THE EFFECT OF PHYSICAL ACTIVITY ON HEALTH-RELATED QUALITY OF LIFE. A CROSS-SECTIONAL INVESTIGATION OF A SAMPLE FROM REGION CENTRAL JUTLAND

BY

LOUISE NÆSER CHRISTENSEN

Master's Thesis in

Medical Market Access

Medicine with Industrial Specialisation

September the 1st 2010 to June the 1st 2011

How can one set of individuals who happen to be researchers know what is important to the quality of life in another set of individuals?

Aalborg University
Department of Health Science and Technology
Fredrik Bajers Vej 7
9220 Aalborg Øst
http://medis.hst.aau.dk/

Title:
The Effect of Physical Activity on Health-Related Quality of Life.
A Cross-Sectional Investigation of a Sample from Region Central Jutland

Author:
Louise Næser Christensen

Project Group:
11gr1022

Project Period:
September the 1st 2010 to June the 1st 2011

Supervisors:
Lars H. Ehlers
Morten B. Jensen

Copies:
6

Report pages:
46

Total Pages:
81

Abstract:

The way of living has changed, particularly over the last three decades, where the necessity of being physically active has declined. This has resulted in an increase in inactivity, which has caused a corresponding increase in different life-style-related diseases.
To take precautionary measures against these diseases and to improve health-related quality of life (HRQOL) among the population the Danish National Board of Health recommends that persons above the age of 18 are active for at least 30 minutes every day at a moderate level.
The purpose of this project is to investigate if these recommendations cause a better health in a sample from Region Central Jutland. HRQOL is measured using physical component summary (PCS) and mental component summary (MCS) scores from the Short-Form Health Survey SF-12. PCS and MCS have been calculated beforehand, and are validated using factor analysis. The effect of physical activity on SF-12 is measured using multiple regression analysis.
The factor analysis shows that the calculated PCS and MCS are valid measures to use in further analysis, because they are highly correlated (98.7 and 99.7 percent respectively) with the actual PCS and MCS from the sample. The multiple regression shows that physical activity has a positive effect both on PCS and MCS, but this variable is not the highest predictor in achieving a high HRQOL. Age and stress have a higher influence on PCS and MCS respectively.
With this it can be concluded that physical activity has a positive effect on HRQOL, but stress is an even higher predictor of how you feel mentally.

Preface

This report constitutes the master's thesis for Louise Næser Christensen, group 10gr922 at the Medical Market Access specialisation at Department of Health Science and Technology. The project has begun September the 1st 2010 and is handed in June the 1st 2011, and has thereby spanned 9th and 10th semester at the education Medicine with Industrial Specialisation.

The project consists of four chapters. Chapter I contains the introduction that leads to the problem formulation and a project demarcation. After this, background knowledge creates a basis for the methods, which are presented in Chapter II together with the results. In Chapter III the results are discussed, and the conclusion of the project completes the first three chapters. In the end, Chapter IV contains the appendices that are referred to in the previous chapters.

The references are referred to using the Vancouver method, where parenthesis surrounds a number, which represents the respective reference e.g. (1) refers to the first reference in the bibliography. The bibliography can be found after Chapter III and before Chapter IV. The quotation on the front page is from source (78).

During the project different abbreviations are used, and the first time the abbreviated word is used it is presented with the respective abbreviation. When the shortening SF-12 is used it means that both the physical and mental component scores are referred to, otherwise they are used separately.

In this project a sample from Region Central Jutland is used, and when referring to this particular sample, the word "sample" is used, but when referring to all persons in Region Central Jutland "population" is used.

A special thank is given to Region Central Jutland and the persons behind the questionnaire survey "How are you?" that was sent out in 2010. These persons have provided a large sample and the data that are used in this project, without them, this project would not be possible.

Aalborg, June the 1st, 2011

Louise Næser Christensen

CONTENTS

Chapter IV: Appendices

CHAPTER I
INTRODUCTION & BACKGROUND

1. THE ACTIVITY LEVEL HAS CHANGED

Many years ago it was a survival criterion to be active, because houses were constructed by hand, and by that muscles were used. Furthermore it was important to have a good physical fitness to hunt down venison, since a poor physical fitness would mean no food. To provide aliment as roots and berries it was necessary to walk, sometimes many kilometres every day. Later it was also necessary to have a good physical fitness to provide food, e.g. farming required a lot of hard physical work.

This changed! During the industrialisation in the 19th century (1) different machines and new technology were developed to improve the production and ease the workload – everything became easier.

Since activity is no longer a survival criterion, the level of physical activity has declined and with that the level of inactivity has increased (2). By developing new technologies (as seen in picture 1.0) it is expected that people will feel better, but different studies have shown that inactivity has a negative influence on health-related quality of life (HRQOL) due to different lifestyle-related diseases such as e.g. cancer (3), diabetes mellitus type-2, and musculoskeletal disorders (2,4) just to mention a few. By being active the physical build gets stronger, since the muscles get larger and bones get stronger due to a release of growth hormones released during exercise (5). Also lifestyle-related diseases have been found to decrease in persons that are active during the day (3,6). Neurotransmitters as endorphins (7), dopamine (8) and serotonin (9) are also released during exercise. These neurotransmitters have a positive effect on mood and behaviour (5,7), and it has been found that inactive persons are more stressed and tense than active persons (4,10).

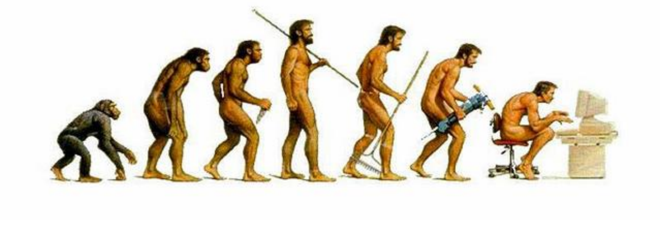

Figure 1.0: The evolution of mankind. Picture from (11)

Based on the preventive effects of exercise, and the decline in HRQOL in persons that are inactive, the Danish National Board of Health has made recommendations about physical activity in the aim to increase the focus among the Danish population of the importance of being physically active in daily life. The primary recommendations are different depending on age. Persons under the age of 18 are recommended to be active at least for 60 minutes every day at a moderate activity. Persons above the age of 18 are recommended to be active for at least 30 minutes every day at a moderate activity level. The "every day" recommendation is made on the basis of the blood lipid levels, which are affected by physical activity (12,13), and by being physically active it is possible to decrease the amount of LDL-cholesterol[1] in the arteries causing a decrease in the risk of developing cardiovascular diseases. The amount of physical activity can be summed, so a person above the age of 18 who has 15 minutes of bicycling to work and 15 minutes home again, will fulfil the primary recommendation. By separating the physical activity into smaller intervals it is possible to incorporate the activity into the daily life by e.g. bicycling to work, to the grocery, or walking more times a day, but it has not yet been determined

[1]LDL-cholesterol: This cholesterol is a big predictor of the risk of cardiovascular diseases (66)

how short the intervals can be (13). The secondary part of the recommendation to all age groups is to be active for at least 20 minutes twice a week to improve the physical fitness, muscle and bone strength (4).

It can be difficult to determine how large an effect exercise has on HRQOL in a general population. Therefore the purpose of this project is to analyse how physical activity affects HRQOL in a randomly selected sample from Region Central Jutland. To do this a HRQOL measure shall be found, and before analysing the effect of physical activity, the HRQOL measure has to be validated with the sample used in this project. The next step is to analyse the effect of physical activity on HRQOL, in a model, which also includes other variables. The purpose will be fulfilled when answering the problem formulation:

Which measure can be used to analyse HRQOL in a sample from Region Central Jutland, and is the HRQOL measure valid? How is your HRQOL affected if you follow the recommendations from the Danish National Board of health?

Hypothesis: The HRQOL measure that is used in this project will represent the sample from Region Central Jutland, and is thereby valid. An increased level of activity causes a better HRQOL.

1.1 Project Demarcation

Physical activity is defined as any activity that increases the energy consumption (14), and the main focus of this project is the recommendations from the Danish National Board of Health, which states that it is healthiest to be active seven days a week for at least 30 minutes for persons above the age of 18. Different studies (15-17) have already shown that following of the recommendations cause fewer unhealthy days and better health. In this study the effect of physical activity is measured on the randomly selected sample from Region Central Jutland, which has answered a questionnaire. The questionnaire can be seen elsewhere (18).

The questionnaire from Region Central Jutland was sent to 52.400 randomly chosen persons over 16 years of age. Due to the random sampling, the answers received from the population are therefore believed to represent both healthy and unhealthy persons, but it is well-known that the prevalence of unhealthy persons is lower than the prevalence of healthy persons that answer questionnaires (19). The results found in this project are therefore believed to represent the persons in the region who have a generally good health. With this the general population is expected to generally score a lower HRQOL than the sample used in this questionnaire. This, however, is not expected to be a problem, since the analyses are expected to show a tendency from physical activity on HRQOL, rather than descriptive statistics of the general population. Since all responses from the questionnaire are included, there is no distinction between the healthy part versus an unhealthy part of the sample, which means that the analyses are carried through on a general sample.

The question from the questionnaire that is used to analyse the effect of physical activity on HRQOL is: "<u>*How many days a week are you active for at least 30 minutes a day?*</u> *Include moderate or strenuous physical activity where you breathe faster, move your muscles, and use your powers - e.g. for exercise or competitive sport, heavy gardening, brisk walking, bicycling at a moderate or fast pace or physical strenuous work. Include both work and leisure time*". With this question, the level of activity may both be moderate or high. The answering options range from zero days a week to every day. Another question that is possible to use is: "*If you look at the past year, what would you say is best suited as a description of your physical activity during leisure time?*". With this question it will be possible to identify, which persons are following the recommendations, but since the question is difficult to answer it is not included in this project. The reason this question is difficult to answer is, that persons that follow the recommendations regarding 30 minutes of physical activity every day, are active less than four hours a week, and cf. the possible answering possibilities in the questionnaire, they are considered sedentary. The recommendation regarding activity twice a week with high intensity is not analysed in this project, since none of the questions from the questionnaire have been considered a good choice. With the question "*Do you participate in sports or do you participate regularly in other activities that provide exercise in your leisure time?*" it is possible to believe, that persons that do participate in sports, do it twice a week, but since the question does not reflect the intensity of the sport it is not possible to make any conclusions from this question. Other questions regarding "exercise in everyday life" ask how many days a person uses a bicycle in daily doings, what the most important means of transportation is, how long the distance is to work, how they judge their physical fitness, and if they want to be more active. Some of these questions are already covered by the primary question of interest, whereas others (the distance to work, judgement of the physical fitness, and if a person wants to be more active) do not relate directly to the purpose of this project. Analysing the secondary recommendations is possible when using a question as e.g. "*How many days a week, are you active for at least 20 minutes at high intensity?*", but this question is not to be found in the questionnaire from Region Central Jutland.

To validate the HRQOL measure, factor analysis is used together with a correlation test, to see if the HRQOL measure correlates with the sample. Afterwards multiple regression analyses are used to analyse the effect of physical activity and other variables on HRQOL. In the end it is analysed if any interaction effects are found. A deeper explanation of the methods and why they are used can be found in section 2.2 and 2.4.

Firstly it is necessary to identify, which measure that shall be used as HRQOL.

2. Background Knowledge As A Basis For The Methods

The purpose of this section is to build up background knowledge, that can be used further in the project to understand and have expectations to the methods and results.

2.1 Which Measure Should Be Used As HRQOL?

The persons from Region Central Jutland have answered a questionnaire that contains questions covering welfare, health and disease in various aspects of daily life. Due to this, Region Central Jutland has chosen to use a 12-item Short-Form Health Status Survey (SF-12) that measures the persons' health status. This survey consists of 12 different items, linked with eight domains, focusing on different functions (20), as seen in figure 2.0. It can be seen that SF-12 is an outcome from a larger 36-item Short-Form Health Status Survey (SF-36[2]), and the use of SF-12 instead of SF-36 has avoided an enlargement of the questionnaire from Region Central Jutland. The items are labelled in relation to their dimension, so items representing e.g. physical functioning (PF) are labelled PF followed by a number, and the same is true for the other items. In SF-12 two items from PF are represented. PF are questions regarding activities in daily life and whether a person is limited in those. The questions regarding PF02 and PF04 are question two and three respectively in Appendix I: The SF-12 Questionnaire. The role-physical (RP) items are represented twice in the questionnaire (question four and five) and ask if a person has had any problems due to the physical health. In figure 2.0 it is seen that one item from bodily pain (BP) and general health (GH) respectively are used in the physical component summary (PCS) score. Those questions relate to the person's general idea of the general health, and if pain has complicated daily life. All six items together represent the physical component summary score (PCS). The same is true for the mental component summary (MCS) score, but the items are different from the ones used to calculate PCS. Two items represent mental health (MH), and these questions are related to the general mood the last four weeks. The two role-emotional (RE) items ask if any emotional problems have disturbed daily activities, and can be seen in question six and seven in Appendix I: The SF-12 Questionnaire. Question 10 and question 12 represents vitality (VT) and social functioning (SF) respectively, and it can be seen that only one item is represented from VT and SF.

The 12 items have been found by using forward-step regression analysis on SF-36, with the purpose to make a shorter questionnaire, which can be answered within two minutes and at the same time correlates at least 95 percent with SF-36 (21). The use of SF-12 instead of SF-36 in this project is considered an advantage, since the use of SF-12 is preferable in a study with large sample sizes (21) and at the same time the two health surveys have shown to be highly correlated both in healthy (22,23) and unhealthy (24-27) persons.

The SF-12 scores from Region Central Jutland are calculated based on answers from the 12 questions related to SF-12 scores in the questionnaire. The algorithm used to calculate PCS and MCS is not publicly available and thus is unknown to me. Therefore it is necessary to analyse if the two component summary scores correlate with the sample from Region Central Jutland.

[2] It shall be noted that the self-reported health transmission is not used to calculate SF-36, and therefore only 35 items are included in the algorithm to calculate the two component summary scores (21).

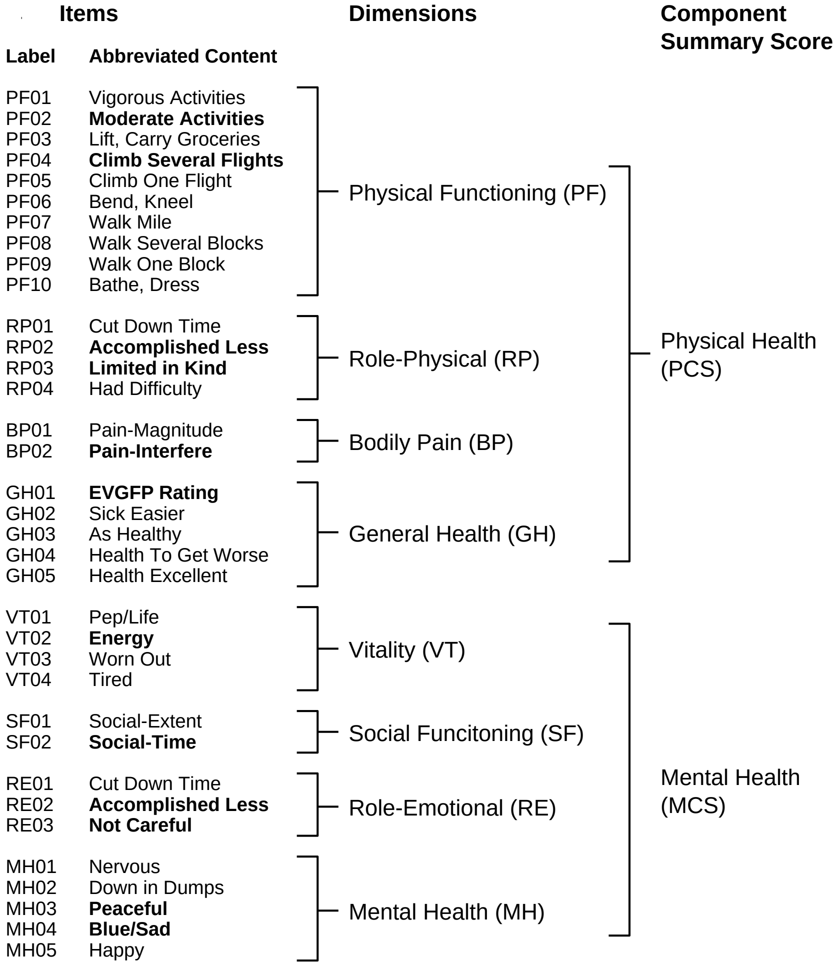

Figure 2.0: The bold items in the figure are the 12 items from SF-36 that forms SF-12. It can be seen that two items are included in the dimensions PF, RP, RE and MH, whereas one item gives the dimensions BP, GH, VT and SF. Modified from (21).

2.2 Factor Analysis Validates The Use Of SF-12 Score As HRQOL Measure

Factor analysis is used because the dataset provided to this project contains answers from the SF-12 part of the questionnaire and the calculated PCS and MCS. The algorithm that has been used in the calculation is unknown and therefore the purpose of the factor analysis is to validate that PCS and MCS can be used on this specific sample, since the sample may not fit perfectly with the algorithm. If the calculated PCS and MCS from Region Central Jutland prove valid it is justifiable to use the calculated PCS and MCS in further analyses.

It is possible to use factor analysis, since this method defines an underlying structure in the correlations of several variables (in this case the items from SF-12) thereby defining new and fewer dimensions, called factors. The factors have already been determined from Region Central Jutland (PCS and MCS), and by using factor analysis on the answers from the SF-12 questions it is possible to analyse if the outcome is two factors. Afterwards it is analysed if the SF-12 in the dataset and the already calculated PCS and MCS correlate. The number of factors (two) and the number of variables (12 items) are known beforehand. Due to this, the answers of the 12 items in SF-12 are used as variables in the factor analysis. Using factor analysis the sample size shall exceed at least five times as many observations as number of factors, and since the sample size is far higher than five times two, the sample size is valid to use.

The answers from the SF-12 questionnaire span 12 dimensions and these 12 dimensions can explain two factors (expressed by vectors in a coordinate system). Six items on each domain will have different loadings meaning that some items have a greater impact on the factor than other items. These loadings are placed on the factors as vectors, where the length of the vectors determine the influence each item has on the factor. As it can be seen in figure 2.0 the items that are used to calculate PCS are different from the items that are used to calculate MCS, which make it reasonable to assume that the component summary measures are not correlated and therefore are orthogonal after rotation. It is taken into consideration that PCS and MCS may correlate with each other, but in the method they are considered orthogonal, because the items belong to each component summary score. The rotation method that is typically used on orthogonal vectors is VARIMAX (28).

