ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics
Department of Economics

HEALTH INEQUALITIES ACROSS DIFFERENT ETHNIC GROUPS IN THE NETHERLANDS

Kim Thu Mai
305492km
kim_mai_@hotmail.com

Supervisor: Teresa Bago D'Uva

- August 2009 -

HEALTH INEQUALITIES ACROSS DIFFERENT ETHNIC GROUPS IN THE NETHERLANDS

Kim Thu Mai*

Erasmus University Rotterdam; August 2009

Abstract

In the Netherlands, the overall health of allochthonous people is comparably lower than the average Dutch citizen. Health inequality minimization has been on the agenda for decades. Various studies have researched the existence of health inequalities and the minimizing necessity, including and excluding the Netherlands in the analysis. However, there is limited empirical evidence available. The aim of this paper is to research and provide evidence on whether nationality or ethnicity has an explanatory role in health inequalities, by using the AVO2003 and the ESE-student survey. After using a logit model for a measure of self-assessed health and odds ratios for interpreting this model, our results indicate differences in perceived health amongst various ethnic groups in the Netherlands. Allochthonous individuals, especially non-western allochthonous people, report a lower level of self-assessed health than autochthonous individuals in the Netherlands. Further research focusing on the root causes of the existing health inequalities will facilitate its minimization.

Keywords: *Self-Assessed Health, Health Inequalities, Logit Model, Nationality, and Ethnicity*

* Kim Thu Mai, author of this Bachelor Thesis, would like to gratefully acknowledge help from and useful conversations with Teresa Bago D'Uva, Professor at the Erasmus University Rotterdam and supervisor for this work.

TABLE OF CONTENT

1. Introduction

The Netherlands had an influx of different nationalities, being immigrants and refugees, since the late sixteenth century. From this moment, the (poor) immigrants came from neighboring countries such as Germany and the Scandinavian countries to find a job, or they were fleeing their country of origin mostly for religious reasons. After WWII there was a rise in new migrants or workers, because of the scarcity of native people in the Dutch labor market interested in the low skilled labor. These migrants initially came from the Mediterranean countries, whereas later, the majority originated from Turkey and Morocco. However, the migrants from Italy and Spain returned to their mother countries, while the Turkish and Moroccan people requested for permanent stay. Moreover, a substantial part of the immigrants emigrated from the former Dutch Colonies: formerly from Indonesia and Moluccas, and later on from Suriname and the Dutch Antilles. More recently, refugees are mainly coming to the Netherlands for political and socioeconomic reasons (Report preliminary inquiry website Migration History of the Netherlands[1], 2006)

Currently, the Netherlands counts 180 different nationalities, where the major ethnic groups are the Turkish, Moroccan, Surinam, and Dutch Antilles. In the Dutch society, 10.8% of the population has a non-western ethnic background (SEGV, 2008). These population subgroups have been regarded as the most important allochthonous groups and have been focused on in most of the research. Even though the overall knowledge regarding the health of allochthonous people in the Netherlands has been scarce, it can be noted that the overall health of these population subgroups are lower compared to the average Dutch citizen (ZonMW, 2009). Furthermore, the World Health Organization (WHO) has prioritized health inequalities between and within countries for many centuries. In the Alma Ata declaration of 1978[2], all nations of the WHO acknowledged that economic, social, and political development is only achievable by minimizing health

[1] This is an English translation. The original Dutch name is "Migranten in Nederland. NL". Report vooronderzoek website migratiegeschiedenis.
[2] Alma Ata declaration was adopted at the international conference on Primary Health Care in USSR, at 6-12 September of 1978.

inequalities (SEGV, 2008). Furthermore, it may be mentioned that from an economic, social, and political perspective the existence of inequalities is unacceptable.
Hence, much information appears to exist regarding the presence of health inequalities and the necessity to minimize them. However, our aim is to examine whether nationality or ethnicity has an explanatory role in health inequalities. In other words, this research focuses on the ability to determine variations in perceived health across different ethnic groups, using nationality and ethnicity as the most important explanatory variables.

The prominence of this paper is to see whether there is a correlation between a person's ethnicity and its self perceived health within the Netherlands while controlling for other factors such as age, gender, household income, household size, and education. Furthermore, we desire to document prevalence of health inequalities and describe the Dutch population with the given data. Moreover, for the purpose of our study we used two available databases AVO2003 and the ESE-student survey. The first dataset represents data from a national-based survey from 2003 and the latter dataset from a student-based survey at ESE[3]. Consequently, the method adopted in this research is to divide the research question into several partial questions. First, we examine whether nationality has a significant effect on self perceived health, with the use of the national-based survey data. Secondly, we test whether ethnicity influences a person's self-assessed health and whether nationality remains to have the similar effect. Lastly, we investigate whether nationality is still an important explanatory variable for self-assessed health when strongly controlled for age and education by using the other dataset, which is based on the ESE student survey. Therefore, by combining the results of the partial research questions at the end, we are able to answer the question whether nationality and/or ethnicity influences self perceived health.

The remainder of this paper is organized as follows. The next section describes the relevant literature. Then, section 3 provides an elaborate description of the applied methodology and the specification of the different models used. In section 4, the dataset

[3] ESE stands for the Erasmus School of Economics, the economics department at the Erasmus University Rotterdam.

as well as the used sample are presented and the different variables are defined in detail. The results are developed in section 5. Section 6 concludes.

2. Related Literature

Health is of vital interest for people as it influences a person's overall capabilities and daily functioning. However, health cannot be simplified to a person's self perception or medical condition, but it is defined as *"a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity"* according to the WHO[4]. Moreover, as medical care appears necessary for sustaining a healthy life, the general concern of its distribution is justified. For this reason, governmental agencies worldwide have strongly emphasized on the equal access to and distribution of medical care. Additionally, one of the main purposes in any health care system is a reduction of inequalities in health and in the access to health services. These concerns stem from the fundamental aim of all citizens being healthy. Despite the fact that the absence of any inequality remains an unattainable goal, health care systems can influence the extent of health inequalities and its relatedness to characteristics such as income, gender, etc. Furthermore, it is a well-known fact that variations in health are largely determined by factors outside the health care system (Wagstaff and Van Doorslaer, 2000). This paper demonstrates that an individual's characteristics, such as socioeconomic status, most likely contribute to health inequalities and not the characteristics of a health care system.
To elaborate on this issue, another feature that has not been researched deeply is ethnicity or nationality and the impact it has on health inequalities. This is exactly the main interest of this paper. Our research focuses on the ability to determine variations in perceived health in the Netherlands with the use of the main variable nationality or ethnicity. Moreover, we want to find an answer to the question whether nationality or ethnic background has an influence on how a person will assess its personal health, by also

[4] WHO definition of health, adopted by the International Health Conference, New York. As a preamble to the Constitution of the World Health Organization.

controlling for other variables. Thus in short, health inequalities in this paper reflect the variations in self-assessed health across different ethnic groups

Previous literature on this specific subject has mostly emphasized on investigating health inequalities related to socioeconomic inequalities and income or ethnic related health inequalities. However, in most of these studies, the Netherlands were omitted from the picture. An overview of the most important papers concerning this particular issue and their main findings and methods is provided below.

Recently, a study by Hernández Quevedo and Jiménez Rubio (2008) exploit the Spanish National Health Survey (SNHS) with the goal of comparing health status and health care utilization patterns between the Spanish autochthonous population and foreigners. The aim of their paper is to provide evidence that inequalities do exist between the two groups. The pooled ordered probit method, for both self-assessed health as well as for several utilization patterns, has been adopted here. The authors conclude that differences in health status and utilization of health care between nationals and immigrants do exist. Their results indicate that immigrants report better levels of self-assessed health compared to Spaniards. However, the health care utilization levels were lower for immigrants. Hence, nationality seems to be statistically significant in explaining health status.

For a case that includes morbidity as dependent variable, Banks et al. (2006) explore the existence of relative health status differences between elderly in England and in the United States and how it varies by socioeconomic status. The methods considered are self-reported measures of health outcomes as well as risk factors, and ordinary least square regression (OLS) models. Similar measures of socioeconomic status (SES) are adopted in our research. This paper uses data from national surveys of both countries, namely the US Health and Retirement Survey and the English Longitudinal Study of Aging, whereas the analysis is limited to non-Hispanic whites in both countries. The authors findings indicate that, based on self-reported illnesses and biological markers of

disease, English residents are healthier than their US counterparts. Moreover, these differences exist in all the layers of the SES distribution.

For the specific case of Canada, Gee et al. (2007) combine both existing literature and statistical data from the Canadian Census and National Population Health Survey in order to examine ethnic inequalities in both economic and health dimensions. In addition, the method applied is a logistic regression analysis. The measures employed in this study are ethnic origin, mother tongue, and immigration status. Their results support the existence of the phenomenon known as the 'healthy immigrant effect'[5] only for chronic conditions and not for self-assessed health. Also, they show that cultural characteristics, such as ethnic origin, are important predictors of health. Moreover, Wang et al. (2000) corroborate this point by concluding that both ethnic origin and immigrant status are factors influencing the self reported prevalence of arthritis in the Canadian population. Thus, it can be concluded that health inequalities are frequently influenced by ethnic inequalities.

