ERASMUS UNIVERSITEIT ROTTERDAM

Erasmus School of Economics

Department of Health Economics

The Relationship Between Occupation and Health

The impact of manual occupation on health in the UK

Liesbeth de Pater

320488

1-7-2011

Supervisor: J.L.W. van Kippersluis
Co-reader: Dr. T.G.M. van Ourti

ACKNOWLEDGE

Dear reader,

In front of you is my bachelor thesis. It all started during the period I attended the major Health Economics, where I became enthusiastic for the field of Health Economics. The dilemma's faced in healthcare are really interesting. Health Economics is getting growing importance, because of the phenomenon of ageing populations and the costs of improved technologies.

During the major I especially got interested in the topic of occupation, when I heard contrary findings about manual jobs. I always thought that, except from some heavy-physically demanding jobs, people with manual jobs would be healthier (if they have similar income/education/etc.) than people with non-physically demanding jobs. However, previous studies suggested that this is not true.

For my thesis I could choose between Germany and Great Britain to take under study. Because my exchange period in Great Britain, I had the preference to study the impact of manual occupation on health in the UK. I was surprised when I then found out that the dataset itself was from the University of Essex, the university where I have studied for some months.

Writing a thesis asks different skills than studying books, and I really enjoyed this period of learning. Beside the skills of writing a clear thesis, I also have learned a lot from statistics which will be useful for my master next year.

With the help of my supervisor J.L.W. van Kippersluis and co-reader dr. T.G.M. van Ourti I have come to the final result of this study, and thus at the end of my bachelor. Without their good advice and quick response I would not have finished my thesis in 9 weeks.

Liesbeth de Pater, July 1, 2011

ABSTRACT

The impact of occupation on health is getting growing interest in the field of Health Economics. Occupation is one factor of socioeconomic status, which may influence health. Other socioeconomic characteristics which do have an impact on health are education and income. The number of years of education and a higher income do have a positive impact on health. Occupation is also discussed as a social determinant of health, although it has received little attention in economic research.

This thesis studies the impact of manual occupation on health, compared to non-manual occupation in the United Kingdom. The dataset consists of 3,347 observations from The British Household Panel Study (BHPS). The panel study makes it possible to examine health deterioration over the period 1991 and 2009.

An ordinal logistic regression is used to study the impact of manual occupation on health, controlling for initial health and several socioeconomic characteristics as education and income. The results suggest that manual occupation has a significant negative impact on health.

Content

ACKNOWLEDGE **2**

ABSTRACT **3**

1. INTRODUCTION **6**

1.1 Research Question 6

1.2 Relevance 7

1.3 Structure 7

2. CONCEPTUAL ISSUES AND PREVIOUS LITERATURE **7**

2.1. Grossman MOdel 8

2.2. Education 10

2.3. Income 10

2.4. Occupation 11

2.4.1. *Reverse causality* 12

2.5 Conceptual model 13

3. DATA AND METHODS **14**

3.1 BHPS 14

3.2 Sample 15

3.3 Estimation method 16

3.3.1. *Ordinal regression* 16

3.3.2. *Model* 17

3.4 Variables 18

3.5 Descriptive statistics 20

4. RESULTS **20**

4.1 Assumptions 21

4.1.1. *Adequate cell count* 21

4.1.2. *Parallel lines* 21

4.2 The models 22

4.2.1. *Goodness-of-Fit* 22

4.2.2. *Model fitting information* 23

4.2.3. *Parameter estimates* 24

4.3 Interaction effect 25

5. CONCLUSION 25

6. DISCUSSION 26

7. REFERENCES 28

APPENDIX 30

1. INTRODUCTION

A long time of life is spent on working. In many countries an extensive legislation on working conditions exists, to diminish hazardous and bad working circumstances. Based on the fact that a large part of life is spent on working, occupation is suggested to have an important impact on health. Although there are many biologic and physiologic studies done on the adverse conditions of occupation likely to harm health, the effect of occupation on health is a relative new topic in the field of Economics. Since the Grossman model by Michael Grossman (1972), new research has started on the influence of socioeconomic determinants on health, like education and income. Since 2000 also research on occupation as main factor in studying socioeconomic status on health has become a relevant topic in Economic research.

This study comes close to the paper of Fletcher et al (2009). In that paper the cumulative impact of physically demanding job characteristics or environmental conditions on health is studied. They use a five year period to measure a *cumulative* impact on health, and control for early and lagged health, years of schooling, labour income, and characteristics like age and gender. They find that individuals working in bad conditions experience a higher deterioration in health.

This thesis will focus on the physical demands of a job in the United Kingdom, omitting other characteristics like mental conditions of a job. A distinction between manual and non-manual occupation will be made, according to the Standard Occupational Classification (SOC). It will be studied whether manual occupation has higher health deterioration during the period 1991-2009, compared to non-manual occupation by using the British Household Panel Survey (BHPS). Five categories of health (excellent, good, fair, poor and very poor health) will be used as the dependent variable, instead of transforming it to a binary variable, to not create considerable information loss.

1.1 RESEARCH QUESTION

By studying the effect of manual occupation on health, the research question is: **what is the impact of manual occupation on health, compared to non-manual occupation in the United Kingdom?** Based on previous literature and the intuition that manual jobs have more hazardous working conditions, the hypothesis is that individuals with manual jobs have worse health than individuals with non-manual demanding jobs.

1.2 RELEVANCE

The impact of manual occupation on health is an interesting topic for both science and society. This study is of scientific relevance, because the linkage between socioeconomic status and health can be expanded and deepened by doing research on occupation. It seems that occupation becomes an important factor in influencing health in later years, because the cumulative aspect of working conditions.

This study is of societal relevance, because a large part of life is spent on working. Therefore it is important to get insight in the effects of working circumstances on health, to be able to realize as healthy working conditions as possible. If there is evidence that physical demanding jobs report a lower health state, reflection on the current working conditions is needed to create healthier circumstances.

1.3 STRUCTURE

In order to answer the research question, an ordinal logistic regression is conducted. In chapter 2, a theoretic framework is given as fundament for the model and interpretation of the results. In chapter 3, the estimation method and variables are described. Chapter 4 contains the results of the three models conducted, plus the results of a fourth model implemented to control for the interaction between occupation and education. A conclusion is found in chapter 5, where after the results will be discussed in chapter 6.