The rotation is used to redistribute the variance and thereby getting a more meaningful pattern. The uncorrelated PCS and MCS will form a coordinate system, where it is decided that PCS will follow the x-axis and MCS will follow the y-axis, as seen in figure 2.1, which illustrates the expected signs of the vectors from the factor analysis on the items. Appendix I: The SF-12 Questionnaire, is used as a supplement to understand the expectations. In the appendix, the outcome of the first question will be one, if you have an excellent health, whereas the outcome will be five, if you have a poor health, which means that a higher value gives a poorer health. The opposite is true in question four, where a higher value means a better health. In the questions where a higher value means a poorer health, the vector goes in the negative direction, meaning that the outcome from the factor analysis will give a negative sign in that item. On the other hand, questions where a higher value means a better health, the vector goes in the positive direction, meaning that the outcome from the factor analysis will give a positive sign in that item. The expected signs in each item can also be seen in figure 2.1.

Item	PCS	MCS
GH01	-	
PF02	+	
PF04	+	
RP02	+	
RP03	+	
RE02		+
RE03		+
BP02	-	
MH03		-
VT02		-
MH04		+
SF02		+

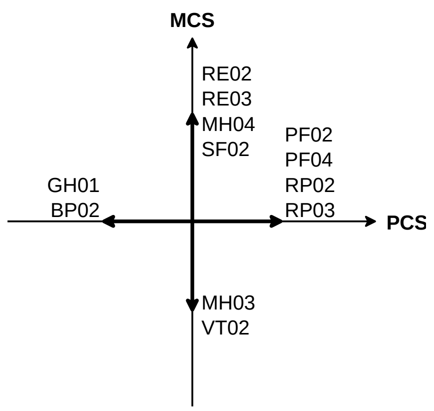

Figure 2.1: The table to the left gives an overview of the expected signs of the items from the factor analysis. The items appear in the same order as in the questionnaire from Region Central Jutland. The picture to the right gives a graphical presentation of the distribution of the factors in a coordinate system, where PCS and MCS are orthogonal.

To analyse the correlation between the two factors and the calculated PCS and MCS from Region Central Jutland, Pearson's correlation analysis is used, because the items can be rated on a scale. This makes it possible to analyse how strong the correlation is and if the correlation is significant.

After the HRQOL measure has been validated, the variables that describe HRQOL are found. This is done by analysing previous studies to see, which variables other authors find important in relation to HRQOL and physical activity. The studies and the variables they have used are presented in the following section together with a solution to which variables this project uses to describe HRQOL.

2.3 The Observations From Previous Studies

By finding studies that have analysed physical activity together with HRQOL it is possible to find the confounders that shall be included in the multiple regression analysis. A confounder is a variable that correlates with both the dependent variable and the independent variable, and if an important confounder is forgotten in a model it may lead to a type I error[3] that may lead to a wrong conclusion of causality between the dependent variable and independent variable. An example is the investigation of the effect of physical activity on HRQOL and a positive correlation between HRQOL and physical activity, but perhaps another variable as e.g. age has a more important effect. With this, the positive correlation found without age, may be caused by a correlation of age in both HRQOL and physical activity. To "eliminate" the risk of forgetting confounders, earlier studies are used as inspiration.

The search to identify papers is computer-assisted, using the search engines Pubmed and Cochrane library that are frequently used by health researchers.

The keywords used to find articles related to HRQOL are: Health-related quality of life, HRQOL, quality of life, QOL. To find articles related to physical activity following keywords are used: Physical activity, activity, exercise, and fitness. The outcomes from the searches are combined with "AND" in every search engine. To match the methods used in this project, with the articles are used to find the confounders of this project, the keywords "questionnaire" and "regression" are also used in combination with the two previous search words. All search words are quoted, so the search is specified with those exact words.

The headlines of every article are read through, and containment of a disease or earlier diseases causes exclusion of the study. This is done to exclude projects, with a focus on an unhealthy population, since this project focuses on a cross-sectional general population and not the change of HRQOL in a longitudinal study in diseased persons. Thus longitudinal studies that analyse HRQOL in a general population are also included in this project. By using studies "similar" to this project, it is possible to use the same confounders, to analyse if the results from the studies match this project. The importance of similarity is also taken into consideration, when analysing the age among the samples used in other articles, and if they use persons under the age of 18 they are excluded from this project, since this project focuses on the recommendations in adults. Ideally studies with about 30,000 persons are included in the project, but since only a few studies are found, every article analysing a general population is included in this study.

In table 2.2 the articles included on the basis of the above is shown. The table presents the purpose of the study, which population and design that is used, how they assess physical activity (abbreviated PA), which variables that are adjusted for in the model, what their results are, and some comments to

[3] Type I error is a false positive error.

the study. The variables that are used in this project are written in **bold**. Only the results that show the connection between HRQOL and physical activity are presented, since this connection is the focus of this project, and further results can be read in the specific source of interest.

The study "Are fitness, activity and fatness associated with health-related quality of life and mood in older persons?" (29) analyses if HRQOL and mood are associated with levels of fitness, habitual activity, and fatness in 38 men and 44 women. They primarily use objective measures, which makes the study quite strong in relationship to response bias. To measure physical activity in leisure time Stanford 7-day Physical Activity Recall Questionnaire[4] is used. The subjective measures used are SF-36, where the primary focus is PCS in correlation with physical activity. Mood is found using Profile Of Mood States[5]. They find that higher levels of aerobic fitness has a positive effect on mood, BP, PF and VT, but that stronger muscles do not provide better HRQOL. The strength of this study is that it provides results by objectively assessed measures. But at the same time the sample size is quite small, which means that the results probably cannot represent a broader population.

In the next presented study "Prospective study of new participants in a community-based mind-body training program" (30) the authors seek to investigate the effect of a three month mind-body intervention on HRQOL. They use SF-36 as the primary outcome, but have chosen not to use PCS and MCS due to limitations in their algorithm. In this project this problem is already considered, and this is why the factor analysis is used in the purpose to analyse if the unknown algorithm fits with the sample of this project. They find that three months training programme has a positive effect in all eight domains of SF-36 and that the persons' mood in general gets better. The strength of this study is that it analyses the effect of exercise over time, and successfully shows a positive effect between HRQOL and physical activity. On the other hand also this study has a low sample size, so the results are difficult to generalise to the entire population of New York City.

The sample size of "Gender differences in effects of physical activity on quality of life and resource utilization" (31) is quite high taking into consideration that the persons are found in a rural town, and therefore the results may be generalised to the population of Naie. They use persons above the age of 20, where this project focuses on persons above the age of 18. This difference in age may affect the amount of exercise between the study and this project, since persons between the age of 18 and 20 may have more leisure time than persons above the age of 20. But since a positive effect is still found with increasing exercise, the difference in age is not considered a problem. The investigation tries to find the effects of activity on HRQOL between genders by controlling the variables: Age, BMI, smoking status, alcohol intake, and co-morbidities, in their regression model. They find that all eight domains in SF-36 increased with activity in both men and women, and that physical activity with high intensity has a better effect in women compared to men.

T'ai Chi Chuan is a traditional form of Kung Fu that is widely practiced in the elderly population in Taiwan. This particular martial art is investigated in "Health-related quality of life in the elderly practicing T'ai Chi Chuan" (32) where the purpose is to investigate, how T'ai Chi Chuan affects HRQOL.

[4] Stanford 7-day Physical Activity Recall questionnaire: Is a structured interview, where the persons estimate the time being moderately, highly or very highly active during the week. The amount of time spent being active is multiplied with the respective metabolic equivalent (MET), thereby obtaining an estimate of the daily physical activity expressed as kilo calories (kcal) per kg per day (67).

[5] Profile Of Mood States: Consists of 65 five-point rating scales, which identify six moods of affective states: fatigue-inertia, depression-dejection, vigor-activity, anger-hostility, tension-anxiety, and confusion-bewilderment (68).

They find that persons participating in this martial art the last year have statistically significant better health in six of the SF-36 domains, and a higher score than the similar control group. Even though the sample size of this group is quite small they use an approximately five times higher control group to analyse the difference. When taking the large difference between the case group and the control group into consideration, the results are quite strong when finding that T'ai Chi Chuan has a positive effect on HRQOL. This project does not include a control group, but due to the large sample size, a control group will be almost impossible to find.

In the last study from 2009 "The relationship between a short measure of health status and physical activity in a workplace population" (33) the authors investigate a general population that does not contain either older persons or persons with a chronic disease. The purpose is to make a study in a general population to analyse, if physical activity has a positive effect on HRQOL. They find that physical activity has a positive effect on HRQOL that is measured by a single-item HRQOL. By analysing the effect of physical activity on HRQOL in a general population this study is the one that is closest to this project. It also emphasises that only few studies investigating a general population rather an old or diseased population exists, which may be the explanation to why only five studies have been found in the used search engines of this project.

Compared to the presented studies, this project contains a large sample size that makes the results reliable and generalisable to the population of Region Central Jutland. Even though the sample size of four of the studies is small, and they still find that activity has a positive effect on HRQOL, which should not be neglected. The results they find are still possible to use as guidance in this project, and when they do not find a negative effect of activity, the expectation of a positive effect of activity on HRQOL in this project is substantiated. Furthermore it is chosen not to use the eight domains of SF-12, but rather the two component summary scores PCS and MCS, which are validated with the sample as described earlier. The other studies use SF-36 rather than SF-12, which means that each domain is supported by more items, than the domains in SF-12, as seen in figure 2.0. By using PCS and MCS the domains are indirectly used, and the outcome still represents both physical health and mental health, but of course fewer outcomes represent HRQOL.

On the other hand the analyses of this project do not include comparison of different groups as the presented studies, but since the purpose is to analyse <u>how</u> HRQOL is affected by physical activity (and the confounders) it is not investigated <u>if</u> physical activity affects HRQOL as in the other studies. The strengths with some of the studies are that their outcomes are measured objectively and therefore response bias is a lesser issue in those studies. In this project the measures are self-reported, which means that not all variables are "reliable" compared to objective measures, which is discussed in Chapter III: Discussion & Conclusion.

Now the studies have provided information regarding HRQOL, physical activity and the variables they have used, and the next step is to analyse, which variables that shall be used in the analyses of this project. This is done in the next section, where the expected effect of each variable is presented on the basis of medical theory. Only the variables, HRQOL and physical activity, that have an effect on both variables are included as confounders of this project.

Purpose/Study	Population/design	PA* assessment	Adjusted for	Summary of results	Comments
To determine if levels of fitness, PA and fatness are associated with HRQOL and mood in older persons (29)	N =38 men and 44 women Cross-sectional study	Aerobic fitness is assessed with maximal oxygen uptake during treadmill testing Muscle strength is measured by a one-repetition maximum measured for seven different exercises on a weight machine Habitual PA is assessed using the Standford 7-day PA recall questionnaire	**Gender** **Age** Height Weight Maximal oxygen uptake Muscle strength **BMI** Percent body fat Daily physical activity	Higher levels of aerobic fitness is associated with better outcomes in Profile Of Mood States anger, and mood disturbances scores SF-36 items BP, PF, and VT are together with PCS, improved with higher levels of aerobic fitness. Muscle strength is not associated with HRQOL or mood	SF-36 is used to analyse HRQOL Mood is analysed using Profile Of Mood States Low sample size
To measure changes in HRQOL as a result of three months of mind-body training (30)	194 are included, and 171 returned the three month follow up Prospective cohort study	A questionnaire including estimated number of classes attending in the last three months, number of classes attended per week, frequency of home exercise completion per week, and whether any other new physical exercise was undertaken	**Gender** **Age** **Education** Ethnicity **Employment** Health insurance Access to a physician Marital status Medications used Medical comorbidity measured by Charlson Comorbidity Index[6] History of hyper-	All eight domains in SF-36 improved statistically significant after three months of community-based mind-body training programme. Furthermore fewer depressive symptoms, less trait anxiety, and greater self-efficacy are reported in a lesser degree in the participants that provided data on secondary outcome measures	SF-36 is the primary outcome measure, but not with PCS and MCS due to known limitations of their algorithms Secondary objective is to identify physical injuries or other adverse experiences associated with the practice

[6] Charlson Comorbidity Index: Is a predictor of mortality for persons that have a range of co-morbid conditions. Each condition is ranged with a score that describes the risk of dying of that disease (69).

			tension or arthritis Paffenbarger Physical Activity and Exercise Index[7] Centers for Epidemiologic Studies Depressive Symptoms Inventory[8] Spielberger Trait Anxiety Inventory[9] Generalized Self-efficacy Scale[10]		
To assess the effects of the amount and maximum intensity of activity on HRQOL between genders (31)	5107 of whom 3529 completed the follow-up after one year Cross-section with one year follow up	Measured with questions on frequency of and time spent on three levels of physical activity: Vigorous activity, moderate activity and light activity. Each level of physical activity is given a MET[11] corresponding to the type of activity. Nine METs for vigous activity, six for moderate activity and three for light physical activity. The METs are	**Gender** **Age** **BMI** **Smoking status** **Alcohol intake** Co-morbidities Marital status **Education** **Income**	All SF-36 domains increased statistically significant in proportion with activity in both men and women The maximum intensity of physical activity has positive effects on most of the HRQOL domains in women, but only in RP, VT, and RE in men	SF-36 is used to measure HRQOL

7 Paffenbarger Physical Activity and Exercise Index: This index gives a score measured by kcal per week, on the basis of different activities during daily life (30,70,71).
8 Centers for Epidemiologic Studies Depressive Symptoms Inventory: Is a 20-item self-report scale that measures the current level of depressive symptoms on a scale from zero to 60, where a higher score indicates more symptoms of depression (72).
9 Spielberger Trait Anxiety Inventory: Is an instrument for measuring "state anxiety" and "trait anxiety" that ranges from 20 to 80, where a higher score indicates more anxiety(73).
10 Generalized Self-efficacy Scale: Is a 10-item psychometric scale that ranges from 10 to 40 and assesses a person's optimistic self-beliefs. A higher score means a better self-belief (74).
11 One MET is defined as 1 kcal/kg/hour, which corresponds with sitting quietly (75).

		used to calculate kcal/week	Medical conditions EuroQOL[12]		
The purpose of this study is to investigate how T'ai Chi Chuan affects HRQOL (32)	140 persons in the case group and 560 in the control group Cross-sectional	Have been practicing T'ai Chi Chuan for at least one year	**Gender** **Age** **Educational level** Marital status **Employment** Duration of practicing T'ai Chi Chuan T'ai Chi Chuan style	The persons in the case group have a higher HRQOL score than the persons in the control group, but only RP, GH, VT, SF, RE and MH are statistically significant higher The multivariate regression model shows that the case group has higher scores in PF, RP, GH, and SF domains compared with the control group	Uses SF-36 to measure HRQOL
To analyse the relationship between HRQOL and physical activity in healthy adults (33)	573 employees Follow-up investigation	Respondents report the weekly frequency and average duration of physical activity on three levels of intensity: light, moderate, and strenuous. Each level of activity is modified to include duration in minutes for each level of intensity thereby calculating MET.	**Gender** **Age** **Smoking status** **Educational level** **BMI** Limitations due to health condition Arm of the original study	Physical activity is positively associated with HRQOL	Points out that earlier studies have focused on elderly and persons with chronic diseases Uses a single-item measure

Table 2.2: This table provides information of the studies used to find the variables that shall be used in further analyses of this project. The results presented are the results that have a connection between HRQOL and physical activity, and the studies may have gotten more results that are not presented in this project.

*PA = Physical activity.

[12] EuroQOL: Is also called EQ-5D, and is a questionnaire containing five questions about your health, each to which three answers exists (76).

2.3.1 Expectations To The Variables Based On Medical Theory

The purpose of this project is to analyse how activity affects HRQOL, and based on the five studies it is expected that activity will affect both PCS and MCS positively. Also the medical theory indicates that physical activity has a good effect on the physical health.

In the bones some specialised cells (osteoblasts) produce new bone matrix, which causes a higher bone mass and with it the osteoblasts causes the bones to get stronger. At the same time other specialised cells osteoclasts break down bone matrix and together with the osteoblasts the matrix of the bones is kept in balance, but this balance is affected by different factors. In cases where the bones are stressed as in e.g. weight lifting and during running the osteoblasts increase their production of new bone matrix, which causes a further increase of the bone mass. On the other hand the osteoblasts are not active, if the bones are not used, but the osteoclasts continue to break down bone matrix, which means that inactivity, after a few weeks, may cause a breakdown of a third of the total bone mass. Due to this it is important to be active to maintain the structure of the bones (7). The use of muscles causes different physiological mechanisms in the body, which result in stronger and bigger muscles, which is beneficial to persons that want to look fit. Being active also has a positive effect on the physical fitness because the blood volume and lung capacity increase causing a higher amount of oxygen in the blood. The heart is also affected by physical activity, because it causes a larger cardiac output causing the cardiac muscle cells to enlarge, which in turn causes a stronger heart (5). Many positive physiological effects are seen with exercise (only a few are mentioned), and it has also been shown that physical activity can prevent breast- and colon cancer (34). During physical activity, mental health is positively affected by endorphins (7), dopamine (8) and serotonin (9), which cause a general better mood (5,7-9).

Also gender and age are controlled for in all five studies, and since there is a difference in HRQOL dependent of gender and age, they are also included in this project. It is well known that men from the side of nature are stronger than women, due to a larger level of testosterone in the body (5), and with this it is believed being male has a positive effect on PCS. It is seen that women more often experience depression (35), but the reason is unclear. With this MCS is believed to be positively associated with being male. With increasing age different diseases occur more frequently as e.g. cancer. Furthermore changes in the cardiovascular system cause more thromboses and in severe cases pulmonary embolism. The heart gets "harder" due to scars caused by damaged muscle cells, the cardiac output decreases, which in turn causing a decrease in physical fitness. Those examples are only few, but overall the picture is a degeneration of the body causes more diseases and worsened physical functions of the body, and as a result of this PCS is believed to decrease with age. MCS is also believed to decrease with age due to the physiological changes that may cause pain and discomfort. Furthermore the social life may be affected negatively, due to hearing loss caused by damages through life (7). In relation to physical activity it is believed that both genders and all age groups benefit from physical activity, but that younger persons are more active than older persons. On the other hand it is believed that physical activity is equal between genders.

An unbalance in the energy consumed and the energy used causes malnutrition, which result in either a low BMI or a high BMI. Since BMI is associated with inactivity (5), and three studies have included BMI, this confounder is also used in this project. A high BMI is associated with higher risks of e.g. different cancers (36) and diabetes mellitus type II (37), and due to this PCS is believed to be negatively affected by a higher BMI. Some persons tend to overeat due to psychological problems as e.g. stressful periods of life (5), and some experience stigma (38), which is why it is believed that a higher BMI has a negative effect on MCS.

Smoking status is represented in two studies, and since smoking causes a decrease in lung function (34) the physical fitness naturally decreases as well. Different studies (39-41) have shown that smoking is related to poor PCS as well as poor MCS, and therefore smoking is also believed to have a negative influence on PCS and MCS in this study. Due to the decrease in physical fitness it is believed that smokers are less active than non-smokers.

