b)

Date	Rating	Duration (h)
16-03-25	Goed	-
16-03-28 06:35:14	-	6.25
16-03-29 05:47:27	-	9.42
16-03-30 06:11:04	Gemiddeld	9.85
16-03-31 05:39:29	Goed	4.25
16-03-31 08:58:53	Goed	1.9
16-04-01 18:09:28	-	0.93
16-04-02	Goed	-
16-04-03	Goed	-
16-04-04 05:33:37	-	9.95
16-04-05 05:58:38	Gemiddeld	9.82
16-04-06	Goed	-
16-04-07 06:22:30	-	7.87
16-04-08 05:46:38	Gemiddeld	8.72
16-04-09 05:53:57	-	7.67
16-04-10 20:19:41	-	1.33
16-04-12 21:37:48	-	0.97
16-04-13 05:39:47	Gemiddeld	7.17
16-04-14 00:21:22	-	3.87
16-04-14 06:33:22	-	5.2
16-04-14 19:42:39	-	1.18
16-04-14 22:25:39	-	1.6
16-04-15 05:51:39	Gemiddeld	7.05
16-04-16 00:45:39	-	0.83
16-04-18 06:31:09	-	9.88
16-04-20 06:17:09	-	10.3
16-04-23 06:19:13	Gemiddeld	5.83
16-04-23 10:16:13	Gemiddeld	1.63
16-04-24 06:42:06	Slecht	8.22
16-04-25	Goed	-
16-04-26	Goed	-
16-04-27 06:08:50	-	10.2
16-04-28 06:23:50	-	7.7
16-04-29 12:50:50	-	1.58
16-04-30 00:11:52	-	1.33
16-04-30 05:19:52	-	4.85
16-04-30 23:01:29	-	0.7
16-05-01 06:20:29	-	6.87
16-05-03 05:36:42	-	6.75
16-05-06 00:27:42	Slecht	1.15
16-05-06 08:45:42	Slecht	8.03
16-05-06 10:17:42	Slecht	0.67
16-05-07	Goed	-
16-05-08 06:45:43	Gemiddeld	6.5
16-05-09 06:09:43	Gemiddeld	6.93
16-05-10 06:30:53	-	9.43
16-05-10 09:12:53	-	1.85
16-05-11 06:33:22	-	5.98
16-05-11 14:28:46	-	1.18
16-05-12 08:05:03	-	9.53
16-05-14 06:57:54	-	7.52
16-05-15	Goed	-
16-05-16 10:37:06	-	0.53

Table A.1: Measurements of the sleep duration of our participants.

(c)

Date	Rating	Duration (h)
16-03-25 00:00:00	-	0.0
16-03-26 00:00:00	Goed	5.68
16-03-27 00:00:00	-	6.75
16-03-28 00:00:00	-	7.88
16-03-29 00:00:00	Gemiddeld	5.97
16-03-30 00:00:00	Gemiddeld	3.6
16-03-31 00:00:00	Goed	5.12
16-04-01 00:00:00	-	6.72
16-04-02 00:00:00	Goed	5.38
16-04-03 00:00:00	Goed	6.67
16-04-04 00:00:00	Goed	6.43
16-04-05 00:00:00	Goed	4.08
16-04-06 00:00:00	Gemiddeld	4.43
16-04-07 00:00:00	Goed	7.22
16-04-08 00:00:00	Goed	6.45
16-04-09 00:00:00	Goed	4.38
16-04-10 00:00:00	Goed	7.7
16-04-11 00:00:00	-	5.72
16-04-12 00:00:00	-	5.78
16-04-13 00:00:00	Gemiddeld	7.77
16-04-14 00:00:00	-	6.9
16-04-15 00:00:00	-	0.0
16-04-16 00:00:00	Goed	5.25
16-04-17 00:00:00	Goed	4.33
16-04-18 00:00:00	Goed	5.03
16-04-19 00:00:00	Gemiddeld	2.77
16-04-20 00:00:00	Goed	6.55
16-04-21 00:00:00	Goed	6.42
16-04-22 00:00:00	Gemiddeld	5.42
16-04-23 00:00:00	Gemiddeld	7.47
16-04-24 00:00:00	Goed	7.5
16-04-25 00:00:00	Goed	6.18
16-04-26 00:00:00	Gemiddeld	8.17
16-04-27 00:00:00	Goed	7.65
16-04-28 00:00:00	Goed	7.53
16-04-29 00:00:00	Goed	5.23
16-04-30 00:00:00	-	8.42
16-05-01 00:00:00	Goed	8.18
16-05-02 00:00:00	Goed	9.43
16-05-03 00:00:00	Goed	8.33
16-05-04 00:00:00	Goed	5.65
16-05-05 00:00:00	Goed	7.08
16-05-06 00:00:00	Goed	8.62
16-05-07 00:00:00	Gemiddeld	1.38
16-05-08 00:00:00	-	0.0
16-05-09 00:00:00	-	0.0
16-05-10	Goed	-
16-05-11	Goed	-
16-05-12	Goed	-
16-05-13	Goed	-
16-05-14	Goed	-
16-05-15	Goed	-
16-05-16	Goed	-