2. CONCEPTUAL ISSUES AND PREVIOUS LITERATURE

This chapter will provide a conceptual model of the effect of socioeconomic status on health, where occupation is highlighted in this study. The socioeconomic status (SES) of an individual represents the person's position in a hierarchical social structure. This combined sociological and economic measure depends on an individual's or household's economic position relative to others, based on (e.g.) education, income and occupation. Socioeconomic status is largely used in studies to explain/predict behaviour.

First, the underlying model for this conceptual framework, the Grossman model, will be explained. After that, the link between health and respectively education, income, and

occupation will be described. In the end, an overview will be given in the form of a conceptual model.

2.1. GROSSMAN MODEL

According Grossman's Health model (1972), individuals are both producers and consumers of health: they produce health by using Health Care, and consume health by living. Health is seen as a consumption good, because it produces direct utility, and an investment good because it produces indirect utility through increased productivity. In the model health is seen as a kind of capital stock, which deteriorates over time. Market goods (e.g. Medical Care), time and human capital can be seen as investment sources for health capital. Investment in health increases healthy days, which is required for activities such as work and leisure. The model makes predictions about the effect of changes in the labour market (e.g. employment and wages), prices of health care and other goods, and technological changes.

Figure 2.1 shows a graph of the Grossman model. The lower left corner represents the budget constraint, determined by income. Given a budget constraint, a consumer must trade off this resource devoted to health against other consumption. It depends on the individuals' preferences (II) and the production function how much an individual will choose to spend on health care and how much on other consumption. The upper left corner, III, shows the production function. The curve flattens, which means that health care has diminishing returns. The first unit spent on health care services creates more healthy days than the next unit does. The production function depends on technological changes, health deterioration and prices in the health care sector. Usually education is seen as human capital, which raises efficiency in production processes, and raises the production function for health. The upper right corner, II, represents the production possibilities derived from I, III, and IV. The production possibility curve represents all production possibilities given all possible points on the budget constraint and the corresponding production function. It depends on the preferences of the individual (represented by the downwards sloping curve) which part of his/her budget to spend on Health Care to maximize his/her utility. The optimal level of investment in health is when the marginal cost equals the marginal benefit. The marginal cost consists of the interest rate faced by the consumer plus the depreciation rate. The marginal benefit is the rate of return from health capital.

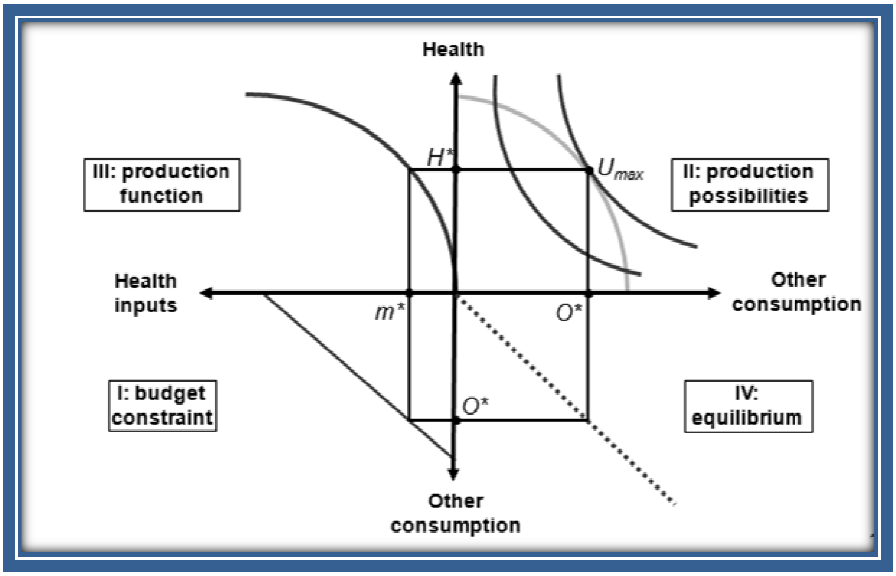


Figure 2.1 The Grossman Model

One prediction of the model is that, if the rate of depreciation increases over the life cycle, the total quantity of health capital demanded would decline with age. The depreciation rate may increase due to deterioration of the body with ones age. When the depreciation rate increases, it becomes more costly to attain the same level of health stock, because more investment is necessary to diminish the effect of a higher depreciation rate. In combination with a lower benefit of health stock with older ages (old people accept being in worse health), the optimal health stock will decrease as individuals gets older. There may be other factors that affect the depreciation rate. It is likely that people with more physically demanding jobs have a higher depreciation rate than people with a less physically demanding job.

Another prediction of the Grossman model is that the consumer's demand for health should be positively correlated with his wage rate. If wage increases, the budget constraint will shift to the left and the optimal point will be at a higher investment level. People with high paid jobs are more able to invest in health, and also have greater incentives to invest due to higher opportunity costs if being in poor health.

A third prediction is that if education raises efficiency in production processes of investment, then a higher educated consumer would demand a larger optimal stock of health. So, the level of education will influence the level of investment, and therefore will influence health.

2.2. EDUCATION

Since the study of Grossman in 1972, the literature of health economics focuses on the role of social determinants affecting health, like education and income. The relationship between education and health has been examined well (Cutler and Lleras-Muney, 2006; Kenkel, 1991). The literature shows that people with a high level of education have better health than lower educated people. Education improves the efficiency in producing health, because a better knowledge of the linkage between health behaviours and health outcomes might help people to choose healthier life-styles (Kenkel, 1991). Because higher educated people have on average a healthier lifestyle, they have lower health deterioration. In combination with a higher efficiency due to education, the marginal costs are lower. This results in a higher production of health.

According cutler and Lleras-Muney (2006) does the effect of education diminish over time, because gradients in behaviour are largest at young ages, although education has a long-term impact on health. However, others state that het impact of education increases with ones age, due to the accumulation of the lagged effect of education related to lifestyle choices (e.g. smoking) on health, and social advantages (e.g. knowledge of health services) (van Doorslaer et al., 2008).

Beside the impact of education on health, education may also have an impact on occupational choice. Higher educated people may choose more often non-manual jobs than low educated people may choose.