Alcohol use is only reported in one study, but since it is proved, that alcohol has an effect on HRQOL this variable is also used in this project. Two studies (42,43) have found that alcohol use has a positive effect on HRQOL, and based on these two studies MCS is believed to be positively affected by alcohol consumption. PCS on the other hand is believed to be negatively affected by alcohol use, even though Chan and Stranges prove otherwise. The reason why alcohol use is believed to have a negative effect on PCS, is that alcohol consumption above the recommended level[13] is a strong risk factor of cancer especially in the oral cavity, oesophagus and the liver (34), and with this it is believed that the incidence of cancer is higher in persons that consumes alcohol, and with this PCS decreases.

Education is included in four of the studies, and since education is correlated to income (44), it is believed that both education and income have a positive effect on both PCS and MCS. Persons that have a high education (and a higher income) are believed to have more knowledge of healthy habits and at the same time they will be able to afford being active in e.g. fitness centres all year. Furthermore persons with a high education and a high income may not experience financial problems, and therefore they have surplus in daily life. In this project Danish Educational Nomenclature (DUN) represents the length of the educational years.

Even though MCS analyses the mental health, it is interesting to investigate if other mental factors have an effect on MCS. In two of the presented studies mental health is investigated. In the questionnaire from Region Central Jutland it is possible to analyse stress level with the use of Choen's perceived stress scale (PSS), which is a questionnaire with ten questions, each with five possible answering possibilities. These answers are given a score that summed up gives a total score between zero and 40, where a higher score expresses a higher level of stress. There is a difference between acute stress and chronic stress, where the acute stress may improve our results (e.g. at an examination), whereas chronic stress has a negative influence due to continuous release of the stress hormones cortisol, epinephrine and norepinephrine. Those hormones are also released during acute stress, where they increases the respiration, heart beat and causes piloerection[14], but with longer periods of stress the influence from the hormones causes hypertension, decreased immune response and in severe cases heart failure (due to overload of the heart). Taking this into consideration, a high level of stress is expected to have a negative influence on both PCS and MCS. At the same time it is thought that persons that experience a high stress level do not feel that they have the time and energy to be physically active.

Employment is controlled for in two of the studies, but the studies do not investigate physical activity at work, and it is believed that physical activity at work will affect both the physical health and the mental health. Activity at work is believed to have a negative effect on physical health, since a working

[13] The Danish National Board of Health recommends that men maximum consume 14 units of alcohol a week, and women maximum consume 7 units of alcohol per week (77).
[14] Piloerection = goosebumps

place with hard work may be back-breaking, and if the work is back-breaking causing pain it is believed to negatively affect your mental health.

In table 2.3 a summary of the expected signs, based on the above, is shown.

	PCS	MCS
Activity	+	+
Male	+	+
Age	-	-
BMI	-	-
Smoking	-	-
Alcohol	-	+
Income	+	+
DUN	+	+
PSS	-	-
WorkPA	-	-

Table 2.3: This table provides information of the expected signs in front of each variable.

2.4 Multiple Regression Analysis

Analysing the causality between HRQOL and activity gives the possibility to use simple regression analysis that forms a straight line thereby examining the relationship between a dependent variable (PCS or MCS) and an independent variable (activity), which gives the following equations:

$$y = \beta_0 + \beta_1 x + u$$

$$PCS = \beta_0 + \beta_1 activity + u$$

$$MCS = \beta_0 + \beta_1 activity + u$$

This equation forms a straight line using ordinary least squares (OLS). This method is described in Appendix II: Multiple Regression Analysis.

The β_0 is the intercept, which tells where the line intersects on the y-axis. β_1 is the slope coefficient, that expresses how much *y* changes with an increasing *x*, holding the other variables fixed. The coefficient shows if the variable has a positive or negative effect on y with an increasing x. *u* is unobserved variables that have an effect on *y*, even though literature has been searched and *u* therefore is expected to be zero, it cannot be certain that no other variables affect *y*, so the *u* is kept in the equation.

The dependent variable may be affected by other variables than the independent variable. The confounders that in addition to activity have shown effect on QOL or HRQOL are found in the previous section and are: Gender, age, BMI, smoking status, alcohol use, income, education, PSS and physical activity at work. Even though those variables are considered confounders they are referred to as variables or independent variables when talking about both independent variables and confounders in general.

To achieve a higher explanation in the causality multiple regression analysis is used, since this method allows to explicitly analyse the ceteris paribus effects of the variables on the dependent variable. By using more variables it is possible to explain more variation in the dependent variable, still using OLS to find the best "straight line" that exists in ten dimensions as ten variables have been found to have an effect on PCS and MCS. The same right side variables are used in both regression models, since this makes it possible to compare the results between the two HRQOL measures. By using the same variables in both PCS and MCS the following equations are found:

$$PCS = \beta_0 + \beta_1 activity + \beta_2 male + \beta_3 age + \beta_4 BMI + \beta_5 smoking + \beta_6 alcohol + \beta_7 income + \beta_8 DUN + \beta_9 PSS + \beta_{10} workPA + u$$

$$MCS = \beta_0 + \beta_1 activity + \beta_2 male + \beta_3 age + \beta_4 BMI + \beta_5 smoking + \beta_6 alcohol + \beta_7 income + \beta_8 DUN + \beta_9 PSS + \beta_{10} workPA + u$$

The multiple regression analysis gives different outcomes, where the outcomes of interests are: $\bar{R}^2$, unstandardised coefficients β, standardised coefficients beta, and significance. The $\bar{R}^2$ and standardised coefficients beta are described in further details in Appendix II: Multiple Regression Analysis.

$\bar{R}^2$ is used to predict future outcomes on the basis of the multiple regression analysis, because it analyses the goodness of fit. It estimates how much variation in y that is explained by x_j. This means that a low $\bar{R}^2$ is an expression of a large variance in the dependent variable explained by x_j, and that a high $\bar{R}^2$ is an expression of a small variance in the dependent variable explained by x_j. In a large sample the $\bar{R}^2$ is expected to be quite low due to the fact that the HRQOL among the population is expected to have a large variance within the independent variable and each confounder. Due to the expected low $\bar{R}^2$ it may be assumed that the model does not fit quite well, but since the sample size is quite large the model gives a reasonably precise estimate of the variance in HRQOL explained by the variables.

The unstandardised coefficient β is the slope coefficient that expresses how much the HRQOL changes if a variable rises by one unit (e.g. one year). This gives that a positive unstandardised β causes an improvement in HRQOL, whereas a negative unstandardised β causes a worsening in HRQOL. The unstandardised β is called the slope coefficient because it gives the slope of the line that describes the change in the dependent variable by a single variable.

It is not possible to compare the arbitrary units of the variables (measurements being: years, kilograms/m^2, Dkr. etc.), but it is possible to compare their standard deviations by using standardised beta. By this a rise of one standard deviation in x_j, the dependent variable changes with $beta_j$ standard deviations. E.g. if the standardised beta is 0.02, a rise of one standard deviation in e.g. age, will lead to an increase of HRQOL by 0.02 standard deviations.

<u>NOTE</u>: In the following chapters both the unstandardised β and the standardised beta are used. The unstandardised β is used as the slope coefficient, which means it is used in the equation of PCS and MCS respectively. The standardised beta on the other hand describes the influence of each variable on the dependent variable. Be aware that two different coefficients, both standardised and unstandardised, are used in the following Chapters.

The significance level used in this project is 0.05, which means that a variable that have a significance level of 0.05 or below, has an effect on the dependent variable that is unlikely to have occurred by chance.

Another way to use multiple regression analysis is to include interaction effects, that may show an effect of exercise AND another variable combined. This is explained in the following section.

2.4.1 INTERACTION EFFECTS: 2+2 = 5?

Analysis of interaction effects is performed to see if there is an even larger effect on PCS or MCS when activity is combined with the other variables. By multiplying the variable with activity it is possible to see if a synergy is achieved.

Let's say it is analysed if there is an effect on the MCS score due to activity 30 min a day, age and gender, the equation will look like this:

$$MCS = \beta_0 + \beta_1 activity + \beta_2 gender + \beta_3 age + \beta_4 gender \cdot activity + \beta_5 age \cdot activity$$

It is important to check for interaction effects because it may reveal that a single variable cannot stand alone and a person may feel much better (or worse) if some of the variables are combined with activity. The first thought may be that activity should be combined with all the variables, but this may lead to e.g. low $\bar{R}^2$ and multicollinearity[15]. Therefore it is important to consider, which variables that should be combined with activity.

It is obvious that gender, education and income are not affected by the level of activity you participate in. Furthermore persons that smoke and consume alcohol will probably continue to do so even though they are active 30 minutes a day. It is believed that activity can affect your health so that persons with increasing age will feel better both physically and mentally, and thereby they will have a better HRQOL due to a reduction in age-related diseases (45) as osteoporosis, cardiovascular diseases, strokes etc. Since a high BMI often is related to lifestyle-related diseases and exercise has proven to reverse those diseases, BMI and activity is believed to have a synergetic effect. Physical condition may be affected by a job that requires heavy lifting and the like, and therefore the time spent working may fulfil the needs of activity. Therefore physical activity both at work and in leisure time may have a higher negative effect on PCS. PSS is believed to be affected in a higher degree with being active, due to the neurotransmitters released having an even more positive effect in persons that are stressed. Taking all this into consideration, will result in these equations:

$$\begin{aligned} PCS = {} & \beta_0 + \beta_1 activity + \beta_2 gender + \beta_3 age + \beta_4 BMI + \beta_5 smoking + \beta_6 alcohol + \beta_7 income + \\ & \beta_8 DUN + \beta_9 PSS + \beta_{10} workPA + \beta_{11} activity \cdot age + \beta_{12} activity \cdot BMI + \beta_{13} activity \cdot workPA + u \\ MCS = {} & \beta_0 + \beta_1 activity + \beta_2 gender + \beta_3 age + \beta_4 BMI + \beta_5 smoking + \beta_6 alcohol + \beta_7 income + \\ & \beta_8 DUN + \beta_9 PSS + \beta_{10} workPA + \beta_{11} activity \cdot age + \beta_{12} activity \cdot BMI + \beta_{13} activity \cdot PSS + u \end{aligned}$$

Even though multiple regression analysis gives a result of how the different variables affect y, it does not analyse if the chosen dependent variable has been calculated correctly. By using factor analysis it is possible to verify that the algotithm used for calculating MCS and PCS is correct.

2.5 SUMMARY OF CHAPTER 1

It has now been concluded that people in general are less active than they should be, and the purpose of the recommendations from the Danish National Board of Health is to improve HRQOL in persons above the age of 18. The HRQOL measures that are used during this project are the two SF-12 outcomes PCS and MCS. The SF-12 is considered a valid outcome in analysing people's health in a larger questionnaire survey, but the validity of the calculated PCS and MCS is still analysed using factor

[15] Multicollinearity: Arises if some of the variables are highly correlated.

analysis and correlation analysis before using them in further analyses. Afterwards the variables that explain PCS and MCS are identified using previous studies, which have investigated the effect of physical activity in different contexts. The variables that are found in the previous studies are used in both PCS and MCS to make it possible to compare the results. The variables that shall be used in multiple regression analysis are: Activity, gender, age, BMI, smoking status, alcohol use, education, income, PSS, and physical activity at work. Interaction effects are used in the aim to analyse if activity has an even higher effect combined with another variable. The variables that are used in combination are: age, BMI, workPA (for PCS) and PSS (for MCS). With this background covered it is possible to continue to the methods that are used to analyse the problem formulation.

CHAPTER II
METHODS & RESULTS

3. Methods In The Analysis Of Data

The purpose of this section is to generate knowledge that is used to fulfil the purpose of the project, which is to investigate how physical activity affects HRQOL.

The programme used through all statistical analyses is PASW Statistics Data Editor Version 18 (SPSS). The syntax of the programming in SPSS and thereby a deeper explanation of the editorial method can be seen in Appendix III: The Editorial Process In SPSS.

3.1 Data Collection Ahead Of This Project Done By Region Central Jutland

The questionnaire contains questions covering welfare, health and disease, and in this project not all questions are included. The variables included in this project are described in table 3.0. The questionnaire is received by 52,400 persons above the age of 16, of which 2,500 questionnaires have been sent to 17 municipalities, 8,200 questionnaires to Aarhus and 1,700 questionnaires to Samsø. This segregation is due to a national agreement that maximum 40 percent of the population in every municipality may receive a questionnaire. The questionnaire has been sent out on the 5th of February 2010 and up to three dunning letters have been sent to persons that did not answer it. 65 percent have answered the questionnaire (46), which correspons to 34,147 answers.

3.2 Inclusion Criterion

This project focuses on the recommendations about exercise from the Danish National Board of Health in the adult population. The adult population includes persons above age of 18, but the questionnaire has been sent out to persons above the age of 16. In the dataset persons under the age of 18, are excluded from the project, by deleting them from the dataset. The questionnaire have been sent out to Danish citizens, which includes both Danes and persons with a different ethnic background that live in Region Central Jutland. The analyses are carried out on every person that answered the questions that are necessary to calculate PCS and MCS score. Only the persons that have answered the six items to PCS and/or answered the six items to MCS are included in the analyses.

3.3 The Editorial Method

Before the analyses are carried through the dataset is edited, and the editorial process can be seen in Appendix III: The Editorial Process In SPSS. The variables are translated to English, to keep them in the language this project is written in. Gender is a binary variable, and are made binary, to distinguish between them. In the case, where the sign of the coefficient is positive, it means that being male has a positive effect on HRQOL (because male has the value 1), whereas a negative sign advantage of being female (because female has the value 0). Some of the variables as age, BMI, and PSS are not divided into categories, since these variables have a logical continuous distribution; an increase of one unit in e.g. age, means an increase of one year no matter if the increase is from 18 to 19 years of age or the increase is from 64 to 65 years of age. Even though activity and alcohol intake also have a logical continuous distribution, they are divided into the same categories as in the questionnaire, but practically makes no difference. This is different in the rest of the variables as smoking status, household income, educational level and physical activity at work, where a change of one unit, has a different magnitude depending on the answer. E.g. if a person answers "no, I have quit" to the question "*do you smoke?*", the person will get a zero, but if the answer is "yes, every day" the person scores seven, even though it is only possible to chose between five answering possibilities. Variables as PCS, MCS, age, BMI, and PSS, have large ranges, and to use them in frequency tables, which provide an overview of the distribution of the population in each variable, they have been categorised, as seen in Appendix III: The Editorial Process In SPSS.

Smoking status is given different values depending of the answers ranging from a zero (meaning that the person does not smoke or have quit) to seven (indicating the person smokes every day). This increasing range ensures the variable smoking status gets a higher value if a person smokes every day, and thereby a larger effect of smoking will be revealed in the multiple regression analysis compared to a person that smokes e.g. less than once per week. The same is accomplished by increasing the value in educational level and physical activity at work; the positive (or negative) effect will be enlarged if a person answers a level higher in the questionnaire. The variables household income and educational level are categorised with the value in the middle of the interval (e.g. 49,500 represents 0-99,000 Dkr. and 13.5 represents 13 to 14 educational years) because the answers are believed to be uniformly distributed within each interval.

A description of the variables and how they are coded can be read in table 3.0, while a deeper explanation of the variables can be seen in Appendix IV: Description Of The Variables.

Variable (question number)	**Variable name**	**Description**	**Categories**
Physical activity (Q60)	activity	Activity account for the person's activity for 30 minutes X days a week.	Sys. mis = missing value 0 = no days a week 1 = one day a week 2 = two days a week 3 = three days a week 4 = four days a week 5 = five days a week 6 = six days a week 7 = every day
Gender (Q1)	male	Male-dummy: Are you male?	0 = no 1 = yes
Age (Q2)	age	Age is calculated as the year 2010 subtracted from the individual's birth year, with no regard to the birth date.	Age is not divided into categories.
BMI (Q68 and Q69)	BMI	$BMI = \frac{weight\,(kg)}{height\ (m) \cdot height\,(m)}$	BMI is not divided into categories.
Smoking status (Q31)	smoking	Smoking is recoded into "frequencies" of smoking per week, which of course is estimated.	Sys. mis = missing value 0 = No, I have never smoked OR No, I have quit 0.5 = Yes, less than once per week 2 = Yes, more than once per week 7= Yes, every day
Alcohol intake (Q47)	alcohol	Alcohol intake is divided into seven groups each representing how many days a week a person consumes alcohol. The categorisation is based on the person's answer in the questionnaire.	Sys. mis = missing value 1 = 0-1 day a week 2 = 2 days a week 3 = 3 days a week 4 = 4 days a week 5 = 5 days a week 6 = 6 days a week 7 = 7 days a week

Household Income (Q109)	income	The deviation of income is related to the middle value within each interval. The middle value is used because it is believed that the answers are uniformly distributed within each interval.	Sys. mis = missing value 49,500 = 0 - 99,000 Dkr. 124,500 = 100,000 - 149,000 Dkr. 199,500 = 150,000 - 249,000 Dkr. 312,000 = 250,000 - 374,000 Dkr. 449,500 = 375,000 - 524,000 Dkr. 612,000 = 525,000 - 699,000 Dkr. 774,500 = 700,000 - 849,000 Dkr. 937,000 = >850,000 Dkr.
Educational level (measured by DUN)	DUN	The DUN number represents the number of educational years the person has gone through. In this variable the middle value is used to express the average years of education.	Sys.mis = system missing or "out of category" 0 = preschool 3.5 = primary school level I (1st to 6th grade) 8.5 = primary school level II (7th to 10th grade) 8.5 = upper secondary level I (10th educational year) 11.5 = upper secondary level II (11th/12th educational year) 13.5 = short higher education (13th/14th educational year) 15.5 = medium length higher education (15th/16th educational year) 17.5 = long higher education (17th/18th educational year) 20.5 = Scientist's level (≥19 educational year)
Cohen's perceived stress scale (Q17)	PSS	This scale tells how stressed a person is on a scale from 0-40, a higher score means a higher level of stress.	PSS is not divided into categories
Physical activity at work (Q104)	workPA	The workload is divided into four categories, each describing the amount of hours each day of working that demands activity. The hours of activity is an estimate.	0 = Primary sedentary work 2 = Standing or walking, not exhausting 4 = Standing or walking, with lifting and bearing work 7 = Heavy or fast, exhausting work

Table 3.0: This table shows each variable with its respective variable name. A short description of each variable is provided, and to the right the categorisation is presented.

3.4 Factor Analysis Validate The Use Of PCS And MCS

In the factor analysis all items from SF-12 are used: GH01, PF02, PF04, RP02, RP03, RE02, RE03, BP02, MH03, VT02, MH04, and SF02. On the basis of figure 2.0 it is expected that a factor analysis of the 12 items gives an outcome of two factors (representing PCS and MCS). To rotate the analysis VARIMAX is used, since this rotation method is used in orthogonal variables, where the purpose is to maximise the sum of variance, thereby achieving the highest explanation of each variable in each factor.