(d)

Date	Rating	Duration (h)
16-03-25 00:00:00	-	0.0
16-03-26 00:00:00	Slecht	6.02
16-03-27 00:00:00	Gemiddeld	7.92
16-03-28 00:00:00	Gemiddeld	8.12
16-03-29 00:00:00	Gemiddeld	5.45
16-03-30 00:00:00	Slecht	9.67
16-03-31 00:00:00	Gemiddeld	9.22
16-04-01 00:00:00	Slecht	8.78
16-04-02 00:00:00	Slecht	7.77
16-04-03 00:00:00	Gemiddeld	9.13
16-04-04 00:00:00	Gemiddeld	7.2
16-04-05 00:00:00	Gemiddeld	7.6
16-04-06 00:00:00	-	9.07
16-04-07 00:00:00	Slecht	9.63
16-04-08 00:00:00	Slecht	11.68
16-04-09 00:00:00	Slecht	8.27
16-04-10 00:00:00	Gemiddeld	10.25
16-04-11 00:00:00	-	9.12
16-04-12 00:00:00	-	8.97
16-04-13 00:00:00	-	6.58
16-04-14 00:00:00	Gemiddeld	5.67
16-04-15 00:00:00	Gemiddeld	8.2
16-04-16 00:00:00	-	7.78
16-04-17 00:00:00	-	6.45
16-04-18 00:00:00	Gemiddeld	4.6
16-04-19 00:00:00	-	8.22
16-04-20 00:00:00	Goed	8.72
16-04-21 00:00:00	Gemiddeld	4.88
16-04-22 00:00:00	Gemiddeld	6.85
16-04-23 00:00:00	Gemiddeld	8.88
16-04-24 00:00:00	Gemiddeld	8.95
16-04-25 00:00:00	Gemiddeld	9.88
16-04-26 00:00:00	Gemiddeld	8.48
16-04-27 00:00:00	-	9.18
16-04-28 00:00:00	-	7.85
16-04-29 00:00:00	-	8.2
16-04-30 00:00:00	-	10.0
16-05-01 00:00:00	-	10.23
16-05-02 00:00:00	-	8.42
16-05-03 00:00:00	Slecht	8.13
16-05-04 00:00:00	Gemiddeld	7.52
16-05-05 00:00:00	Slecht	8.5
16-05-06 00:00:00	Slecht	7.2
16-05-07 00:00:00	Slecht	8.48
16-05-08 00:00:00	Gemiddeld	8.92
16-05-09 00:00:00	Slecht	8.98
16-05-10	Gemiddeld	-
16-05-11	Gemiddeld	-
16-05-12	Gemiddeld	-
16-05-13	Gemiddeld	-
16-05-14	Gemiddeld	-
16-05-15	Gemiddeld	-
16-05-16	Gemiddeld	-