2.3. INCOME

There is positive relationship between income and life expectancy across countries. Countries with a higher GDP have a higher life expectancy than lower income countries. Also the relation between income and health within countries has found to be strong and positive. Between 1979 and 1989 white men with a family income above $25,000 had a life expectancy of 6.6 years more than white men with a family income below $10,000 (Smith, 1999).

Ferrie (2001) has found a strong and negative relationship between household wealth and mortality in 1850 and 1860. Despite the U.S. population was largely rural and agricultural, and therefore had similar occupation, different income levels showed different life expectancies. Individuals with a higher-paid job are able to finance more health investments, and therefore have a better health. Furthermore, they have more incentives to invest in health, due the fact that they face higher opportunity costs.

2.4. OCCUPATION

In previous literature, studies focussing on socioeconomic status did often ignore occupation, or did only control for occupation, but they did not study specifically the effect of occupation on health. Recently, occupation is getting growing interest in the research of the effects of social economic determinants on health. Several studies are done by using the US Panel Study of Income (PSID) to examine the role of occupational characteristics on health. Cumulative exposure to low-decision jobs and passive work are found to have a significant positive effect on mortality.

Occupation might affect health in several ways. It may affect health in a direct way by physical conditions of a job. People doing heavy physical work may have a higher deterioration than people in less manual occupations. Fletcher and Sindelar (2009) examined the cumulative impact of physically demanding or environmentally hazardous job characteristics on health, by using the US Panel Study of Income Dynamics (PSID) and historical Census data. They show that manual workers have a significantly lower health at older ages than non-manual workers.

Health might also be affected by mental conditions of a job. Karasek (1979) has developed the 'job-strain model', which says that a combination of both a high psychosocial workload and low decision latitude in a job will lead to a higher risk of cardiovascular diseases. Figure 2.2 shows the job-strain model. 'Psychological demands' is defined by questions like 'working very hard' and 'not enough time to get the job done'. 'Decision latitude' is defined as both the ability of making job decisions for the worker and the ability to use skills in work. A combination of low decision latitude and low psychological demands might lead to passivity and lack of participations of workers. The figure shows two diagonal lines. The arrow from the upper left corner to the lower right corner represents the increasing risk of heart diseases. Following the arrow from the lower left corner to the upper right corner indicates increased activity, participation, motivation to learn, etc.

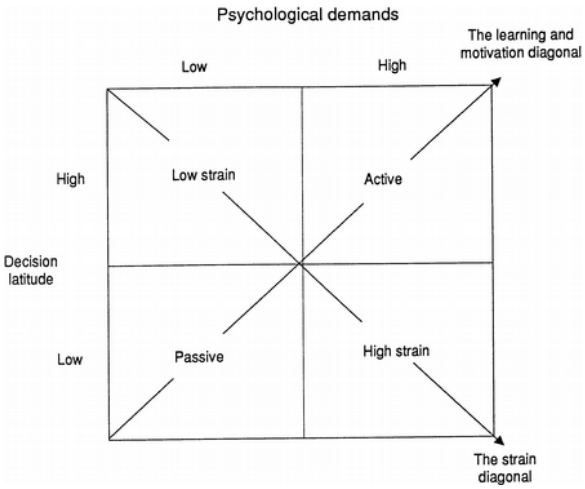


Figure 2.2 The job-strain model

Murphy (1991) has found that individuals in hazardous work situations (required vigilance and responsibility for others) have a higher risk for cardiovascular diseases, by using two U.S. data sets. Jobs included in the study are traffic controllers, airline pilots, bus drivers, preschool teachers, craftsmen and truck drivers.

McEwan (2000) has introduced the 'alostatic load', which shows that the accumulation of high stress and other disadvantages (and also advantages) has an impact on health. Brief periods of exposure to bad job conditions do not have an impact on health. When a person is confronted with stress, the adrenaline level will increase, so that the body can achieve a higher level of performance. Accumulated adrenalin may cause high blood pressure and heart rate, which does not cause problems in first instance. However, after repeated cumulative episodes of stress, it is hardly possible that the body returns to its normal state. This results in a constant functioning at high levels of adrenalin, which may lead to diseases (Smith, 1999). A set of studies on British civil servants focus on social position, occupational stress, and job control. They find that lower occupational status (which is associated with low decision latitude and high psychological demand) is associated with worse health (Fletcher et al., 2009).

Education might affect health indirectly through occupation. Education is an important factor in occupational choice. Individuals with low education choose more physical jobs. It then may seem that occupation do influence health, while actually it is education that causes the health difference across occupational choices.

Another indirect effect is that individuals stick together with other individuals in the same occupation. Health habits and behaviour of other member might affect one's own health habits and behaviour (e.g. smoking, consumption of high fat, lack of exercise), and could differ across occupations.

Also, workers in some occupations may have differential access to information related to health behaviours or methods of illuminating health problems.

2.4.1. REVERSE CAUSALITY

In the paragraphs above the causality between occupation and health is assumed from occupation to health. However, it might also be that health affects occupational choice. Sindelar and Fletcher (2007) control for health before entering the labour market to address the potential for reverse causality.

2.5 CONCEPTUAL MODEL

Figure 2.3 gives an overview of all factors described above in a conceptual model. This model gives insight into the relations between the several socioeconomic characteristics, as explained in the paragraphs above.