For a case that includes the Netherlands, but also 10 other countries, Mackenbach et al. (1997) analyse the socioeconomic inequalities in morbidity as well as mortality in Western Europe. Several data sources from 11 different countries were taken into account, ranging from longitudinal to cross-sectional studies and national surveys. Odds ratios are used to measure morbidity and rate ratios for mortality. The aim of this paper is to draw conclusions on income-related inequalities between the different western European countries and whether they can provide evidence to support or contradict the conventional views on the health inequalities of between-country patterns in western European countries. On the one hand, morbidity is measured using four health indicators to compare the different countries, which are perceived general health, chronic conditions, long-term disabilities, and long-standing health problem. On the other hand, mortality was measured by age at death. To conclude, for morbidity the study suggests that welfare policies may affect one dimension of socioeconomic inequalities in health, but countries with an egalitarian influence do not have the lowest inequalities in health.

[5] The phenomenon that indicates a superior health status of recent immigrants compared to that of the native population of the immigrant receiving countries.

The results of the study also show that inequalities in health are dynamic and determined by various factors during a person's lifetime.

Nevertheless, in the last few years extensive research has been preformed on ethnic differences in risk factors, health, and health care in the Netherlands due to the increasing number of Dutch people with a non-western ethnic background during the last decennia. However, the various studies in the last 10 years have shown remarkable differences in methodology, focus, as well as results. Conversely, most of the studies had a descriptive function of several health situations, whereas only a limited amount of studies actually attempted to research the root causes and underlying factors of these health situations (Kunst et al., 2008). According to this paper, these previous studies mainly aimed at determining which specific diseases influence the health of Dutch people with a non-western ethnic background. Furthermore, it is determined that non-western allochthonous people have a 22% higher burden of disease, measured with DALY's[6], compared to all the people in the Netherlands. However, the data for most of the studies have shown problems in internal validity and representativity, lack of usage of information from Dutch databases and questionnaires, incomplete methods of research, and limitations in representing results. Nonetheless, this paper aims to contribute to the research literature on health inequalities and ethnic differences in the Netherlands. Moreover, we aspire to strengthen this general conclusion with the argument that due to the higher burden of disease of non-western allochthonous people, nationality or ethnicity of an individual will most likely have an effect on their self-assessed health. The data used in this paper corresponds to the AVO2003 National Household Survey and the Erasmus School of Economics student survey 2008.

[6] DALY, according to WHO, is short for Disability Adjusted Life Years. It is a time based measure that summarizes two effects; one representing the effect on expected life years lost and the other representing the effect of years lost living in other states than perfect health. The latter also takes into account the severity of the disease.

3. Methodology and Hypotheses

In the introduction, the main research question and some sub questions were mentioned briefly. In this section, they are worked out more elaborately and the empirical specification of the models is provided in detail.

In this study, we focus on the following research question:

'Does an individual's nationality or ethnic background have a significant impact on how a person assesses its own health?'

We opted to use the logit model[7] to examine this question, with self-assessed health (SAH) being the dependent variable (Field, 2005). Advantages of using SAH are that it is the most extensively used measure of health in the existing literature (Kakwani, et al.,1997) and it is commonly known to be a good predictor of other outcomes, especially with respect to mortality and illness. However, there are concerns about its validity when comparing different population subgroups on their health levels and their way of measuring health inequalities, whereby each cultural group may have different norms and expectations. According to Lindenboom and Van Doorslaer (2004), homogeneous reporting is needed to ensure that the responses to health questions reflect the true health differences. Nevertheless, we use SAH since it is the most frequently used measure.

The basic formula of the logit model, that is an expression for the probability of the respons variable being equal to one, is shown by equation (0.1). This model looks complicated because of its non-linear shape, but it can be rewritten simpler as in equation (0.2). Here, the logit is defined as the natural log of the odds.

$$P(Y=1) = \frac{1}{1+e^{-\text{logit (p)}}} = \frac{1}{1+e^{-(\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_nX_n)}} \tag{0.1}$$

$$\text{logit }(p) = \ln(odds) = \ln\left[\frac{p}{1-p}\right] = \beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_nX_n \tag{0.2}$$

[7] Logit model is also known as a logistic regression model.

The motivation behind this model is that it is most appropriate for a binary dependant variable and numeric or categorical predictor variables. Instead of using a linear regression, this model overcomes the problem of violating the assumption of linearity due to the logarithmic transformation (Wooldridge, 2006). In the end, this model can be used to determine the probability that a person will assess him- or herself as being healthy by filling in the individual's values for all predictor variables into the estimated equation.
The interpretation of the coefficients' estimates in the logit model is more difficult than in a linear regression model since the influence of an increase or decrease of one of the explanatory variables on the dependent variable is dependent on the value of that explanatory variable. Therefore, we use odds ratios for the interpretation of the model. The odds of an event, in our case a person assessing him- or herself as being healthy, are defined as the probability that the event occurs divided by the probability that the event does not occur. Moreover, the use of qualitative information in regression analysis is executed by creating dummy variables. The estimated coefficients for all different dummy subgroups are interpreted by comparing the dummy with the reference group.

After estimating the *β coefficients of the logit model,* the odds ratios can be derived as the value of exp β (e^{β}). Thus there is a direct relationship between the coefficients and the odds ratios (UCLA Academic Technology Services). The coefficients are in terms of log odds, whereby a one unit change in the independent variables results in a β-value change in the log of the odds. Moreover, the exp β indicates the proportional change in odds resulting from a unit change in the predictor or independent variable. For example, odds for people with a foreign or double nationality see equation 0.3.

$$\Delta\text{odds} = \frac{\text{odds for people with Foreign or double nationality}}{\text{odds for people with Dutch nationality}} \tag{0.3}$$

Furthermore, in addition to the main research question mentioned above, several null and alternative hypotheses are formulated and tested in three models that are described subsequently.

First of all, the most important objective of the first two models is to explain whether a person's nationality has a significant effect on its self-assessed health. Afterwards, the continuation of this research is to use the specific variable ethnicity. In an attempt to test whether ethnic background significantly influences an individual's SAH. For the latter, the individuals are divided into autochthonous, western allochthonous, and Non-western allochthonous people. Consequently, to investigate the possible correlation between nationality and ethnicity, and SAH, we first look at nationality and in the second model, we add ethnicity into the equation. The third and final model then tests whether nationality remains to have a significant effect on SAH, even while it is based on a select group of individuals, namely students. It should be remarked lastly that the first two models utilize the AVO2003 National Household Survey and the third model purely uses student-based data from the Erasmus School of Economics student survey 2008, as will be discussed in more detail in the next section. Hence, the possibility arises to further control for age and education in our respective models.

Model 1:

H_{01} People with a double or foreign nationality have the same level of health as (only) Dutch people, controlling for all other variables.

H_{A1} People with a double or foreign nationality do not have the same level of health as (only) Dutch people, controlling for all other variables.

We expect a negative association between people with multiple nationalities compared to people with only a Dutch nationality, regarding their self-assessed health. According to the related literature (Kunst et al., 2008), the higher burden of disease of non-western allochthonous, compared to the Dutch population, will most logically have a negative impact on the self perceived health of allochthonous people.

The first model is given by the following expression:

$$\ln\left(\frac{P(SAH=1)}{1-P(SAH=1)}\right) = \beta_0 + \beta_1 \text{nationality} + \beta_2 \text{age} + \beta_3 \text{gender} + \beta_4 \text{education} + \beta_5 \text{income} + \beta_6 \text{householdsize} \quad (0.4)$$

In equation 0.4, the $\ln\left(\frac{P(SAH=1)}{1-P(SAH=1)}\right)$ represents the latent level of self-assessed health. After estimating this equation and filling in the values of the predictor variables for a certain individual, it can be used to work out expression 0.1. This then results in a value between 0 and 1 indicating the probability that the individual assesses him or herself as being healthy. The interpretation of this value is that the nearer it comes to one, the higher the probability that he or she feels healthy. Moreover, the different predictor variables (nationality, age, gender, education, income and household size) are all categorical variables except household size. How most of these different variables are coded as dummies is explained extensively in the next section. As mentioned before, the main focus of the first model is on the significance of the nationality variable, while the significance of the other independent variables is of secondary importance.

Model 2:

H_{02} Allochthonous people or people with a double or foreign nationality have the same level of health as people with only a Dutch nationality and ethnicity, ceteris paribus.

H_{A2} Allochthonous people or people with a double or foreign nationality do not have the same level of health as people with only a Dutch nationality and ethnicity, ceteris paribus.

Our expectation for the variable nationality is a weakened effect of the odds compared to the previous model since the ethnicity variable may possibly include overlapping information. It is suggested that ethnicity has the same relationship with SAH as nationality, however even more strongly. To elaborate further, we expect that the ethnicity variable incorporates the effect that people with a foreign or double nationality have probably a lower self-assessed health compared to people with merely a Dutch nationality, as was mentioned before, and combines this with incorporating the effects of a different ehtnic background. It is expected that allochthonous people assess themselves as being less healthy compared to autochthonous people.

The second model is given by the following expression:

$$(0.5)\qquad \ln\left(\frac{P(SAH=1)}{1-P(SAH=1)}\right)=\beta_0+\beta_1\text{nationality}+\beta_2\text{ethnicity}+\beta_3\text{age}+\beta_4\text{gender}+\beta_5\text{education}+\beta_6\text{income}+\beta_7\text{householdsize}$$

This model is almost equivalent to the previous model. However this model includes the independent variable ethnicity.

The motivation for the third model, which is based purely on the student-focused survey, is that it gives the opportunity to control in a more extensive way for the variables age and education. After all, the survey was completed by individuals of the same age and the same level of education. This survey enables to strengthen the previous model (based on Dutch national survey data) by predicting the possible correlation between an individual's nationality and his or her self perceived health. In addition, this model does not include the education variable since all the respondents have the same level of education (post secondary education). Nonetheless, we have decided to also analyze the same model including the education levels of the parents. This can be found in section 5.3. Conversely, the age variable remains to be included, since there are some differences in age of the individuals who have completed this survey.