The outcomes are saved in SPSS with the purpose of making a comparison with the calculated PCS and MCS given by Region Central Jutland. To compare the two factors with the ones from Region Central Jutland, Pearson's correlation is used. Ware and colleagues (21) found that the correlation between SF-36 and SF-12 should be minimum 95 percent with a significance level below 0.05, which is considered a fair result. Hence the criteria for a successful correlation test in this project is a correlation of at least 95 percent at a significance level of 0.05, between the factors found in the factor analysis and the calculated PCS and MCS.

3.5 Influences Of The Variables Are Found Using Multiple Regression Analysis

The causality between the dependent variable and the independent variables is found using multiple regression analysis. Firstly it is analysed if age and being male has an effect on PCS or MCS, to get an idea if they affect PCS and MCS without activity. The dependent variables are the two SF-12 scores: PCS and MCS. The focus of this project is to find the effect of physical activity on HRQOL, which is why physical activity is included as an independent variable. The confounders (that also are independent variables) used in this project have been found using inspiration from previous studies and are: Gender, age, BMI, smoking status, alcohol intake, household income, educational level, Cohen's perceived stress scale, and physical activity at work. The expectations to each variable are supported by medical theory in section 2.3.1: Expectations To The Variables Based On Medical Theory.

The same independent variable and confounders are used on both dependent variables, because this makes it possible to analyse if the effect of the variables is different in the two component summary scores and because it is believed that all variables affect both physical and mental health.

This leads to the multiple regression analyses analysing activity 30 minutes a day:

$$\begin{aligned} PCS &= \beta_0 + \beta_1 activity + \beta_2 male + \beta_3 age + \beta_4 BMI + \beta_5 smoking + \\ &\quad \beta_6 alcohol + \beta_7 income + \beta_8 DUN + \beta_9 PSS + \beta_{10} workPA + u \\ MCS &= \beta_0 + \beta_1 activity + \beta_2 male + \beta_3 age + \beta_4 BMI + \beta_5 smoking + \\ &\quad \beta_6 alcohol + \beta_7 income + \beta_8 DUN + \beta_9 PSS + \beta_{10} workPA + u \end{aligned}$$

These equations give an outcome that is analysed to see, which variables that affect PCS and MCS, especially to see how activity affects HRQOL.

After the multiple regression analysis it is analysed if there is any interaction effects.

3.5.1 The Product Of Variables May Show Interaction Effects

The analyses regarding interaction effects are made with the original multiple regression analyses whereto the product of activity and some of the other variables are added, which gives the following equations:

$$\begin{aligned} PCS &= \beta_0 + \beta_1 activity + \beta_2 male + \beta_3 age + \beta_4 BMI + \beta_5 smoking + \beta_6 alcohol + \beta_7 income + \\ &\quad \beta_8 DUN + \beta_9 PSS + \beta_{10} workPA + \beta_{11} activity \cdot age + \beta_{12} activity \cdot BMI + \beta_{13} activity \cdot workPA + u \\ MCS &= \beta_0 + \beta_1 activity + \beta_2 male + \beta_3 age + \beta_4 BMI + \beta_5 smoking + \beta_6 alcohol + \beta_7 income + \\ &\quad \beta_8 DUN + \beta_9 PSS + \beta_{10} workPA + \beta_{11} activity \cdot age + \beta_{12} activity \cdot BMI + \beta_{13} activity \cdot PSS + u \end{aligned}$$

By analysing these equations it will be possible to see if the assumptions about activity combined with other variables are fulfilled.

4. Results

4.1 Factor Analysis

In table 4.0 the rotated factor matrix is shown. It can be seen that the 12 items from the SF-12 questionnaire produce two factors, and each factor consists of six items as expected. The items linked with PCS are linked with factor 1, and the six items linked to MCS are linked to factor 2. The picture shows how each item affect factor 1 and factor 2 respectively, and the signs are as expected in table 2.1.

	Factor	
Item	**1**	**2**
GH01	-0.647	
PF02	0.826	
PF04	0.776	
RP02	0.809	
RP03	0.854	
RE02		0.772
RE03		0.736
BP02	-0.764	
MH03		-0.714
VT02		-0.636
MH04		0.771
SF02		0.663

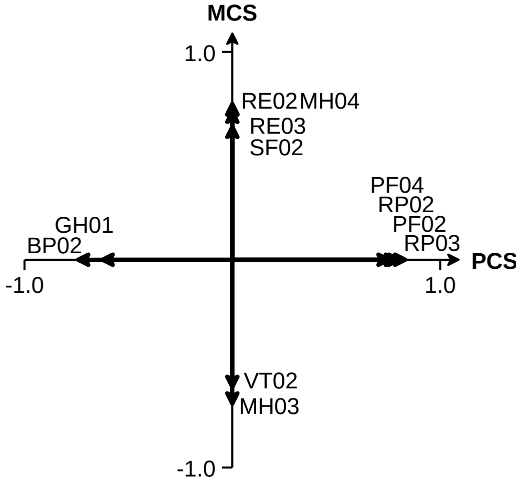

Table 4.0: The rotated factor matrix from the factor analysis. In the picture each item is shown in a system of coordinates.

Some of the questions in the SF-12 questionnaire are designed so a higher score means a better health, which is expressed as a positive sign in the item that is connected to that question. Other questions have a reverse scoring, which means that a higher score means a poorer health, and if the item is connected to a question with reverse scoring it means that the item has a negative sign. The positive signs in the items PF02, PF04, RP02, and RP03 in factor 1 and PF02, PF04, RP02, and RP03 in factor 2 mean that a higher score in those items means a better health. The reverse scoring is expressed by the negative signs in GH01, BP02, MH03, and VT02 expresses a poorer health with a higher score.

		MCS	PCS	Factor 1	Factor 2
MCS	Pearson correlation	1	-0.018	-0.018	0.997
	Significance		0.002	0.002	0.000
PCS	Pearson correlation	-0.018	1	0.987	0.002
	Significance	0.002		0.000	0.668
Factor 1	Pearson correlation	-0.018	0.987	1	0.000
	Significance	0.002	0.000		1.000
Factor 2	Pearson correlation	0.997	0.002	0.000	1
	Significance	0.000	0.668	1.000	

Table 4.1: The results from the correlation test between the two factors from above with PCS and MCS. The number of cases included for the analyses is the number of persons, who have answered the items for PCS and MCS, which is 29,992 and 29,989 for PCS and MCS respectively.

The two factors found in the factor analysis are saved in SPSS and a Pearson's correlation test is run on factor 1 versus PCS and factor 2 versus MCS. The result from the correlation test can be seen in table 4.1. The correlation between MCS and PCS is close to zero (-0.018) with a high significance (0.002), which means that they are significantly correlated, but since the correlation score is very low, it practically means that no correlation exists between PCS and MCS. The correlation test between the two factors found in the factor analysis show that the two factors do not correlate, and no statistical significance is found (1.000). This missing significance emerges, since the factors are orthogonal by construction.

The correlation test PCS and factor 1 shows a high correlation (98.7 percent) with statistical significance (>0.000). This means that the calculated PCS is a good measure to find the physical HRQOL in the sample from Region Central Jutland, because the correlation between factor 1 and the calculated PCS is statistically significant high.

The correlation between MCS and factor 2 is statistical significant (>0.000) too, and they are almost identical with a correlation of 99.7 percent. This means it is valid to use the calculated MCS as mental HRQOL outcome for the sample from Region Central Jutland.

In summary, even though an unknown algorithm is used to calculate PCS and MCS in the sample from Region Central Jutland they are statistically significant correlated with the actual PCS and MCS, which means that the calculated PCS and MCS are valid to use in further analyses.

4.2 Multiple Regression Analyses

Firstly it is analysed if gender and age affect PCS and MCS with no regards to the remaining variables, and the outcome can be seen in table 4.2. It can be seen that the $\bar{R}^2$ is quite low for both PCS (13.4 percent) and MCS (4.6 percent), and with this it seems the variables explain little of the variance in PCS and MCS (47). To analyse if the sample size can justify the low $\bar{R}^2$, descriptive statistics is made. In Appendix V: Frequency Tables, it is seen that the answering percent in each variable is above 80 percent, which means that minimum 27,000 persons have answered the question. Only household income and physical activity at work have lower answer percent with 72 percent and 52.2 percent respectively, but even with an answer percent of 52.2 percent 17,840 persons have answered the question, which is still considered a large sample size.

With the large sample size, the low $\bar{R}^2$ is not considered a problem, because the model is believed to give a reasonably precise estimate of the effect of the variables on PCS and MCS. Both variables used in this model are found to be statistically significant. Being male has a positive effect on both PCS and MCS, where the largest effect of being male is seen in MCS (0.102). This is quite interesting since the general idea is that men are stronger than women (which this test also supports), but this analysis shows men in Region Central Jutland generally have a better mental health compared to physical health, and the mental health is stronger than women's.

The outcomes also show aging has a positive effect on mental health, which is different from the expectation, so even though the physical health gets poorer with age persons will mentally feel better.

	PCS $\bar{R}^2$=0.134			MCS $\bar{R}^2$=0.046		
Variables	**Unstandardised β**	**Standardised beta**	**Sig.**	**Unstandardised β**	**Standardised beta**	**Sig.**
(Constant)	60.315		0.000	45.039		0.000
Male	1.035	0.055	0.000	1.886	0.102	0.000
Age	-0.199	-0.365	0.000	0.100	0.185	0.000

Table 4.2: This table shows the outcome of a multiple regression analysis on the effect of male and age on PCS and MCS. Both variables have a statistical significant effect on both PCS and MCS, and the only variable that has a negative effect is the effect of age on PCS.

The results from the multiple regression analyses with all the variables can be seen in table 4.3, where the $\bar{R}^2$ is quite low for PCS (11.4 percent), but better for MCS (50 percent). Also in this analysis the variables seem to explain little of the variance in PCS and MCS, but again the sample size is large and therefore it is believed that the variables give a reasonably estimate of the effect of the variables on PCS (47). On the other hand the variables describe 50 percent of the variance in MCS, which is quite high taking the sample size into consideration, and the model describing MCS is therefore believed to be quite strong.

Since alcohol and income show no statistical significance in PCS and MCS respectively, they are removed from the model using backward selection. By removing the variables from each health outcome all variables become statistically significant as seen in table 4.4. The $\bar{R}^2$ increases a little when removing the variables, which indicates that alcohol and income in PCS and MCS respectively degraded the models.

	PCS $\bar{R}^2$=0.114			MCS $\bar{R}^2$=0.500		
Variables	**Unstandardised β**	**Standardised beta**	**Sig.**	**Unstandardised β**	**Standardised beta**	**Sig.**
(Constant)	64.701		0.000	58.756		0.000
Activity	0.174	0.056	0.000	0.148	0.040	0.000
Male	0.735	0.053	0.000	0.279	0.017	0.008
Age	-0.132	-0.198	0.000	0.078	0.098	0.000
BMI	-0.264	-0.162	0.000	0.052	0.027	0.000
Smoking	-0.126	-0.048	0.000	-0.091	-0.030	0.000
Alcohol	-0.039	-0.010	0.261	-0.063	-0.014	0.042
Income	1.557E-6	0.049	0.000	3.297E-8	0.001	0.896
DUN	0.154	0.058	0.000	-0.229	-0.073	0.000
PSS	-0.142	-0.119	0.000	-0.961	-0.679	0.000
WorkPA	-0.333	-0.090	0.000	0.112	0.025	0.000

Table 4.3: This table provides results from the multiple regression analysis, with the dependent variables PCS and MCS with the chosen variables: Activity, male, age, BMI, smoking, alcohol, income, education, PSS and workPA. The slopes for the variables in PCS are as expected, but not as expected in MCS. It can be seen that the only variable that has no statistically significant effect on PCS and MCS is alcohol and income respectively.

Analysing PCS more thoroughly in table 4.4 it can be seen that the scope coefficients (described by unstandardised β) are as beforehand expected. Taking a look at standardised beta the variable with most positive effect is education (0.057) closely followed by activity (0.054). The small difference of course has relevance, but it still means that activity has (almost) the same influence on PCS as education. The variable that has the largest negative effect on PCS is age (-0.200), which means that the physical health decreases the most with age compared to the other variables.

The slopes of the variables in MCS are not as expected. The variables that have a negative slope, where it is expected to have a positive slope are: Alcohol use, and education. The variables that have a positive slope, where it was expected to have a negative slope are: Age, BMI, and physical activity at work. The standardised beta coefficients show that the variable with the largest positive effect on MCS is age (0.098), and thereby has almost triple effect on MCS compared to activity (0.036). The variable that has the largest negative effect on MCS is, not surprisingly, PSS that is as high as (-0.684). This means that if you score one standard deviation higher in PSS, the MCS will fall by (0.684) standard deviations. All the variables, except income, have a statistically significant effect on MCS.

Activity has almost the same positive effect on PCS and MCS, with a difference of only (0.018), which may indicate that even though physical health gets better with increasing level of activity, the mental health has almost the same benefit from activity. Some of the variables have different slope signs depending which dependent variable that is analysed: PCS or MCS. The variables that differ in βs are: Age, BMI, education, and physical activity at work. Meaning if the variable has a positive effect on PCS it will have a negative effect on MCS and vice versa.

	PCS $\bar{R}^2$=0.118			**MCS** $\bar{R}^2$=0.505		
Variables	**Unstandardised β**	**Standardised beta**	**Sig.**	**Unstandardised β**	**Standardised beta**	**Sig.**
(Constant)	64.761		0.000	59.247		0.000
Activity	0.170	0.054	0.000	0.135	0.036	0.000
Male	0.785	0.056	0.000	0.212	0.013	0.036
Age	-0.136	-0.200	0.000	0.078	0.098	0.000
BMI	-0.265	-0.161	0.000	0.046	0.023	0.000
Smoking	-0.132	-0.050	0.000	-0.095	-0.031	0.000
Alcohol	-	-	-	-0.068	-0.015	0.019
Income	1.726E-6	0.053	0.000	-	-	-
DUN	0.154	0.057	0.000	-0.238	-0.075	0.000
PSS	-0.154	-0.126	0.000	-0.976	-0.684	0.000
WorkPA	-0.344	-0.091	0.000	0.124	0.028	0.000

Table 4.4: Backward selection of the variables in the multiple regression analysis. The variables that are excluded are alcohol and income for PCS and MCS respectively. Now all the variables in the model have a statistically significant impact on PCS and MCS.

From table 4.4 it is possible to find the equation that can be used to calculate PCS and MCS for a random individual in Region Central Jutland, and the result should be close to the PCS and MCS given from Region Central Jutland:

$$PCS = 64.761 + 0.170 \cdot activity + 0.785 \cdot male - 0.136 \cdot age - 0.265 \cdot BMI - 0.132 \cdot smoking + 1.726 \cdot 10^{-6} \cdot income + 0.154 \cdot DUN - 0.154 \cdot PSS - 0.344 \cdot workPA + u$$

$$MCS = 59.247 + 0.135 \cdot activity + 0.212 \cdot male + 0.078 \cdot age + 0.046 \cdot BMI - 0.095 \cdot smoking - 0.068 \cdot alcohol - 0.238 \cdot DUN - 0.976 \cdot PSS + 0.124 \cdot workPA + u$$

In figure 4.5 the effect of each variable as a result of standardised beta can be seen in PCS and MCS respectively. The higher standardised beta, the steeper gradient. Due to the method used to calculate standardised beta, 95 percent of the answers lie between (-1.96) and (1.96), and because the interval is relatively small, a high standardised beta as e.g. PSS causes a quite steep gradient compared to e.g. alcohol in MCS.

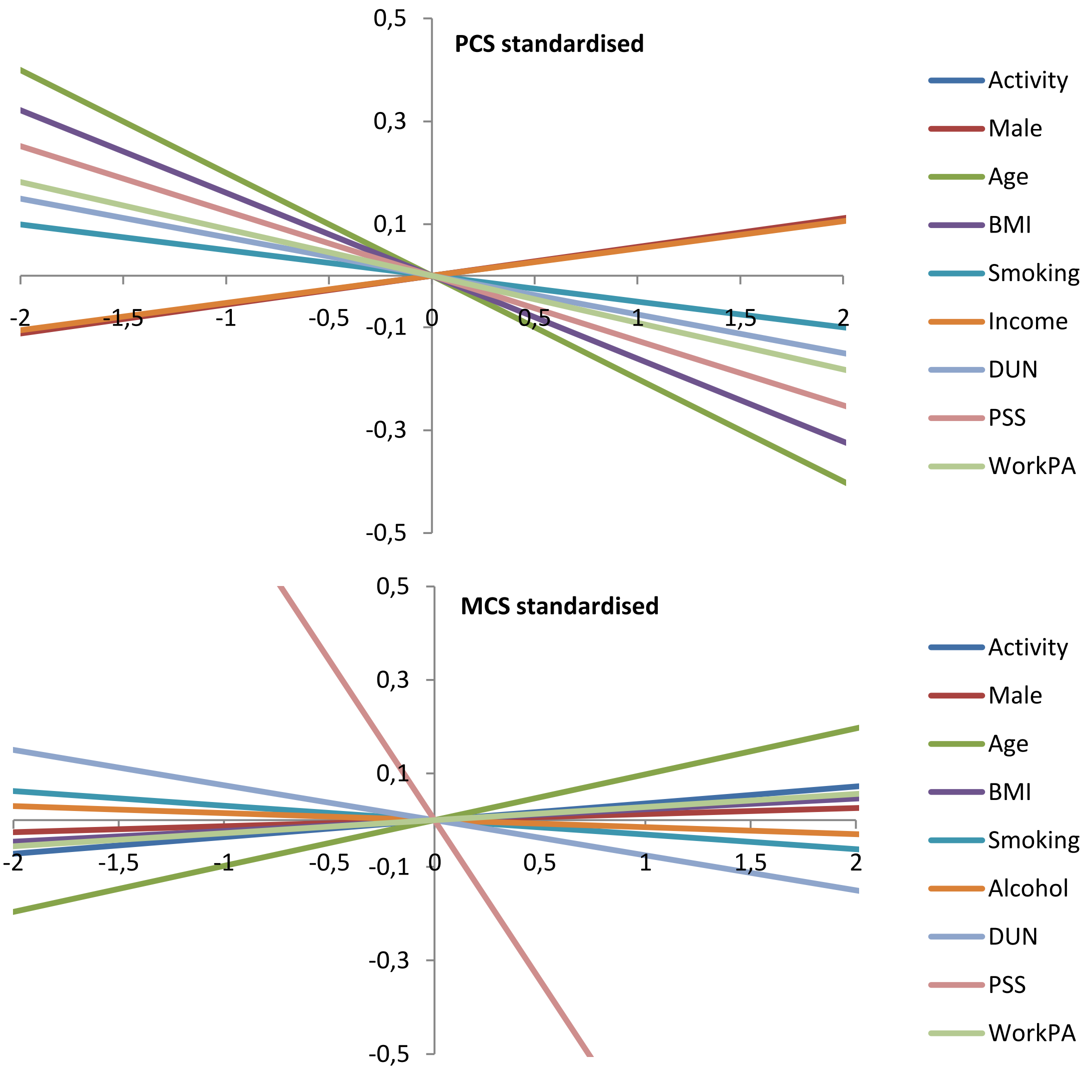

Figure 4.5: In the figures the standardised betas effect on PCS and MCS respectively is seen. The higher the standardised beta coefficients the steeper gradient. A "steeper" gradient causes a higher influence on HRQOL, since an increase of one standard deviation causes a higher increase/decrease in PCS or MCS.