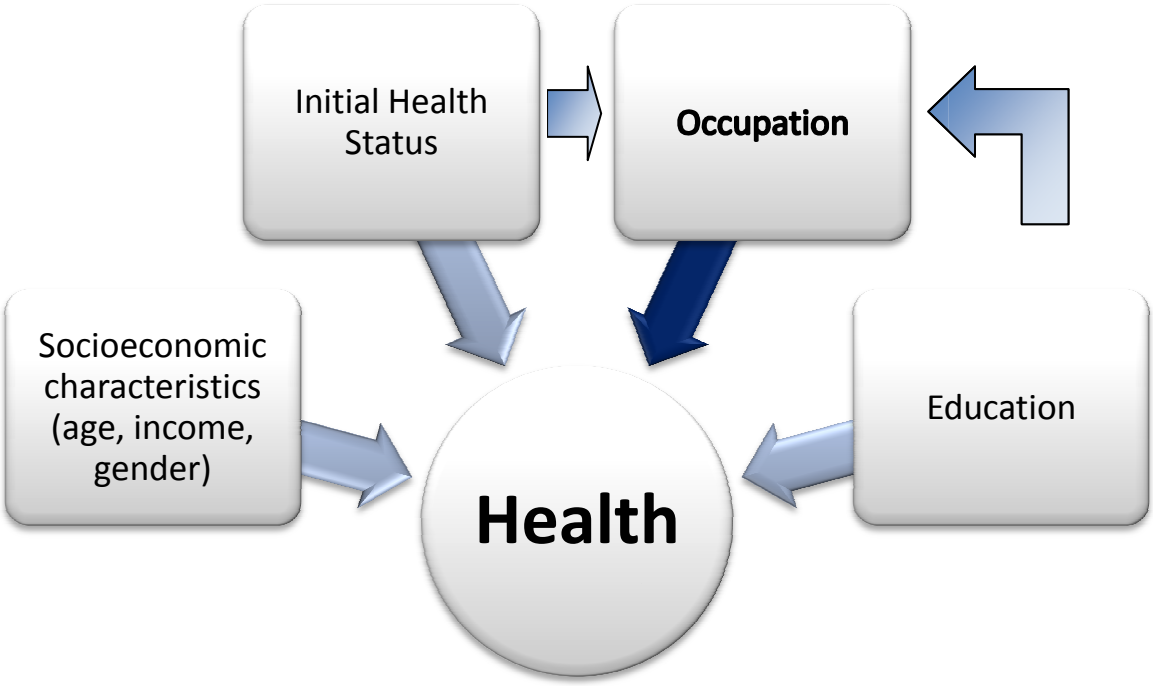


Figure 2.3 Conceptual model

3. DATA AND METHODS

In the previous chapter a conceptual model is given with the socioeconomic characteristics influencing health. The aim of this thesis is to determine the effect of occupation on health, so this arrow is highlighted. Education, initial health status and several socioeconomic characteristics have an impact on health as well. The arrows from initial health status and education to occupation implicate their effect on occupational choice[1]. The following chapter describes the dataset used to examine the relationship between occupation and health. Furthermore, it explains the estimation method and the variables used.

3.1 BHPS

The British Household Panel Survey (BHPS) contains data from 1991 onwards. It is an initiative from the ESRC UK Longitudinal Studies Centre with the Institute for Social and Economic Research at the University of Essex. BHPS can be used for studying causes and consequences of social and economic changes at the individual and household level in Great Britain, by using a set of socio-economic variables. Major topics are household organization, employment, income and wealth, housing, health and socio-economic values. A representative sample of more than 5,000 households (including 10,000 individuals of 16 years and older) are interviewed annually. These individuals are being re-interviewed every year. If members of a household are split-off, all adults from this new household are interviewed as well.

One advantage of the BHPS is that it is a panel dataset, which allows studies of changes in socioeconomic environment and their consequences. Another advantage is that both households and individuals can be followed[2]. Many of the questions asked in the first wave (1991) have been repeated in all waves. Some other questions are asked in alternating waves, while others are asked once during the panel study. This should be taken into account by choosing variables and waves. The BHPS consists of several files. There are different files for data belonging to individuals and data belonging to households. Furthermore, the files are repeated for subsequent waves. Some linking operations may be necessary to match (individual in subsequent waves) or distribute (household-level information to individual-level records) data to

[1] The relation between the characteristics are restricted to the described conceptual model. It could be that initial health has an impact on education, that gender has an impact on occupation and/or education, etc. It is expected that these relations are not as important as the relations described in the model for this study. For keeping the study clear, this restriction is needed.

[2] In this thesis income is taken at the household level. The other variables are taken at the individual level.

obtain the right dataset for study. The records are mostly provided with a combination of a household-number and a within-household person-number. This identification can be used to link information from different levels within one wave. To connect data across waves, the cross-wave personal identity number PID is included.

3.2 SAMPLE

The original dataset is divided into subsequent years and categories. In order to create a useful dataset the relevant variables of six files are matched. All variables are taken from 1991, except from the variable *Health*, which is taken from 2009. The questions corresponding to the variables are reported in table 1 in the appendix. The dataset in 1991 includes 26,932 individuals. However, 16,671 individuals do have missing values for the variable *Occupation*, leaving 10,261 observations. Furthermore, 4,110 individuals do not have a job, and therefore are not useful for this study. 6,151 individuals with a valid occupational status were included in BHPS. The sample then is restricted to those who have answered the question concerning self-reported health in 2009, resulting in 3,421 observations. Finally, one individual then did not answer the question concerning self-reported health in 1991, and 73 individuals did not have a valid education level, leaving 3,347 observations available for the analysis, as shown in table 3.1. All other control variables did not have missing values. The table also shows that from the dataset only 12.4 percent has valid observations for this study. While the original dataset was representative for the British society, it is not sure whether the dataset used is representative for the British working force.

The period 1991-2009 is taken to be able to examine whether health deterioration takes place through occupation.

Table 3.1 Sample		
	Frequency	Valid Percentage
number of individuals in 1991	26,932	100
missing occupational observations	16,671	
Total	**10,261**	**38**
Inapplicable occupational observations	4110	
Total	**6151**	**22.8**
Whereof:		
Manual workers	3085	50.2
Non-manual workers	3066	48.8
Total	**6151**	
missing SRH in 2009	2730	
Total	**3421**	**12.7**
missing SRH in 1991	1	
Total	**3420**	**12.7**
missing education in 1991	73	
Total	3347	**12.4**

3.3 ESTIMATION METHOD

In order to answer the main question 'what is the impact of manual occupation on health, compared to non-manual occupation in the UK' an Ordinal Logistic Regression Model is conducted. The hypotheses are:

H_0: manual occupation has no impact on health, compared to non-manual occupation.

H_1: manual occupation does have an impact on health, compared to non-manual occupation.

3.3.1. ORDINAL REGRESSION

Because health is an ordinal variable (excellent, good, fair, poor, and very poor) an Ordinal Logistic Model will be used. Ordinal regression is, like the linear regression model, a method to predict the impact of independent variables on a dependent variable. Using the ordinal logit model has the advantage that it takes into account that the distance between excellent and good health might be different to the distance between good and fair health.