Model 3:

H_{03} Nationality has no effect on the odds of feeling healthy, controlling for all other variables especially for education and age.

H_{A3} Nationality does have an effect on the odds of feeling healthy, controlling for all other variables especially for education and age.

We expect that the people with multiple nationalities will feel less healthy compared to people with only a Dutch nationality and moreover we expect this to be more significantly so than in the previous models. This assumption is derived from the fact that this model is run on student-based survey data and this gives the ability to control even more for age as well as education, since all respondents belong to the same age group and have similar education levels.

Once more, the third model is given by the following expression:

$$\ln\left(\frac{P(SAH=1)}{1-P(SAH=1)}\right) = \beta_0 + \beta_1 \text{nationality} + \beta_2 \text{age} + \beta_3 \text{gender} + \beta_4 \text{income} + \beta_5 \text{householdsize} \tag{0.6}$$

It can be remarked that this model is very similar to the previous models. The main difference is that it is based on another dataset, what will be discussed elaborately in the next section. The model can be interpreted in the same way as explained above for the first model.

Furthermore, we estimate these binary regression models for self-assessed health using the statistical program SPSS 16.0. The data provided by AVO2003 is collected on an independent and random basis to make the results representative for the Dutch population. Besides, we have conducted robustness tests for each model, taking into account a different specification of the dependent variable self-assessed health, in the effort to strengthen the results of our main models. The details are provided in section 5.

4. The data

For the purpose of our study, we use two different datasets. The first one is a national-based household survey from 2003 and the latter one is a student-based health, education, and behavior survey from the Erasmus School of Economics 2008. In the following section, we elaborate on each dataset as well as on all the variables used for our models.

4.1. The AV02003 survey

The AVO2003 survey (Aanvullend Voorzieningengebruik Onderzoek) is executed by GFK Panel Services Benelux in 2003-2004 under the authority of the Netherlands Institute for Social Research in Den Haag. The AVO2003 is the seventh research of a four-yearly series, which started in 1979. The main goal is to obtain insights into the

extent to which people in the Dutch households from the age of six use social and cultural provisions of varying nature.

The research started on the 15th of September 2003 and ended on the 5th of March 2004 and both oral interviews and written surveys were used to collect the data. A representative and geographically distributed sample of addresses was taken out of the TPG-Post database. Furthermore, the gross sample consisted out of 10.680 addresses of which 6.404 provided successful surveys of a complete household. The net response ratio was thus approximately 60%. The completed surveys were processed manually resulting in data for 13.776 persons of over 6 years old living in a Dutch household. From this database, only the variables that are closely related to self-assessed health are considered in this research. An elaborate description of those variables is provided further. After filtering the data to fit it accordingly to the logistic model, 10.102 observations remained.

4.2. The ESE-student survey

The student survey is an internet based survey which was completed by 511 out of the 600 students that were enrolled in the second bachelor course 'Methods, Techniques and Research Project' during the academic year 2008-2009 at the Erasmus University Rotterdam. Since students were granted one bonus point in their final grade when they responded to the survey, we anticipate that the students took the survey seriously and sincerely completed all questions. Hence, this corrects for the possible invalidity concern. After adjusting for missing data, 470 participants were taken into account in the sample. Only the variables that are closely related to self-assessed health are considered in this research. An elaborate description of those variables is provided in the next section.

4.3. Specification of the used variables

Dependant variable: *Self-Assessed Health (SAH)*

The dependant variable in this model is self-assessed health (SAH). The scales used in the two surveys (AVO2003 and ESE-student survey) to asses this variable are similar.

The given options include excellent, very good, good, fair and poor. The options *I don't know* or *I don't want to answer* are accounted for as missing values. Furthermore, a definite limitation of this variable is that it can be considerably biased since it does not take into account the possibility of misrepresented self perception. It is suggested here that the timing of the completion of the survey may influence the interviewee's judgment of self perceived health, due to its mood of the day. The interpretation of questions and answer categories also depend on expectations, optimism, etc. Moreover, as was mentioned previously, interpretation of questions and health status may differ per person or population subgroups. However, this model does control for other socioeconomic variables, such as age, gender, household income, household size, education, nationality and ethnicity.

The dependant variable, which consists of five previously mentioned subgroups is transformed and recoded into a binary one such that a logit model can be executed. Similarly to the procedure from the 1997 article by Mackenbach et al., the combination of the options into a dichotomous variable is done by grouping the options good, very good and excellent together and giving them the value one (1), while the options fair and poor are both turned into the value zero (0).

Independent variables:

Nationality and Ethnicity

As mentioned before, the impact of these two variables on the dependent variable is the main focus of this research. Firstly, regarding the nationality variable, the AVO2003 database includes three options, namely *the Dutch nationality*, *both the Dutch and a foreign nationality*, and *a foreign nationality*. A person's nationality is not limited to the country of birth, since it can also reflect the place of residence. Therefore, it is possible for a person to have multiple nationalities. Hence, the model includes two dummy variables for which the reference group consists of people with a Dutch nationality. The first dummy thus allows a comparison between people with a Dutch nationality and people with more than only the Dutch nationality, while the second dummy makes it possible to compare people with either a Dutch or a foreign nationality.

Moreover, concerning the ESE-student survey, the same dummy variables and reference group are used. However, the subgroup including those with a foreign nationality also encompasses students with foreign nationalities that are studying in the Netherlands as part of an exchange program.

Furthermore, it is common in the Netherlands to make a distinction between a person's nationality and its ethnicity. Due to the rich information provided in the AVO2003 survey, it is possible to take the ethnicity into account. However, the exact numbers about the sheer size of each ethnicity have been difficult to obtain. Due to the definition of ethnicity used in the Netherlands, this variable cannot be obtained immediately from the dataset provided. Besides, individuals in the Netherlands are classified either as allochthonous or autochthonous people and this distinction only occurs in the Netherlands. The exact definition of an allochthonous person according to the Statistics Netherlands[8] is an individual who is either born abroad him- or herself, or of which at least one of the parents' country of birth is not the Netherlands (den Heeten and Verweij, 1993).

Additionally, the Dutch government (Department of Internal Affairs, in collaboration with VNG[9]) has developed three criteria to be able to identify a person's ethnicity. These are the native country of the individual and those of both parents. To be as concise as possible, the countries of one's birth are clustered into three groups:

A1 the Netherlands

A2 all other western native countries, such as Northern and West Europe, North America, Oceania, Japan, Israeli/Palestine, and Indonesia

B all other remaining native countries, such as Eastern and Southern Europe, Asia, Latin America and the Caribbean, and Africa.

The next step is to work according these identification criteria such that an individual's ethnicity can be specified, whereby the priority is given to the least developed nations (B

[8] In Dutch this is called CBS (Centraal Bureau voor de Statistiek).
[9] From the same paper of Heeten and Verweij, 1993; VNG means in Dutch: Vereniging Nederlandse Gemeentes

has priority over A2 and A2 over A1) as well as to the persons own country of birth (own country of birth is superior to the mother's native country, which is superior to the father's country of origin). To determine the ethnicity, the applied rule is to first look at the person's own native country to see whether or not it is categorized as a B-country. If this is not the case, the mother's country of origin is examined to check if it is a B-country. When this is the case, the individual's ethnicity is classified as B. If not, you continue to look at the father's country of birth and check for a B-country. Further, when none of these countries are B-countries, this procedure is continued for the A2 countries. Finally, the option arises to assign the individual with an unknown ethnicity or simply a Dutch ethnicity (category A1). The following example clarifies the procedure used to classify one's ethnicity: Philip is born in the Netherlands, but his mother is born in France and his father in Nigeria. Philip's ethnicity is Nigerian, whereas his nationality is Dutch.

Besides, this procedure is consistent with the definition of Statistics Netherlands. They divide ethnicity globally into *western allochthonous* and *non-western allochthonous* people. Furthermore, the countries of birth are divided into nine subgroups, which can be further categorized into the following three categories.

- Autochthonous: the Netherlands.
- Western allochthonous: former Dutch East Indies and Indonesia, Europe, other Western countries.
- Non-western allochthonous: Turkey, Morocco, Surinam, Dutch Antilles and Aruba, other non-Western countries.

It can thus be noticed that this division is closely in line with the rules that were applied in this research.

From what has been stated above, the possibility arises to test whether the self-assessed health varies significantly for a western or non-western allochthonous individual compared to the reference group, which consists of autochthonous people.

Summarizing, to include the ethnicity variable into the model, a specification of a person's ethnicity is necessary. This specification is performed by combining the information regarding the country of origin for the individual him- or herself as well as

for both parents with the criteria of the Dutch government. Consequently, the ethnicity variable is encoded into the following dummy variables: autochthonous people (reference category), western allochthonous people, and non-western allochthonous people.

Finally, the ethnicity variable is not included in the model based upon the ESE-student survey since no such data was provided in the database. The analysis is thus limited to the nationality variable.