4.2.1 Interaction Effects

The analysis of interaction effects can be seen in table 4.6. The $\bar{R}^2$ is 11.4 percent for PCS and 50.1 percent for MCS. None of the expected interaction effects are found to have a statistically significant effect on PCS. At the same time the statistical significance of activity has disappeared, which indicates multicollinearity. The only interaction that is found is activityPSS in MCS, and the statistical significance disappears from activity. Therefore backward selection is used, to remove the variables that show the least significance (meaning the highest number in the significance column) one by one.

	PCS $\bar{R}^2$=0.114			**MCS** $\bar{R}^2$=0.501		
Variables	**Unstandardised β**	**Standardised beta**	**Sig.**	**Unstandardised β**	**Standardised beta**	**Sig.**
(Constant)	64.556		0.000	59.524		0.000
Activity	0.218	0.070	0.245	-0.044	-0.012	0.800
Male	0.730	0.053	0.000	0.272	0.017	0.010
Age	-0.141	-0.211	0.000	0.075	0.094	0.000
BMI	-0.239	-0.146	0.000	0.064	0.033	0.004
Smoking	-0.126	-0.048	0.000	-0.093	-0.030	0.000
Alcohol	-0.038	-0.010	0.272	-0.060	-0.013	0.052
Income	1.559E-6	0.049	0.000	1.240E-8	0.000	0.961
DUN	0.155	0.059	0.000	-0.231	-0.074	0.000
PSS	-0.142	-0.119	0.000	-1.044	-0.737	0.000
WorkPA	-0.403	-0.109	0.000	0.110	0.025	0.000
ActivityAge	0.002	0.038	0.370	0.001	0.011	0.740
ActivityBMI	-0.007	-0.059	0.226	-0.003	-0.021	0.561
ActivityWorkPA	0.017	0.026	0.222	-	-	-
ActivityPSS	-	-	-	0.022	0.090	0.000

Table 4.6: The outcome from the multiple regression analysis with the expected interaction effects. The only interaction effect is found in activityPSS, which shows a statistically significant impact on MCS.

In PCS the variables that are removed are activityAge, alcohol, activityBMI, and activityWorkPA. The variables are not removed all at once, which means that the level of significance changes, and therefore alcohol is the second variable removed. The order in which the variables are removed can be seen in Appendix III: The Editorial Process In SPSS. The same is done for MCS, where income firstly is removed followed by removing activityAge. No further variables are removed, since the next variable that shall be removed is activity. The outcome can be seen in table 4.7.

The model for PCS is the same as in table 4.4, and the results have therefore already been presented.

The results for MCS show that the next variable that shall be removed is activity, but since this is the primary focus of this project it is not removed. It seems like all the statistical significance has manifested in activityPSS, that has a positive effect on MCS. This indicates that stressed persons will benefit even more from physical activity compared to persons that are "only" active and with this an even higher health profit is gained when being active if you are stressed.

It is not possible to predict what will happen to the significance, if activity and activityBMI are removed from the model, but removing activity is not appropriate in a project where the primary purpose is to analyse the effect of activity on HRQOL.

	PCS $\bar{R}^2$=0.118			MCS $\bar{R}^2$=0.506		
Variables	**Unstandardised β**	**Standardised beta**	**Sig.**	**Unstandardised β**	**Standardised beta**	**Sig.**
(Constant)	64.761		0.000	59.666		0.000
Activity	0.170	0.070	0.245	0.035	0.009	0.793
Male	0.785	0.053	0.000	0.210	0.013	0.037
Age	-0.136	-0.211	0.000	0.078	0.098	0.000
BMI	-0.265	-0.146	0.000	0.064	0.033	0.003
Smoking	-0.132	-0.048	0.000	-0.096	-0.031	0.000
Alcohol	-	-	-	-0.067	-0.014	0.022
Income	1.726E-6	0.053	0.000	-	-	-
DUN	0.154	0.057	0.000	-0.240	-0.076	0.000
PSS	-0.154	-0.126	0.000	-1.056	-0.740	0.000
WorkPA	-0.344	-0.091	0.000	0.124	0.028	0.000
ActivityAge	-	-	-	-	-	-
ActivityBMI	-	-	-	-0.005	-0.035	0.319
ActivityWorkPA	-	-	-	-	-	-
ActivityPSS	-	-	-	0.021	0.086	0.000

Table 4.7: Backward selection outcome for interaction effects. The outcomes for PCS are the same as in table 4.4. The outcomes for MCS show statistical significance in all variables except from activity and activityBMI.

4.3 Summary of Chapter II

Firstly the inclusion criteria are determined. Persons under the age of 18 have been excluded in the dataset since this project focuses on the recommendations to the adult population. The variables are categorised preparing them to use in multiple regression analysis. The factor analysis shows, that the 12 items forms two factors, each representing PCS and MCS respectively, as expected. Furthermore the correlation test shows that PCS is highly correlated with factor 1 at a statistically high level. The same is true for MCS and factor 2. With this test PCS and MCS have been validated for the use in the further analyses.

The multiple regression analysis results regarding PCS are as expected, since activity has a positive effect together with being male, having a high income and a higher education. At the same time the variables age, BMI, smoking, PSS and workPA have a negative effect on PCS. No interaction effects are found.

The results regarding MCS on the other hand are not as expected since age, BMI, alcohol, DUN and workPA have other signs than expected. This means that activity, being male, increasing age and BMI and physical activity at work have positive effect on mental health, whereas smoking, alcohol, DUN and PSS have negative effects. The interaction effects are only present in activityPSS.

Even though the results from the factor analysis show the use of PCS and MCS as valid, other questions arise about using SF-12 in the outcome for HRQOL, and why do the results from multiple regression analysis and interaction effects give the outcome presented? Those topics are treated in Chapter III: Discussion & Conclusion.

Chapter III
Discussion & Conclusion

5. Discussion

This section discusses the whole project from theory to methods and results. The results referred to can be found in Chapter II: Methods & Results. The tables that are used in the discussion of the results are table 4.0, 4.1, 4.4, 4.5 and 4.7.

The purpose of this project was to investigate how physical activity affected HRQOL in a sample from Region Central Jutland, and if the HRQOL measure was valid. Firstly the validity of HRQOL was determined followed by multiple regression analysis to show how physical activity affects HRQOL.

5.1 The Chosen HRQOL Measure

Firstly it was necessary to identify the outcome, which should be used as a measure for HRQOL. Since the questionnaire contained questions from SF-12, this was the chosen measure. The SF-12 was a good choice in this particular questionnaire, because it was a long questionnaire containing many questions, and by using the SF-12 questions an additional elongation of the questionnaire was avoided. The purpose of the whole questionnaire was to cover welfare, health and disease, but had the purpose of the questionnaire been to cover HRQOL, it might been beneficial to use SF-36 instead of SF-12, since SF-36 covers more aspects of HRQOL. Even though the SF-36 would cover more health-related aspects than SF-12, the correlation between the two short-form health surveys was found to be highly correlated both in healthy (22,23) and unhealthy (24-27) persons. On this basis, SF-12 was considered a valid choice of measure in this project.

In the questionnaire the formulations of some of the SF-12 questions were reverse scored. This reverse scoring was taken into consideration in the design of the questionnaire, because the length of the questionnaire might have caused the person to mark the same box through the whole questionnaire and the reverse scoring made the person aware of answering each question correctly. Furthermore the questions had been arranged differently from each other where some of the options were arranged horizontally and others vertically, which also contributed to the awareness of the answering possibilities. The questionnaire from Region Central Jutland contained five answering possibilities (ranging from "all the time" to "none of the time"), which was different from the original SF-12 questionnaire, where some of the questions only gave only two answering possibilities as yes/no. The use of a scale in the questionnaire from Region Central Jutland would cover more aspects of an individual's HRQOL, and therefore this scale was considered beneficial in the aim to cover HRQOL in this project.

In the questionnaire some dimensions were represented twice (PF, RP, RE, and MH), since those items were found to have the highest influence on SF-36 using regression models, whereas others (GH, BP, VT, SF) were only represented once. This gave a total of 12 items, six items representing each SF-12 score. The factor analysis of this project confirmed that that the 12 items represented two factors, where the items from PCS represented factor 1 and the items from MCS represented factor 2. It was also analysed if the reverse scoring was expressed in the factor analysis, and as expected some of the items had a negative effect on the component summary scores. The objectives for Ware and colleagues (21) were to reproduce PCS-36 and MCS-36 with a score explaining at least 90 percent of the variance in PCS-36 and MCS-36 in a shorter form, that only took two minutes to answer and fitted in one single page. To fulfil the objectives only 12 items were necessary to use, and the remaining 24 items were excluded. In the questionnaire some questions regarding health were inserted between the SF-12 questions, and the typing was large to provide easier reading, therefore the SF-12 questions did not fit in one single page, but this was only considered an advantage, since persons in all age groups would be able to read the questions.

Some authors (48-50) have criticised the rotation method of which the scores in SF-36 are made. The authors postulated that PCS and MCS were not orthogonal, from which it might be concluded, due to the high correlation between SF-36 and SF-12 (22-27), that the PCS and MCS in SF-12 were not orthogonal either. In the results of this project it was found that the component summary scores were statistically significantly correlated, but with a very low value (-0.018), which practically means that no correlation was present between PCS and MCS in the sample. By that the rotation method was not considered a problem. The criticising authors might have experienced that their calculated factor scores did not fit into the unknown algorithm, and therefore the calculated PCS and MCS could not be used on their sample. The purpose of using factor analysis was to find two factors and to analyse if they correlated with the calculated PCS and MCS, because it was not known if the sample fitted into the unknown algorithm. Through this project, without knowing the algorithm, the 12 items gave two factors that satisfied the demand of six items in each factor when using factor analysis. The use of the calculated SF-12 scores in the sample in this project were according to the results a very good choice, since a statistically high correlation between the factors and the component summary scores (98.7 percent ($p<0.000$) and 99.7 percent ($p=0.002$) for PCS and MCS respectively) was found. Therefore, the SF-12 measures were considered valid and useable in this project and with the sample.

5.2 The Questionnaire

The basis of the analyses was a dataset from Region Central Jutland. A consideration of analysing the questionnaire was how the responses represented the whole population, and if the responses misrepresented the population making selection bias. Persons, who had difficulties physically or mentally, were probably not represented in the statistics, since it was expected that this group in a lesser degree answered the questionnaire. Persons physically disabled or in another way physically injured might not have answered the questionnaire simply because they were unable to write or hold a pen. In some circumstances a helper might have assisted those persons to answer the questionnaire, which would have been beneficial for the investigation, but if no one could have helped persons, who would score a low PCS, they were not represented in the analyses. Therefore the PCS in the sample might be higher than the PCS in the population. To analyse if the PCS was higher in the sample compared to the general population it would have been necessary to have an even higher sample size representing the population of Region Central Jutland. The PCS level in the sample was not used in the analyses and since the focus of this project was to analyse how HRQOL was affected by physical activity, it was possible to assume that also persons with a low PCS would benefit from an increasing level of activity. This assumption was supported by (51,52), who also found that PCS would increase with physical activity. The questionnaire consisted of 23 pages with a total of 109 questions, where some of the questions consisted of more than one. Since the questionnaire would take more than the recommended length of ten to twenty minutes (53) to complete, it was not expected that persons with mental difficulties or mental disorders would respond in the same extend as they were present in the whole population. The MCS in the population was therefore expected to be lower compared to the MCS in the sample. It was not possible to compare the MCS of the sample with the MCS of the population, but in the same way as with the PCS this was not considered a problem, and it was also expected that persons with mental difficulties or mental disorders would experience a better mental health if they were active 30 minutes a day.

The questionnaire was written in Danish, which might be a problem for the randomly selected persons, who could neither read nor understand Danish. They would not be able to answer the questionnaire unless they were helped by e.g. an interpreter to answer the questionnaire. This might have caused persons with another ethnical background to be represented in a lesser degree in the sample compared to the representation in the population. Since persons with another ethnical background were expected to score the same PCS and MCS as Danes it would not affect the results. By sending questionnaires to randomly selected persons in the Region trying to reach all social stratums the difference in PCS and MCS were attempted to be avoided, but the SF-12 scores in this analysis were still considered to be better than the scores would have been if every person in the Region had answered the questionnaire.

5.3 Basis of The Project

To describe HRQOL it was necessary to find variables with a connection to it. By systemically searching the literature, it was possible to identify variables included in studies analysing the connection between HRQOL and physical activity. The articles were found by using different search words in the chosen search engines. Only five studies that analysed the effect of activity on HRQOL in a general population were found, and they were all included in this project. The following variables were found on the basis on the five studies: Activity, gender, age, BMI, smoking status, alcohol use, household income, education, PSS and physical activity at work. A variable that was not included in this project was marital status that earlier was found to have a positive effect on both HRQOL and physical activity (31), and if this variable was included in this project, it was believed to have a positive effect on PCS and MCS as well. With the absence of marital status the results still showed that being physically active would improve HRQOL, and it was not believed that marital status would change this result.

The methods used to analyse the effect of the chosen variables on PCS and MCS was firstly multiple regression analysis and secondly multiple regression analysis with interaction effects. The multiple regression analysis was used because the data was cross sectional, and by using cross-sectional data, it was not possible to establish cause and effect, but rather to analyse how physical activity affected HRQOL. The assumption when using multiple regression analysis was that the variables had a linear connection to PCS and MCS, but this might not have been the case for all the variables. Some of the effect from the variables might have had a decreasing or increasing effect with an increasing variable, which could have been investigated by using the product of the variable multiplied with the square of the variable. The studies used in this project, did not analyse if the effect of the variables were even more increasing with an increasing x, and neither was done in this project.

5.4 The Multiple Regression Analyses

The use of a questionnaire made it possible to reach a large amount of the population. 64 percent of 52,400 persons answered the questionnaire, which is a very large sample that (taking section 5.2: The Questionnaire into account) represented the better part of Region Central Jutland. In the dataset, 29,992 persons answered all PCS questions, which correspond to 87.8 percent. 29,989 persons answered all questions regarding MCS, which also corresponded to 87.8 percent.

Even though the $\bar{R}^2$ was quite low for PCS (0.118) and higher for MCS (0.505) in the multiple regression analyses, the amount of persons, who answered the questions in the variables was high. The variable that was represented less was physical activity at work, where "only" 17840 persons answered this question. Other variables were represented in a higher degree, which meant that both models made a quite precise estimate of the effect of the variables on PCS and MCS in Region Central Jutland. Furthermore the questionnaire was self administered, which might cause some biases since people might not answer correctly either because they did not know the right answer to the question or because they thought their score was different from the actual objective score. Persons might not have answered wrong intentionally, but they wanted to answer so the outcome would make them look best (54).

The primary outcome of interest was how the level of activity affected HRQOL. To analyse this it was possible to use three out of eight different questions regarding "exercise in everyday life". The question used in this project was: "*<u>How many days a week are you active for at least 30 minutes a day?</u> Include moderate or strenuous physical activity where you breathe faster, move your muscles and use your powers - e.g. for exercise or competitive sport, heavy gardening, brisk walking, bicycling at a moderate or fast pace or physical strenuous work. Include both work and leisure time.*". The possible answering possibilities stretch from “no days” to “every day”. This question made it possible to analyse if the first part of the recommendations, which was to be active for at least 30 minutes a day seven days a week at a moderate level, were fulfilled. The question was elaborated, so the persons understood, what was meant with “active”, and it was easy to understand, which type and what kind of harshness of physical activity the question was directed at. One thing that might have caused problems in the persons that answered the questionnaire, was that their level of activity might be different from week to week, which could make it difficult to answer, if the answer should be every day or six days a week. This was however not believed to be a problem, since it was expected that the persons normally would choose the answer that generally fitted the best, and persons who were generally active every day would probably choose “every day”. In the categorisation of activity as a variable the order from “no days” to “every day” was logically increasing, and therefore each day got the value they had meaning that “no days” was zero, “one day” was one and so on. The categorisation could have been made so “one day” was 30, “two days” were 60 and so on, thereby using the number of minutes each day, but this would not be appropriate since the question states “at least 30 minutes”, some persons might be active for more than 30 minutes a day, which is why this method was not used.

Another question that could have been used in the analyses of exercise was the question "*If you look at the past year, what would you say is best suited as a description of your physical activity during leisure time?*". This question gave four possible answers: "Hard training and competitive sports regularly and several times a week"; "Do exercising sports or performing heavy gardening or similar at least four hours per week"; and "Reads, watch television or other sedentary activity". The problem with this question was that persons, who were active for e.g. 3.5 hours a week, thereby fulfilling the recommendation regarding physical activity 30 minutes a day, had to describe themselves as sedentary cf. the questionnaire. This question particularly could have been used if the phrasing of the answering possibilities were different so the boundary between “sedentary” and physical activity four hours a week so persons who were active less than four hours a week did not feel that their level of activity were neglected. This might have led to dishonest answers, and therefore this question was not used.

The second part of the recommendations was to be active for at least 20 minutes twice a week at a more demanding level could have been analysed by using the question: "*Do you participate in sports or do you participate regularly in other activities that provide exercise in your leisure time?*" This question left two answering possibilities "no" and "yes" with the sub question: "*if yes, what kind of sport*". The formulation of this question was quite vague because "regularly" was subjective, so one person might think that "regularly" was once a week whereas others might think it was four times a week. This problem could have been avoided with an additional question that asked: "*If yes, how many times a week do you do sports?*". With the question no length or regularity of the exercise was answered leaving little options of analysing in this project.

The multiple regression analyses showed a positive effect of physical activity in both PCS and MCS as expected. In PCS physical activity was the variable with almost the largest positive effect, when looking at the standardised coefficient beta (0.054), but DUN was the one with the largest positive effect (0.057) closely followed by gender (0.056). Analysing standardised beta, age (0.098) had a higher influence than activity (0.036) on MCS, but activity still had a positive effect as expected. Physical activity had a larger effect on PCS compared with MCS, which made good sense, since physical activity would increase many functions in the body including increasing blood volume (13,55), increasing physical fitness (5,13), and increasing strength in muscles, bones and tendons (5,7,13). At the same time, neurotransmitters would be released during exercise, which would result in a generally better mood (5,7-9), and the results showed that even though the general mood would increase (0.036), the effect of physical activity had the largest impact on the physical part of the body.