The model can be expressed in the following way:

$$Ln\left(\frac{prob(event)}{1-prob(event)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

, where the left side of the equation is called a logit. A logit is the log of the odds that an event occurs. Because both probabilities have the same denominator, only the number of events divided by the number of non-events is left. The coefficients on the right side of the equation show how much the logit changes based on the values of the predictor variables.

If there are four scores given, the following odds can be defined:

$$\theta_1 = prob(score\ of\ 1)/prob(score\ greater\ than\ 1)$$

$$\theta_2 = prob(score\ of\ 1\ or\ 2)/prob(score\ greater\ than\ 2)$$

$$\theta_3 = prob(score\ of\ 1, 2, or\ 3)/prob(score\ greater\ than\ 3)$$

Three logits instead of four logits need to be defined. The last score does not need an odds associated with it, since that probability is 1.

A general equation of the odds can be written as:

$$\theta_j = prob(score\ \leq j)/1 - prob(score > j)$$

The ordinal logistic model is then:

$$Ln(\theta_j) = \alpha_j - \beta X$$

3.3.2. MODEL

The following odds can be specified to examine the association between occupation and health:

$$\theta_1 = prob(score\ of\ excellent\ health)/prob(score\ worse\ than\ excellent\ health)$$

$$\theta_2 = prob(excellent\ or\ good\ health)/prob(worse\ than\ good\ health)$$

$$\theta_3 = prob(excellent, good, or\ fair\ health)/prob(worse\ than\ fair\ health)$$

$$\theta_4 = prob(excellent, good, fair, or\ poor\ health)/prob(worse\ than\ poor\ health)$$

The fourth odds do not need to be estimated, because that probability is 1. The following model can be specified:

$$Ln(Health_j) = \alpha_j + \beta InitialHealth + \gamma OCC + X$$

, where *Health* is self-reported health in 2009, and is an ordinal variable (5 scores). $H_{Initial}$ represents the initial self-reported health in 1991 and *OCC* represents whether a person has a manual job (1) or a non-manual job (0). X_t denotes the personal characteristics age, gender, income and education.

3.4 VARIABLES

Dependent variable

Ill-health

Self-reported health (SRH) is used as a measurement of health. Idler et al. (1997) have found that self-ratings are found to be very valuable in measuring health status. Although SRH is a subjective measure, it represents the prediction of mortality better than several objective measures of health status. One shortcoming, which is relevant for this study is that people with a lower labour market status may report worse health to justifice for their reduced labour supply. However, McFadden et al. (2009) show that the relationship between SRH and mortality was similar in both manual and non-manual classes.

SRH is an ordinal variable, ranked from excellent to very poor health. BHPS has subdevided self reported health into excellent, good, fair, poor and very poor health. Approximately 70 percent of the individuals under study report excellent or good health (table 3.2). Converting *Health* into a binary variable would impose important information losses. A deterioration in health from excellent to good health for example would not be noticed. Furthermore, the distance between excellent and good health do not seem to be the same as the distance between fair and poor health. By using an ordinal logistic model the information of all five categories will be preserved.

Health - 1991	
Health	1991 (%)
Excellent	20.0
Good	49.8
Fair	22.5
Poor	6.1
Very Poor	1.6
Total	100

Independent variables

Initial good health

To measure initial health, the same SRH-scale is used. Initial health is reported in the year 1991, approximately eighteen years before the 2009-wave. The period between the years is taken this long to be possible to determine health deterioration due to occupation. In first instance, initial health was also not transposed to a binary variable, but remained an ordinal variable to keep relevant information in the analysis (table 3.3). However, because this variable causes serious problems for the parallel

Health	
Health	2009 (%)
Excellent	33.3
Good	48.7
Fair	14.1
Poor	3.5
Very Poor	0.4
Total	100

lines test (as will be explained in chapter 4), *Initial Health* is transformed to a binary variable where excellent and good health are set to '1', and fair, poor and very poor health are set to '0'.

Occupation

In the BHPS database occupation is split up according the 1990 *Standard Occupational Classification* (SOC) as developed by Peter Elias. The SOC is devided into three areas: the major groups (the main categories of occupational classification), subdevided into the minor groups, and the unit groups. The groups are classified with a 3 digit classification. The first number shows the major group (of which there are nine), the second number the minor group (of which there are 77) and the third number shows the unit group (of which there are 374). The table below shows the 9 major groups. The major groups cluster together the occupational groups who are similar in terms of experience, qualifications, skills and training. The major group can be split up in two groups of physical demanding and non-physical demanding work, which is used for the analysis. The minor and major groups fall into these categories, and do not need further specification for this study.

Table 3.3 Standard Occupational Classification - Major groups	
Non-physical demanding	**1** Managers & administrators **2** Professional occupations **3** Associate professional & technical occupations **4** Clerical & secretarial occupations
Physical demanding	**5** Craft & related occupations **6** Personal & protective service occupations **7** Sales occupations **8** Plant & machine operatives **9** Other occupations

Control variables

Education

In the BHPS the *International Standard Classification of Education* (ISCED) is included as a variable. ISCED is organized by UNESCO (United Nations Educational, Scientific and Cultural Organization) and organizes information on education and training, so that it is suitable for internatonally education comparisons. ISCED is devided into two cross-classification variables: levels and fields of education. The control variable *Education* includes the six levels of education (as shown in table 3.4), where zero is excluded.

Table 3.4 Education levels		
level		BHPS codes
0	pre-primary education	not included in BHPS
1	Primary education	1
2	Lower secondary education	2
3	Upper secondary education	3 + 4
4	Post-secondary non-tertiary education	5
5	First stage of tertiary education	6
6	Second stage of tertiary education	7

Income, age and gender

Three other socio-demographic variables are included in the analysis. *Income* and *Age* are continuous variables. The variable *Male* is created by recoding gender into a binary variable where male is one, and female zero.