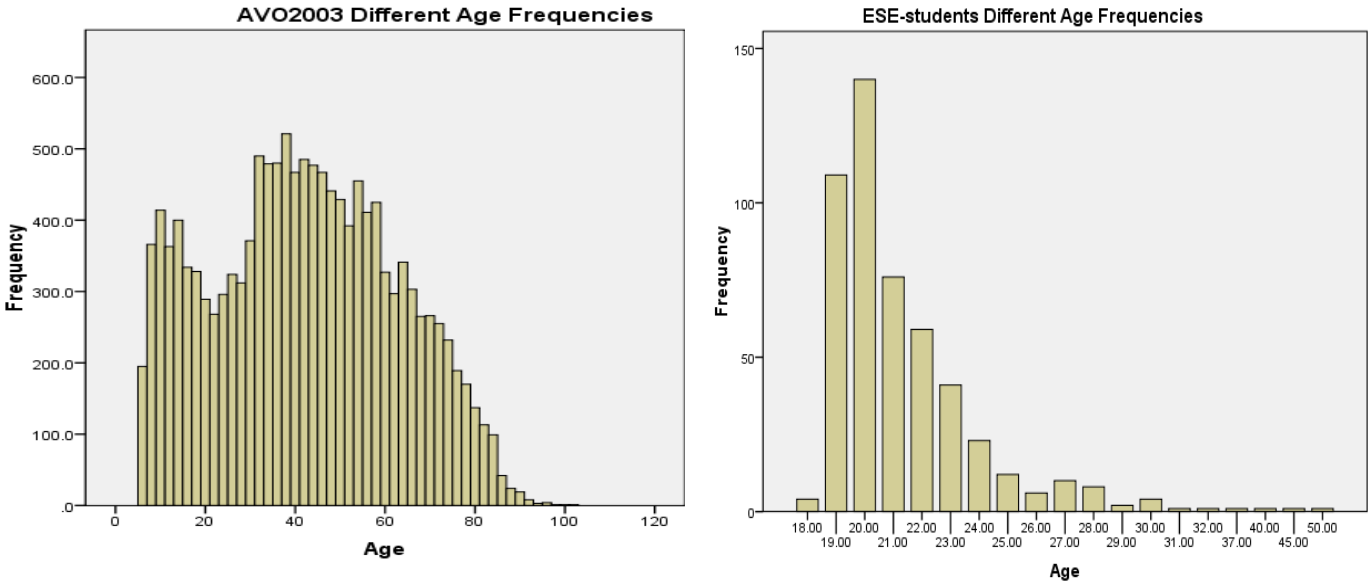
Age

The reason to include the age variable into the logit model is that health is perceived to be strongly influenced by a person's age. According to Fertig (2009), the Grossman model explains the use of age as a variable perfectly. Individuals are perceived to have similar characteristics as capital, since health depreciates through time with age, but it can be increased through investments of time, effort, knowledge and seeking health care. In short, health is both produced and demanded by individuals (Morris, Devlin and Parkin, 2007). By investing time and money in health the two main effects that occur are an increased productivity and happiness. Regarding age, it is thus expected that health depreciates with time, so the older a person becomes, the more likely that he or she is less healthy. Moreover, the relationship between age and health is usually described in scientific articles as being convex shaped and thus not linear.

In both surveys, the age variable is provided as a numerical variable. However, the Dutch AVO2003 survey includes all persons within a household starting from an age of 6, while the ESE-student survey includes data from the age of 18 onwards. Due to this difference, the final outcome of the Dutch survey is expected to reflect a less reliable picture of reality. After all, questions arise such as: How is a child of the age of 7 able to rank his or her self-assessed health? How objective are the results when the parents are most likely to fill in the survey on behave of their children? How old must a person be to reach an age where he or she could make objective decisions? Therefore, to deal with this anticipated bias, the model based upon the AVO2003 survey includes only people of the age of 16

and older. Two graphs for both the ESE-student and the AVO2003 survey are presented on the following page, displaying the frequency of each age.

Graph 1 and 2: AVO2003 and ESE-student different age frequencies



Since the impact of a one year age difference on a person's self-assessed health is expected to be negligible, the usual approach is to cluster the different age groups together. For our research, we subdivide the age variable into 5 intervals all encompassing an age difference between maximum and minimum around 15 years. The groups thus range from 16 to 29 years old for group 1, 30 to 44 for group 2, 45 to 59 for group 3, 60 to 74 for group 4, and 75+ for group 5. Hence, four dummy variables are created with group 1 (16 to 29 years old) as the reference group.

Gender

The most obvious differences between male and female individuals commonly originate from socially constructed roles as well as biological differences. Therefore, it is predicted that a person's gender has an influence on its health and its health perception (Standing, 1998). However, with the recent changes in social behavior, implying that the traditional roles are dissolving, these differences are expected to diminish. Even this can be observed with the latest shifts in women labor positions (positive discrimination) and education levels (more female graduates).

In both databases, gender is categorized as male (value 1) or female (value 2). However, by recoding the variable into a dichotomous one, the male receives the value zero (0), while the female has the value (1). The male gender is thus taken as the reference group.

Socioeconomic status

Similarly to the paper by Mackenbach et al. (1997), we conclude that the two most commonly used indicators of socioeconomic status that are related to perceived health are (household) income and education. Furthermore, this model also takes into account the variable household size to correct for the differences in household income.

Education

As the age variable, the use of the education variable can be justified by the Grossman model. It is stated by Fertig (2009) that the Grossman implies that education makes it easier to install health and that it also encourages individuals to aim at a higher level of health since educated people realize how important a good health is. Hence, the Grossman model argues that a higher level of education results in more investments in health.

Here the education variable is measured as the highest obtained degree. In the AVO2003 database, the variable as22 categorizes the individual's education into 8 groups. However, using the OECD guidelines[10], we reclassify the original individual data into four classes: no and primary, lower, upper, and post secondary. The latter is used as the reference category. Since the ESE-student survey only contains information on the education of the parents and all respondents are university students and thus categorized in the same subgroup, the education variable is not included in the logistic model based upon the ESE-student survey.

[10] OECD. 1990. "Education in OECD Countries: A Compendium of Statistical Information, 1988-89 and 1989-90", special edition. Paris.

Income

It is understandable that a person's income has an impact on his or her health, because the richer somebody is, the more he or she can spend on medical care. Besides, previous research (for example Kunst and Mackenbach, 1994, and Humphrie and van Doorslaer, 2000) has led to the conclusion that a lower SES is related to poorer health.

In our model, the income variable is specified as the net household income. In the AVO2003, income is represented by either net monthly household income or net yearly household income and categorized into 15 possible response categories. About 90 percent of the participants preferred to provide the monthly household income, while the remaining 10 percent filled in their yearly household income. Furthermore, the 15 categories are grouped together by creating approximate quintiles such that around one fifth of the participants are represented in each group. For the monthly household income this leads to the following 5 subgroups: less than 1.400 euros, 1.400 till 1.800 euros, 1.800 till 2.400 euros, 2.400 till 3100 euros and more than 3100 euros. The same procedure for the annual income results into the next 5 subgroups: less than 11.000 euros, 11.000 till 20.000 euros, 20.000 till 31.000 euros, 31.000 till 40.000 euros and more than 40.000 euros. For both, the lowest income group is considered as the reference group.

As can be noticed, the respective categories of the monthly and yearly income do not correspond when simply multiplying the monthly income intervals by twelve. Nevertheless, the respective categories are combined together such that the 10% of the cases for which the annual income is provided are not omitted out of the equation. Since this is only a rough approximation, a robustness test is provided using only the observations for which a monthly household income is available.

The picture is quite different for the ESE-student survey. The students were asked to make an estimation of the gross yearly income of their parents or guardians, while being able to choose from 8 different categories. Reclassifying the data in approximate quintiles yields the following 5 categories: less than 30.600 Euros, 30.600 till 40.800 Euros, 40.800 till 51.000 Euros, 51.000 till 71.400 Euros and more than 71.400 Euros.

Again, the lowest income group is taken as the reference category. However, about 25% of the participating students indicated that they do not know the gross family income or

that they refuse to answer. Hence, to not lose one fourth of the observations, an extra dummy variable is created for which all respondents who did not provide an income figure receive the value one (1) and all others the value zero (0).

Household size

The last variable included into the models is the numerical variable household size. As mentioned, this variable is mainly taken into account because it allows correcting for differences in household income. In the AVO2003 database, the household size is given as a separate variable. However, the student survey does not contain a specific variable that indicates household size, but it includes questions concerning the parents and the brothers and sisters. Hence, household size can be calculated using the following approximation: household size = (1 if mother still alive) + (1 if father still alive) – (1 if parents NOT together) + (number of brothers and sisters + 1).

5. Results

5.1 Descriptive statistics

The descriptive statistics for the independent and dependent variables used in the model based on the ESE-student survey and the model based on the AVO2003 survey are presented in table1 and table 2 respectively.

As explained in detail above, most of the used variables are dummy variables. The sample size is provided to display how many observations are included in every dummy subgroup. For instance, out of table 1 we can derive that 65 of the students that completed the survey have a foreign nationality only, while 53 have both a foreign and a Dutch nationality. It should also be noted that 429 out of 470 students perceive themselves as being healthy. This can be derived from the sample size of the SAH variable. Furthermore, the mean and standard deviation are given for all variables. We can for example notice that the students are approximately 21 years old on average and that the

average household includes about 5 members. However, the mean and standard deviation are less meaningful for the dummy variables.

Table 1: Descriptive statistics for the variables of the ESE-Student survey

	Mean	Std. Deviation	Sample size
Self assesed health (SAH)	.9147	.27961	429
Both a Dutch and a foreign nationality	.1130	.31694	53
Only a foreign nationality	.1386	.34589	65
Age	21.2964	2.79109	500
Gender Female	.2854	.45207	145
Income not given	.1898	.39253	89
Income between €30601and €40800	.1087	.31165	51
Income between €40801 and €51000	.0917	.28889	43
Income between €51001 and €71400	.1557	.36291	73
Income from €71401	.2836	.45122	133
Household size	4.7036	1.28253	472

In table 2, the most important is the column displaying the sample size. It shows for instance that almost 80% (7952 out of 10102) of the interviewees perceives themselves as being healthy. Moreover, only a small fraction of the respondents has a foreign nationality or is allochthonous.