Table A.1

A.7.2 Heart Rate

(a)

Date	Rating	$\tilde{x}$	μ	σ
16-03-25	Goed	85.0	88.0	12.84
16-03-26	Slecht	82.0	87.0	16.12
16-03-27	Gemiddeld	86.0	88.0	15.02
16-03-28	Gemiddeld	84.0	88.0	13.02
16-03-29	Gemiddeld	83.0	86.0	15.32
16-03-30	Goed	87.0	92.0	14.96
16-03-31	Goed	84.0	87.0	13.15
16-04-01	Goed	85.0	88.0	13.21
16-04-02	Goed	84.0	89.0	14.45
16-04-03	Goed	82.0	85.0	14.74
16-04-05	Goed	80.0	82.0	12.68
16-04-07	Goed	81.0	86.0	15.81
16-04-08	Gemiddeld	82.0	85.0	16.4
16-04-09	Gemiddeld	79.0	82.0	13.9
16-04-10	Goed	81.0	84.0	13.51
16-04-14	Gemiddeld	78.0	82.0	18.07
16-04-15	Goed	81.0	84.0	17.17
16-04-18	Goed	79.0	84.0	17.69
16-04-20	Gemiddeld	89.0	91.0	16.48
16-04-21	Gemiddeld	80.0	87.0	17.84
16-04-23	Gemiddeld	80.0	85.0	17.64
16-04-24	Gemiddeld	82.0	86.0	13.59
16-04-25	Goed	81.0	86.0	16.18
16-04-26	Goed	81.0	84.0	13.82
16-05-02	Gemiddeld	84.0	87.0	15.58
16-05-03	Slecht	80.0	81.0	12.0
16-05-04	Gemiddeld	80.0	85.0	17.69
16-05-05	Gemiddeld	88.0	89.0	18.38
16-05-06	Slecht	88.0	89.0	15.63
16-05-07	Slecht	85.0	87.0	13.71
16-05-08	Gemiddeld	91.0	93.0	17.18
16-05-09	Gemiddeld	79.0	80.0	12.45
16-05-10	Goed	81.0	83.0	15.81
16-05-11	Gemiddeld	84.0	87.0	18.22
16-05-12	Slecht	81.0	86.0	21.18
16-05-13	Gemiddeld	90.0	93.0	18.24
16-05-14	Slecht	86.0	89.0	16.41
16-05-15	Goed	86.0	88.0	14.05
16-05-16	Gemiddeld	80.0	86.0	16.76

(b)

Date	Rating	$\tilde{x}$	μ	σ
16-03-25	Goed	100.0	99.0	15.44
16-03-30	Gemiddeld	91.0	95.0	16.47
16-03-31	Goed	85.0	90.0	12.85
16-04-02	Gemiddeld	86.0	90.0	11.13
16-04-03	Gemiddeld	97.0	97.0	10.14
16-04-05	Gemiddeld	88.0	91.0	13.29
16-04-06	Gemiddeld	80.0	86.0	16.38
16-04-08	Gemiddeld	88.0	90.0	14.42
16-04-13	Gemiddeld	85.0	89.0	17.33
16-04-15	Gemiddeld	98.0	98.0	16.36
16-04-23	Gemiddeld	79.0	83.0	15.98
16-04-24	Slecht	76.0	80.0	13.83
16-04-25	Gemiddeld	78.0	81.0	12.43
16-04-26	Gemiddeld	79.0	84.0	13.03
16-05-06	Slecht	93.0	96.0	17.28
16-05-07	Gemiddeld	82.0	86.0	8.9
16-05-08	Gemiddeld	92.0	94.0	13.98
16-05-09	Gemiddeld	87.0	89.0	12.8
16-05-15	Gemiddeld	85.0	88.0	12.25

Table A.2: Measurements of the heart rates of our participants. For each day the median ($\tilde{x}$) and the mean (μ) along with the standard deviation (σ) are shown.