3.5 DESCRIPTIVE STATISTICS

Table 2 in the appendix shows the descriptive statistics of the variables used in the regression model. It shows how the variables are divided into excellent, good, fair, poor, and very poor health in 2009. The table shows that poor and very poor health are underrepresented. During the period between 1991 and 2009, people being in poor health has increased from 602 to 1,008 persons. People being in good health has decreased from 2,745 in 1991 to 2,339 in 2009. The sample contains slightly more non-manual workers (1,760) than manual workers (1,587). The minimum age is 15. The sample contains more young people than old people. However, the study covers 18 years of study, so people falling in the category 31-45 in 1991, will belong to the category 45-60 or even >60 in 2009. The model includes 1,702 females and 1,645 males. Almost half of the sample has upper secondary education. Only 63 individuals out of 3,347 have got a higher degree, and 637 individuals only have primary education. 65.8% of the sample falls into the category with an annual household income of 10,000-30,000 a year.

4. RESULTS

Three models are created to study the effects of (non)manual occupation on health. In the first model, only the main variable *Occupation* is included in the model. In the second model, also *Initial health* is included, to see whether the effect of *Occupation* changes. In the third model, several socioeconomic control variables are included, which might also affect health.
In this chapter first several assumptions of the ordinal logistic model 1, 2 and 3 will be tested. After that, some overall model tests will be done to see whether the models are appropriate for interpretation. Next, the interpretation of the three models will be given. At the end, a fourth model is included, to examine the interaction effect of *Education* and *Occupation*.

4.1 ASSUMPTIONS

The ordinal logistic model contains three assumptions: the one dependent variable, adequate cell count, and parallel lines assumption. The first assumption, one dependent variable, is taken into account by using only the dependent variable *Health*. The other two assumptions will be explained and tested below.

4.1.1. ADEQUATE CELL COUNT

In ordinal logistic regression there is the rule of thumb that 80 percent of the cells in the crosstab must have more than 5 counts, and no cell should have zero count. Greater amounts of cells with less count results in a much less reliable chi-square test. In de first model, no cell has zero count and has at least 25 cell counts. In the second model this is similar, although the smallest cell has 17 counts now. In the third model this assumption causes some problems. 860 cells do have zero counts, which stays for 51.8 percent. This might cause difficulties with interpreting the Chi-square.

4.1.2. PARALLEL LINES

This assumption, also called the proportional odds assumption, means that if an independent variable affects the likelihood of a person being in the ordered categories, the coefficients linking the independent variable value to the different dependent outcomes will be the same across all the outcomes. So, manual occupation will affect the likelihood of an individual being in excellent health exactly the same as it will affect the likelihood of that individual being in poor health. If the assumption of parallel lines does not hold, the ordered logistic regression is not longer appropriate.

Table 4.1 shows the tests of parallel lines for the three models estimated. Only the first model holds. In both model two and three (only model three has been shown) the test of parallel lines is significant, which suggest that lines are not parallel. By excluding initial health from the model, the test is not significant anymore and the assumption of parallel lines holds.

However, initial health is an important control variable for the ordinal logistic model, so that exclusion is no option. To keep *initial health* in the model, the variable is transformed to a binary variable, where excellent and good health are set to '1' (good health) and fair, poor and very poor health to '0' (bad health). Now the assumption of parallel lines is significant for all three models[3].

Table 4.1 Tests of Parallel Lines*

		-2 Log Likelihood	Chi-Square	df	Sig.
Model 3 (initial health 1-5)	Null Hypothesis	3086,666			
	General	3031,401	55,266	27	.001
Model 1	Null Hypothesis	58,532			
	General	52,521	6,011	3	.111
Model 2	Null Hypothesis	99,666			
	General	92,436	7,230	6	.300
Model 3	Null Hypothesis	2328,382			
	General	2309,365	19,017	18	.391

* The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

4.2 THE MODELS

Before interpreting the individual parameters some overall tests will be done to examine the usefulness of the models.

4.2.1. GOODNESS-OF-FIT

The Pearson and Deviance Chi-Square test is computed to see how many predicted cell frequencies differ from observed frequencies. The null hypothesis states that the observed and expected cell counts are similar. The models 1 and 2 fit well, as shown by the large significance

[3] It has to be mentioned that by transforming the variable *Initial Health* to a binary variable information have been lost. E.g. the model does not say whether a person is in fair or in very poor health.

level for especially the second model (table 4.2). However, the third model shows an ambivalent result: the Pearson goodness-of-fit test is significant, while the Deviance goodness-of-fit test is not. As mentioned in the paragraph 'adequate cell count' this might be caused by the great amount of zero cell counts. The Chi-Square therefore is not appropriate.

Table 4.2 Goodness-of-Fit statistics

		Chi-Square	df	Sig.
Model 1	Pearson	6,035	3	.066
	Deviance	6,011	3	.067
Model 2	Pearson	9,190	10	.514
	Deviance	10,299	10	.506
Model 3	Pearson	1607,616	1318	.00
	Deviance	1230,633	1318	.958

4.2.2. MODEL FITTING INFORMATION

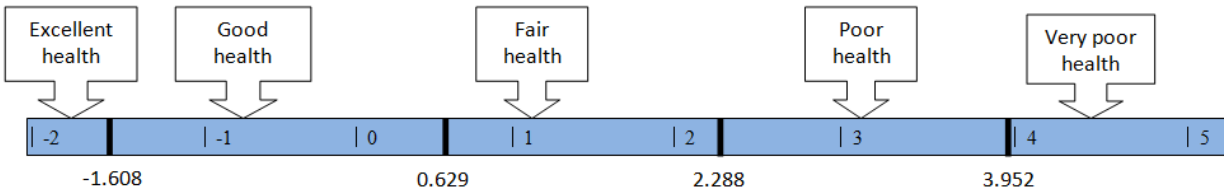
Before interpreting the parameters, an overall test is done to test whether the model without predictors is as good as the models with the predictors. The null hypothesis states that the location coefficients for one of more variables in the model are 0. The results are shown in table 4.3. The likelihoods of the 'intercept only' and the complete model are given. Because the small number, the natural log of the probability is given, multiplying by -2 (which makes the result positive). The -2LogLikelihood is a badness-of-fit indicator: large numbers mean poor fit of the model to the data. For model 1, the initial model has a value of 8302.761. By including occupation, the Likelihood is reduced to 8262.142, which represents a loss of 40.619. This results in a Chi-square of 40.618, which is significant with 1 degree of freedom. For all three models the Chi-square is significant, which means that for all three models the null hypothesis that the model without predictors is as good as the model with predictors can be rejected.