Information regarding gender was assessed with the question: "Are you", with the options "male" and "female". Not every person answered this question (0.3 percent missing data, see table C in Appendix V: Frequency Tables), which might be because of misunderstanding the question or because the person was unable to answer due to e.g. transsexuality. It would have been a good idea to make an extra box that stated "none of the above", "other" or "do not know", so every person could have answered. Due to the low percentage of missing data it was not thought that this question would have changed the results. Gender was made a dummy variable, where the question was "Are you male?" where 1 indicated "yes" and 0 indicated "no". This was done because no logical order between the genders existed. As expected the results showed a benefit of being male in both PCS and MCS, but also that the effect of being a male was higher in PCS (0.056) compared to MCS (0.013) when looking at standardised beta. This larger effect in PCS than MCS in males might relate to the paragraph above, where physical superiority would have higher effect on HRQOL than mental superiority.

Age was found by using the last four numbers of the birth date (DD-MM-YYYY[16]) addressed by the person when fulfilling the question "*When were you born?*", thereby asking about date, month and year. The only number used in this project was the birth year that was subtracted from 2010 (the year of the investigation), and if the birth year was stated wrong or if the number was read wrong by the machine the age would off course also be wrong. Due to the logical continuous distribution, age was not divided into categories for the use in the multiple regression analysis, which made it possible to see the effect of increasing age on HRQOL. On the other hand age was categorised to see the distribution of age groups. The age group that was represented the most was persons between the age of 50 and 67, and 98.3 percent answered the question as seen in Appendix V: Frequency Tables. The unstandardised β showed that age had a negative effect on PCS (-0.136) as expected, possibly due to a general degeneration of the body with following diseases. Most surprisingly increasing age meant a better mental health with a slope coefficient of (0.078), in fact, age was the variable that had the largest effect on mental health when looking at standardised beta (0.098).

Some questions typically answered less honestly were information regarding height and weight (56) simply because people scored their heights larger and their weight lower than it was. By increasing the height and lowering the weight, a lower BMI was achieved in the calculation, and the BMI in the sam-

[16] DD-MM-YYYY= date, month, year

ple was therefore believed to be lower than it would be if the sample were measured objectively with a weight and measuring tape. It was however not believed to affect the results from the multiple regression analysis, since it only analysed how HRQOL was affected by BMI, and thereby the results would only be affected if they included descriptive statistics. Thus an increase in BMI would still affect HRQOL, even though the BMI was believed to be practically higher. The questions regarding height and weight sounded: "*What is your height (without shoes)?*" and "*How much do you weigh in whole kilograms (without clothes)?*". The questions were plain and easy to understand, but the height question might have included "whole centimetres", so the question revealed the unit the answer should be stated in. On the other hand, an example was presented, which might have simplified the understanding of the question. The BMI was calculated from weight and height as showed in Appendix IV: Description Of The Variables, using the international equation used to calculate BMI. The BMI was not categorised for the use in the multiple regression analyses, since there was a logical continuous order in a scale variable, but it was categorised in Appendix V: Frequency Tables, where the distribution of the BMI categories can be seen. Even though the BMI was thought to be lower in the sample compared to the population, it was not expected that the results regarding the slope would have been different. PCS was statistically significantly negatively affected with an increasing BMI (-0.265), which might be related to e.g. too much load causing pain and discomfort in tendons, more incidences of lifestyle-related diseases and the fact that overweight and obesity was related to inactivity (6) that might decrease the muscle tonus. Analysing the standardised beta, BMI had a higher negative effect (-0.161) than activity had a positive effect (0.054) on PCS, which meant it was better to be normal weight than being active. This was different than the result from Blair and colleagues, who found it was beneficial to mortality and morbidity to be obese and active compared to be normal weight and inactive (6). The difference between this project and Blair and colleagues was that this project did not investigate the effect of inactivity, and perhaps the negative effect of inactivity would be higher than the positive effect of activity, as seen with BMI in this project. The MCS was positively (0.046) affected by BMI, which was not expected. This finding was different from Kolotkin, Meter and Williams (57) that found that persons with high BMI experienced a lower QOL than persons with a normal BMI, and the question to why BMI was associated with a better mental health in Region Central Jutland was left open.

Smoking was believed to interact with HRQOL and therefore this variable was included in the analyses. The question that was used in this project to clarify the effect of smoking was: "*Do you smoke?*". To this question, it was possible to choose five solutions: "Yes, every day"; "Yes, more than once per week"; "Yes, more rarely than once per week"; "No, I have stopped smoking"; and "No, I have never smoked". These five questions were divided into four groups as described in table 3.0. This procedure made it possible to enhance the effect of smoking in the multiple regression analysis. The amount of smoking was generally underestimated in self-reports (58) and this would possibly have been a problem if the question had been: "How much do you smoke every day?". Since it was possible for the persons who answered the smoking question to estimate the amount of smoking instead of giving the precise amount, the answers were believed to be true. Smoking was statistical significant negatively associated with both PCS (-0.132) and MCS (-0.095) as the unstandardised β showed. It was seen that smoking had a higher negative effect on PCS (-0.050) than MCS (-0.031), which might be due to the physical side effects as e.g. poorer physical fitness due to a decrease in lung function and decreasing bone mass (34), which might turn into osteoporosis. It was expected that smoking would have a positive effect on MCS, due to a relaxing effect, but this was not the case in the sample of this project. Perhaps smoking would correlate with other factors as e.g. stress and thereby smoking would have a negative effect since persons that were stressed would smoke, but an analysis of the connection between smoking and other variables were not carried through in this project.

The use of alcohol was assessed with the question: "*How many times a week do you consume alcohol*?". To answer this question it was possible to give seven answers ranging from zero/one day per week to seven days per week. The outcomes were kept in the categories as they were in the questionnaire where zero to one time per week was 1, two times a week was 2 and so on. The use of alcohol did not affect PCS with statistical significance, but it was still believed to some degree the physical health was affected if a person did drink more than the recommended units per week. Alcohol consumption had a statistically significant negative effect on MCS (-0.068), which is supported by Saarni and colleagues (59), that also found that alcohol was negatively associated with QOL.

The household income was assessed with the question: "*How high was your and your household income in 2009 before taxes and other deductions? Gross income*". To this question it was possible to answer in eight intervals, which were also used in this project, but with the middle value of each interval to get a stringent distribution. The middle value was used since it was believed that the distribution within each interval was uniform, and therefore it was possible to use the middle value. Income was found to have a statistical significant positive effect on PCS ($1.726E^{-6}$) as expected. The very low unstandardised β was understandable since this low value should be multiplied with a very large number (e.g. 312,000) when using the equation found in Chapter II: Methods & Results. No statistically significant effect was found on MCS.

The educational level expressed by DUN was found through public authority, and was therefore considered valid and objective. It was divided into the same groups as they already were except from group 9 (out of category) that was made "missing data", because persons in that category might have been educated in other countries than Denmark. To get a stringent distribution of DUN, the middle value of years of education was used to express an increasing effect of education. In the analyses, only persons over the age of 18 were included, and the descriptive statistics showed that most of the persons (44.3 percent) were in group 11.5, which meant that they have had 11 or 12 educational years. Unstandardised β showed that DUN was positively associated with PCS (0.154) as expected but negatively associated with MCS (-0.238), which was not expected. Other studies (31,60,61) found that education had a positive effect on HRQOL, thus the finding in this project might be an expression of correlation between education and another variable as e.g. stress. The standardised beta showed that the effect of DUN was more negative in MCS (-0.075) than positive in PCS (0.057), which meant that education would have a larger negative effect on your mental health than positive effect on the physical health.

A variable as stress was assessed via ten questions, where the person should answer different questions regarding experiences within the last month e.g. the question "*How often have you felt nervous and stressed?*" and then rate it from "never" to "very often". A never would give a score of zero and a "very often" would give a score of four, and in between these options three other options existed, and in some of the questions would be reversed so a "never" would score four. This reverse scoring made the person aware of the formulation of each question. The ten answers were summed, which gave each person a score between zero and 40, the higher the score the higher level of experienced stress. The problem with the ten questions was that they were very personal and sensitive, and some persons might not have answered honestly because they would chose the option that made them look best (54). Furthermore the rating as "almost never", "sometimes", "often" and "very often" was difficult to distinguish between, because it might be difficult to determine if "often" is eight times within a month or 12 times within a month. By this PSS was a very subjective measurement, because some persons might only have felt "sometimes" stressed, where others might have experienced the same level of stress as "often" stressed, but since the purpose of using PSS was to measure experienced stress, the difference in perceived stress was expressed in the total PSS score. As expected an increasing level of stress was associated with a lower HRQOL both physically (-0.154) and especially mentally (-0.976), and the standardised beta showed that stress was the variable with the highest negative impact on mental health (-0.684). This finding was quite interesting since the focus of the Danish National Board of Health was to make people quit smoking, achieving a normal BMI, to eat healthy, to drink less and to be more active, but it seemed that the persons in Region Central Jutland would achieve an even better mental health if they were less stressed.

No literature regarding answering the physical activity at work was found, but taking the argument from Donaldson and Grant-Vallone into consideration (54) some persons might have thought their activity at work was higher than it really was. Physical activity at work affected PCS statistically significantly negatively (-0.344), which might be due to the assumption that physical activity at work might be backbreaking. Even though it was expected that physical activity at work would have a negative effect on MCS this was not the case, thus physical activity at work showed a statistically significant positive effect (0.124) on your mental health. This positive effect on MCS might have been experienced due to the neurotransmitters released when being active and perhaps some persons experienced a positive social effect of being active at work. The standardised beta showed that workPA was not the factor that had the highest impact on neither PCS (-0.091) nor MCS (0.028), and persons who had an active job would possible keep it even though it physically had a negative effect.

5.4.1 Interaction Effects

The interaction effects on PCS were not as expected and the outcome showed the same as the backward selection of the multiple regression analysis. By this the effect of the interaction effects were not as expected, which meant that age, BMI and workPA would not have a better effect with activity on PCS. The results regarding MCS showed that activityPSS had a statistically significant connection to MCS, and at the same time, the significance was not found in activity. This might mean that the effect of physical activity would be even stronger if a person experienced stress. It could also mean that physical activity had a preventive role in relation to stress.

5.5 The Recommendations

The recommendation from the Danish National Board of Health states:

"***That all adults are physically active at least 30 minutes of moderate intensity, preferably all days. The 30 minutes can be divided into smaller periods, e.g. 15 minutes in the morning and 15 minutes later or 3 times 10 minutes during the day***."(62)

This recommendation has been the focus of this project, and it is shown that activity in general has a positive effect on HRQOL if you are from Region Central Jutland.

Another part of the recommendation regarding activity is:

"***The Danish National Board of Health further recommends that all adults at least twice a week, promote and maintain their fitness, muscular strength, flexibility and bone health. Training must be of high intensity of 20-30 minutes duration.***" (62)

This issue is not investigated in this project, which already has been discussed.

Reading further in the homepage for the Danish National Board of Health, they state that it is of importance that physical activity is a natural part of the daily life instead of being a hard task to overcome. Even though the analyses of this project do not analyse the intensity of the physical activity, this suggest that the intensity is not as important as being active. The accumulation and the intensity of the exercise has not been investigated to a great extend (63), and it is believed that incorporating activity may be difficult to comprehend every day, compared to fewer days in the week. At the same time World Health Organisation (WHO) has other recommendations regarding exercise than the Danish National Board of Health, and WHO state that:

- *"Adults aged 18–64 should do at least 150 minutes of moderate-intensity aerobic physical activity throughout the week or do at least 75 minutes of vigorous-intensity aerobic physical activity throughout the week or an equivalent combination of moderate- and vigorous-intensity activity.*
- *Aerobic activity should be performed in bouts of at least 10 minutes duration.*
- *For additional health benefits, adults should increase their moderate-intensity aerobic physical activity to 300 minutes per week, or engage in 150 minutes of vigorous-intensity aerobic physical activity per week, or an equivalent combination of moderate- and vigorous-intensity activity.*
- *Muscle-strengthening activities should be done involving major muscle groups on 2 or more days a week."* (64)

The recommendation of 150 minutes can be converted into physical activity at a moderate level in 30 minutes **five** days a week, which is different from the Danish recommendations. This difference may be because purpose of the recommendations from the Danish National Board of Health, is to incorporate activity in the daily life among the Danish population. Some persons may think that physical activity five days a week is more manageable than seven days a week, because this will give the opportunity to fulfil the recommendations, if they have 15 to 20 minutes of bicycling time to work every day. Furthermore WHO recommends strengthening of muscles in major parts of the body on two or more days a week, and the Danish National Board of Health has a similar approach in their recommendations in the secondary part of the recommendation. Even though the recommendations from the Danish National Board of Health are different than the ones from WHO, they still resemble each other, and the aim of all recommendations are to increase the HRQOL internationally.

No investigation (that I am aware of) has yet shown if the effect on HRQOL is the same if you are active 30 minutes at a time, if the effect is the same with accumulated minutes distributed more times a day, and if exercise twice a week with high intensity is better for your HRQOL. Thus, investigation in activity is necessary to improve the knowledge about its effect in a general population and not only in persons already experiencing low HRQOL due to current diseases.

Even though activity in the daily life is no longer a survival criterion the aim of living is not just to be alive, but also to feel well with as many healthy days as possible, because this will cause the HRQOL to increase and by that, persons feel better. Increasing the level of activity seems to be a good place to begin, if a person wants to increase the HRQOL.

CONCLUSION

Because physical activity is no longer a survival criterion, it is important to have a good HRQOL both physically and mentally. To achieve a better HRQOL this project has shown that physical activity may be one of the solutions to how this can be achieved. The PCS and MCS that have been calculated with an unknown algorithm are statistically significantly correlated with the two outcomes from the factor analysis. This means that PCS and MCS in this project are validated as outcome measures and that these outcomes can be used in future analyses in Region Central Jutland. The recommendations regarding exercise 30 minutes a day have been analysed and it is found that activity has a statistically significant positive effect on both PCS and MCS, and even though persons with low HRQOL possibly have not been included in this study to a great extent, they are also believed to benefit from an increasing level of physical activity. By this, I state that every person, with no regard to HRQOL, will benefit from an increasing level of activity. Nevertheless, the most important issue is to incorporate physical activity in daily living, so the physical activity does not become a necessary evil.

Researchers cannot predict what is important for other individuals, but they can try to investigate how HRQOL may be affected by other factors in daily life, and from the findings make recommendations in the aim to increase HRQOL among the population.

"*One of the greatest challenges facing health care providers and researchers is the need to ensure that the outcomes measured reflect the medical services' goals and that the measures are relevant to the population group* " (65)

Bibliography

(1) Wikipedia contributors. Industrialiseringen i Danmark , 2011.

(2) Indenrigs- og Sundhedsministeriet. Sund hele livet.ashx (application/pdf Objekt) Available at: http://www.sum.dk/Aktuelt/Publikationer/Publikationer/~/media/Filer - Publikationer_i_pdf/2002/Sund hele livet.ashx. Accessed 4/12/2011, 2011.

(3) Wiggins MS, Simonavice EM. Cancer prevention, aerobic capacity, and physical functioning in survivors related to physical activity: a recent review. Cancer Manag Res 2010 Jun 9;2:157-164.

(4) Sundhedsstyrelsen. Fysisk aktivitet og sundhed - En litteraturgennemgang - Resume og anbefalinger Available at: http://www.sst.dk/publ/publ2001/fysisk_aktivitet_sundhed/kap02.htm. Accessed 4/12/2011, 2011.

(5) Guyton AC, Hall JE. Textbook of Medical Physiology: With STUDENT CONSULT Online Access (Guyton Physiology) 11th ed.: Saunders; 2005.

(6) Blair SN, Brodney S. Effects of physical inactivity and obesity on morbidity and mortality: current evidence and research issues. Med Sci Sports Exerc 1999 Nov;31(11 Suppl):S646-62.

(7) Martini FH. Fundamentals of Anatomy & Physiology (7th Edition) (MyA&P Series) 7th ed.: Benjamin Cummings; 2005.

(8) Foley TE, Fleshner M. Neuroplasticity of dopamine circuits after exercise: implications for central fatigue. Neuromolecular Med 2008;10(2):67-80.

(9) Christie BR, Eadie BD, Kannangara TS, Robillard JM, Shin J, Titterness AK. Exercising our brains: how physical activity impacts synaptic plasticity in the dentate gyrus. Neuromolecular Med 2008;10(2):47-58.

(10) Jiang Y, Hesser JE. A comparison of depression and mental distress indicators, Rhode Island Behavioral Risk Factor Surveillance System, 2006. Prev Chronic Dis 2011 Mar;8(2):A37.

(11) do the evolution — The League of Ordinary Gentlemen Available at: http://ordinary-gentlemen.com/blog/2009/08/27/do-the-evolution/. Accessed 5/10/2011, 2011.

(12) Shono N, Urata H, Saltin B, Mizuno M, Harada T, Shindo M, et al. Effects of low intensity aerobic training on skeletal muscle capillary and blood lipoprotein profiles. J Atheroscler Thromb 2002;9(1):78-85.

(13) Sundhedsstyrelsen. Fysisk aktivitet – håndbog om forebyggelse og behandling 2004; Available at: http://www.sst.dk/publ/publ2005/cff/Fysisk_aktivitet_haandbog2udg/fysisk_aktivitet_haandbog2udg.pdf. Accessed 4/29/2011, 2011.

(14) Sundhedsstyrelsen. aktivi.pdf (application/pdf Objekt) Available at: http://www.sst.dk/publ/publ2001/fysisk_aktivitet_sundhed/aktivi.pdf. Accessed 4/12/2011, 2011.

(15) Brown DW, Balluz LS, Heath GW, Moriarty DG, Ford ES, Giles WH, et al. Associations between recommended levels of physical activity and health-related quality of life. Findings from the 2001 Behavioral Risk Factor Surveillance System (BRFSS) survey. Prev Med 2003 Nov;37(5):520-528.

(16) Brown DW, Brown DR, Heath GW, Balluz L, Giles WH, Ford ES, et al. Associations between physical activity dose and health-related quality of life. Med Sci Sports Exerc 2004 May;36(5):890-896.

(17) Andersen LB, Schnohr P, Schroll M, Hein HO. All-cause mortality associated with physical activity during leisure time, work, sports, and cycling to work. Arch Intern Med 2000 Jun 12;160(11):1621-1628.

(18) Region Midtjylland. VOKSEN Spørgeskema 2010 (web).pdf (application/pdf Objekt) Available at: http://www.centerforfolkesundhed.dk/files/Sundhed/Folkesundhed/Sundhedsfremme Århus/HVORDAN HAR DU DET/HVORDAN HAR DU DET_2010/VOKSEN Spørgeskema 2010 (web).pdf. Accessed 4/12/2011, 2011.