(c)

Date	Rating	$\tilde{x}$	μ	σ
16-03-26	Goed	79.0	81.0	10.62
16-03-29	Gemiddeld	78.0	79.0	6.88
16-03-30	Gemiddeld	77.0	79.0	9.56
16-03-31	Goed	78.0	80.0	10.2
16-04-02	Goed	81.0	83.0	9.85
16-04-03	Goed	80.0	82.0	8.23
16-04-04	Goed	79.0	81.0	8.84
16-04-05	Goed	84.0	84.0	8.06
16-04-06	Gemiddeld	82.0	83.0	9.83
16-04-07	Goed	84.0	86.0	8.95
16-04-08	Goed	82.0	84.0	9.12
16-04-09	Goed	80.0	82.0	10.76
16-04-10	Goed	81.0	83.0	9.25
16-04-13	Gemiddeld	81.0	85.0	11.08
16-04-15	Goed	82.0	84.0	9.29
16-04-16	Goed	83.0	84.0	8.73
16-04-17	Goed	89.0	91.0	11.39
16-04-18	Goed	86.0	89.0	11.06
16-04-19	Gemiddeld	88.0	89.0	9.74
16-04-20	Goed	88.0	89.0	9.53
16-04-21	Goed	85.0	88.0	10.41
16-04-22	Gemiddeld	86.0	89.0	10.32
16-04-23	Gemiddeld	83.0	87.0	10.15
16-04-24	Goed	87.0	88.0	8.78
16-04-25	Goed	83.0	86.0	10.39
16-04-26	Gemiddeld	86.0	88.0	11.49
16-04-27	Goed	90.0	90.0	10.42
16-04-28	Goed	90.0	90.0	8.34
16-04-29	Goed	85.0	88.0	9.66
16-05-01	Goed	92.0	94.0	8.12
16-05-02	Goed	89.0	91.0	11.44
16-05-03	Goed	82.5	84.0	8.6
16-05-04	Goed	80.0	82.0	8.01
16-05-05	Goed	82.0	85.0	7.48
16-05-06	Goed	87.0	87.0	10.57
16-05-07	Gemiddeld	84.0	84.0	10.24
16-05-08	Slecht	84.0	85.0	9.88
16-05-09	Gemiddeld	85.0	86.0	8.34
16-05-10	Goed	86.0	86.0	7.55
16-05-11	Slecht	84.0	88.0	12.61
16-05-12	Gemiddeld	83.0	85.0	9.15
16-05-13	Gemiddeld	83.0	84.0	10.79
16-05-14	Gemiddeld	75.0	78.0	6.61

(d)

Date	Rating	$\tilde{x}$	μ	σ
16-03-25	Gemiddeld	65.0	67.0	5.26
16-03-26	Slecht	66.0	67.0	6.4
16-03-27	Gemiddeld	63.0	64.0	5.49
16-03-28	Gemiddeld	65.0	66.0	6.21
16-03-29	Gemiddeld	62.0	63.0	8.35
16-03-30	Slecht	66.0	68.0	7.88
16-03-31	Gemiddeld	62.0	63.0	6.31
16-04-01	Slecht	63.0	64.0	6.44
16-04-02	Slecht	65.0	68.0	9.96
16-04-03	Gemiddeld	64.0	66.0	7.53
16-04-04	Gemiddeld	61.0	63.0	6.48
16-04-05	Gemiddeld	63.0	65.0	6.98
16-04-07	Slecht	61.0	63.0	7.23
16-04-08	Slecht	65.0	66.0	5.81
16-04-09	Slecht	62.0	63.0	5.66
16-05-05	Slecht	60.0	62.0	8.55
16-05-06	Slecht	61.0	62.0	6.68
16-05-07	Slecht	63.0	64.0	8.19
16-05-08	Gemiddeld	66.0	68.0	9.26
16-05-09	Slecht	67.0	68.0	7.68
16-05-10	Gemiddeld	67.0	69.0	8.96
16-05-11	Gemiddeld	65.0	68.0	10.42
16-05-12	Gemiddeld	66.0	67.0	7.13
16-05-13	Gemiddeld	66.0	69.0	9.15
16-05-14	Gemiddeld	67.0	71.0	10.69
16-05-15	Slecht	62.0	63.0	6.0
16-05-16	Slecht	60.0	61.0	6.53