Table 4.3 Model Fitting Information					
		-2 Log Likelihood	Chi-Square	df	Sig.
Model 1	Intercept only	8302,761			
	Final	8262,142	40,618	1	,000
Model 2	Intercept only	8302,761			
	Final	8105,358	197,403	2	,000
Model 3	Intercept only	8302,761			
	Final	7983,210	319,550	6	,000

4.2.3. PARAMETER ESTIMATES

Model 1

Table 3 of the appendix shows that in model 1 all parameters are significant at a 95% significance level. The thresholds of health in 2009 represent the intercept α_j and indicate where the latent variable is cut to mark the border between the categories. The absolute values of the thresholds are not relevant for interpretation, but the distance between them, which are shown in the timeline below. It confirms the thought that the distance between different health states are not the same. The decrease in health from excellent health to fair health is bigger than moving from good health to poor health.



The effect of non-manual occupation on health is -0.424 and significant. This means that manual occupation has a significant negative impact on good health, compared to non-manual occupation.

Model 2

Next, the variable *Initial Health* is included into the model. The threshold values have slightly changed, though the differences in distance have remained. Only the decrease in health from fair to very poor health has relatively become bigger, compared to the other distances. All parameters stay significant.

The effect of non-manual occupation on health has increased to a -0,381 and is still significant at a 99% significance level. The effect of *Initial Health* on health is 1.044 and significant. That means that being in poor health has a negative impact on good health, as expected. For example, a person with initial good health with manual occupation has a $e^{-1.608} = 0.200$ times bigger probability of being in excellent health in 2009 than the probability of being in a worse health state. For a person with initial poor health this probability is $e^{(-1.608-1.044)} = 0.071$.

Model 3

In the third model several socioeconomic variables are included. The thresholds only change slightly and remain significant. The effect of non-manual occupation on health remain positive (although slightly less positive), and is still significant. Initial health, education, income and age are significant, and gender is not (0.544). Non-manual occupation, education and income have a positive effect on good health. Poor initial health and age have a negative effect on good health.

4.3 INTERACTION EFFECT

Because of the suggested interaction between occupational choice and education, an interaction term was added to the model. The several tests done are still significant (tables 4-7 in the appendix). However, was non-physical demanding job significant in the previous three models, now it is not. Also education is not significant anymore. Instead, the interaction effect of education and non-physical demanding job is significant at a 95% level and is positive. The other variables do not have remarkable changes.

5. CONCLUSION

The main objective of this study is to investigate whether manual occupation has a significant impact on health. Using an ordinal logistic regression, the results assume a relationship between the two factors. First a model is estimated with only the main variable *Occupation*. It is found that manual occupation has a significant negative impact on good health. Including initial health into the second model, and education, age, gender and income in the third model, this effect remains, although the effect in absolute sense is slightly less than in the first model. According to the three models can be concluded that manuals jobs have a significant negative impact on health. When adding an interaction term for the linkage between occupation and education, both non-manual occupation and education are not significant anymore. This stresses the relation between occupation and education.

To answer the main question, what is the impact of physical demanding jobs on health, compared to non-physical demanding jobs in the UK, can be concluded that physical demanding jobs do have a negative impact on health. Therefore the null hypothesis that physical demanding jobs have no impact on health, compared to non-manual occupation, is rejected.

6. DISCUSSION

This study has several limitations which should be taken into consideration. First, only the effect of occupation on health is examined. However, there are much more relations possible. For example, the model controls for initial health, but initial health might also influence occupational choice. Individuals with physical health problems may choose non-manual jobs more than manual jobs.

Next, the waves 1991 and 2009 are studied, without examining events during that period. If a person had a severe accident in 1991 and reports bad health in 1991 but is recovered in 1992 and is in good health again until 2009, the model conceives bad health as a 'long-term illness'. The opposite could also be that in 1991 an individual reports good health, but reports worse health in 2009 due to an accident. This short-term events may influence the results.

Third, initial occupation is not taken into consideration. It could be that a person has started as a manual worker, but some time is changed to a non-manual job (or vice versa). The initial job may have influenced health, but this is not represented in the model. It would be better to include more waves into the model, to diminish these effects. However, this study is restricted to two waves to prevent a too extensive research.

Fourth, the theoretical framework has described a paragraph about to the effects of mental conditions on health. The job-strain model states that a combination of both a high psychosocial workload and low decision latitude in a job will lead to a higher risk of cardiovascular diseases. This study only studied the effects of manual occupation on health, but did not take into account factors as mental conditions and therefore is limited in that sense. It could be that mental conditions have a significant influence on health. However, it is not clear what the exact mental conditions of manual and non-manual jobs are, and it can even differ within the two categories.

Fifth, to meet the assumption of parallel lines, the variable Initial Health was transformed to a binary variable. This may have caused information loss.

Sixth, because half of the cells in the third model have zero count, the results may have diverged. However, it is supposed that this fact has not created serious problems for the interpretation.

Seventh, the categories poor and very poor health are underrepresented in this thesis, which may limit the precision of the analysis.

This study has focussed on Great Britain. It cannot be generalized to other countries, due to different working conditions, health insurance legislation and other factors which might have an influence across countries.

7. REFERENCES

Cutler D.M. and A. Lleras-Muney. (2006). "Education and Health: Evaluating Theories and Evidence". Working Paper no. 12352, National Bureau of Economic Research, Cambridge.

Doorslaer, E. van, H. Van Kippersluis, O. O'Donnell and T. Van Ourti. (2008). Socioeconomic Differences in Health over the Life Cycle: Evidence and Explanations. Netspar Panel Papers, no. 12.

Ferrie, J.P. (2001). 'The Poor and the Dead: Socioeconomic Status and Mortality in the U.S., 1850-1860'. Historical paper 135, National Bureau of Economic Research, Cambridge.

Fletcher, J.M., J.L. Sindelar and S. Yamaguchi. (2009). "Cumulative Effects of Job Characteristics on Health". Working Paper no. 15121, National Bureau of Economic Research, Cambridge.