Table 2: Descriptive statistics for the variables of the AVO2003 survey

	Mean	Std. Deviation	Sample size
SAH	.8681	.33838	7952
Both a Dutch and a foreign nationality	.0213	.14435	195
Only a foreign nationality	.0188	.13575	172
Western allochthonous people	.0955	.29395	875
Non-western allochthonous people	.0335	.17999	307
Age between 30 and 44 years	.3133	.46387	2870
Age between 45 and 59 years	.2639	.44075	2417
Age between 60 and 74 years	.1706	.37621	1563
Age from 75 years	.0081	.08952	74
Gender Female	.52	.500	4718
No or primary education	.1529	.35996	1401
lower education	.5467	.49784	5008
upper education	.2384	.42615	2184
Income monthy and yearly 1	.1576	.36442	1444
Income monthy and yearly 2	.2227	.41609	2040
Income monthy and yearly 3	.2028	.40214	1858
Income monthy and yearly 4	.1838	.38738	1684
Household size	2.71	1.295	13776

For the income and age variable, the respondents are fairly equally divided between the different subgroups, while for the education variable, almost half of the people that completed the survey have only a lower educational degree. Finally, remark that the average household size over the entire sample is approximately 3.

5.2 Regression results for the AVO2003 – survey logistic model

The table below shows the logistic regression coefficients, the Wald test statistics, the significance level and the odds ratio for each of the predictors used in the Dutch National survey AVO2003-based model. Regarding the variables in this model, five out of fifteen are insignificant, employing a 0.05 criterion of statistical significance. It can also be noted that the overall model is significant and correctly classifies about 87% of all observations. The results for each variable separately are provided in more detail below.

Table 3: MODEL 1: Dependent variables output for the AVO2003 model including the combined income dummies and the nationality variable, but excluding the ethnicity variable.

Sample size	10102
(-2) Log Likelihood	7045

	Beta	Wald Statistic	P-Value	Odds ratio
Both a Dutch and a foreign nationality	-.446	5.304	.021	.640
Only a foreign nationality	-.178	.631	.427	.837
Age between 30 and 44 years	.072	.539	.463	1.075
Age between 45 and 59 years	-.539	38.332	.000	.583
Age between 60 and 74 years	-.562	37.520	.000	.570
Age from 75 years	-.798	8.877	.003	.450
Gender Female	-.249	15.575	.000	.780
No or primary education	-1.037	32.003	.000	.355
Lower education	-.387	4.853	.028	.679
Upper education	-.240	1.730	.188	.787
Income monthly and yearly 1	.527	35.166	.000	1.695
Income monthly and yearly 2	.715	62.711	.000	2.044
Income monthly and yearly 3	.972	87.119	.000	2.645
Income monthly and yearly 4	1.212	100.410	.000	3.360
Household size	.183	38.862	.000	1.201

Nationality (see table 3)

The nationality variable was dummy coded into 3 groups, using the Dutch nationality as the reference group. According to the results, only the group including the Dutch and foreign nationality appears to be significantly different from the reference group. The odds of feeling healthy are 36% lower compared to the individuals with a Dutch nationality. However, for people with only a foreign nationality no clear-cut conclusions can be drawn due to the insignificance.

Ethnicity (see table 4)

The block entry model (Model 2) includes the variable ethnicity together with all other variables that are used in the previously mentioned model, which excludes ethnicity from the equation. The variable ethnicity is included to see whether a person's ethnic background has an influence on his or her self-assessed health. In general, it can be noted that allochthonous people have lower odds of perceived good health compared to autochthonous people. To be more specific, the odds of feeling healthy for non-western allochthonous people are 54% lower, whereas western allochthonous people odds are 18% lower compared to the reference group, which consists of only autochthonous people. Nonetheless, this discrepancy is only significant for the non-western allochthonous segment on the 0.05 significance level. The western allochthonous subgroup significantly varies from the reference group on a 0.10 significance level only. Moreover, when adding the ethnicity variable to the equation, the previously mentioned results for the nationality variable change. No significant difference can be observed anymore between the several nationality groups. A possible explanation for this shift in results is a strong correlation between the variables nationality and ethnicity. Therefore, the nationality variable may incorporate the effects of ethnicity.

To sum up, autochthonous people are shown to feel healthier than allochthonous people in general. This support the findings that allochthonous people are in general less healthy compared to the entire Dutch population. A possible explanation might be that allochthonous individuals have to adapt to a different culture and a new set of norms and

values, hereby laying a burden on the individual's self-assessed health. Moreover, as was previously mentioned in the introduction, ethnicity most likely correlates with physical, mental, social characteristics; and which affect a person's lifestyle, living environment, and (perception and experience of) illness, health, and health care demand. Hence, it can be concluded that a possible reason for the lower self-assessed health of allochthonous people is their different lifestyle.

Table 4: MODEL 2: Dependent variables output for the AVO2003 model including the combined income dummies and both the nationality and the ethnicity variable.

Sample size	10102
(-2) Log Likelihood	7021

	Beta	Wald Statistic	P-Value	Odds ratio
Both a Dutch and a foreign nationality	-.079	.139	.710	.924
Only a foreign nationality	.086	.128	.721	1.090
Western allochthonous people	-.193	2.927	.087	.824
Non-western allochthonous people	-.782	25.889	.000	.458
Age between 30 and 44 years	.063	.408	.523	1.065
Age between 45 and 59 years	-.552	39.850	.000	.576
Age between 60 and 74 years	-.590	40.861	.000	.555
Age from 75 years	-.837	9.805	.002	.433
Gender Female	-.245	15.107	.000	.783
No or primary education	-1.040	32.038	.000	.353
Lower education	-.388	4.862	.027	.679
Upper education	-.231	1.606	.205	.794
Income monthly and yearly 1	.507	32.278	.000	1.660
Income monthly and yearly 2	.689	57.820	.000	1.991
Income monthly and yearly 3	.940	80.796	.000	2.560
Income monthly and yearly 4	1.176	94.032	.000	3.242
Household size	.194	43.216	.000	1.214

Further, the results for the remaining control variables using the model including the ethnicity variable (Model 2) are briefly presented in table 4. The outcomes of this model are similar to the model excluding the ethnicity variable. Therefore, the results for the model including the nationality variable and not the ethnicity variable (Model 1) are presented below since it may also offer a better base for a comparison with the ESE-Student survey model, for which the results are provided in the next section.

Age (see table 3)

Overall, the different subgroups of the age variable appear to have significantly lower odds of feeling healthy compared to the base group. The latter represents all individuals of the age between 16 and 29 years old in the sample. Moreover, while the age in the different subgroups increases, the odd ratio decreases. For instance, the odds of feeling healthy for the oldest segment are only 0.450 times the odds for the reference group, whereas the preceding segment shows an odd ratio of 0.570. It thus seems that the older you are, the less likely it is that you assess yourself as being healthy.

Gender (see table 3)

The results show that women have lower odds of assessing themselves as healthy, compared to their counterparts. To elaborate, the odds of feeling healthy for the female gender are approximately 22% lower compared to the male gender. Overall, men thus seem to feel healthier than women.

Education (see table 3)

According to the results, the differences in odd ratio between the reference group, which is post secondary education, and the other lower groups are coherent with the general expectations. The odds ratio of the different highest obtained degree variable dummies is significantly declining with decreasing education levels. For instance, the odds of feeling healthy for the segment with the lowest level of obtained degree (i.e. no or only primary education) are only 0.355 times the odds of the reference group. Or to put it differently, the odds of feeling healthy for individuals with no or primary education degree is roughly 66% lower compared to the post secondary education group. However, the upper education segment reveals an insignificant difference in odds of feeling healthy compared to the post secondary education segment. It does not seem to have any substantial impact on a person's perceived health whether he or she has obtained a secondary or a post secondary degree.

Household size (see table 3)

The odds ratio of 1.201 for household size indicates that the odds of feeling healthy increases by approximately 20% for an increase of household size by one member. Furthermore, this relationship is statistically significant according to the model. Therefore, it can be stated that people living together with more family members are likely to assess their own health more positively.

Income (see table 3)

In general, the results show that the higher the level of household income, the greater the odds of feeling healthy increase compared to the odds of the reference group, which consists of the lowest income group. For example, the odds of feeling healthy for the highest household income group are 3.36 times as high as the reference group. Hence, it can be concluded that the higher the household income is, the more likely a person is to perceive him- or herself as being healthy. This conclusion is closely in line with what was expected in the related literature, since people with a greater income are capable of spending a higher amount of money on their personal health. Moreover, it supports the general assumption that the less money a person possess (or in general a lower SES), the poorer health he or she has.

We also performed a robustness test by running the model that only includes the cases for which the monthly household income was provided, instead of using a combination of the observations for which either a monthly or a yearly income was available. However, the results are similar between the two models and the same conclusions can thus be drawn. Therefore, our preferences lie with the model that includes the most observations, which is obviously the model that combines both the monthly and yearly household income data as a variable in the equation. Table A1 and A2 in the appendix display the results for the model including only the observations for which a monthly income was provided.