Table A.2

A.7.3 Walking Test

(a)

Date	Distance Walked (m)	Rating
16-03-25	-	Goed
16-03-26	-	Slecht
16-03-27	-	Gemiddeld
16-03-28	-	Gemiddeld
16-03-29	-	Gemiddeld
16-03-30	174.0	Goed
16-03-31	212.0	Goed
16-04-01	-	Goed
16-04-02	-	Goed
16-04-03	171.0	Goed
16-04-05	-	Goed
16-04-07	-	Goed
16-04-08	-	Gemiddeld
16-04-09	174.0	Gemiddeld
16-04-10	-	Goed
16-04-14	-	Gemiddeld
16-04-15	-	Goed
16-04-18	-	Goed
16-04-20	-	Gemiddeld
16-04-21	-	Gemiddeld
16-04-23	-	Gemiddeld
16-04-24	-	Gemiddeld
16-04-25	-	Goed
16-04-26	-	Goed
16-05-02	-	Gemiddeld
16-05-03	-	Slecht
16-05-04	-	Gemiddeld
16-05-05	-	Gemiddeld
16-05-06	-	Slecht
16-05-07	-	Slecht
16-05-08	-	Gemiddeld
16-05-09	-	Gemiddeld
16-05-10	-	Goed
16-05-11	-	Gemiddeld
16-05-12	140.0	Slecht
16-05-13	-	Gemiddeld
16-05-14	-	Slecht
16-05-15	-	Goed
16-05-16	-	Gemiddeld

(b)

Date	Distance Walked (m)	Rating
16-03-25	-	Goed
16-03-30	-	Gemiddeld
16-03-31	186.0	Goed
16-04-02	*	Gemiddeld
16-04-03	-	Gemiddeld
16-04-05	234.0	Gemiddeld
16-04-06	-	Gemiddeld
16-04-08	-	Gemiddeld
16-04-13	-	Gemiddeld
16-04-15	-	Gemiddeld
16-04-23	-	Gemiddeld
16-04-24	-	Slecht
16-04-25	-	Gemiddeld
16-04-26	-	Gemiddeld
16-05-06	-	Slecht
16-05-07	-	Gemiddeld
16-05-08	-	Gemiddeld
16-05-09	-	Gemiddeld
16-05-15	-	Gemiddeld

Table A.3: Results of the walking tests done by our participants. The distance walked is measured in meters. A value of '-' in the column Distance Walked means that the participant did not perform a walking test that day. A value of '*' means that the quality of the GPS track was insufficient to perform DBSCAN clustering on the corresponding data.

(c)

Date	Distance Walked (m)	Rating
16-03-25	-	Goed
16-03-26	-	Goed
16-03-29	20.0	Gemiddeld
16-03-30	-	Gemiddeld
16-03-31	-	Goed
16-04-02	-	Goed
16-04-03	-	Goed
16-04-04	-	Goed
16-04-05	-	Goed
16-04-06	-	Gemiddeld
16-04-07	-	Goed
16-04-08	-	Goed
16-04-09	-	Goed
16-04-10	-	Goed
16-04-13	-	Gemiddeld
16-04-15	-	Goed
16-04-16	-	Goed
16-04-17	-	Goed
16-04-18	-	Goed
16-04-19	-	Gemiddeld
16-04-20	-	Goed
16-04-21	-	Goed
16-04-22	-	Gemiddeld
16-04-23	-	Gemiddeld
16-04-24	-	Goed
16-04-25	-	Goed
16-04-26	-	Gemiddeld
16-04-27	-	Goed
16-04-28	-	Goed
16-04-29	-	Goed
16-05-01	-	Goed
16-05-02	-	Goed
16-05-03	-	Goed
16-05-04	-	Goed
16-05-05	-	Goed
16-05-06	-	Goed
16-05-07	-	Gemiddeld
16-05-08	-	Slecht
16-05-09	-	Gemiddeld
16-05-10	-	Goed
16-05-11	-	Slecht
16-05-12	15.0	Gemiddeld
16-05-13	-	Gemiddeld
16-05-14	-	Gemiddeld
16-05-15	-	Goed
16-05-16	-	Gemiddeld