Grossman, M. (1972). On the concept of health capital and the demand for health. *The Journal of Political Economy, 80*(2), pp. 223-255.

Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior, 38*(1), pp. 21-37.

Karasek, R.A. (1979). Job Demands, Job Decision Latitude, and Mental Strain: Implications for Job Redesign. *Administrative Science Quarterly*, 24(2), pp. 285-308.

Kenkel, D.S. (1991). Health behavior, health knowledge, and schooling. *The Journal of Political Economy, 99*(2), pp. 287-305.

Lammers, J., B. Pelzer, J. Hendrickx and R. Eisinga. (2007). *Categorische Data Analyse met SPSS. Inleiding in loglineaire analysetechnieken*, first edition, Koninklijke Van Gorcum BV, 2007.

McEwen, B.S. (2000). Allostasis and Allostatic Load: Implications for Neuropsychopharmacology. *Neuropsychopharmacology,* 22(2), pp. 108-124.

McFadden, E., R. Luben, S. Bingham, N. Wareham, A. Kinmonth and K. Khaw. (2009). Does the association between self-rated health and mortality vary by social class? *Social Science & Medicine, 68*(2), 275-280.

Murphy, L.R. (1991). Job Dimensions Associated with Severe Disability Due to Cardiovascular Disease. *Journal of Clinical Epidemiology*, 44(2), pp. 155-166.

Smith, J.P. (1999). Healthy Bodies and Thick Wallets: The Dual Relation between Health and Economic Status. *The Journal of Economic Perspectives,* Vol. 13(2), pp. 145-166.

Sindelar, J.L., J. Fletcher, T. Falba, P. Keenan and W.T. Gallo. (2007). "Impact of First Occupation on Health at Older Ages". Working paper no. 13715, National Bureau of Economic Research, Cambridge.

APPENDIX

Table 1 Variables - dataset		
Variable	Question	Year
Self-reported ill-health (RHLSTAT)	'Please think back over the last 12 months about how your health has been. Compared to people of your own age, would you say that your health has on the whole been...' (excellent, very good, good, fair, poor)	2009
Initial ill-health (AHLSTAT)	'Please think back over the last 12 months about how your health has been. Compared to people of your own age, would you say that your health has on the whole been...' (excellent, very good, good, fair, poor)	1991
Occupation (AJBSEG)	What was your (main) job last week? Please tell me the exact job title and describe fully the sort of work you do.	1991
Income (AFIHHYR)	Annual household income (1.9.90-1.9.91)	1991
Education (AISCED)	ISCED levels	1991
Age (AAGE)	Age at date of interview	1991
Gender (AHGSEX)	Gender	1991

Table 2 Crosstabs							
		Health status - 2009					
		Excellent	Good	Fair	Poor	Very Poor	Total
Initial health - 1991	Good	621	1418	544	129	33	2745
	Poor	50	250	207	76	19	602
Total		**671**	**1668**	**751**	**205**	**52**	**3347**
Occupation	Non-manual	402	911	325	95	27	1760
	Manual	269	757	426	110	25	1587
Total		**671**	**1668**	**751**	**205**	**52**	**3347**
Gender	Female	342	852	371	110	27	1702
	Male	329	816	380	95	25	1645
Total		**671**	**1668**	**751**	**205**	**52**	**3347**
Education	Primary	76	281	202	64	14	637
	Lower secondary	24	76	28	7	1	136
	Upper secondary	316	788	348	74	21	1547
	Post-secondary non-tertiary	150	322	112	47	10	641
	First stage of tertiary	95	162	52	9	5	323
	Second stage of tertiary	10	39	9	4	1	63

Total		**671**	**1668**	**751**	**205**	**52**	**3347**
Annual household income	< 10,000	51	155	106	32	15	359
	10,000 - 30,000	430	1092	517	137	26	2202
	>30,000	190	421	128	36	11	786
Total		**671**	**1668**	**751**	**205**	**52**	**3347**
Age - 1991	15 - 30	279	539	185	48	13	1064
	31 - 45	266	671	316	92	22	1367
	46 - 60	114	414	209	60	15	812
	> 60	12	44	41	5	2	104
Total		**671**	**1668**	**751**	**205**	**52**	**3347**

Table 3 Parameter Estimates Ordinal Logistic Regression

		Model 1	Model 2	Model 3
Threshold	Excellent health (2009)	-1,608	-1,443	-1,802
	Good health (2009)	,629	,865	,586
	Fair health (2009)	2,288	2,577	2,331
	Poor health (2009)	3,952	4,258	4,020
Location	Non-manual job	-,424	-,381	-,190
	Poor health (1991)		1,044	1,047
	Gender (female)			-,040
	Education			-,121
	Income			-,301
	Age			,310

* significant at a 99% significance level

** significant at a 95% significance level

Table 4 Parameter Estimates Ordinal Logistic Regression			
		Model 3	Model 4
Threshold	Excellent health (2009)	-1,802	-1,59
	Good health (2009)	,586	0,801
	Fair health (2009)	2,331	2,547
	Poor health (2009)	4,020	4,236
Location	Non-manual job	-,190	0,195
	Poor health (1991)	1,047	1,049
	Gender (female)	-,040	-0,038
	Education	-,121	-0,048
	Income	-,301	-0,294
	Age	,310	0,32
	Non-manual job * Education		-0,133
* significant at a 99% significance level ** significant at a 95% significance level			

Table 5 Tests of Parallel Lines*					
		-2 Log Likelihood	Chi-Square	df	Sig.
Model 3	Null Hypothesis	2328,382			
	General	2309,365	19,017	18	,391
Model 4	Null Hypothesis	2323,283			
	General	2301,544	21,740	21	,415
* The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.					

Table 6 Model Fitting Information					
		-2 Log Likelihood	Chi-Square	df	Sig.
Model 3	Intercept only	2647,932			
	Final	2328,382	319,550	6	,000
Model 4	Intercept only	2647,932			
	Final	2323,283	324,649	7	,000

Table 7 Goodness-of-Fit statistics				
		Chi-Square	df	Sig.
Model 3	Pearson	1607,616	1318	,000
	Deviance	1230,633	1318	,958
Model 4	Pearson	1584,149	1317	,000
	Deviance	1225,534	1317	,965