Moreover, an extra check is executed to see whether there is a significant difference in the odds of feeling healthy between respondents providing monthly data or respondents providing annual income data. This test is performed by introducing a new dummy variable having a value one when the person gave its monthly income. The reference

group consists thus of participants filling in their annual income. Since the p-value is clearly above 0.05 there is no significant difference between both groups. Therefore, it does not make a difference when people with annual income are included or excluded. The details can be found in table A3 and A4 of the appendix.

5.3 Regression results for the ESE-student survey logistic model

The results for the ESE-student survey based model (Model 3) are substantially different from the previously mentioned models. The beta coefficients, Wald statistics, significance levels and odds ratios are presented in the table below (table 5). In this logistic model, only 2 out of the 10 variables are statistically significant with a p-value below 0.05 as the statistic significance criterion. Nevertheless, it can be noticed that the model is overall significant and correctly predicts the self-assessed health for 91.5 percent of all observations. The detailed results for each variable are provided below.

Table 5: MODEL 3: Dependent variables output for the ESE-Student model including a dummy for observations with missing income figures

Sample size	470
(-2) Log Likelihood	249

	Beta	Wald Statistic	P-Value	Odds ratio
Both a Dutch and a foreign nationality	-.899	4.682	.030	.407
Only a foreign nationality	.831	1.669	.196	2.297
Age	-.048	.916	.339	.953
Gender: Female	-1.017	8.214	.004	.362
Income not given	.551	1.151	.283	1.734
Income between €30601and €40800	-.023	.002	.966	.977
Income between €40801 and €51000	1.680	2.426	.119	5.367
Income between €51001 and €71400	.463	.691	.406	1.588
Income from €71401	.807	2.290	.130	2.240
Household size	-.007	.020	.888	.993

Nationality

The results reveal that an individual with multiple nationalities has approximately a 59% lower odd of feeling healthy compared to an individual with only a Dutch nationality. However, it is not possible to yield any statistically significant conclusions for

individuals who only have a foreign nationality. A possible explanation for this relationship is that individuals who filled in the option of foreign nationality are only international or exchange students, which have a high average socioeconomic status and hence also a higher self perceived health. Furthermore, they are not likely to differ considerably in their health perception from students with only a Dutch nationality. In comparison with the previous model, the odds ratio of individuals with both a Dutch and a foreign nationality compared to individuals with a Dutch nationality is higher than in this model. Here, the odds of feeling healthy for people with a double nationality are only 0.4 times the odds of people with the Dutch nationality. Again, the results of the dummy variable representing individuals with only a foreign nationality do not allow to draw any conclusions, because of the insignificance.

Age

According to the results, the odds ratio of 0.953 for age indicates that the odds of feeling healthy decreases with approximately 5% for a 1 year increase in age. As in the previous model it seems that people are more likely to asses themselves as being healthy when they are younger. However, dissimilar to the model from section 5.2, the relationship regarding age is statistically insignificant.

Gender

Compared to the results of the previous model, this model shows that women have an even lower odd of assessing themselves as healthy, compared to their counterparts. Here, the odds of feeling healthy for women is approximately 64% lower that the odds for men. Again, it can be noticed that the difference between the two genders regarding their self-assessed health is highly significant. In conclusion, female students seem to be more likely to assess their health negatively than male students.

Household size

The relationship between the household size variable and self-assessed health is insignificant in this model. Therefore, no straightforward conclusions can be drawn. It

can thus be concluded that the household size is of less importance for students than for people in general, regarding the impact on the self perceived health.

Income

The results in this model show an insignificant relationship for all the different income categories, when compared with the reference group, which again consists of the lowest incomes. Overall, the household income variable does not seem to have any relevant effect on the perceived health of a student.

Education

As mentioned before, it is suggested to omit the education variable from the equation since all students are categorized within the same subgroup. However, data is provided regarding the education of the parents. Since the income variable is not significant, it might be useful to include another variable representing the socioeconomic status such as the education level of the parents. Therefore, a model is executed that includes both a dummy variable for the highest obtained degree of the father and for the highest obtained degree of the mother. The same dummy subgroups were used as in the AVO2003 model: no or only primary education, lower education, upper education, and post-secondary education as the reference group. The outcome is shown in table 6.

A clear relationship can be noticed: a higher level of education for either one of the parents implies higher odds of feeling healthy for the student. However, only for the situation were the father had no or only primary education, the odds of feeling healthy are significantly lower compared to the case were the father would have had a post-secondary education.

Table 6: Dependent variables output for the ESE-Student model including dummies for the education level of the parents

Sample size	448
(-2) Log Likelihood	225

	Beta	Wald Statistic	P-Value	Odds ratio
Both a Dutch and a foreign nationality	-1.016	3.961	.047	.362
Only a foreign nationality	.579	.653	.419	1.784
Age	-.088	4.364	.037	.916
Gender: Female	-1.010	7.054	.008	.364
Income not given	1.146	4.337	.037	3.146
Income between €30601and €40800	.714	1.362	.243	2.042
Income between €40801 and €51000	2.132	3.732	.053	8.428
Income between €51001 and €71400	.771	1.583	.208	2.162
Income from €71401	.981	2.598	.107	2.667
Household size	.055	.171	.679	1.056
Father: no or primary education	-1.797	4.874	.027	.166
Father: lower education	-.863	2.547	.110	.422
Father: upper education	-.125	.038	.845	.883
Mother: no or primary education	.992	1.081	.299	2.698
Mother: lower education	-.255	.170	.680	.775
Mother: upper education	-.163	.070	.791	.850

Finally, it can be noted that personal factors such as age, gender and nationality have a greater influence on a student's self perceived health than impersonal factors such as household size, household income and the education level of the parents.

5.4 Robustness test: different specification of the dependent variable SAH

In this section we discuss some additional tests with a different specification of the dependent variable self-assessed health. As was previously mentioned, the motivation for these robustness tests is to try to clarify or strengthen our main model results. Moreover, the method adopted here is through redefining the variable SAH by altering the two values 1 and 0. To elaborate, the value one (1) was initially given to the options *good, very good* and *excellent*, while the value zero (0) consisted of the options *fair* and *poor*. However, the new SAH is redefined by shifting the option *good* to the group with the value zero (0), and thus only *excellent* and *very good* are given the value one (1).
The results of the performed robustness tests are found in the tables 7 to 10. We have performed four models in total. The first two are based on the AVO2003 dataset and the last two on the ESE-student survey dataset.

Table 7: Dependent variables output for the AVO2003 model with different SAH specification including the combined income dummies and the nationality variable, but excluding the ethnicity variable.

Sample size	10102
(-2) Log Likelihood	12862

	Beta	Wald Statistic	P-Value	Odds ratio
Both a Dutch and a foreign nationality	-.030	.043	.835	.970
Only a foreign nationality	-.254	2.578	.108	.776
Age between 30 and 44 years	-.161	7.963	.005	.851
Age between 45 and 59 years	-.614	106.929	.000	.541
Age between 60 and 74 years	-.709	97.698	.000	.492
Age from 75 years	-.791	7.961	.005	.453
Gender Female	-.341	64.868	.000	.711
No or primary education	-.868	67.346	.000	.420
Lower education	-.536	36.154	.000	.585
Upper education	-.278	9.129	.003	.757
Income monthly and yearly 1	.160	4.908	.027	1.173
Income monthly and yearly 2	.376	31.825	.000	1.457
Income monthly and yearly 3	.510	54.796	.000	1.666
Income monthly and yearly 4	.704	93.614	.000	2.022
Household size	.146	67.125	.000	1.157

Table 7 shows the dependant variable output of the AVO2003 model with different SAH specification including the combined income dummies and the nationality variable, but excluding the ethnicity variable. The most prominent difference with the main results discussed in section 5.2. concerns the nationality variable. Neither the dummy variable for having a foreign nationality nor the dummy regarding a double nationality is significant in this model. Hence, no clear-cut conclusion can be made. Furthermore, it appears that most of the dummy variables show a lower odds ratio, except education. Therefore, it may be concluded that with the new SAH, the overall odds of feeling healthy for each individual dummy variable is lower compared to the odds of their respective reference groups.

Table 8: Dependent variables output for the AVO2003 model with different SAH specification including the combined income dummies and both the nationality and ethnicity variable

Sample size	10102
(-2) Log Likelihood	12849

	Beta	Wald Statistic	P-Value	Odds ratio
Both a Dutch and a foreign nationality	.191	1.442	.230	1.211
Only a foreign nationality	-.060	.121	.728	.942
Western allochthonous people	-.203	6.220	.013	.816
Non-western allochthonous people	-.375	8.374	.004	.687
Age between 30 and 44 years	-.163	8.174	.004	.849
Age between 45 and 59 years	-.616	107.514	.000	.540
Age between 60 and 74 years	-.715	99.149	.000	.489
Age from 75 years	-.796	8.081	.004	.451
Gender Female	-.338	63.824	.000	.713
No or primary education	-.879	68.584	.000	.415
Lower education	-.547	37.376	.000	.579
Upper education	-.283	9.401	.002	.754
Income monthly and yearly 1	.147	4.166	.041	1.159
Income monthly and yearly 2	.363	29.571	.000	1.438
Income monthly and yearly 3	.495	51.302	.000	1.641
Income monthly and yearly 4	.690	89.447	.000	1.994
Household size	.148	68.636	.000	1.160

Table 8 depicts the dependant variable output of the AVO2003 model with different SAH specification including the combined income dummies and both the nationality variable and ethnicity variable. This model is coherent with the previously listed results. The nationality variable remains insignificant. The most significant result is that non-western allochthonous people have lower odds of feeling healthy compared to autochthonous people in the Netherlands.