(19) Rådet for Socialt Udsatte. socialt_udsattes_sundhed_litteraturgennemgang.pdf (application/pdf Objekt) Available at: http://www.si-folkesundhed.dk/upload/socialt_udsattes_sundhed_litteraturgennemgang.pdf. Accessed 4/12/2011, 2011.

(20) Amir M, Lewin-Epstein N, Becker G, Buskila D. Psychometric properties of the SF-12 (Hebrew version) in a primary care population in Israel. Med Care 2002 Oct;40(10):918-928.

(21) Ware J,Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233.

(22) Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE, et al. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol 1998 Nov;51(11):1171-1178.

(23) Wee CC, Davis RB, Hamel MB. Comparing the SF-12 and SF-36 health status questionnaires in patients with and without obesity. Health Qual Life Outcomes 2008 Jan 30;6:11.

(24) Iraurgi Castillo I. A comparative analysis of SF-12 with the SF-36 among patients in methadone treatment]. Adicciones 2007;19(1):59-67.

(25) Lacson E,Jr, Xu J, Lin SF, Dean SG, Lazarus JM, Hakim RM. A comparison of SF-36 and SF-12 composite scores and subsequent hospitalization and mortality risks in long-term dialysis patients. Clin J Am Soc Nephrol 2010 Feb;5(2):252-260.

(26) Muller-Nordhorn J, Roll S, Willich SN. Comparison of the short form (SF)-12 health status instrument with the SF-36 in patients with coronary heart disease. Heart 2004 May;90(5):523-527.

(27) Pickard AS, Johnson JA, Penn A, Lau F, Noseworthy T. Replicability of SF-36 summary scores by the SF-12 in stroke patients. Stroke 1999 Jun;30(6):1213-1217.

(28) Joseph F. Hair, Ronald L. Tatham, Rolph E. Anderson, William Black. Multivariate Data Analysis . 5th ed.: AbeBooks.

(29) Stewart KJ, Turner KL, Bacher AC, DeRegis JR, Sung J, Tayback M, et al. Are fitness, activity, and fatness associated with health-related quality of life and mood in older persons? J Cardiopulm Rehabil 2003 Mar-Apr;23(2):115-121.

(30) Lee SW, Mancuso CA, Charlson ME. Prospective study of new participants in a community-based mind-body training program. J Gen Intern Med 2004 Jul;19(7):760-765.

(31) Morimoto T, Oguma Y, Yamazaki S, Sokejima S, Nakayama T, Fukuhara S. Gender differences in effects of physical activity on quality of life and resource utilization. Qual Life Res 2006 Apr;15(3):537-546.

(32) Ho TJ, Liang WM, Lien CH, Ma TC, Kuo HW, Chu BC, et al. Health-related quality of life in the elderly practicing T'ai Chi Chuan. J Altern Complement Med 2007 Dec;13(10):1077-1083.

(33) Bize R, Plotnikoff RC. The relationship between a short measure of health status and physical activity in a workplace population. Psychol Health Med 2009 Jan;14(1):53-61.

(34) McCance KL, Clayton F. Parkinson, Huether SE. Pathophysiology: the biologic basis for disease in adults and children St. Louis, Mo.: Elsevier Mosby; 2006.

(35) WHO | Gender and women's mental health Available at: http://www.who.int/mental_health/prevention/genderwomen/en/. Accessed 5/19/2011, 2011.

(36) Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. Lancet 2008 Feb 16;371(9612):569-578.

(37) Abdullah A, Peeters A, de Courten M, Stoelwinder J. The magnitude of association between overweight and obesity and the risk of diabetes: a meta-analysis of prospective cohort studies. Diabetes Res Clin Pract 2010 Sep;89(3):309-319.

(38) Rebecca Puhl P. Stigma of Obesity.pdf (application/pdf Objekt) Available at: http://www.obesityaction.org/magazine/oacnews3/Stigma of Obesity.pdf. Accessed 5/19/2011, 2011.

(39) Mody RR, Smith MJ. Smoking status and health-related quality of life: as findings from the 2001 Behavioral Risk Factor Surveillance System data. Am J Health Promot 2006 Mar-Apr;20(4):251-258.

(40) de Miguel Diez J, Esteban y Pena MM, Puente Maestu L, Hernandez Barrera V, Carrasco Garrido P, Alvarez-Sala Walther LA, et al. Relationship between tobacco consumption and health-related quality of life in adults living in a large metropolitan area. Lung 2010 Oct;188(5):393-399.

(41) Wilson D, Parsons J, Wakefield M. The health-related quality-of-life of never smokers, ex-smokers, and light, moderate, and heavy smokers. Prev Med 1999 Sep;29(3):139-144.

(42) Chan AM, von Muhlen D, Kritz-Silverstein D, Barrett-Connor E. Regular alcohol consumption is associated with increasing quality of life and mood in older men and women: the Rancho Bernardo Study. Maturitas 2009 Mar 20;62(3):294-300.

(43) Stranges S, Notaro J, Freudenheim JL, Calogero RM, Muti P, Farinaro E, et al. Alcohol drinking pattern and subjective health in a population-based study. Addiction 2006 Sep;101(9):1265-1276.

(44) Martin Jeppesen. privat.pdf (application/pdf Objekt) Available at: http://www.dst.dk/pukora/epub/upload/9605/privat.pdf. Accessed 5/29/2011, 2011.

(45) Ludlow AT, Roth SM. Physical activity and telomere biology: exploring the link with aging-related disease prevention. J Aging Res 2011 Feb 21;2011:790378.

(46) Larsen F.B., Ankersen P.V. & Poulsen S. editor. Hvordan har du det? 2010. Sundhedsprofil for region og kommuner. Voksne http://www.centerforfolkesundhed.dk/files/Sundhed/Folkesundhed/Konferencer/Hvordan har du det 2010_hovedrapport_samlet_til CFF.pdf ed. Aarhus: Center for Folkesundhed; 2011.

(47) Wooldridge J. Introductory Econometrics International ed of 4th revised: South Western College; 2008.

(48) Simon GE, Revicki DA, Grothaus L, Vonkorff M. SF-36 summary scores: are physical and mental health truly distinct? Med Care 1998 Apr;36(4):567-572.

(49) Taft C, Karlsson J, Sullivan M. Do SF-36 summary component scores accurately summarize subscale scores? Qual Life Res 2001;10(5):395-404.

(50) Tucker G, Adams R, Wilson D. New Australian population scoring coefficients for the old version of the SF-36 and SF-12 health status questionnaires. Quality of Life Research 2010;19(7):1069-1076.

(51) Li WC, Chen YC, Yang RS, Tsauo JY. Effects of exercise programmes on quality of life in osteoporotic and osteopenic postmenopausal women: a systematic review and meta-analysis. Clin Rehabil 2009 Oct;23(10):888-896.

(52) Penedo FJ, Dahn JR. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. Curr Opin Psychiatry 2005 Mar;18(2):189-193.

(53) Choi BC, Pak AW. A catalog of biases in questionnaires. Prev Chronic Dis 2005 Jan;2(1):A13.

(54) Donaldson SI, Grant-Vallone EJ. Understanding Self-Report Bias in Organizational Behavior Research. Journal of Business and Psychology 2002;17(2):245-260.

(55) Samspillet mellem motion og vægttab Available at: http://www.motion-online.dk/sundhed_og_vaegt/vaegt_og_fedtprocent/samspillet_mellem_motion_og_vaegttab/. Accessed 4/29/2011, 2011.

(56) Pernille Due, Berit L. Heitmann and Thorkild I.A. Sørensen. Adipositasepidemien i Danmark 2006; Available at: http://www.ugeskriftet.dk/LF/UFL/2006/2/pdf/VP46220.pdf. Accessed 4/15/2011, 2011.

(57) Kolotkin RL, Meter K, Williams GR. Quality of life and obesity. Obes Rev 2001 Nov;2(4):219-229.

(58) Gorber SC, Schofield-Hurwitz S, Hardt J, Levasseur G, Tremblay M. The accuracy of self-reported smoking: a systematic review of the relationship between self-reported and cotinine-assessed smoking status. Nicotine Tob Res 2009 Jan;11(1):12-24.

(59) Saarni SI, Joutsenniemi K, Koskinen S, Suvisaari J, Pirkola S, Sintonen H, et al. Alcohol consumption, abstaining, health utility, and quality of life--a general population survey in Finland. Alcohol Alcohol 2008 May-Jun;43(3):376-386.

(60) Chapman B, Duberstein P, Lyness JM. Personality traits, education, and health-related quality of life among older adult primary care patients. J Gerontol B Psychol Sci Soc Sci 2007 Nov;62(6):P343-52.

(61) Konig HH, Heider D, Lehnert T, Riedel-Heller SG, Angermeyer MC, Matschinger H, et al. Health status of the advanced elderly in six European countries: results from a representative survey using EQ-5D and SF-12. Health Qual Life Outcomes 2010 Nov 29;8:143.

(62) Sundhedsstyrelsen. Anbefalinger til voksne Available at: http://www.sst.dk/Sundhed og forebyggelse/Fysisk aktivitet/Anbefalinger til voksne.aspx. Accessed 5/6/2011, 2011.

(63) Motions- og Ernæringsrådet. 13.06.07 7767_fysisk_inaktivitet_webny2.pdf (application/pdf Objekt) 2007; Available at: http://www.meraadet.dk/gfx/uploads/rapporter_pdf/13.06.07 7767_fysisk_inaktivitet_webny2.pdf. Accessed 5/6/2011, 2011.

(64) WHO | Physical Activity and Adults Available at: http://www.who.int/dietphysicalactivity/factsheet_adults/en/index.html. Accessed 5/20/2011, 2011.

(65) Bentur N, King Y. The challenge of validating SF-12 for its use with community-dwelling elderly in Israel. Quality of Life Research 2010;19(1):91-95.

(66) Wikipedia contributors. Low-density lipoprotein , 2011.

(67) Blair SN, Haskell WL, Ho P, Paffenbarger RS,Jr, Vranizan KM, Farquhar JW, et al. Assessment of habitual physical activity by a seven-day recall in a community survey and controlled experiments. Am J Epidemiol 1985 Nov;122(5):794-804.

(68) Maurice Lorr, Ph.D. Douglas M. McNair, Ph.D. Leo F. Droppleman, Ph.D. POMS - Profile of Mood States Available at: http://www.mhs.com/product.aspx?gr=cli&prod=poms&id=overview. Accessed 5/16/2011, 2011.

(69) Wikipedia contributors. Comorbidity , 2011.

(70) Paffenbarger RS,Jr, Wing AL, Hyde RT. Physical activity as an index of heart attack risk in college alumni. 1978. Am J Epidemiol 1995 Nov 1;142(9):889-903; discussion 887-8.

(71) New England Research institute
CSM_PASE.pdf (application/pdf Objekt) Available at: http://www.sswahs.nsw.gov.au/balmain/strong/Questionnaires/CSM_PASE.pdf. Accessed 5/16/2011, 2011.

(72) IDACC - Questionnaires Available at: http://idacc.healthbase.info/questionnaires.html. Accessed 5/16/2011, 2011.

(73) Charles D. Spielberger. State-Trait Anxiety Inventory for Adults by Charles D. Spielberger - Mind Garden, Inc. Available at: http://www.mindgarden.com/products/staisad.htm. Accessed 5/16/2011, 2011.

(74) Ralf Schwarzer & Matthias Jerusalem. Self-Efficacy Available at: http://userpage.fu-berlin.de/~health/selfscal.htm. Accessed 5/16/2011, 2011.

(75) Wikipedia contributors. Metabolic equivalent , 2011.

(76) EuroQol - Home Available at: http://www.euroqol.org/. Accessed 5/17/2011, 2011.

(77) Sundhedsstyrelsen. Alkohol Available at: http://www.sst.dk/Sundhed og forebyggel-se/Alkohol.aspx. Accessed 5/19/2011, 2011.

(78) J. Ogden. Health Psychology, a textbook. 4th edition. Open University Press; 2007.

Chapter IV
Appendices

Appendix I: The SF-12 Questionnaire

In this appendix the questions from the SF-12 questionnaire are presented. This will help understand how the expectations to the factor analysis have evolved. The persons that have answered the questionnaire have marked their answer with a cross, and it can be seen, what value a mark in the respective space gives as a score. After each question, it is marked, which item the question represents.

1. In general, would you say your health is (this question covers the GH item):
____ Excellent [1]
____ Very Good [2]
____ Good [3]
____ Fair [4]
____ Poor [5]

The following two questions are about activities you might do during a typical day. Does YOUR HEALTH NOW LIMIT YOU in these activities? If so, how much? (This question covers the PF items)

	Yes, limited a lot [1]	Yes, limited a little [2]	No, not limited at all [3]
2. MODERATE ACTIVITIES, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf			
3. Climbing SEVERAL flights of stairs			

During the PAST 4 WEEKS have you had any of the following problems with your work or other regular activities AS A RESULT OF YOUR PHYSICAL HEALTH? (This question covers the RP items)

	All of the time [1]	Most of the time [2]	A good bit of the time [3]	Little of the time [4]	None of the time [5]
4. ACCOMPLISHED LESS than you would like					
5. Were limited in the KIND of work or other activities					

During the PAST 4 WEEKS, were you limited in the kind of work you do or other regular activities AS A RESULT OF ANY EMOTIONAL PROBLEMS (such as feeling depressed or anxious)? (This question covers the RE items)

	All of the time [1]	Most of the time [2]	A good bit of the time [3]	Little of the time [4]	None of the time [5]
6. ACCOMPLISHED LESS than you would like					
7. Didn't do work or other activities as CAREFULLY as usual					

8. During the PAST 4 WEEKS, how much did PAIN interfere with your normal work (including both work outside the home and housework)? (This question covers the BP item)
_____ Not At All [1]
_____ A Little Bit [2]
_____ Moderately [3]
_____ Quite A Bit [4]
_____ Extremely [5]

The next three questions are about how you feel and how things have been DURING THE PAST 4 WEEKS. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the PAST 4 WEEKS – (This questions covers VT and MH)

	All of the time [1]	Most of the time [2]	A good bit of the time [3]	Little of the time [4]	None of the time [5]
9. Have you felt calm and peaceful?					
10. Did you have a lot of energy?					
11. Have you felt downhearted and blue?					

12. During the PAST 4 WEEKS, how much of the time has your PHYSICAL HEALTH OR EMOTIONAL PROBLEMS interfered with your social activities (like visiting with friends, relatives, etc.)? (This question covers the SF item)
_____ All of the time [1]
_____ Most of the time [2]
_____ A Good Bit of the time [3]
_____ Some of the time [4]
_____ A Little of the time [5]
_____ None of the time [6]

APPENDIX II: MULTIPLE REGRESSION ANALYSIS

In the aim to explain how multiple regression analysis can be used, simple regression analysis is firstly explained.

The simple regression analysis is an analysis, where "y" is explained in terms of "x", which can be analysed with the following equation:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

y_i is the dependent variable, x_i is the independent variable and u_i represents unobserved variables other than x_i that explains y_i. If u is held fixed (u=0) the equation forms a straight line, which gives that the change in y_i is β_1 multiplied by the change in x_i. With this β_1 is the slope coefficient that shows how y_i and x_i are correlated: if the sign in front of β_1 is positive, they have a positive correlation and a negative sign means they are negatively correlated. β_0 is a constant term, which simply explains where on the y-axis the line intersects.

The ordinary least squares (OLS) is a method, where the straight line is estimated. This estimated straight line is written:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Where the "hat" means the variables are estimates of the "true" value. The difference between y_i and $\hat{y}_i$ is called the residual, and can be seen in figure A.

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

Some of the residuals may be negative, which is why all residuals are squared, and the following sum of squared residuals (SSR) is minimized by OLS in order to obtain β_0 and β_1.

$$SSR = \sum_{i=1}^{n} \hat{u}_i^{\,2} = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

As an example the sample of this project has answered a question about physical activity, and I use this in relation to PCS, which gives the following equation (holding u fixed as 0):

$$PCS = \beta_0 + \beta_1 activity$$

The sample will be distributed on the x-axis into the eight categories, and on the y-axis depending on which PCS score the individual has from the SF-12 questionnaire. A simplified example can be seen in figure A, where some individuals have answered the question about physical activity (in the figure, x corresponds the number of days), and are given a PCS score from the questions related to the items used to calculate PCS (y is PCS). Each dot represents an individual, which have answered the activity question with a given PCS. The y_i are different from $\hat{y}$, and the difference between $\hat{y}$ (that is the line) and y_i is called the residual $\hat{u}_i$. The residuals are squared (to make potentially negative residuals positive), and the OLS tries to minimize the sum of squared residuals in the aim to estimate the best straight line. $\bar{y}$ is the average, which is used in later equations.

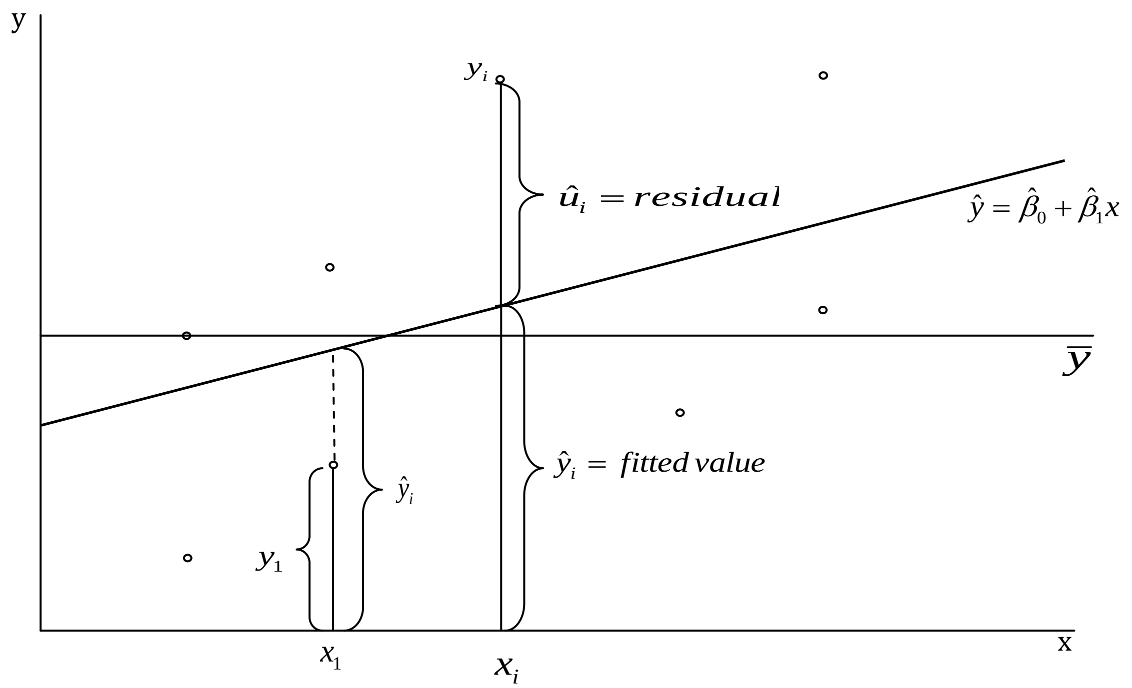

Figure A: Here the principle of ordinary least squares is seen. The line ŷ is the estimate made from OLS. The residual value $\hat{u}_i$ are squared to make negative values, positive. Also the average ȳ is presented. Modified from (47).