(d)

Date	Distance Walked (m)	Rating
16-03-25	-	Goed
16-03-26	-	Gemiddeld
16-03-27	-	Gemiddeld
16-03-28	-	Gemiddeld
16-03-30	76.0	Gemiddeld
16-03-31	-	Goed
16-04-02	-	Gemiddeld
16-04-03	-	Gemiddeld
16-04-04	-	Gemiddeld
16-04-05	-	Gemiddeld
16-04-06	77.0	Gemiddeld
16-04-07	53.0	Gemiddeld
16-04-08	75.0	Gemiddeld
16-04-09	49.0	Goed
16-04-10	56.0	Gemiddeld
16-04-11	-	Goed
16-04-13	52.0	Goed
16-04-14	-	Gemiddeld
16-04-15	-	Goed
16-04-16	-	Gemiddeld
16-04-17	-	Gemiddeld
16-04-18	57.0	Goed
16-04-19	-	Goed
16-04-20	-	Gemiddeld
16-04-21	-	Gemiddeld
16-04-22	-	Gemiddeld
16-04-23	-	Gemiddeld
16-04-24	-	Gemiddeld
16-04-25	-	Goed
16-04-26	-	Goed
16-05-02	-	Gemiddeld
16-05-03	-	Gemiddeld
16-05-09	91.0	Goed
16-05-10	-	Goed
16-05-11	-	Gemiddeld
16-05-12	-	Gemiddeld
16-05-13	66.0	Goed
16-05-14	-	Goed
16-05-15	60.0	Gemiddeld
16-05-16	100.0	Gemiddeld

Table A.3

(e)

Date	Distance Walked (m)	Rating
16-03-25	-	Gemiddeld
16-03-26	-	Slecht
16-03-27	-	Gemiddeld
16-03-28	-	Gemiddeld
16-03-29	-	Gemiddeld
16-03-30	-	Slecht
16-03-31	-	Gemiddeld
16-04-01	-	Slecht
16-04-02	-	Slecht
16-04-03	-	Gemiddeld
16-04-04	-	Gemiddeld
16-04-05	-	Gemiddeld
16-04-07	-	Slecht
16-04-08	-	Slecht
16-04-09	-	Slecht
16-04-10	-	Gemiddeld
16-04-14	-	Gemiddeld
16-04-15	-	Gemiddeld
16-04-18	-	Gemiddeld
16-04-20	-	Goed
16-04-21	-	Gemiddeld
16-04-22	-	Gemiddeld
16-04-23	-	Gemiddeld
16-04-24	-	Gemiddeld
16-04-25	-	Gemiddeld
16-04-26	-	Gemiddeld
16-05-03	-	Slecht
16-05-04	-	Gemiddeld
16-05-05	-	Slecht
16-05-06	-	Slecht
16-05-07	-	Slecht
16-05-08	-	Gemiddeld
16-05-09	-	Slecht
16-05-10	-	Gemiddeld
16-05-11	-	Gemiddeld
16-05-12	-	Gemiddeld
16-05-13	-	Gemiddeld
16-05-14	-	Gemiddeld
16-05-15	-	Slecht
16-05-16	-	Slecht

Table A.3