Table 9: Dependent variables output for the ESE-Student model with different SAH specification including a dummy for observations with missing income figures

Sample size	464
(-2) Log Likelihood	623

	Beta	Wald Statistic	P-Value	Odds ratio
Both a Dutch and a foreign nationality	-.053	.029	.864	.949
Only a foreign nationality	.854	8.000	.005	2.349
Age	-.030	.885	.347	.971
Gender: Female	-.354	2.715	.099	.702
Income not given	.787	5.138	.023	2.197
Income between €30601and €40800	.775	3.587	.058	2.171
Income between €40801 and €51000	.657	2.306	.129	1.929
Income between €51001 and €71400	1.088	8.098	.004	2.967
Income from €71401	.704	4.143	.042	2.022
Household size	-.022	.085	.771	.978

From table 9 it may be concluded that there are significant differences in results compared to the original model. In the original model (see table 5), the dummy variable *both a Dutch and a foreign nationality* was significant, however in this model the dummy variable *only a foreign nationality* is significant. Nevertheless, the general conclusion still holds. The reason why there is a higher odds ratio for individuals with only a foreign nationality compared to people with only a Dutch nationality most likely lies in the fact that people who filled in that option are all exchange or international students, who probably come from a high-average socioeconomic status and hence have a higher SAH. Furthermore, it can be noticed that the odds ratios for the variables age and female gender are not as low as before.

Table 10: Dependent variables output for the ESE-Student model with different SAH specification including dummies for the education level of the parents

Sample size	448
(-2) Log Likelihood	599

	Beta	Wald Statistic	P-Value	Odds ratio
Both a Dutch and a foreign nationality	.019	.003	.956	1.019
Only a foreign nationality	.819	5.946	.015	2.268
Age	-.025	.620	.431	.975
Gender: Female	-.350	2.510	.113	.704
Income not given	.875	5.525	.019	2.399
Income between €30601and €40800	.801	3.490	.062	2.227
Income between €40801 and €51000	.719	2.521	.112	2.052
Income between €51001 and €71400	1.124	7.526	.006	3.077
Income from €71401	.698	3.191	.074	2.011
Household size	-.023	.091	.763	.977
Father: no or primary education	-.517	.790	.374	.597
Father: lower education	-.050	.030	.862	.951
Father: upper education	-.243	.705	.401	.784
Mother: no or primary education	.203	.125	.724	1.225
Mother: lower education	-.032	.010	.919	.969
Mother: upper education	.089	.095	.758	1.093

As is shown in table 10, for the dependent variables output for the ESE-Student model with different SAH specification including dummies for the education level of the parents, the same conclusions can be drawn. Again, the dummy variable *only a foreign nationality* appears to be significant. However, none of the education dummy variables of the parents are significant. Therefore, it is not possible to derive any conclusions about

SAH regarding education. Nevertheless, the highest income dummy variable is significant and this is similar to the results of the main model.

To conclude, the additional robustness tests, which are based upon a different specification of the dependent variable self-assessed health, do in fact strengthen our main model results. The first two tests that exploit the AVO2003 dataset have not shown any contradiction with the results of the main models. Ethnicity appears to have a significant affect on a person's odds of feeling healthy. Furthermore, the last two tests, which are based on the ESE-student survey dataset, do not show any inconsistency with the main model results. Personal factors such as age, gender and nationality still have a greater influence on a student's self perceived health than impersonal factors such as household size, household income and the education level of the parents.

6. Discussion and Conclusion

Health inequalities minimization has been on the agenda as one of the top priorities for nations worldwide. This is not limited to an inter-country comparison, but it also includes an intra-country minimization. For the Netherlands, a considerable amount of research exists on health inequalities, which have been measured by emphasizing on various aspects, such as income, socioeconomic status and ethnicity. The existing literature on the Netherlands indicates a correlation between certain specific diseases and ethnic background. Moreover, it has been determined that non-western allochthonous people have a 22% higher burden of disease compared to the entire Dutch population. However, while most papers and policy documents focus on minorities having less healthy lifestyles and lower health care, our paper aims to contribute to the literature on health inequalities and ethnic differences in the Netherlands. Furthermore, we aspire to strengthen the generally accepted conclusion with the argument that an individual's nationality or ethnicity most likely has an effect on the self-assessed health due to the higher burden of disease for non-western allochthonous people. The data used in this

paper corresponds to the AVO2003 National Household Survey and the Erasmus School of Economics Student Survey 2008.

In the analysis of health inequalities or variations in health outcomes between the different ethnic groups, special attention is drawn towards the measurement of the self-assessed health (SAH) that is available in both surveys as it is used as the dependent variable in all the applied models. The main objective is to find a correlation between a person's nationality or ethnicity and its self perceived health within the Netherlands, while controlling for other factors such as age, gender, household income, household size, and education. The main results show that a worse self reported health is related to a person's ethnicity, age, the amount of education, gender, family size, and even the amount of household income. For the particular case of allochthonous people, we found evidence that autochthonous people tend to feel healthier than foreigners in the Netherlands. In particular, individuals from a non-western ethnic background are more likely to report a lower level of self-assessed health.

Additionally, although the ethnicity variable shows significant explanatory power, nationality remains inconclusive. A possible explanation might be that allochthonous individuals have to adapt to a different culture and a new set of norms and values, hereby laying a burden on the individual's self-assessed health. Moreover, it can be concluded that the lifestyles of the (most) allochthonous people are less healthy and this in turn may affect their SAH. Besides, the other control variables: age, gender, education, household income and household size have supported our previously mentioned expectations. The evidence shows that the older a person becomes, the less likely he or she will assess themselves as healthy. Regarding education, the lower the level of education a person has obtained, the lower the odds of feeling healthy compared to individuals with a post-secondary education. Furthermore, an increasing household size is shown to have a positive effect on an individual's self-assessed health, although there is no evidence from the student-based model supporting this. Additionally, our findings indicate that gender has a significant effect on SAH. More specific, women tend to assess their health condition worse than men do. This relation has proven to be strongest when based upon

the ESE-student data. Also, we find evidence of income-related health inequalities, whereby a higher household income results into a higher personal SAH and visa versa. However, this relationship is not significant in the ESE-student based model.

Beside the main results, it is of equal importance to recognize that the results of the additional robustness tests strengthen our main results. Neither the first two tests based upon the AVO2003 dataset nor the last two tests based on the ESE-student survey have pointed out any contradiction with the results of the main models. Even when the SAH variable is slightly adjusted, the most significant result reveals that in the Netherlands non-western allochthonous people have lower odds of feeling healthy compared to autochthonous people. Concerning nationality however, the robustness tests provide evidence of individuals with only a foreign nationality having higher odds of feeling healthy than people with only a Dutch nationality. An acceptable reason why this is the case is the fact that people with only a foreign nationality who filled in the survey are all exchange or international students, which most likely have a high socioeconomic status and hence a higher SAH. Finally, the results demonstrate that personal factors have a greater influence on a student's self perceived health than impersonal factors.

Overall, our results imply that nationality and especially ethnicity do have a significant effect on self-assessed health. A general interpretation of our results indicates higher self perceived health of autochthonous people compared to allochthonous people in the Netherlands, especially compared to non-western allochthonous people. Even though nationality seems to have a large impact on self-assessed health, this effect appears to weaken when ethnicity is taken into account. Moreover, the relationship between an individual's nationality and SAH appears to be strengthened when the data is more strongly controlled for age and education. Finally, it can be noted that personal factors such as age, gender and nationality have a greater influence on a student's self perceived health than impersonal factors such as household size, household income and the education level of the parents.

7. Limitations and Recommendations

The limitations of our research are mainly attributable to the validity of the data and the variables. Regarding the data from both surveys, questions are raised concerning whether the used data is representative for the ethnic minorities in the Netherlands, since this is the main priority for this research. The data shows a relative low percentage of non-Dutch nationalities in both surveys. In the AVO2003 survey, merely 4.3% of all individuals do not have a Dutch nationality, whereas official data indicates that 10.8% of the Dutch society has a non-western ethnicity (SEGV, 2008). Additionally, this number does not yet account for all allochthonous people (western and non-western) in the Netherlands. Moreover, in the ESE-student survey 76% of all participants have only a Dutch nationality. The remainder represents mostly international and exchange students. Also, the ESE-student survey is limited to 500 students, which might influence the significance of the results.

Furthermore, our list of comparable control variables might be considered as incomplete and a fuller set may offer additional explanatory power. Our research only included five control variables: age, gender, education, household income, and household size, next to the main variables nationality and ethnicity. Previous research has demonstrated that health inequalities are dynamic and determined by various factors during a person's lifetime. Therefore, by extending and adding other factors in the model, the variations in health are more likely to be accurately predicted. Concerning self-assessed health, additional questions can be raised about the validity of comparisons between various population subgroups with respect to health levels and the measurement of inequalities, since each cultural group has different norms and expectations (Lindenboom et al., 2004). The SAH variable measures how the respondent perceives his or hers own health. However, when assessing your own health, the outcome can be biased by for example the mood of the day. To get a clearer picture, it could be advisable to deal with other kinds of measurements of a person's health level. The health assessment could for instance be executed by an objective outsider such as a doctor or a nurse. An obvious disadvantage of

the latter method is that it is very time consuming and hence less people would be willing to participate in a survey based on this methodology.
To sum up, the main limitations of our research are validity concerns regarding both the data and the variables used, since invalidity may affect the final analysis and the interpretations of the results.