In this project however, more than one variable is used to explain PCS and MCS. With more variables to explain y_i, the equation will look like this:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u_i$$

The variables mean the same as in the simple regression model, but the difference is that more variables now explain y_i. Also in this equation the OLS estimates the "line" by minimizing the sum of squared residuals (SSR), but now with *k* variables instead of just one.

$$SSR = \sum_{i=1}^{n} \hat{u}_i^{\,2} = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik})^2$$

A value that is important in this project is R^2 (actually $\bar{R}^2$, but more about that later), which describes the proportion of the sample variation in y_i explained by the model. To find R^2 it is necessary to know the total sum of squares (SST) that is calculated using the sample average ($\bar{y}$) of y_i.

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$R^2 = 1 - \frac{SSR}{SST}$$

R^2 is a number between zero and one, and is explained by the OLS regression line. If R^2 is "close" to zero it means that the $\bar{y}$ can predict the dependent variable "as good as" the estimated model, whereas a "high" R^2 means that the dependent variable fits well with the model.

The reason $\bar{R}^2$ is used instead of R^2 is that $\bar{R}^2$ imposes a penalty for adding additional independent variables (or confounders) to the regression model, and $\bar{R}^2$ will help to decide if a certain independent variable belongs in the model. In this project backward selection is used, and if the $\bar{R}^2$ increases when removing a variable, it means that the particular variable degrades the model.

To find $\bar{R}^2$ the equation is changed, revealing that R^2 is an estimate:

$$R^2 = 1 - \frac{SSR/n}{SST/n}$$

According to (47) SSR/n is a biased estimator for σ_u^2, which is the error variance. In the same way SST/n is a biased estimator for σ_y^2. Instead of using biased estimators it is possible to use unbiased estimators, that adjust for the number of independent variables (*k*) in the model. By taking the independent variables into account, the $\bar{R}^2$ becomes:

$$\bar{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)}$$

The standardised coefficients beta are derived from the original equation. Before finding the standardised beta, the original equation with estimated $\hat{\beta}$, is used. <u>NOTE</u> the difference between standardised beta and unstandardised $\hat{\beta}$.

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_k x_{ik} + \hat{u}_i$$

To standardise all variables, both the dependent and independent, the mean of each variable is subtracted from the variable itself. This gives the following equation:

$$y_i - \bar{y} = \hat{\beta}_1 (x_{i1} - \bar{x}_1) + \hat{\beta}_2 (x_{i2} - \bar{x}_2) + \ldots + \hat{\beta}_k (x_{ik} - \bar{x}_k) + \hat{u}_i$$

The difference between all y_i in the sample and the $\bar{y}$ is used to calculate an estimate of the standard deviation for the dependent variable. The estimate of the standard deviation is called $\hat{\sigma}_y$.

$$\hat{\sigma}_y = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2}$$

In the same way the standard deviations for the independent variables are calculated:

$$\hat{\sigma}_j = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}, \; j = 1, \ldots, k$$

By subtracting the average from the actual variable, and dividing with the standard deviation, it is possible to calculate the standardised variables z_y and z_j:

$$z_y = \frac{(y - \bar{y})}{\hat{\sigma}_y}$$

$$z_j = \frac{(x_j - \bar{x}_j)}{\hat{\sigma}_j}$$

These equations "move" the average of both the dependent variable and the independent variables into the origin and scales them such that 95 percent of the answers lie between (-1.96) and 1.96 on both the x-axis and y-axis.

The last step in deriving standardised coefficients beta (below presented as b_j (with hat)) is to divide with the standard deviations and multiply with $\hat{\beta}_j$:

$$\hat{b}_j = \hat{\beta}_j \frac{\hat{\sigma}_j}{\hat{\sigma}_y}$$

The final standardised equation now becomes:

$$z_y = \hat{b}_1 z_1 + \hat{b}_2 z_2 + \ldots + \hat{b}_k z_k + error$$

This equation is the outcome from the standardised betas, when using multiple regression analysis.

APPENDIX III: THE EDITORIAL PROCESS IN SPSS

In this appendix the editorial process in SPSS of this project is presented.

RENAME VARIABLES (aktivdag = activity) (sex=male) (ryger_10=smoking) (alkodage=alcohol) (husindk_10=Houseincome) (fÃ¸ddato=birthdate) (hoejde=height) (vaegt=weight) (dunkode=DUN) (jobbelast=workPA).
EXECUTE.

RECODE PCS (0 thru 25=1) (25 thru 50=2) (50 thru 75=3) (75 thru 100=4) INTO PCScat.
EXECUTE.

RECODE MCS (0 thru 25=1) (25 thru 50=2) (50 thru 75=3) (75 thru 100=4) INTO MCScat.
EXECUTE.

COMPUTE MaleDummy=male.
EXECUTE.

RECODE male (1=1) (2=0) (2147483622=SYSMIS).
EXECUTE.

Age is inserted manually using Visual Studio to remove the first four numbers of the birth date (DD-MM-YYYY) and retracting the birth year from 2010 (which was the year of investigation)

RECODE Age (18 thru 35=1) (35 thru 50=2) (50 thru 67=3) (67 thru 110=4) INTO AgeCat.
EXECUTE.

COMPUTE BMI=Weight / (Height * Height * 0.0001).
EXECUTE.

RECODE BMI (SYSMIS=SYSMIS) (0 thru 18.5=1) (18.5 thru 25=2) (25 thru 30=3) (Lowest thru 30=4) INTO BMIcat.
EXECUTE.

RECODE activity (1=0)(2=1)(3=2)(4=3)(5=4)(6=5)(7=6)(8=7).
EXECUTE.

RECODE Houseincome (1=49500) (2=124500) (3=199500) (4=312000) (5=449500) (6=612000) (7=774500)
(8=937000) (SYSMIS=SYSMIS) INTO HouseIncomeCoded.
EXECUTE.

RENAME VARIABLES (Houseincome=HouseUncoded) (HouseIncomeCoded=Income).
RECODE smoking (SYSMIS=SYSMIS) (1=7) (2=2) (3=0.5) (4 THRU 5 =0) INTO smoking2.
EXECUTE.

RENAME VARIABLES (smoking =smokingUncoded) (smoking2=smoking).

RECODE DUN (SYSMIS=SYSMIS)(0=0)(1=3.5)(2 THRU 3=8.5)(4=11.5)(5=13.5)(6=15.5)(7=17.5)(8=20.5)(9=SYSMIS).
EXECUTE.

```
RECODE workPA (SYSMIS=SYSMIS)(1=0)(2=2)(3=4)(4=7).
EXECUTE.

RECODE PSS (0 thru 10=1) (10 thru 20=2) (20 thru 30=3) (30 thru 40=4) INTO PSSCategories.
EXECUTE.

FREQUENCIES VARIABLES=PCScat MCScat activity male AgeCat BMIcat smoking alcohol Income DUN
PSSCategories workPA
/ORDER=ANALYSIS.
FACTOR
/VARIABLES GH01 PF02 PF04 RP02 RP03 RE02 RE03 BP02 MH03 VT02 MH04 SF02
/MISSING LISTWISE
/ANALYSIS GH01 PF02 PF04 RP02 RP03 RE02 RE03 BP02 MH03 VT02 MH04 SF02
/PRINT INITIAL CORRELATION SIG KMO AIC EXTRACTION ROTATION
/FORMAT BLANK(.40)
/CRITERIA MINEIGEN(1) ITERATE(25)
/EXTRACTION PC
/CRITERIA ITERATE(25)
/ROTATION VARIMAX
/SAVE REG(ALL)
/METHOD=CORRELATION.
CORRELATIONS
/VARIABLES=MCS PCS FAC1_1 FAC2_1
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE.

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT PCS
/METHOD=ENTER male Age.

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT MCS
/METHOD=ENTER male Age.

DATASET ACTIVATE DataSet1.
DESCRIPTIVES VARIABLES=PCS MCS
/STATISTICS=MEAN STDDEV MIN MAX.

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT PCS
```

```
/METHOD=ENTER activity male Age BMI smoking alcohol income DUN PSS workPA.
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT MCS
/METHOD=ENTER activity male Age BMI smoking alcohol income DUN PSS workPA.

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT PCS
/METHOD=ENTER activity male Age BMI smoking income DUN PSS workPA.
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT MCS
/METHOD=ENTER activity male Age BMI smoking alcohol DUN PSS workPA.

DATASET ACTIVATE DataSet1.
COMPUTE activityAge=activity * Age.
EXECUTE.
DATASET ACTIVATE DataSet1.
COMPUTE activityBMI=activity * BMI.
EXECUTE.

DATASET ACTIVATE DataSet1.
COMPUTE activityPSS=activity * PSS.
EXECUTE.
DATASET ACTIVATE DataSet1.
COMPUTE activityWorkPA=activity * workPA.
EXECUTE.

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT PCS
/METHOD=ENTER activity male Age BMI smoking alcohol income DUN PSS workPA activityAge activ-
ityBMI activityWorkPA.
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT MCS
```

```
/METHOD=ENTER activity male Age BMI smoking alcohol income DUN PSS workPA activityAge activityBMI activityPSS.

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT PCS
/METHOD=ENTER activity male Age BMI smoking alcohol income DUN PSS workPA activityBMI activityWorkPA.
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT MCS
/METHOD=ENTER activity male Age BMI smoking alcohol DUN PSS workPA activityAge activityBMI activityPSS.

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT PCS
/METHOD=ENTER activity male Age BMI smoking income DUN PSS workPA activityBMI activityWorkPA.

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT MCS
/METHOD=ENTER activity male Age BMI smoking alcohol DUN PSS workPA activityBMI activityPSS.

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT PCS
/METHOD=ENTER activity male Age BMI smoking income DUN PSS workPA activityWorkPA.

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT PCS
/METHOD=ENTER activity male Age BMI smoking income DUN PSS workPA.
```

APPENDIX IV: DESCRIPTION OF THE VARIABLES

In this appendix the variables and the questions they are derived from, are described a little more into detail. The answer to each question is only approved if the person only has answered the question once; meaning that only one cross is present in each question in the questionnaire.

Activity: To analyse activity seven days a week (question 60): "*How many days a week are you active for at least 30 minutes a day? Include moderate or strenuous physical activity where you breathe faster, move your muscles and use your powers - e.g. for exercise or competitive sport, heavy gardening, brisk walking, bicycling at a moderate or fast pace or physical strenuous work. Include both work and leisure time".* The answering possibilities range from 0 days a week to seven days a week.

Male: Gender is based on the sentence: "*Are you:*" and the options male or female in question one. This variable is made a dummy variable with the question: "Are you male?", where the value 0 is "no" and the value 1 is "yes". By making gender a dummy variable it is possible to statistically see a magnitude between males and females.

Age: By asking the question "*When were you born?*" in question 2, the birth year is possible to use to calculate how old a person is, by subtracting 2010 from the birth year. This way no regards to the birth data is taken into consideration. The age is not divided since there is a logical continuous ranging in this variable.

BMI: BMI is calculated from the formula: $BMI = \frac{weight\,(kg)}{height\;(m) \cdot height\,(m)}$, based on the person's answer regarding height and weight in question 68 and 69 respectively. Since the BMI is a continuous variable, it is not divided into categories.

Smoking: If the person is a smoker is asked in question 31 "*Do you smoke?*", and the questionnaire gives six options. These options are recoded so that the given number represents how many times in a week the person smokes: The persons that answer "no" or "no, I have stopped smoking" to the answer will get a 0; persons that have answered "yes, less than every week" are scored 0.5; persons that have answered "yes, at least once per week" scores 2; and persons that answer "yes, every day" will score 7. The numbers are an estimate of how many days a week a person smokes.

Alcohol: The use of alcohol is asked in question 47: "*How many days a week do you consume alcohol?*". This question is divided into seven categories ranging from zero/one day a week to seven days a week, which gives seven categories used in the analyses.

Income: Question 109 asks about both personal and household income. In this analysis the household income is used, because a high household income will make it economically possible to afford a healthy lifestyle and thereby obtain a higher HRQOL. As with education, household income is the average of each economical level, because it is assumed that a uniform distribution exists within each category.

DUN: Data regarding education has been extracted from "Denmark's Statistics" by the DUN. DUN is divided into categories each representing the length of the person's education. Persons that are out of category (originally category 9) are made system missing, since those persons may be persons that have completed an education outside Denmark and that now are Danish citizens. Scientist's level is difficult to categorise since the middle value cannot be found from the number 8, but it is believed that scientists have an average of 20.5 educational years.

PSS: Question 17 concerning PSS asks about stressing situations the last month and consists of ten questions. Five answering opportunities give a score, which summed together expresses how much stress a person feels on a scale from zero to 40. A higher score expresses a higher level of stress. This variable is not categorised since there is a logical continuous ranging.

WorkPA: Physical activity at work is assessed from question 104, which sounds: *"How would you describe your strain in your daily work?"*. It is possible to answer four different categories, ranging from sedentary work to heavy lifting. Each category is given a number that expresses how many hours a day the person is active on their job.

Appendix V: Frequency Tables

In this appendix it is possible to see the distribution within the variables.

PCS divided into categories	Frequency	Percent
0-25	713	2.1
25-50	8717	25.5
50-75	20562	60.2
75-100	0	0
Number of persons that have answered the question	29992	87.8
System missing	4155	12.2
Total	34147	100.0

Table B: PCS is divided into four categories to get a rough picture of the physical health in the sample. 29992 persons have answered the items related to PCS.

MCS divided into categories	Frequency	Percent
0-25	538	1.6
25-50	10656	31.2
50-75	18794	55.0
75-100	1	0.0
Number of persons that have answered the question	29989	87.8
System missing	4158	12.2
Total	34147	100.0

Table C: As with physical health MCS is roughly divided into four categories. It can be seen that 29989 persons have answered all items related to MCS.

Activity: How many days a week are you active for at least 30 minutes a day?	Frequency	Percent
No days a week	2848	8.3
One day a week	3063	9.0
Two days a week	3733	10.9
Three days a week	4950	14.5
Four days a week	3597	10.5
Five days a week	3922	11.5
Six days a week	2544	7.5
Every day	8848	25.9
Number of persons that have answered the question	33505	98.1
System missing	642	1.9
Total	34147	100.0

Table D: This table shows the frequency table of the independent variable activity. It can be seen that 25.9 percent of the sample is active for at least 30 minutes every day.

Gender: Are you male?	Frequency	Percent
No	18032	52.8
Yes	16013	46.9
Number of persons that have answered the question	34045	99.7
System missing	102	0.3
Total	34147	100.0

Table E: The distribution of gender. It can be seen that 0.3 percent has not answered the question.

Age divided into categories	Frequency	Percent
18 to 35 years of age	6927	20.3
35 to 50 years of age	9064	26.5
50 to 67 years of age	11340	33.2
67 to 110 years of age	6239	18.3
Number of persons that have answered the question	33570	98.3
System missing	577	1.7
Total	34147	100.0

Table F: Even though age has not been divided into categories in the project, it is done in this table to analyse how age is distributed. It can be seen that most persons are between the ages 50 and 67.

BMI divided into categories	Frequency	Percent
Underweight (BMI≤18.5)	748	2.2
Normal weight (BMI 18.5-25)	16246	47.6
Overweight (BMI 25-30)	11539	33.8
Obesity (BMI≥30)	4789	14.0
Number of persons that have answered the questions	28533	83.6
System missing	825	2.4
Total	34147	100.0

Table G: Show the distribution of BMI in the sample. It can be seen that it is more "normal" to be overweight or obese than having a normal weight.

Smoking: Do you smoke?	Frequency	Percent
No I have never smoked OR no, I have quit	25592	74.9
Yes, less than once per week	677	2.0
Yes, more than once per week	481	1.4
Yes, every day	6666	19.5
Number of persons that have answered the question	33416	97.9
System missing	731	21
Total	34147	100.0

Table H: In this table it can be seen that 74,9 percent of the sample do not smoke or have stopped to do so, but that almost 20 percent still smoke every day.

Alcohol: How many days a week do you consume alcohol?	Frequency	Percent
0-1 days per week	15551	45.5
Two days a week	5363	15.7
Three days a week	2927	8.6
Four days a week	1442	4.2
Five days a week	1143	3.3
Six days a week	736	2.2
Every day	2879	8.4
Number of persons that have answered the question	30041	88.0
System missing	4106	12.0
Total	34147	100.0

Table I: This table shows the outcome from the question "How many days a week do you consume alcohol?

Income: How large was your household income in 2009 before taxes?	Frequency	Percent
0-99,000 kr.	537	1.6
100,000-149,000 kr.	1177	3.4
150,000-249,000 kr.	2977	8.7
250,000-374,000 kr.	4011	11.7
375,000-524,000 kr.	4802	14.1
525,000-699,000 kr.	5615	16.4
700,000-849,000 kr.	3026	8.9
>850,000 kr.	2454	7.2
Number of persons that have answered the question	24599	72.0
System missing	9548	28.0
Total	34147	100.0

Table J: This table show the household income of the sample. It can be seen that almost a third of the sample has not answered this question.

Education expressed by DUN	Frequency	Percent
Primary school level II (7th to 10th grade) AND upper secondary level I (10th educational year)	6137	18,0
Upper secondary level II (11th/12th educational year)	15125	44.3
Short higher education (13th/14th educational year)	2264	6.6
Medium length higher education (15th/16th educational year)	5777	16.9
Long higher education (17th/18th educational year)	2510	7.4
Scientist's level (≥19 educational year)	145	0.4
Number of persons that have an education	31958	93.6
System missing	2189	6.4
Total	34147	100.0

Table K: Shows the distribution of the education of the sample. It can be seen that most of the sample have had 11 to 12 years of education.

PSS divided into categories	Frequency	Percent
Not stressed (0-10)	14988	43.9
Little stressed (10-20)	13717	40.2
Often stressed (20-30)	2747	8.0
Severe stressed (30-40)	248	0.7
Number of persons that have answered ten questions about stress	31700	92.8
System missing	2447	7.2
Total	34147	100.0

Table L: Cohen's perceived stress scale has been divided into four categories, in the aim to analyse the "general" stress level.

WorkPA: How would you describe your strain in your daily work?	Frequency	Percent
Primary sedentary work	7202	21.1
Standing or walking, not exhausting	5078	14.9
Standing or walking with lifting	4941	14.5
Heavy or fast, exhausting	619	1.8
Number of persons that have answered the question	17840	52.2
System missing	16307	47.8
Total	34147	100.0

Table M: This table shows how the sample thinks of their daily strain at work. A high amount of persons have not answered the question, and the main parts of the ones that have, think they have a sedentary work.