Another limitation of our research is related to the applied logit model. For all the models that were used in this research, we only allowed the dependent variable, self-assessed health, to have two discrete outcomes. However, by applying a multinomial logit model, the dependent variable could have more than two values. Such an analysis could be valuable and lead to interesting conclusions since none of the different possible outcomes: excellent, very good, good, fair and poor, would have to be grouped together.

Finally, our results imply further research, both quantitative and qualitative, regarding health inequalities in the Netherlands. Our results indicate that allochthonous people have a different self-assessed health than autochthonous people. Moreover, our findings give supporting evidence for the inequalities in health that exist nowadays.
However, in order to resolve the possible validity concerns, further investigation should be directed towards homogenous reporting of health levels and towards watertight methods that assure representative data of ethnic minorities. Health inequalities remain to be important as well as the interest about its origins and the exact measures that are needed to minimize it. Hence, a suggestion for further research is a greater emphasis on the root causes of health inequalities across different ethnic groups. Qualifications of and recognizing these root causes will help to minimize the occurrence of inequalities. Moreover, this investigation will evidently help the worldwide goal of minimizing any existing health inequalities between and within countries.

8. Appendix

Table A1: Dependent variables output for the AVO2003 model including the dummies for monthly income and the nationality variable, but excluding the ethnicity variable.

Sample size	9160
(-2) Log Likelihood	6423

	Beta	Wald Statistic	P-Value	Odds ratio
Both a Dutch and a foreign nationality	-.399	3.876	.049	.671
Only a foreign nationality	-.101	.181	.670	.904
Age between 30 and 44 years	.073	.508	.476	1.076
Age between 45 and 59 years	-.531	33.713	.000	.588
Age between 60 and 74 years	-.557	34.443	.000	.573
Age from 75 years	-.837	9.451	.002	.433
Gender Female	-.245	13.803	.000	.783
No or primary education	-1.054	27.931	.000	.349
Lower education	-.407	4.509	.034	.665
Upper education	-.292	2.158	.142	.747
Income monthly 1	.540	33.776	.000	1.716
Income monthly 2	.746	62.241	.000	2.109
Income monthly 3	1.036	89.872	.000	2.817
Income monthly 4	1.349	106.999	.000	3.853
Household size	.179	33.399	.000	1.196

Table A2: Dependent variables output for the AVO2003 model including the dummies for monthly income and both the nationality and ethnicity variable.

Sample size	9160
(-2) Log Likelihood	6397

	Beta	Wald Statistic	P-Value	Odds ratio
Both a Dutch and a foreign nationality	-.045	.042	.838	.956
Only a foreign nationality	.134	.278	.598	1.144
Western allochtonous people	-.136	1.306	.253	.873
Non-western allochtonous people	-.830	27.835	.000	.436
Age between 30 and 44 years	.062	.362	.547	1.064
Age between 45 and 59 years	-.544	35.084	.000	.581
Age between 60 and 74 years	-.588	37.984	.000	.556
Age from 75 years	-.877	10.415	.001	.416
Gender Female	-.242	13.447	.000	.785
No or primary education	-1.047	27.445	.000	.351
Lower education	-.397	4.265	.039	.672
Upper education	-.273	1.887	.170	.761
Income monthly 1	.519	31.028	.000	1.680
Income monthly 2	.719	57.408	.000	2.053
Income monthly 3	1.001	83.389	.000	2.721
Income monthly 4	1.310	100.349	.000	3.707
Household size	.192	37.919	.000	1.212

Table A3: Dependent variables output for the AVO2003 model including combined income dummies and an extra dummy indicating whether monthly or annual income data is available

Sample size	10102
(-2) Log Likelihood	7019

	Beta	Wald Statistic	P-Value	Odds ratio
Both a Dutch and a foreign nationality	-.079	.138	.710	.924
Only a foreign nationality	.086	.129	.719	1.090
Western allochthonous people	-.191	2.863	.091	.826
Non-western allochthonous people	-.778	25.634	.000	.459
Age between 30 and 44 years	.065	.426	.514	1.067
Age between 45 and 59 years	-.557	40.458	.000	.573
Age between 60 and 74 years	-.590	40.897	.000	.554
Age from 75 years	-.834	9.734	.002	.434
Gender Female	-.245	15.053	.000	.783
No or primary education	-1.030	31.347	.000	.357
Lower education	-.379	4.644	.031	.684
Upper education	-.225	1.517	.218	.799
Income monthly and yearly 1	.504	31.919	.000	1.655
Income monthly and yearly 2	.690	57.898	.000	1.993
Income monthly and yearly 3	.944	81.340	.000	2.570
Income monthly and yearly 4	1.179	94.378	.000	3.251
Monthly income provided	-.146	1.630	.202	.864
Household size	.191	41.817	.000	1.211

Table A4: Dependent variables output for the AVO2003 model including a dummy indicating whether monthly or annual income data is available

Sample size	10102
(-2) Log Likelihood	7164

	Beta	Wald Statistic	P-Value	Odds ratio
Both a Dutch and a foreign nationality	-.176	.707	.400	.839
Only a foreign nationality	-.003	.000	.989	.997
Western allochthonous people	-.209	3.501	.061	.812
Non-western allochthonous people	-.949	39.251	.000	.387
Age between 30 and 44 years	.161	2.723	.099	1.175
Age between 45 and 59 years	-.426	24.638	.000	.653
Age between 60 and 74 years	-.533	34.407	.000	.587
Age from 75 years	-.820	9.703	.002	.441
Gender Female	-.269	18.608	.000	.764
No or primary education	-1.504	71.470	.000	.222
Lower education	-.690	16.137	.000	.502
Upper education	-.356	3.877	.049	.701
Monthly income provided	-.119	1.086	.297	.888
Household size	.301	105.666	.000	1.351

References

- Banks, J. et al. 2006. "Disease and Disadvantage in the United States and in England", American Medical Association.

- Eijl, C. van. 2006. "Migranten in Nederland. NL". Report vooronderzoek website migratiegeschiedenis. Amsterdam.

- Deekes, J. 1996. *What is an odds ratio?* Bandolier, Evidence-Based Health Care.

- Fertig, A. 2009. *Demand for health and health care*. Lecture slides.

- Field, Andy. 2005. *Discovering Statistics Using SPSS*, 2nd edn. SAGE Publications.

- Heeten, J. den, A.O. Verweij. 1993. "Identificatie en registratie van etnische herkomst; een handleiding voor registratie en beleid", Vereniging van Nederlandse Gemeenten (VNG).

- Hernández Quevedo, C., Jimenez Rubio, D. 1997. "A comparison of the health status and health care utilisation patterns between foreigners and the national population in Spain: new evidence from the Spanish National Health Survey". York, HEDG.

- Humphries, K.H. and van Doorslaer, E. 2000. "Income-related health inequalities in Canada" Social Science & Medicine 50. p. 663-671.

- Kunst, A., Mackenbach, J., Lamkaddem, M., Rademaker, J., Devillé, W. 2008. "Overzicht en evaluatie van resultaten van wetenschappelijk onderzoek naar

etnische verschillen in gezondheid, gezondheidsrisico's en zorggebruik in Nederland". Erasmus MC Rotterdam en NIVEL.

- Kunst, A.E., Mackenbach, J.P., 1994. *The size of mortality differences associated with educational level in nine industrialized countries.* American Journal of Public Health 84. p. 932-937.

- Kakwani, N., Wagstaff, A., van Doorslaer, E. 1997. "Socioeconomic inequalities in health: measurement, computation, and statistical inference". Journal of Econometrics, vol. 77. p. 87-103.

- Mackenbach, J., Kunst, A., Cavelaars, A., Groenhof, F., Geurts, J., the EU Working Group on Socioeconomic Inequalities in Health. "Socioeconomic inequalities in morbidity and mortality in western Europe". The Lancet, vol 349. June 7 1997. p. 1655-1659.

- Morris, S., Devlin, N., Parkin, D. 2007. *Economic Analysis in Health Care.* John Wiley & Sons, Ltd. p. 40-47.

- OECD. 1990. "Education in OECD Countries: A Compendium of Statistical Information, 1988-89 and 1989-90", special edition. Paris.

- SEGV plan van aanpak / gezond leven. 2008. "Naar een weerbare samenleving - Beleidsplan aanpak gezondheidsverschillen op basis van sociaaleconomische achtergronden". Bilthoven: RIVM

- Standing, Hillary. 1998. "Guidelines for the analysis of Gender and Health", chp2. Gender and Health Group at the Liverpool School of Tropical Medicine

- Wagstaff, A, Doorslaer, E. van. 2000. *Handbook of Health Economics*, volume 1B. Elsevier Science B.V. p. 1846.

- Wang, P., Elsbett-Koeppen, R., Geng, G., and Badley, E. 2000. "Arthritis Prevalence and Place of Birth: Findings from the 1994 Canadian National Population Health Survey". American Journal of Epidemiology, 152: 442-445.

- World Health Organization. 1948. "WHO definition of health", WHO, http://www.who.int/about/definition/en/print.html (10-06-2009)

- World Health Organization. 2004. "The global burden disease", WHO, http://www.who.int/healthinfo/global_burden_disease/en/index.html (20-07-2009)

- Wooldridge, Jeffrey M. 2006. *Introduction to Econometrics, a modern approach*, 3rd edn. Michigane: Thomson South-Western

- UCLA Academic Technology Services. "How do I interpret odds ratio in Logistic Regression", http://www.ats.ucla.edu/stat/Spss/faq/oratio.htm (21-07-2009)